농학박사학위논문

Identification of Adaptive Signatures in the Cattle Genome

소전장유전체에서 적응적 흔적 발굴

2018년 2월

서울대학교 대학원

농생명공학부

Mengistie Taye Terefe

Identification of Adaptive Signatures in the Cattle Genome

By

Mengistie Taye Terefe

Supervisor: Professor Heebal Kim, Ph.D

February 2018

Department of Agricultural Biotechnology

Seoul National University

Identification of Adaptive Signatures in the Cattle Genome

Academic advisor: Professor Heebal Kim (Ph.D.)

Submitting a doctoral thesis of Agricultural Biotechnology

December 2017

Graduate School of Seoul National University

Agricultural Biotechnology Major

Mengistie Taye Terefe

Confirming the doctoral thesis written by Mengistie Taye Terefe

January 2018

Chairperson	Cheol-Heui Yun	(Seal)
Vice Chairperson	Heebal Kim	(Seal)
Examiner	Ki-Duk Song	(Seal)
Examiner	Sook Hee Yoon	(Seal)
Examiner	Seoae Cho	(Seal)

Abstract

Identification of Adaptive Signatures in the Cattle Genome

Mengistie Taye Terefe

Department of Agricultural Biotechnology

The Graduate School

Seoul National University

Cattle are one of the most common and numerous domestic ungulates. The genomes of domesticated cattle breeds harbor the history of domestication and breed formation due to the combined effect of natural and artificial selection forces. Deciphering the footprints of these selection forces in the genome of cattle breeds is of great interest from the perspective of evolutionary biology seeking to understand the key adaptive features that have generated enormous morphological and production phenotypic variations currently observed within and between populations. Recently, professionals from molecular population and evolutionary genetics have shown growing interest in distinguishing neutral molecular variations from variations that are subject to selection, particularly positive selection, in the genomes of multiple organisms including cattle. The building of the bovine reference genome and the accumulation of single nucleotide polymorphism (SNP) data from geographically and biologically diverse cattle breeds – due to the emergence of low cost and high throughput Next

Generation Sequencing (NGS) technologies – has created unprecedented opportunities and facilitated efforts to uncover and understand this variation.

In this doctoral dissertation, the whole genome NGS SNP data from African, N'Dama, Ankole, Holstein, Hanwoo, and Angus cattle breeds were used to elucidate the footprints of natural and artificial selection forces that have contributed to the major phenotypes of the respective breeds. The cross-population extended haplotype homozygosity (XP-EHH) and cross-population composite likelihood ratio (XP-CLR) statistical methods were used to search for the genes/gene regions affected due to selection. The reference genome of cattle (UMD3.1) was used to annotate genes in outlier regions under selection from these analyses. I used the Database for Annotation, Visualization, and Integrated Discovery (DAVID) gene ontology and annotation tool for gene enrichment analysis to understand the biological functions and pathways of genes identified under selection.

In Chapter 1, I introduced the variations in cattle breeds with special emphasis to African cattle breeds, the principles behind signature of positive selection, and the objectives and methods of identification of signature of positive selection. In addition, previously reported results of studies on selection signatures from genetically diverse cattle breeds were reviewed.

In Chapter 2, the genome of African cattle breeds was compared with the genome of Commercial Asian-European taurine cattle breeds to reveal genomic regions under selection in African cattle in relation to tropical environment adaptation traits. African cattle breeds have evolved in a hot tropical climate for millennia, which helped them to develop an inherent superior thermotolerance ability. The study revealed several genes/gene regions under selection that are overrepresented in different biological process (BP) terms and pathways in a gene enrichment analysis. In relation to heat stress response, "angiogenesis" and "regeneration" BP terms were enriched. Moreover, several selected genes were involved in anatomical structures, and physi-

ological and/or molecular functions that are associated with heat tolerance mechanisms. These genes are involved in oxidative stress response, osmotic stress response, heat shock response, hair and skin properties, sweat gland development and sweating, feed intake and metabolism, and reproduction functions. Therefore, the genes and BP terms identified here directly and/or indirectly contribute to the superior heat tolerance mechanisms of African cattle populations. The high tropical temperature where these cattle breeds have evolved for millennia could be a selective pressure for the development of these thermotolerance mechanisms.

In Chapter 3, the genomes of Holstein, Hanwoo, and N'Dama cattle breeds were explored in order to decipher genomic regions affected due to divergent selection for milk traits, meat production and quality traits, and environmental adaptation traits, respectively. Artificial and natural selection for a particular trait in cattle have significantly modified the cattle genome. Due to this, several cattle breeds have been developed with a mosaic of morphological, productivity, and environmental adaptation characteristics. Holstein cattle are evolved as dairy cattle and Hanwoo cattle are evolved as beef cattle under artificial selection, whereas N'Dama cattle are evolved as a general-purpose breed – a breed that does not artificially selected for a particular purpose under natural selection. Identifying genomic regions affected due to artificial and natural selection forces in cattle would give an insight into the history of selection for economically important traits and genetic adaptation to specific environments of populations under consideration. From this study, genes/gene regions that are related to milk traits (e.g., *CSN3*, *PAPPA2*, and *ADIPOQ*), meat production and quality traits (e.g., *NCOA2*, and *PITPN3*), and environmental adaptation traits (e.g., *SLC40A1*, *STOM*, and *COMMD1*) were found under positive selection from the genomes of Holstein, Hanwoo and N'Dama cattle breeds, respectively. Moreover, significant functional annotation cluster terms including milk protein and thyroid hormone signaling pathway, histone acetyltransferase activity, and renin secretion were enriched from gene lists identified under selection in Holstein, Hanwoo, and N'Dama cattle breeds, respectively.

In Chapter 4, the genome of Ankole cattle (African Sanga cattle) was explored in order to identify genes and genomic regions under positive selection in relation to meat quality traits. African Sanga cattle are an intermediate type of cattle resulting from interbreeding between *B. taurus* and *B. indicus* sub-species. Recently, experimental evidence on the potential of African Sanga cattle breeds for superior beef quality traits over their indicine counterparts has emerged. In this study, the whole genome SNP data of Ankole (Sanga cattle) was compared with the genome of indicine cattle breeds using XP-EHH and XP-CLR statistical methods. As a result, several genes including those affecting beef quality traits such as tenderness, intramuscular fat (IMF) content, and meat color were found under positive selection. The genes identified are involved in BP terms and KEGG pathways that affect muscle structure and metabolism, adipose metabolism, and adipogenesis – which in turn affects meat quality traits. This study asserted that Ankole cattle have the potential for higher meat production and quality traits under the prevailing tropical environmental conditions. These results provide a basis for further research on the genomic characteristics of Ankole and other Sanga cattle breeds for better quality beef in tropical Africa.

In Chapter 5, the genetic blueprint behind the superior beef quality characteristics and other associated phenotypes of Angus cattle were elucidated. Angus cattle have been intensively selected for superior beef quality characteristics for decades. Annotating genomic regions under selection in the genomes of Angus cattle resulted in several genes including those associated with beef quality traits and coat color. In addition, putative genes that potentially cause genetic disorders in Angus cattle were identified. The results from this study will help to further improve Angus cattle beef quality, and take a precaution on the associated genetic disorders which ultimately reduce production and productivity.

In conclusion, from these studies, a catalog of genes were identified under positive selection from African, N'Dama, Ankole, Holstein, Hanwoo, and Angus cattle breeds in relation to the major economic and adaptation traits of the respective

breeds to which they have been selected for. The findings in this dissertation will help us to better understand the adaptive events that have generated the enormous phenotypic variation observed between cattle breeds prevailing today. Molecular markers that contribute to local environmental adaptations (e.g., thermotolerance mechanisms - markers that are difficult to identify with other laboratory experimental methods) were revealed in addition to those affecting production traits such as milk production and quality, beef production and quality, reproduction and other associated traits. The markers identified in these studies help to understand the genetic merit of the breeds and can be used in genomic selection and breeding programs to further improve the respective breeds.

Keywords: African cattle, biological process, bio-marker, KEGG pathways, signature of positive selection, XP-CLR, XP-EHH

Student number: 2014-31453

Table of Contents

List of Tables

List of Figures

Abbreviations

DAVID	Database for Annotation, Visualization, and Integrated Discovery
FCR	Feed Conversion Ratio
FDR	False Discovery Rate
F_{ST}	Fixation index based on Wright's F-statistics
GO-BP	Gene Ontology Biological Processes
IMF	Intramuscular Fat
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
ML	Maximum Likelihood
NGS	Next Generation Sequencing
NJ	Neighbor-joining
PCR	Polymerase Chain Reaction
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphism
Ts/Tv	Transition/Transversion ratio
XP-CLR	Cross-Population Composite Likelihood Ratio
XP-EHH	Cross-Population Extended Haplotype Homozygosity

Chapter 1. General Introduction

1.1 The Genetic Resource of Cattle

Cattle are the most numerous and common types of large domesticated ungulates to date - domestication dates back to about 10,500 years ago in the near east (Loftus et al. 1994; Bollongino et al. 2012). Since domestication, being the economic and cultural heritage, cattle have been migrating to every part of the world with their owners contributing a lot for the civilization of human beings (Ajmone-Marsan et al. 2010; Decker et al. 2014; Wright 2015). Because of this, cattle have been introduced to new environments and diets, insults of disease and parasites, and different climatic conditions (Mirkena et al. 2010). These days' modern cattle breeds can be grouped mainly into *Bos taurus* (taurine) and *Bos indicus* (indicine) sub-species. Interbreeding between them, however, has resulted in many other groups with intermediate morphological characteristics. Currently, there are a large number of cattle breeds in the world (>1200) with considerable intra- and inter-population variation in their production (milk yield and quality, meat production and quality), morphology (coat color, presence/absence of horns) and adaptation (disease resistance, heat tolerance) characteristics (The Bovine HapMap Consortium 2009; Rothammer et al. 2013; Decker et al. 2014).

Together with domestication, natural and artificial selection for breed development contributed a lot for the differentiation of cattle breeds prevailing today. Domestication is the first step of selection that is thought to be responsible for the significant change in the behavioral, coat color and morphological characteristics of cattle breeds – for example, large horns that were necessary for fighting with predators in the wilderness had been replaced with the emergence of short-horned and even hornless cattle after captivity (Ajmone-Marsan et al. 2010; Mirkena et al. 2010; Wright 2015). Tameness, defined as the quality of an animal being welcoming towards the presence of humans, is a common feature of domestic animals as a result of domestication

(Albert et al. 2009; Wright 2015). Wright (2015), has reviewed genes related to domes-domestication in different livestock species. Moreover, natural selection and human artificial selection towards a desired type and level of production and environmental adaptation traits resulted in several diverse general purpose and specialized cattle breeds (Rothammer et al. 2013). Humans have modified the genotypes of cattle for their own benefit more than any other species of domestic livestock. A centuries effort of intensive selective breeding for milk production have evolved dairy cows (e.g., Holstein, Jersey) to become the most efficient biological machines in the world (Stella et al. 2010; Höglund et al. 2015). Similarly, the development of cattle breeds with superior beef quality and production potential (e.g., Angus, Hanwoo) have been possible through intensive artificial selection (Arthur et al. 2001; Chambaz et al. 2003; Albertí et al. 2008; McClure et al. 2010; Porto-Neto et al. 2014). Molecular breeding and genetics methods, together with the development of reproductive technologies (Artificial Insemination and Embryo Transfer), and statistical methods for estimation of breeding values facilitated the development of breeds with the required specialty (Meuwissen et al. 2016).

Considering cattle genetic resources in Africa, in addition to indicine and taurine groups, interbreeding between the two subspecies resulted into two additional genetically well-established groups – the Sanga and Zenga groups. The African taurine cattle are the result of the first introduction of cattle to Africa that happened in two moves – the humpless Longhorns were introduced around 6000 BC followed by the humpless Shorthorn, 2500 years later (Rege 1999). Currently, African taurine cattle are found mainly distributed in the wet-humid parts of West and Central Africa where the prevalence of trypanosomiasis (a disease caused by a blood parasite) hinders the introduction of other cattle groups (Yaro et al. 2016). Sheko cattle are the only remnants of taurine cattle found in East Africa, Ethiopia (Rege and Tawah 1999). Because of the stringency of the environment where these cattle breeds have evolved for millennia, African taurine breeds are smaller and less productive as compared to

African indicine cattle breeds. However, they are known for their hardiness to adapt to the disease (especially trypanosomiasis) and high-temperature challenges of the region. In this regard, the most known and well-characterized breed of this group is the trypanotolerant N'Dama cattle (Rege and Tawah 1999; Okeyo et al. 2015; Yaro et al. 2016).

African indicine cattle (*B. indicus*) were introduced from Asia to Africa later around 1500 BC (Rege 1999; Ajmone-Marsan et al. 2010). They were introduced through the Horn and spread to the north and southern parts of Africa (Okeyo et al. 2015). These group of cattle are larger in size and are better in production traits than African taurine cattle groups. They are found widely distributed in the Eastern and drier parts of West Africa (Rege and Tawah 1999). The third group, the Sanga cattle, is an intermediate cattle which is believed to be a result of interbreeding between taurine and indicine cattle sub-species in Africa (Rege and Tawah 1999). Archaeological evidence by Grigson (1991), however, argue that African Sanga cattle are an ancient autochthonous origin and have come to be mixed with taurine and humped cattle probably only in the last few hundred years, which is why they share a mosaic of characters with the other two taxa. African Sanga cattle breeds can be grouped as Sanga of eastern and southern Africa based on their geographical distribution (Rege and Tawah 1999). The fourth group, the Zenga cattle of east Africa, is a result of interbreeding between Zebu (indicine) and Sanga of Africa (Rege and Tawah 1999). Zenga cattle are found in the highlands of East Africa where a high concentration of Zebu cattle are found – this gave an opportunity for interbreeding with Sanga cattle (Rege 1999; Okeyo et al. 2015). Despite the general belief that African cattle are a result of an introduction from east Asia, Decker et al. (2014) hypothesized the possibility of a genetic contribution of African Aurochs to the genetics of African cattle breeds.

In general, in addition to many uncharacterized, sub-Saharan Africa harbors more than 150 genetically diverse indigenous cattle breeds (Rege 1999; Okeyo et al. 2015). These cattle have evolved to adapt to the harsh environmental conditions prevailing in the continent vis. high prevalence of diseases, high temperature, low and seasonally varied quality and availability of feed and water, and risk of predators that resulted in breeds of different production and morphological characteristics (Mirkena et al. 2010). The current geographical distribution of African cattle breeds is geared to their environmental adaptation characteristics (Okeyo et al. 2015). Despite the prevailing diverse cattle genetic resources, there is a dearth of information on both phenotypic and genotypic characteristics of African cattle breeds. At the same time, this diversity of African cattle is fading away that lots of them are already extinct and others are found at different level of extinction and admixture due to the factors like stringency of the environment and indiscriminate crossbreeding with European breeds for productivity reasons (Rege 1999; Hanotte et al. 2010; Okeyo et al. 2015). Therefore, "*We need to better understand and exploit the genetic diversity of Africa's indigenous livestock breeds—before they fade away"* (Hanotte et al. 2010).

1.2 Positive Selection Signature

1.2.1 Definition and principles of positive selection

According to the theory of neutral evolution, most of the molecular variations within and between species are selectively neutral that does not affect the fitness of the organism. But, when a variant (either a newly arisen variant or standing variant) give a fitness advantage to the carrier individual relative to other members of the population, its carrier is more likely to thrive and leave more offspring than non-carriers, causing the frequency of the beneficial allele to increase in the population

(Utsunomiya et al. 2015). This causes other linked neutral variants to be carried along with the selected variant through a process called hitchhiking resulting in a selective sweep (Biswas and Akey 2006; Vitti et al. 2013; Gouveia et al. 2014; Utsunomiya et al. 2015) – means that a haplotype carrying the beneficial allele will spread in the population quickly. Figure 1.1, adapted from Vitti et al. (2013), clearly demonstrates the effect of an advantageous allele on the genome of the carrier and the population at large. A selective sweep in a genomic region results in reduced genetic variation, skewed site frequency spectra, elevated linkage disequilibrium, and reduced levels of inter-population differentiation and elevated rates of inter-species divergence at and around the selected loci (Vitti et al. 2013; Gouveia et al. 2014; Jensen et al. 2016). On the other hand, variants that do not provide a fitness advantage to the organism carrying them may increase in frequency in the population due to random factors, i.e. by random drift (Gouveia et al. 2014).

Natural and artificial selection forces affected the cattle genome generating an enormous effect on genetic diversity. The principle behind natural and artificial selection forces on the genome of organisms is almost the same – by natural selection, the fittest will survive and reproduce under the prevailing conditions increasing its frequency in the population. In artificial selection, those individuals with the preferred trait (e.g., coat color, milk yield and quality, and beef quality) will be selected by the breeder to produce the next generation. Therefore, in both cases, the frequency of a variant that gives a fitness advantage for the carrier individual or affecting a trait of interest by the breeder together with the hitchhiking neighboring neutral variants will increase in the population. Figure 1.1a illustrates the reduced levels of genetic diversity at a genomic region under selection. Other panels of Figure 1.1(b-d) describes how a selective sweep and hitchhiking of neighboring neutral variants affect frequency spectrum, linkage disequilibrium and population differentiation at a selected loci (Vitti et al. 2013).

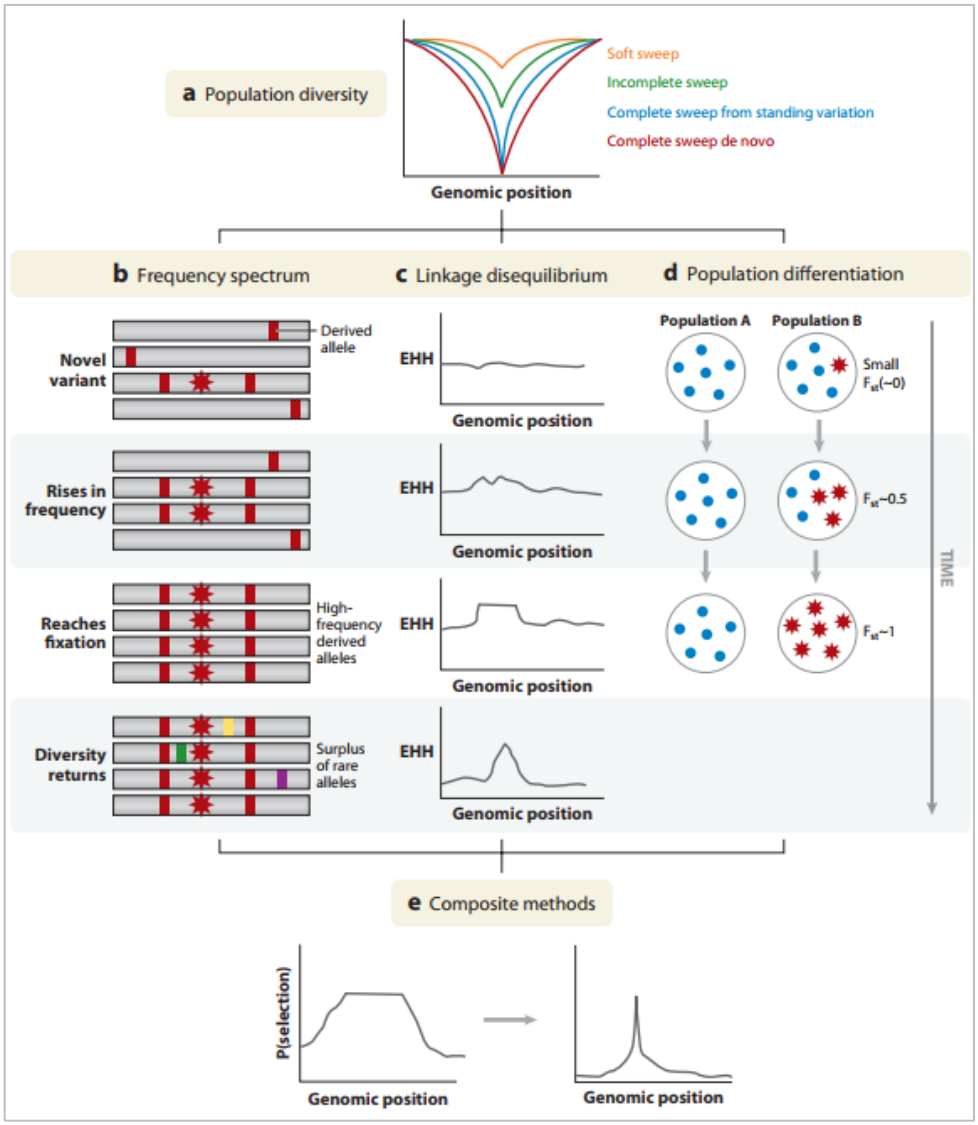


Figure 1.1 Illustration of the effect of an advantageous genetic variant in the genome causing a selective sweep. a) effect of selective sweep on population diversity, b) occurrence of a novel variant and its effect on frequency spectrum, c) effect of a selective sweep on linkage disequilibrium, d) effect of a selective sweep on population differentiation, and e) method for detecting selective sweeps. Adapted from Vitti et al. (2013).

Deciphering genomic regions affected due to natural and artificial selection forces, especially positive selection, is of paramount importance in genomics and population genetics perspectives as well as livestock breeding. The principal objective of identification of signature of selection in cattle is to gain knowledge about the evolutionary processes that are shaping the genome and functional information about genes/genomic regions. It enables to scrutinize the key adaptive events that have generated the enormous phenotypic variation observed between cattle breeds prevailing today (The Bovine HapMap Consortium 2009; Utsunomiya et al. 2015). In addition, it helps to identify and prove causal mutations in regions previously identified by QTL mapping experiments and can reveal genes related to ecological traits (*e.g.*, genes related to tropical adaptation) that are difficult to identify experimentally (Gouveia et al. 2014). Moreover, it helps us understand the biological functions of genes affected by selection that contribute to the differences in adaptation and production traits and help in designing breeding programs to further improve these traits, and design conservation and use mechanisms for endangered breeds.

Recently, with the advent of genomics and bioinformatics tools, revealing the effect of domestication and selection forces on the genome of livestock became possible. In cattle, the sequencing of the whole genome of a female Hereford cattle has significantly facilitated and contributed a lot to the efforts sought to uncover and understand the genetic landscape of several cattle breeds with specific and general traits (Elsik et al. 2009; The Bovine HapMap Consortium 2009).

1.2.2 Methods to identify signature of positive selection in livestock genomes

In order to identify positive selection signature in the genome of livestock resulted due to natural and artificial selection forces, several methods have been developed

and used by different scholars. These methods are based on probing the site frequency spectrum or allele frequency differentiation - e.g., Tajima's D, Nucleotide diversity (Raymond and Rousset 1995; Oleksyk et al. 2010), haplotype length and linkage disequilibrium - e.g., EHH, *iHS*, XP-EHH (Sabeti et al. 2007; Oleksyk et al. 2010; Vitti et al. 2013), or the level of within and/or between population allele frequency differentiation - e.g., F_{ST}, PBS-population branch statistics, XP-CLR (Holsinger and Weir 2009; Chen et al. 2010; Vitti et al. 2013). Methods based on allele frequency differentiation assess the level of DNA polymorphism for a genome-wide set of loci within a population. For instance, Tajima's D explores the distortion in allele frequency of a genomic region as compared to the rest of the genome under neutrality (Vitti et al. 2013). Similarly, methods based on population differentiation (e.g., Wright's *F*-Statistics, F_{ST}), detects an increase or decrease in population differentiation in genomic regions under selection relative to the rest of the genome (Holsinger and Weir 2009). F_{ST} is among the most widely used statistics in population and evolutionary genetics that provide important insights into the evolutionary processes that influence the structure of genetic variation within and among populations (Holsinger and Weir 2009). Whereas, methods that detect positive selection based on differences in haplotype lengths search for extended regions of strong LD (long haplotypes) relative to their prevalence within or between populations (Vitti et al. 2013).

So far, various methods have been applied to identify the signature of positive selection in various livestock species. In cattle, The Bovine HapMap Consortium (2009) used F_{ST} statistics; Makina et al. (2015) used approaches based on haplotype structures and allele frequency differences between populations (F_{ST}); Bahbahani et al. (2015) used inter-population genome-wide F_{ST} analysis and extended haplotype homozygosity (EHH)-derived statistics (iHS and *Rsb*); Kim et al. (2017a) and Lee et al. (2014a) used cross-population extended haplotype homozygosity (XP-EHH) and cross-population composite likelihood ratio (XP-CLR) methods, and revealed several genomic regions under selection due to domestication, natural and intensive artificial

selection forces. Qanbari et al. (2011) and Zhao et al. (2015) used site frequency (iHS) and F_{ST} methods to identify signature due to decades of intensive artificial selection for traits of economic importance in modern cattle breeds. Rothammer et al. (2013), applied XP-EHH statistics to search for signature of artificial selection in ten beef and dairy cattle breeds and identified genes associated with known QTL for beef or dairy traits.

On other species, studies used XP-EHH (Sabeti et al. 2007; Pickrell et al. 2009), XP-CLR (Chen et al. 2010), and iHS (Voight et al. 2006) statistical methods to identify and characterize positive selection signature in human populations. Rubin et al. (2012) and Lee et al. (2017) also applied heterozygosity statistics (ZHp) in the domestic pig genome.

From literature, it is understood that almost every individual method has its own strengths and weaknesses (Qanbari and Simianer 2014; Ma et al. 2015a; Utsunomiya et al. 2015; Vatsiou et al. 2016). Genomic regions identified by one method may not be identified by other methods even using the same data set due to the differences in the methods that they focus and target the signal left by selection and the time scale on which selection can act (Oleksyk et al. 2010; Qanbari and Simianer 2014). However, it has been stated that the success of one test and failure of the other does not exclude the region of interest from having been subjected to selection (Gouveia et al. 2014). Because of this, the use of a combination of approaches for scanning the signals of selection has been suggested that increase the reliability, power, and resolution of the studies (Vitti et al. 2013; Gouveia et al. 2014; Ma et al. 2015b; Utsunomiya et al. 2015; Vatsiou et al. 2016). In this dissertation, I used XP-EHH, XP-CLR, F_{ST} and Tajima's D statistical methods in different chapters together and/or separately as indicated in Table 1.1. Characteristics of each of the methods, based on previous studies, is presented in Table 1.1.

Table 1.1 Statistical methods used for identification of signature of selection in this thesis and their characteristics

Parameter	XP-CLR	XP-EHH	Tajima's D	F_{ST}
Chapters used	▪ Chapters 2, 3, 4 & 5	▪ Chapters 2, 3, & 4	▪ Chapters 3 & 4	▪ Chapters 4
Purpose	▪ Between population comparison	▪ Between population comparison	▪ Within-population statistics	▪ Between population statistics
Characteristics	▪ It detects SNPs that are under selection in one population (test) but not in the other population (reference) ▪ Uses allele frequency differentiation between populations ▪ Robust for ascertainment bias ▪ Do not need phasing of data ▪ Detect recent and ongoing selection sweep regions	▪ It detects genomic regions that are under selection in one population (test) but not in the other population (reference) ▪ A haplotype-based method - assesses haplotype differences between two populations ▪ Follows a standard normal distribution ▪ Need phasing of data ▪ Detect recent, fixed or nearly fixed sweep regions	▪ Compares the difference between the mean pairwise difference and the number of segregating sites in nucleotide polymorphism data to detect selection signatures ▪ It detects SNPs selected in a genomic region as compared to the rest of the genome ▪ Detects selective sweep regions going to fixation in the population that makes rare alleles in excess, which results in a negative Tajima's D value ▪ Detect old layers of selective sweep regions	▪ It detects genomic regions differentially selected between two populations ▪ Uses divergence of allele frequencies among populations ▪ Detect old layers of selective sweep regions

1.3 Signature of selection in the cattle genome

The cattle genome has 30 pairs of chromosomes and a size of 2.65 Gbps comprising 22,000 genes (Elsik et al. 2009). The genetic variability and diversity found in several traits of cattle breeds is a reflection of differences in DNA sequences (polymorphisms) across the functional DNA regions and are a result of natural and artificial selection forces. Identification of signature of natural and artificial selection forces in the genomes of cattle enables us to understand the biological mechanisms that differentiate breeds artificially selected for different traits and those adaptation mechanisms to a local environment. So far, a catalog of genes have been identified from several cattle breeds of different demographic history, and these efforts will hopefully continue to identify additional more genes that might be related to local adaptation traits of different local breeds.

The Bovine HapMap Consortium (2009), scanned the genome of nineteen geographically and biologically diverse cattle breeds and identified genomic regions affected due to domestication and artificial selection. In the study, various genomic regions affecting milk yield and composition, meat quality, and feed conversion efficiency were revealed under selection in different breeds (The Bovine HapMap Consortium 2009). Other studies exposed the signature of genomic regions affecting coat color (*MCR1, KIT*), and body size and stature (*PLAG1*) in several beef and dairy cattle breeds (McClure et al. 2010; Lee et al. 2013b; Lee et al. 2014a; O'brien et al. 2014; Porto-Neto et al. 2014; Kim et al. 2017a). Similarly, Rothammer et al. (2013), scanned the genome of ten beef and dairy cattle breeds and revealed several genes (*TG, ABCG2, DGAT1, GH1, GHR* and the Casein Cluster) that are strongly associated with QTL regions of beef and dairy traits. A study by Lee et al. (2014a), exploring the genome of Holstein cattle, identified genomic regions affected due to selection that are related to milk yield and composition, and those associated with genetic disorders (cardiovascular disease). Artificial selection for milk and beef traits has significantly

affected the genome of dairy and beef cattle breeds, respectively, that genes and gene regions related to milk and beef traits have been found under positive selection in the respective breeds.

Despite the currently available methods and technologies, the genomes of African indigenous cattle breeds are relatively less intensively studied. However, recently, due to the emerging efforts here and there using the whole genome and SNP CHIP data of cattle breeds, the situation is expected to change rapidly (Okeyo et al. 2015). Bahbahani et al. (2015), analyzed the genome of small East African Shorthorn Zebu (EASZ) cattle for signature of positive selection and identified 24 candidate genomic regions under selection that included 409 annotated genes – these genes are involved in biological pathways of immunity, reproduction, development and heat tolerance. They identified the signature of several genes associated with heat shock protein family *(HSPB9, DNAJC7, DNAJC8, DNAJC14,* and *DNAJC18*), and heat stress response (*PPP1R10*) genes (Bahbahani et al. 2015). Similarly, Makina et al. (2015) has performed signature of selection analysis in six South African cattle breeds and identified genes associated with adaptation to tropical environments (*KRT222, KRT24, KRT25, KRT26, KRT27,* and *HSPB9*), immune response (*CYM, CDC6,* and *CDK10*), nervous system development (*WNT5B, FMOD, PRELP,* and *ATP2B*), production (*MTPN, IGFBP4, TGFB1,* and *AJAP1*), and reproduction performances (*ADIPOR2, OVOS2,* and *RBBP8*) to be under selection. In West African cattle, 42 strong candidate genes whose physiological function is mainly related to immune response (*MHC, CD79A, CXCR4, DLK1, RFX3, SEMA4A, TICAM1* and *TRIM21*), nervous system (*NEUROD6, OLFM2, MAGI1, SEMA4A,* and *HTR4*) and skin and hair properties (*EDNRB, TRSP1,* and *KRTAP8-1*) were identified (Gautier et al. 2009).

Natural selection forces that drive positive selection in the genome of African cattle are diverse and numerous. The high environmental temperature of the continent might be the possible driver for the development of superior thermotolerance mechanisms including enhanced thermoregulation, higher fertility, and growth rate of

African cattle than European cattle breeds under tropical environments (Hansen 2004; Paula-Lopes et al. 2013; Bahbahani et al. 2015). The mechanisms of thermotolerance are various that include molecular and cellular functions (Belhadj Slimen et al. 2015), physical structures of hair and skin (Jian et al. 2014), thermal sweating (Jian et al. 2014; Lenis Sanin et al. 2016) and physiological mechanisms (Hansen 2009).

Higher disease prevalence of the African environment could be a driving force for the development of disease and parasite resistance in African cattle breeds (Okeyo et al. 2015). The humid and subhumid West African region is known for its infestation with tsetse fly, a vector for trypanosomiasis (Berthier et al. 2015; Yaro et al. 2016). Cattle breeds that prevail in these regions have developed adaptive characteristics in response to the selective pressure due to trypanosome challenges. In an experiment that compared different West African taurine cattle breeds with indicine cattle breeds for a trypanosome challenge, shorthorn taurine breeds displayed superior anemia control indicating their local adaptation to trypanosomiasis (Berthier et al. 2015). In addition, genetic signature analysis in West African N'Dama cattle revealed resistance to trypanosomiasis as compared to other African zebu and commercial cattle breeds (Kim et al. 2017a; Kim et al. 2017b).

In this dissertation, the whole genome SNP data of African cattle breeds (Boran, Ogaden, Kenana, Ankole, and N'Dama), Holstein, Hanwoo, Jersey, and Angus cattle breeds were investigated for signature of selection in relation to major phenotypes of the breeds. African cattle breeds, evolved under local tropical environment, are known for their superior heat tolerance ability as compared to cattle breeds evolved in temperate regions. Comparing the genomes of African cattle with the genome of Asian-European Commercial cattle breeds helps to unravel the biological mechanisms behind these adaptation characteristics. Similarly, identifying the genomic foot-print of artificial selection in intensively selected beef (Hanwoo and Angus), and dairy (Holstein and Jersey) cattle breeds help us to decipher genomic regions affected in relation to milk, and beef traits, respectively.

This chapter was published in *Animal Science Journal*

in partial fulfillment of Mengistie Taye Terefe's Ph. D program

Chapter 2. Whole Genome Detection of Signature of Positive Selection in African Cattle Reveals Selection for Thermotolerance

2.1 Abstract

As African indigenous cattle are evolved in a hot tropical climate, they have developed an inherent thermotolerance; survival mechanisms include a light-colored and shiny coat, increased sweating, and cellular and molecular mechanisms to cope with high environmental temperature. Here, several genes that are under positive selection in African cattle breeds which contribute to their superior heat tolerance mechanisms are reported. To identify the positive selection of genes, the genomes of five indigenous African cattle breeds were compared with the genomes of four Commercial cattle breeds using cross-population composite likelihood ratio (XP-CLR) and cross-population extended haplotype homozygosity (XP-EHH) statistical methods. As a result, 296 (XP-EHH) and 327 (XP-CLR) genes were found under selection. Gene ontology analysis, using all the genes from both methods, resulted in 41 biological process terms and six KEGG pathways. Several of the genes and pathways were found involved in molecular functions associated with heat tolerance mechanisms including oxidative stress response, osmotic stress response, heat shock response, hair and skin properties, sweat gland development and sweating, feed intake and metabolism, and reproduction functions. The genes and pathways identified directly or indirectly contribute to the superior heat tolerance mechanisms of African cattle populations. The result will improve our understanding of the biological mechanisms of heat tolerance in African cattle breeds and opens an avenue for further study.

2.2 Introduction

African cattle have been reared in the continent since 6000 BC when the first humpless longhorn *Bos taurus* cattle were introduced; they were soon followed by the humpless shorthorn *B. taurus* 2500 years later. *Bos indicus* were introduced to Africa later, around 1500 BC (Rege 1999; Ajmone-Marsan et al. 2010). The Sanga cattle of Africa, *Bos Africanus*, is a cross between Hamitic longhorn *B. taurus* and *B. indicus* which was developed around 800 AD. These cattle sub-species have developed into many African cattle breeds through a multitude of human and natural selection in order to adapt to different agro-climatic and sociocultural conditions in the continent (Rege 1999; Mwai et al. 2015).

Thermal stress is an important problem compromising animal production and productivity in Africa and other tropical and subtropical regions; this is a particularly significant problem, as most of the world's meat and milk is produced in areas with such climates. Thermal stress affects animal production, reproduction, and health through its effect on feed intake, metabolism, and physiology. It induces heat shock, oxidative stress and osmotic stress which are deleterious to normal cellular functions (Sunil et al. 2011; Paula-Lopes et al. 2013; Belhadj Slimen et al. 2015).

African cattle, like other tropical cattle, are more tolerant to increased atmospheric temperature than breeds evolved in temperate regions (Paula-Lopes et al. 2013; Mwai et al. 2015). They have developed different thermotolerance mechanisms of lowered metabolic rates and an increased capacity to lose heat as a result of their experience on chronic heat stress for a long period (Hansen 2004). Thermotolerance is the ability of thermotolerant animals to regulate body temperature (Paula-Lopes et al. 2013) by efficiently balancing heat gain and heat loss (Sunil et al. 2011). This adaptive mechanism is derived from changes in gene expression and activity of biochemical molecules that control cellular functions against stress (Paula-Lopes et al.

2013) and involves the physiological integration of many organs and systems (Sunil et al. 2011). The anatomical, physiological, biochemical, and molecular components thermotolerance have been reviewed previously (Gupta et al. 2013; Paula-Lopes et al. 2013).

Understanding the genetic control of thermotolerance and other environmental adaptation traits in cattle could help to design sustainable breed selection programs (Hansen 2004). Heat tolerance is a heritable trait with moderate to low heritability (Dikmen et al. 2012) and found to be moderately positively correlated with direct genomic values in beef cattle (Howard et al. 2014). However, improving heat tolerance among productive animals is difficult because of its negative correlation with the heritability of other performance traits; it is favorable only with fertility (Dikmen et al. 2012; Nguyen et al. 2016). Therefore, investigating the underlying cellular and molecular processes may provide ways to select heat-tolerant animals with high production potential (Belhadj Slimen et al. 2015).

Several genes/gene regions have previously been identified to contribute to thermotolerance in different species. The well-known and widely-studied genes in relation to heat stress response are heat shock proteins (HSPs), whose overexpression protects the cell against hyperthermia (Collier et al. 2008; Gupta et al. 2013). Genes involved in antioxidant production have also been reported to contribute to thermotolerance (Belhadj Slimen et al. 2015). The importance of sweating in heat tolerance, and the involvement of genes in sweat gland development and sweating, has also been reported (Wilke et al. 2007; Akers and Denbow 2008). However, information on genes contributing to the heat tolerance mechanisms of African cattle is very scarce.

Several methods have been used to identify the positive selection of genes in relation to different adaptation traits. Bahbahani and his colleagues used F_{ST}, *i*HS and R*sb* analysis to detect signatures of positive selection in African Shorthorn Zebu cattle (Bahbahani et al. 2015). Makina et al. (2015), also used haplotype and allele frequency

based methods to scan for selection signatures in six cattle breeds of South Africa. Here, the genomes of five African cattle breeds were compared with four Asian-European Commercial cattle breeds using cross-population extended haplotype homozygosity (XP-EHH) which assess haplotype differences between two populations (Sabeti et al. 2007), and cross-population composite likelihood ratio (XP-CLR), a method based on allele frequency differentiation (Chen et al. 2010), to detect genomic regions under positive selection in African cattle breeds that are related to thermotolerance mechanisms.

2.3 Materials and Methods

2.3.1 Data description and whole genome re-sequencing

The data used in this chapter is obtained from a previously published paper where detail information about sampling and sequencing is available (Kim et al. 2017a). DNA was extracted from a whole blood sample of five African cattle breeds (10 Ankole, 9 Boran, 9 Ogaden, 10 N'Dama and 10 Kenana) using G-DEXTMIIb Genomic DNA Extraction Kit (iNtRoN Biotechnology, Seoul, Korea) based on the manufacturer's protocol. About ~300 bp inserts were generated by randomly shearing 3 µg of genomic DNA using Covaris System. Library was constructed using the TruSeq DNA Sample Prep. Kit (Illumina, San Diego, CA) and whole genome sequencing was performed using the Illumina HiSeq 2000 platform. In addition, previously published data from four Commercial cattle breeds (10 Holstein, 10 Angus, 10 Jersey and 11 Hanwoo) were used and a per-base sequence quality check was performed using the fastQC software (Andrews 2010). Using Bowtie2 (Langmead and Salzberg 2012), pair-end sequence reads were mapped to the reference bovine genome

(UMD 3.1) with an overall alignment rate of 98.50% and average read depth of 10.8x. On average across the whole samples, the reads covered 98.51% of the genome.

Picard tools (http://picard.sourceforge.net) was used to filter potential PCR duplicates. SAMtools was used to create index files for reference and bam files (Li et al. 2009). Genome analysis toolkit 1.4 (GATK) performed local realignment of reads (McKenna et al. 2010). The "UnifiedGenotyper" and "SelectVariants" arguments of GATK was used to call candidate SNPs. To filter variants and avoid possible false positives, the "VariantFiltration" argument of the same software was adopted using options: 1) SNPs with a phred-scaled quality score of less than 30 were filtered; 2) SNPs with MQ0 (mapping quality zero; total count across all samples of mapping quality zero reads) > 4 and quality depth (unfiltered depth of non-reference samples; low scores are indicative of false positives and artifacts) < 5 were filtered; and 3) SNPs with FS (Phred-scaled P-value using Fisher's exact test) > 200 were filtered since FS represents variation on either the forward or the reverse strand, which is indicative of false positive calls. BEAGLE (Browning and Browning 2007) was used to infer the haplotype phase and impute missing alleles for the entire set of cattle populations simultaneously. After all the filtering processes, a total of ~37 million SNPs were retained and used for further analysis. Sequences used in this study are available from GenBank with the Bioproject accession numbers of PRJNA312138 (African cattle), PRJNA210521 (Holstein), PRJNA318089 (Jersey), PRJNA318087 (Angus), and PRJNA210523 (Hanwoo).

2.3.2 Population structure

STRUCTURE software (Evanno et al. 2005) was used to identify groups of individuals corresponding to the uppermost hierarchical levels using genome data from five African (10 Ankole, 9 Boran, 9 Ogaden, 10 N'Dama and 10 Kenana) and four

Commercial (10 Holstein, 10 Angus, 10 Jersey and 11 Hanwoo) cattle breeds. STRUCTURE software uses Bayesian algorithms to detect the true number of clusters, *K* (the number of ancestral populations). PLINK (Purcell et al. 2007) was used to generate STRUCTURE input files using -thin option. The number of loci used for structure analysis were 7231 with options of Length of Burnin Period of 2000, Number of MCMC Reps after Burnin period of 100000 and MAF of 0.05.

2.3.3 Detection of signals of positive selection

Positive selection signature in the genome of animals can be detected using several methods. In this study, I used cross-population comparison statistical methods of XP-CLR (Chen et al. 2010) and XP-EHH (Sabeti et al. 2007) as described previously. XP-EHH statistic assesses haplotype differences between two populations since it is designed to detect alleles that have increased in frequency to the point of fixation or near-fixation in one of the populations (Sabeti et al. 2007; Pickrell et al. 2009). The genomes of five African cattle populations (Boran, Ogaden, Kenana, Ankole, and N'Dama) grouping together as a test population were compared with the genomes of four Commercial cattle breeds (Holstein, Hanwoo, Jersey and Angus - used as a reference population) to detect positive selection signatures in African cattle populations. XP-EHH was used to compare the integrated EHH from African cattle populations with Commercial cattle populations. It determines the direction of selection with extreme values indicating selection in the African cattle genome. Input files used for XP-EHH analysis (.hap and .map) were prepared from the resequencing (VCF) file using an in-house python code. In this analysis, the maximum XP-EHH score was computed for each 50 kb non-overlapping genomic segments to compare genomic regions across populations. Then empirical *P*-values were defined based on the genomic windows binned in increments of 500 SNPs (combining all windows ≥ 1000

SNPs into one) according to the method used previously (Pickrell et al. 2009). The regions with empirical *P*-values less than 0.01 (1%) were considered strong signals of selection in African cattle population. A low *P*-value indicates that a locus is an outlier with respect to the rest of the genome.

In addition, an XP-CLR test, which has a potential to identify genomic regions differentially selected between the two populations (Chen et al. 2010) was employed. It is a likelihood method for detecting selective sweeps that model the multilocus allele frequency differentiation between two populations. The script available online freely (http://genepath.med.harvard.edu/reich, accessed on June 2015) was used to calculate XP-CLR scores. In the script, non-overlapping sliding windows of 50 kb and a maximum number of 600 SNPs within each window were used to calculate the XP-CLR scores. As per the recommendation of the software, a weighted CLR scheme was used to estimate XP-CLR that the pairwise correlation coefficients (r^2) of SNPs from the reference population are used to give the weights. When r^2 is greater than 0.95, CLR scores for these two SNPs are down-weighted. The regions with XP-CLR values in the top 1% (0.01) of the empirical distribution were designated as candidate sweeps and the genes that span the window regions were defined as candidate genes (Lee et al. 2014a). Genes located in the significant genomic regions identified from XP-EHH and XP-CLR methods were annotated based on UMD 3.1.

2.3.4 Characterization of candidate genes under selection

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) gene ontology and annotation tool for gene enrichment analysis was used to further understand the biological functions and pathways of genes identified under selection (Huang et al. 2009). In DAVID gene enrichment analysis, GO terms represented by more number of genes than expected are considered overrepresented and the terms

provide insights into the functional characteristics of annotated genes. Within DAVID, the default settings were used to search for Gene Ontology Biological Process (GO-BP) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, which clusters genes of similar biological functions and related pathways. The Manhattan plot of the adjusted XP-CLR scores and XP-EHH values were drawn using R software (version 3.2.1).

The linkage equilibria (LD) and haplotype structure of selected candidate gene regions were visualized using haploview (Barrett et al. 2005). PLINK (Purcell et al. 2007) was used to prepare input files (.map and .ped) of candidate genes used by haploview. The minor allele frequency (MAF) cutoff was set to be below 5%. Blocks were defined based on confidence interval and haplotypes above 2% are displayed. SNP position is displayed along the top of the LD diagram. Colors represent D′ values (dark red=high inter-SNP D′; blue=statistically ambiguous D′; white=low inter-SNP D′), and r^2 values are contained within blocks (bold=high r^2).

2.4 Result and Discussion

2.4.1 Population structure and description

Population structure was created to detect the true number of clusters in the individual samples using STRUCTURE (Evanno et al. 2005) at different population assumptions (Figure 2.1). At K = 2 population assumptions, there was a clear differentiation between African groups and the Commercial cattle breeds with N'Dama and Ankole breeds in between the two groups with different level of admixture. When K is set to 3 to 5, the N'Dama and Ankole breeds became separate groups from others and each other, and at K = 6-9, all the breeds tend to be admixed (Figure 2.1).

African sample breeds are composed of taurine (N'Dama), Sanga (Ankole), and indicine (Boran, Ogaden, and Kenana) cattle breeds. N'Dama cattle are among the humpless longhorn *B. taurus* widely reared in the tsetse-infested humid areas of West Africa where other breeds suffer to survive from trypanosomiasis. N'Dama cattle, like other African taurine breeds, have a genetic contribution from the African aurochs that makes them different from other out of Africa taurine breeds (Decker et al. 2014). Ankole cattle are among the typical Sanga of Africa, a well-established breed resulted from interbreeding between *B. taurus* and *B. indicus*, reared in the eastern and central African countries (Mwai et al. 2015). The Kenana are nomadic cattle of northern Sudan descended from the first zebu introduction to Africa. Ogaden cattle are among the Small East African Zebu found in the Ogaden area of Somali region in eastern Ethiopia. Ogaden cattle share similar characteristics with Boran cattle, which are grouped in the Large East African Zebu and found distributed in Ethiopia and Kenya. All the African sample breeds are known for their heat tolerance and harsh environmental condition adaptations ((Rege 1999; Mwai et al. 2015), http://dagris.ilri.cgiar.org). Commercial cattle breeds are taurine cattle to which the Holsteins (Stella et al. 2010) and Jersey (Kim et al. 2015a) are intensively artificially selected for milk yield and milk related traits, and the Angus (http://www.dpi.nsw.gov.au /animals-and-livestock/beef-cattle/breeding/beef-cattle-breeds/angus) and Hanwoo (Porto-Neto et al. 2014) are selected for meat production and quality traits. From this, it is possible to understand that these two groups of cattle (African and Commercial) have different demographic and selection histories. Therefore, comparing African breeds with Commercial cattle breeds may help to identify genomic regions affecting tropical environmental adaptations in African cattle breeds. It is known that artificial selection for production traits reduce heat tolerance and other environmental adaptation traits (Belhadj Slimen et al. 2015).

2.4.2 Positive selection signature in African cattle populations

By annotating the top 1% outlier regions from XP-EHH and XP-CLR analysis, 296 (XP-EHH; Table 2.1) and 327 (XP-CLR; Table 2.2) genes were found under positive selection, of which 35 of the genes were common for both of the statistics. The Manhattan plot of the $-\log_{10}$ transformed XP-EHH values and XP-CLR scores *P*-values are presented in Figure 2.2. Among the genes identified, those with the top XP-CLR scores and XP-EHH values include tumor necrosis factor and receptor (*TNFRSF10D, TNFAIP2*), *CD1A*, transcription factor (*T, L3MBTL2,* and *EP300*), dehydratases (*ALAD, CA5A*), *HDHD3*, and *FTO* genes.

The DAVID gene ontology analysis, using the default settings, resulted in 41 significant GO-BP terms ($p<0.05$; Figure 2.3) and four KEGG pathways ($p<0.05$; Table 2.3). The BP terms overrepresented were related to cell adhesion, protein modification, hormone stimulus and biosynthetic processes. Hedgehog, Wnt, hematopoietic cell lineage and T cell receptor signaling pathways were also enriched in the KEGG pathways. Even though the genes identified as outlier regions might be associated with several different phenotypes and selection forces, this chapter selectively focuses and discusses the genes and pathways that putatively contribute to the thermotolerance mechanisms of African cattle. The genes and pathways associated with thermotolerance mechanisms are manually selected based on their biological function (genecards) and from the literature search. Concordant with this result, immune response genes, heat shock and heat stress response genes, and genes related to hair and skin structure have been reported to be under selection in African cattle in response to parasite infestation and high environmental temperature (Bahbahani et al. 2015; Makina et al. 2015).

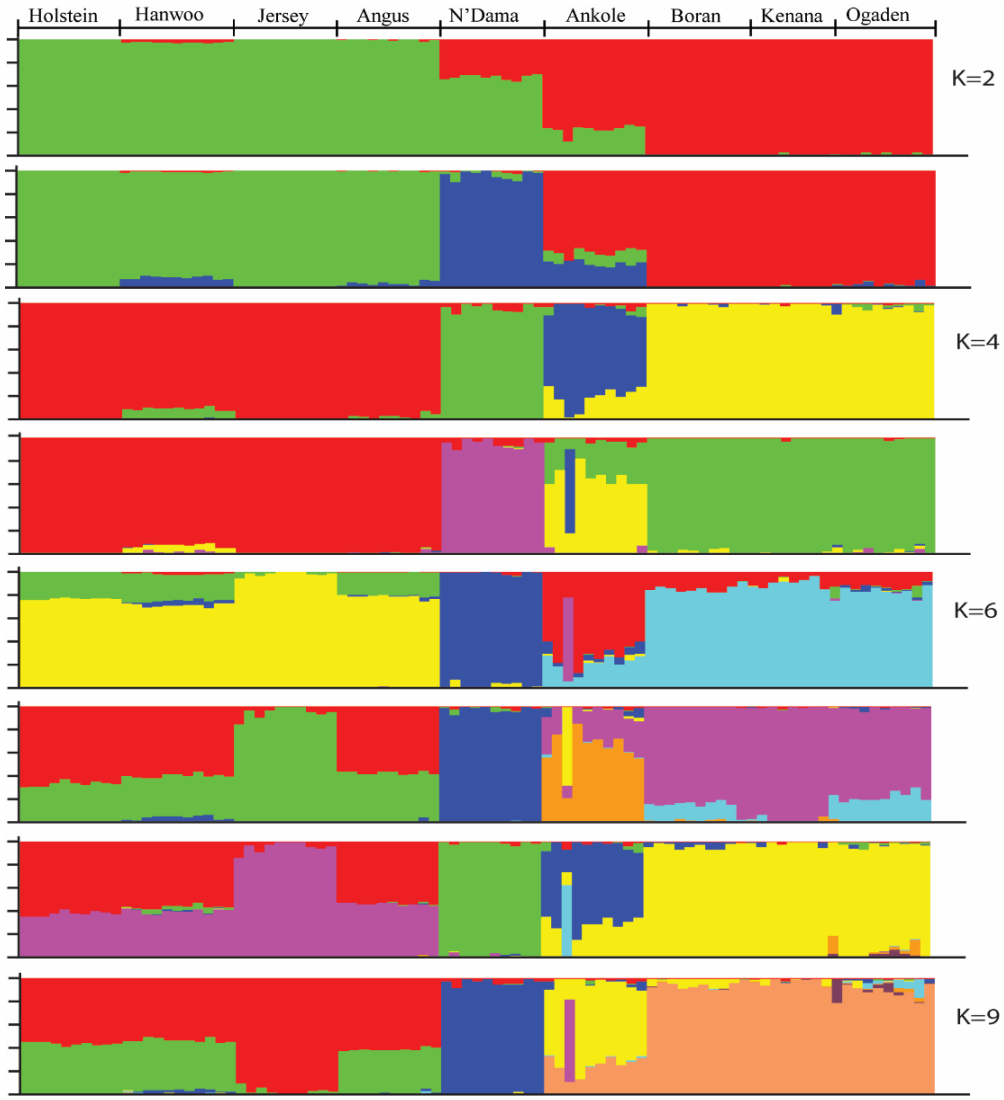


Figure 2.1 Population structure at *K* = 2-9 population assumptions in nine African and Commercial cattle breeds. At the top of the figure is the name of breeds considered and at the right side is the number of population assumptions.

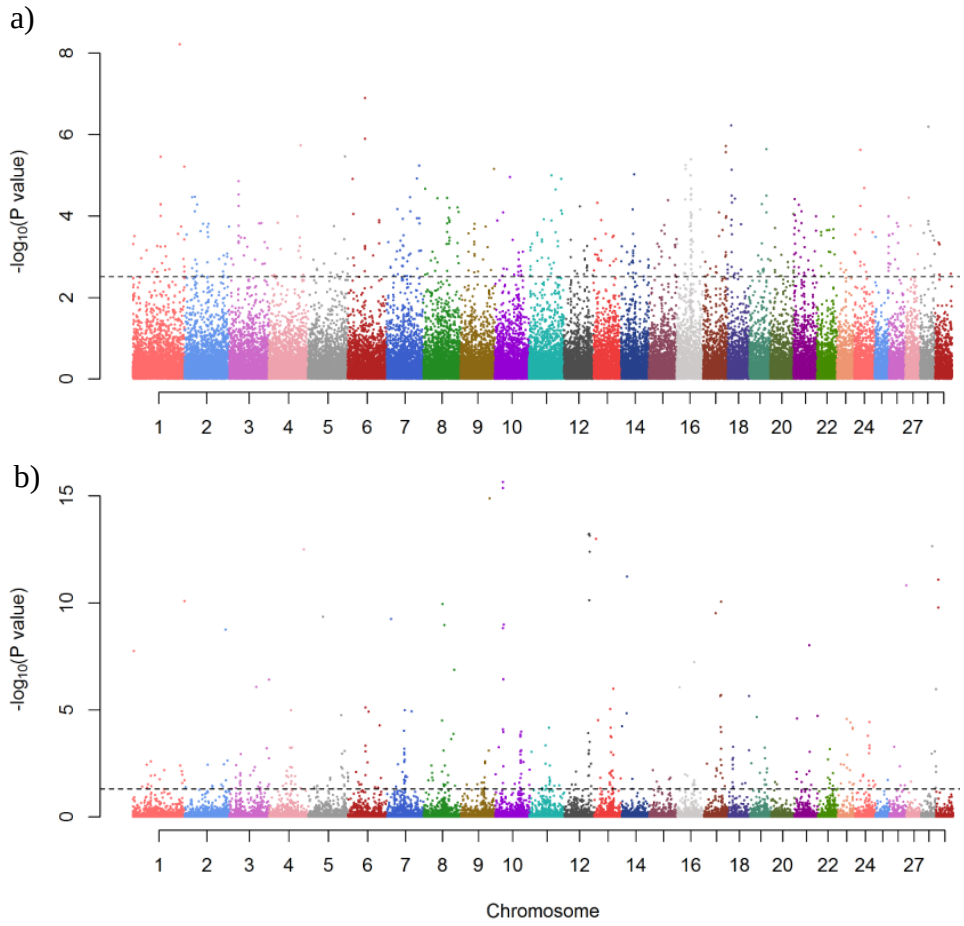


Figure 2.2 Manhattan plot of the $-\log_{10}$ transformed XP-EHH values (a) and the $-\log_{10}$ transformed XP-CLR scores P-values (b). The *y*-axis shows the $-\log_{10}$(*P*-value) of XP-EHH values and XP-CLR scores, and *x*-axis shows chromosomal positions. The horizontal dotted lines represent the 1% (0.01) XP-EHH values and XP-CLR scores outlier regions, which resulted in 296 (XP-EHH), and 327 (XP-CLR) genes to be annotated.

Table 2.1 Summary of genes identified under selection in African cattle as detected by XP-EHH statistics. Common genes between XP-EHH and XP-CLR are indicated with **bold** and *italic* text.

Chr.	Start (kb)	XP-EHH	P-value	Genes
1	4.85-4.9	1.470	0.0010	KRTAP27-1
1	43.65-43.7	0.990	0.0100	COL8A1
1	50.5-50.55	1.030	0.0080	ALCAM
1	82.25-82.3	0.990	0.0090	***5S_rRNA***,LIPH,TMEM41A
1	90.6-90.65	1.310	0.0020	TBL1XR1
1	122.9-122.95	0.990	0.0100	PLSCR2
1	143.05-143.1	1.120	0.0050	FAM3B
1	143.1-143.15	1.040	0.0070	FAM3B,MX2
1	146.7-146.75	0.990	0.0090	AGPAT3
1	152.3-152.35	0.990	0.0090	KCNJ15
1	153.2-153.25	0.990	0.0100	PIK3R4
1	155.35-155.4	1.130	0.0040	DAZL
1	155.4-155.45	1.060	0.0070	DAZL
2	54.6-54.65	1.230	0.0040	***5S_rRNA,5S_rRNA***
3	6.55-6.6	1.200	0.0030	C1orf110
3	6.65-6.7	1.090	0.0060	DDR2,HSD17B7
3	7.25-7.3	1.340	0.0020	NOS1AP
3	7.3-7.35	1.110	0.0070	NOS1AP
3	36.4-36.45	1.460	0.0010	PRMT6
3	48.6-48.65	1.070	0.0070	ALG14
3	60-60.05	1.210	0.0040	SAMD13
3	60.05-60.1	1.200	0.0030	PRKACB
3	113.85-113.9	1.190	0.0030	USP40
3	117.6-117.65	1.170	0.0040	***MLPH***,PRLH
3	117.65-117.7	1.400	0.0010	RAB17
3	117.8-117.85	1.050	0.0070	LRRFIP1
3	120.9-120.95	1.010	0.0080	HDLBP
4	12.6-12.65	1.060	0.0070	ASB4
4	30.6-30.65	1.020	0.0080	DNAH11
4	79.65-79.7	1.140	0.0040	***GLI3***

4	89.15-89.2	1.030	0.0080	***5S_rRNA***
4	103.35-103.4	1.210	0.0030	KIAA1549
4	105.1-105.15	1.240	0.0030	MRPS33
5	28.4-28.45	1.270	0.0020	***U6***
5	48.6-48.65	1.310	0.0030	MSRB3
5	63.25-63.3	0.990	0.0090	ANKS1B
5	65.05-65.1	1.200	0.0030	ANO4
5	79.2-79.25	1.520	0.0010	CAPRIN2
5	83.9-83.95	1.010	0.0080	ITPR2
5	83.95-84	1.130	0.0040	ITPR2
5	97.1-97.15	1.210	0.0030	EMP1
5	101.6-101.65	1.120	0.0070	MFAP5,RIMKLB
5	103.85-103.9	1.110	0.0070	***ATN1,C5H12orf57***,EMG1,***ENO2***,PHB 2,PTPN6,***U7***,bta-mir-141,bta-mir-200c,snoU89
5	104.05-104.1	1.020	0.0080	COPS7A,MLF2
5	112.85-112.9	1.860	0.0000	EP300,L3MBTL2,bta-mir-2441
5	113-113.05	1.080	0.0080	TEF,ZC3H7B
5	113.15-113.2	1.000	0.0090	CSDC2,PMM1,POLR3H
5	113.25-113.3	1.000	0.0090	C22orf46,MEI1,NHP2L1,XRCC6
5	120.85-120.9	1.240	0.0030	***BRD1***
5	120.9-120.95	1.200	0.0030	***ALG12***,CRELD2
5	121.05-121.1	1.110	0.0050	MLC1,MOV10L1,***TTLL8***
5	121.1-121.15	1.340	0.0030	MOV10L1
6	14.55-14.6	1.200	0.0030	C6H4orf32
6	27.3-27.35	1.090	0.0060	TSPAN5,***U6***
6	104-104.05	1.210	0.0030	HSD17B11
6	105.25-105.3	1.020	0.0080	EVC,EVC2
6	108.65-108.7	1.090	0.0060	POLN
6	111.4-111.45	1.020	0.0080	HS3ST1
7	28.15-28.2	1.330	0.0020	C5orf63
7	31.55-31.6	1.380	0.0010	CSNK1G3
7	32.7-32.75	0.990	0.0090	SNCAIP
7	33.05-33.1	1.610	0.0000	LOX
7	33.1-33.15	1.000	0.0090	LOX,SRFBP1
7	41.6-41.65	1.390	0.0010	BTNL9

7	48.85-48.9	1.560	0.0000	SLC25A48
7	48.9-48.95	1.300	0.0020	IL9,SLC25A48
7	48.95-49	1.340	0.0030	FBXL21,LECT2
7	51.85-51.9	1.470	0.0020	CTNNA1,***SIL1***
7	55.9-55.95	1.110	0.0070	***ARHGAP26***
7	62.6-62.65	1.210	0.0050	ABLIM3
7	62.65-62.7	1.230	0.0040	AFAP1L1
7	62.7-62.75	1.150	0.0060	AFAP1L1,GRPEL2
7	62.8-62.85	1.320	0.0030	bta-mir-143,bta-mir-145
7	62.85-62.9	1.070	0.0090	CSNK1A1
7	98.9-98.95	1.220	0.0030	LIX1
7	98.95-99	1.180	0.0050	LIX1
7	99-99.05	1.380	0.0020	LIX1,RIOK2
7	111.6-111.65	1.040	0.0070	MAN2A1
8	10.15-10.2	1.010	0.0080	ZNF395
8	18.75-18.8	1.130	0.0050	***7SK***
8	23.05-23.1	1.330	0.0020	IFNAH
8	23.15-23.2	1.090	0.0060	***IFNB3***
8	24.9-24.95	1.100	0.0080	***DENND4C,RPS6***
8	25.2-25.25	1.120	0.0050	FAM154A
8	26.9-26.95	1.110	0.0050	***CNTLN***
8	28.65-28.7	1.030	0.0080	CCDC171
8	44.6-44.65	1.140	0.0060	FOXD4L1
8	45.75-45.8	1.180	0.0040	FAM189A2
8	48.45-48.5	1.080	0.0080	C8H9ORF85,FAM108B1
8	60-60.05	1.180	0.0060	***UNC13B***
8	76.3-76.35	1.570	0.0000	SPINK4
8	104.1-104.15	1.350	0.0010	FKBP15,SLC31A2
8	104.15-104.2	1.430	0.0020	FKBP15,SLC31A1
8	104.3-104.35	1.620	0.0000	ALAD,BSPRY,HDHD3
9	14.95-15	1.060	0.0070	COL12A1
9	33.85-33.9	1.010	0.0080	ROS1
9	69.2-69.25	1.330	0.0020	***L3MBTL3***
9	87.85-87.9	1.120	0.0070	PPIL4,ZC3H12D
9	88.55-88.6	1.100	0.0050	IYD
9	98.05-98.1	1.150	0.0040	PLG

9	98.45-98.5	1.150	0.0050	PARK2
9	102.65-102.7	1.800	0.0000	T
10	4.4-4.45	0.990	0.0100	TMED7
10	7.65-7.7	1.550	0.0000	IQGAP2
10	15.05-15.1	1.200	0.0030	FEM1B
10	23.05-23.1	1.120	0.0050	***TRAC,TRAV29DV5***
10	23.1-23.15	1.230	0.0020	***TRAC***
10	25-25.05	1.040	0.0070	TRAV16,TRAV17
10	26.75-26.8	1.280	0.0020	PARP2,TEP1
10	42.75-42.8	1.120	0.0070	KLHDC1,KLHDC2
10	44.9-44.95	1.050	0.0070	NID2
10	59.1-59.15	1.070	0.0090	GLDN
10	59.5-59.55	0.990	0.0090	TNFAIP8L3
10	68.25-68.3	1.080	0.0060	KTN1
11	4.4-4.45	1.290	0.0020	TXNDC9
11	4.45-4.5	1.230	0.0030	EIF5B
11	4.5-4.55	1.270	0.0020	EIF5B,REV1
11	4.55-4.6	1.090	0.0060	REV1
11	4.6-4.65	1.390	0.0010	AFF3
11	4.65-4.7	1.410	0.0010	AFF3
11	5.45-5.5	1.130	0.0040	CHST10,LONRF2
11	5.5-5.55	1.100	0.0080	NMS
11	7.2-7.25	1.000	0.0090	IL18RAP,SLC9A4
11	9.65-9.7	1.430	0.0010	POLE4,***U6***
11	21.05-21.1	1.180	0.0040	GALM
11	46.1-46.15	0.990	0.0100	POLR1B,TTL
11	48.35-48.4	1.100	0.0050	REEP1
11	105.2-105.25	1.130	0.0060	CACNA1B
12	29.65-29.7	1.000	0.0090	B3GALTL,bta-mir-2299
12	32.2-32.25	1.000	0.0090	FLT3
13	0.85-0.9	1.130	0.0040	PLCB1
13	47.55-47.6	1.040	0.0070	***SLC23A2,U6***
13	53.3-53.35	1.020	0.0080	TGM3
13	57.9-57.95	1.460	0.0020	NELFCD
13	62.65-62.7	1.170	0.0060	COMMD7
13	64.5-64.55	1.090	0.0060	PIGU

13	64.55-64.6	1.070	0.0060	PIGU
13	64.6-64.65	1.010	0.0080	NCOA6,TP53INP2
13	74.45-74.5	1.030	0.0080	DBNDD2,PIGT
14	0.55-0.6	1.320	0.0020	***5S_rRNA,5S_rRNA***
14	28.15-28.2	1.070	0.0090	CHD7
14	48.45-48.5	1.160	0.0040	EXT1
14	57.25-57.3	0.990	0.0090	NUDCD1
14	57.45-57.5	1.080	0.0060	TRHR
14	83.15-83.2	1.330	0.0020	SNX16
14	83.2-83.25	1.280	0.0020	CHMP4C,U4,ZFAND1
14	84.35-84.4	1.300	0.0020	SNTB1
14	84.4-84.45	1.010	0.0090	SNTB1
14	84.6-84.65	1.100	0.0080	***U6***
15	39.75-39.8	1.060	0.0090	BTBD10
15	46.65-46.7	1.140	0.0050	***U6***
15	81-81.05	1.110	0.0070	OR5M3
15	81.85-81.9	1.410	0.0010	P2RX3,PRG3
15	81.9-81.95	1.420	0.0010	PRG3
16	2.1-2.15	1.430	0.0010	MDM4
16	20.2-20.25	1.080	0.0060	USH2A
16	33.5-33.55	1.210	0.0030	ADSS
16	70.4-70.45	1.180	0.0040	CENPF
17	5.3-5.35	1.230	0.0030	FBXW7
17	14.3-14.35	1.100	0.0050	GYPB
17	22.05-22.1	1.050	0.0100	SNORA25
17	39.05-39.1	1.100	0.0050	***5S_rRNA***
17	40.8-40.85	1.010	0.0080	TNIP3
17	41.2-41.25	1.250	0.0030	PPID
17	65.9-65.95	1.070	0.0060	UBE3B
17	65.95-66	1.120	0.0050	KCTD10,MYO1H
17	66-66.05	0.990	0.0090	MYO1H
17	72.15-72.2	0.990	0.0100	LIMK2,PIK3IP1
17	72.4-72.45	1.060	0.0070	PISD,SFI1
18	13.35-13.4	1.480	0.0010	CA5A,LAT
18	13.4-13.45	1.640	0.0000	BANP,CA5A
18	14-14.05	1.080	0.0060	***PIEZO1,bta-mir-2327***

18	18.55-18.6	1.240	0.0030	CNEP1R1,HEATR3
18	22.3-22.35	1.660	0.0000	FTO
18	22.35-22.4	1.250	0.0040	FTO
18	51.2-51.25	1.120	0.0070	CNFN,HSL,MEGF8
18	63.2-63.25	1.150	0.0040	***LILRA4***
19	9.6-9.65	1.200	0.0030	HSF5,RNF43
19	12.5-12.55	1.100	0.0080	BCAS3
19	22.15-22.2	1.010	0.0080	BHLHA9,TUSC5
19	27-27.05	1.360	0.0020	INCA1,KIF1C
19	27.05-27.1	1.260	0.0040	CAMTA2,ENO3,GP1BA,INCA1,PFN1, RNF167,SLC25A11,SPAG7
19	36.05-36.1	1.080	0.0060	UTP18
19	41.05-41.1	1.270	0.0040	CASC3,MSL1
19	41.15-41.2	1.050	0.0100	CDC6,WIPF2
19	42.1-42.15	1.050	0.0070	KRTAP9-1
19	42.15-42.2	1.260	0.0020	KRT33A
19	43.9-43.95	1.270	0.0020	ARL4D,U2,U2,U2,U2
19	54.5-54.55	1.000	0.0090	TMEM235
19	57.35-57.4	1.340	0.0030	RAB37
19	63.45-63.5	1.110	0.0050	bta-mir-2284p
20	18.15-18.2	1.140	0.0040	U1
20	39.8-39.85	1.180	0.0050	AMACR,SLC45A2
20	39.85-39.9	1.210	0.0030	RXFP3,SLC45A2
20	39.9-39.95	1.120	0.0070	ADAMTS12
21	26.05-26.1	1.000	0.0090	***5S_rRNA***
21	58.85-58.9	1.090	0.0060	UNC79
21	59.1-59.15	1.270	0.0020	ASB2,FAM181A
21	59.15-59.2	1.100	0.0080	ASB2
21	59.3-59.35	1.000	0.0090	IFI27,IFI27L2,ISG12(B)
21	59.45-59.5	1.180	0.0050	PPP4R4,SERPINA10
21	59.5-59.55	1.040	0.0070	SERPINA10,SERPINA6
21	70.85-70.9	1.010	0.0080	***ADSSL1***,AKT1,***SIVA1***
22	30.2-30.25	1.500	0.0010	***bta-mir-1284***
22	36-36.05	1.380	0.0010	***MAGI1***
22	36.05-36.1	1.250	0.0030	***MAGI1***
22	37.9-37.95	1.500	0.0010	SYNPR

22	41.75-41.8	0.990	0.0090	FHIT
22	59.8-59.85	1.090	0.0060	COPG1,H1FX,HMCES,bta-mir-2374
23	22.2-22.25	1.250	0.0030	CRISP1
23	27.35-27.4	1.120	0.0050	CLIC1,DDAH2,MGC151586,MSH5,VARS,VWA7
23	27.6-27.65	1.190	0.0030	***BOLA***
23	27.75-27.8	1.270	0.0020	CCHCR1,POU5F1,PSORS1C2,TCF19
23	27.85-27.9	1.220	0.0020	***BOLA***
23	43.4-43.45	1.210	0.0030	PHACTR1
23	44.45-44.5	1.100	0.0050	***C23H6ORF105***
24	3.8-3.85	1.030	0.0080	ZNF407
24	3.95-4	1.160	0.0040	ZNF407
24	4.05-4.1	1.040	0.0070	ZNF407
24	4.35-4.4	1.480	0.0010	C18orf63
24	4.4-4.45	1.370	0.0010	CYB5A
25	41.1-41.15	1.110	0.0050	GNA12
25	41.15-41.2	1.080	0.0060	AMZ1,GNA12
26	25.05-25.1	1.540	0.0000	WDR96
26	25.15-25.2	1.050	0.0070	CCDC147,ITPRIP
26	33.85-33.9	1.200	0.0050	TCFL2
26	35-35.05	1.000	0.0090	VWA2
26	42.75-42.8	1.100	0.0050	DMBT1
26	45.85-45.9	1.100	0.0050	***ADAM12***
27	7.05-7.1	1.140	0.0040	VEGFC
27	28.45-28.5	1.110	0.0050	RNF122
28	2.95-3	1.070	0.0070	SNORD65
28	44.3-44.35	1.370	0.0010	***PARG***

Table 2.2 Summary of genes identified under selection in African cattle as detected by XP-CLR statistics.

Chr	Window (Mpb)	XP-CLR	Genes
1	3.08-3.13	186.33	SCAF4,SOD1
1	42.43-42.48	99.82	CLDND1,GPR15
1	43.88-43.93	72.06	CMSS1
1	46.08-46.13	76.62	IMPG2
1	56.33-56.38	74.25	PVRL3
1	67.18-67.23	88.42	CD86
1	81.88-81.93	83.00	TRA2B
1	95.88-95.93	93.89	FNDC3B
1	145.03-145.08	99.26	PTTG1IP,SUMO3
1	150.03-150.08	88.83	SETD4
2	85.83-85.88	72.43	PGAP1
2	111.78-111.83	76.26	ACSL3,***RPS6***
2	116.28-116.33	73.59	COL4A3,MFF
2	120.58-120.63	100.45	DIS3L2
2	124.78-124.83	198.37	PTPRU
2	125.88-125.93	81.57	EYA3
2	131.28-131.33	104.43	WNT4
3	11.98-12.03	336.18	CD1A
3	19.58-19.63	71.56	PIP5K1A,PSMD4,***U6***
3	19.63-19.68	85.51	PIP5K1A,VPS72
3	21.18-21.23	99.34	U1,U1
3	33.08-33.13	80.81	PROK1
3	33.13-33.18	90.79	LAMTOR5,SLC16A4
3	33.43-33.48	81.36	ALX3,UBL4B
3	33.58-33.63	75.07	CSF1
3	33.83-33.88	88.05	GSTM2,GSTM4
3	34.43-34.48	110.64	TAF13,WDR47
3	74.48-74.53	93.63	PTGER3
3	77.88-77.93	97.51	GNG12
3	94.33-94.38	78.81	GPX7
3	94.68-94.73	70.77	ZFYVE9
3	95.58-95.63	71.76	RNF11,TTC39A
3	113.38-113.43	116.32	INPP5D

3	117.58-117.63	73.65	***MLPH***
3	120.33-120.38	76.85	CAPN10
3	120.78-120.83	106.84	MTERFD2,PASK,SNED1
4	49.03-49.08	69.97	CBLL1,SLC26A3
4	51.23-51.28	77.89	CFTR
4	64.33-64.38	97.42	AVL9
4	66.63-66.68	147.62	C4H7orf41
4	68.63-68.68	80.89	JAZF1
4	68.68-68.73	69.85	JAZF1
4	68.73-68.78	97.16	JAZF1
4	68.93-68.98	116.89	HIBADH,SNORD56
4	79.43-79.48	86.93	***GLI3***
4	79.68-79.73	76.87	***GLI3***
4	95.93-95.98	71.29	MKLN1
4	106.43-106.48	536.21	TRB@
4	120.73-120.78	712.69	***5S_rRNA***
5	24.13-24.18	70.18	PLXNC1
5	25.98-26.03	74.16	SMUG1
5	26.23-26.28	74.72	HOXC12,HOXC13
5	66.53-66.58	85.34	IGF-I
5	99.73-99.78	97.28	NKG2A
5	103.38-103.43	488.19	CD163L1
5	103.88-103.93	87.88	***ATN1,C5H12orf57,ENO2***,LRRC23,***U7***
5	107.98-108.03	85.79	***5S_rRNA***,NINJ2
5	109.58-109.63	74.22	ATP6V1E1,BCL2L13
5	110.83-110.88	113.60	GTPBP1,JOSD1,TOMM22,UNC84B
5	119.53-119.58	91.10	FAM19A5
5	120.83-120.88	80.87	***BRD1***
5	120.88-120.93	84.58	***ALG12***
5	120.98-121.03	86.55	IL17REL,***TTLL8***
6	24.78-24.83	91.99	PPP3CA
6	35.23-35.28	76.43	CCSER1
6	62.63-62.68	146.39	BEND4
6	91.68-91.73	72.15	DKFZP564O0823
6	96.53-96.58	135.99	ANTXR2
6	99.43-99.48	84.60	SEC31A

6	116.23-116.28	70.12	TAPT1
7	9.73-9.78	349.53	***U6***
7	33.03-33.08	100.18	ZNF474
7	46.43-46.48	91.76	FSTL4
7	51.43-51.48	82.76	EGR1,ETF1,REEP2
7	51.98-52.03	131.66	***SIL1***
7	52.13-52.18	82.92	***SIL1***
7	52.18-52.23	103.45	SNORA74,SNORA74
7	52.23-52.28	68.92	PAIP2,SLC23A1
7	52.43-52.48	115.72	UBE2D2
7	52.68-52.73	112.13	NRG2,PSD2
7	53.03-53.08	110.39	CYSTM1
7	53.08-53.13	102.63	CYSTM1,PFDN1
7	53.93-53.98	105.16	PCDHB6,PCDHB7
7	53.98-54.03	147.43	PCDHB10,PCDHB11,PCDHB14,PCDHB16,PCDHB7
7	54.03-54.08	80.43	PCDHB15
7	54.58-54.63	85.96	PCDH1
7	54.68-54.73	70.11	KIAA0141,PCDH12,RNF14
7	55.93-55.98	87.38	***ARHGAP26***
7	60.68-60.73	85.96	DPYSL3
8	23.18-23.23	99.17	IFNB1,***IFNB3***
8	24.93-24.98	85.50	***DENND4C,RPS6***
8	24.98-25.03	91.15	***DENND4C***
8	25.68-25.73	90.83	ADAMTSL1
8	26.93-26.98	81.80	***CNTLN***
8	52.68-52.73	68.68	PCSK5
8	54.93-54.98	139.63	RAN
8	55.93-55.98	92.34	TLE4
8	59.93-59.98	114.11	***UNC13B***
8	60.23-60.28	76.90	ARHGEF39,CA9,CCDC107,RNase_MRP,SIT1,TPM2
8	61.63-61.68	74.23	ZCCHC7
8	62.58-62.63	200.83	SHB,bta-mir-2474
8	70.88-70.93	336.34	RHOBTB2
8	70.93-70.98	572.60	TNFRSF10D

8	73.18-73.23	78.96	***5S_rRNA,***NEFL
8	73.68-73.73	88.25	DOCK5,GNRH1,KCTD9
8	83.23-83.28	124.30	FANCC,TSPY-M2
8	92.38-92.43	174.86	LPPR1
9	49.98-50.03	76.57	ASCC3
9	63.28-63.33	78.87	***5S_rRNA***,RARS2,SLC35A1
9	63.33-63.38	70.49	SLC35A1
9	69.18-69.23	69.07	***L3MBTL3***
9	73.08-73.13	71.99	TBPL1
9	73.13-73.18	103.41	SLC2A12,TBPL1
9	73.28-73.33	102.40	SGK1
9	88.23-88.28	260.75	RAET1G
9	88.28-88.33	308.10	RAET1G
10	10.08-10.13	117.23	BHMT2
10	22.38-22.43	132.79	***TRAC***
10	22.48-22.53	265.31	***TRAC***
10	22.63-22.68	199.17	***TRAC***
10	22.93-22.98	268.48	***TRAC***
10	23.03-23.08	78.29	***TRAC***
10	23.08-23.13	336.12	***TRAC,TRAV29DV5***
10	72.58-72.63	79.02	DHRS7,PCNXL4
10	75.33-75.38	71.85	KCNH5
10	76.63-76.68	128.41	SYNE2
10	76.68-76.73	126.74	ESR2,SYNE2
10	76.73-76.78	110.41	ESR2,***U6***
10	76.83-76.88	84.66	MTHFD1
10	77.28-77.33	101.84	***5S_rRNA***
10	77.33-77.38	103.09	CHURC1
10	77.43-77.48	104.81	FNTB
10	78.13-78.18	114.51	FUT8
10	78.98-79.03	130.88	GPHN
10	85.38-85.43	102.76	MGC160092
10	103.48-103.53	75.64	RPS6KA5
10	103.98-104.03	94.73	EIF3J,SPG11,***U6***
11	5.98-6.03	112.84	NPAS2,RPL31,TBC1D8
11	49.48-49.53	101.12	RETSAT,TGOLN2

11	49.83-49.88	74.69	***U6***
11	59.93-59.98	134.02	USP34
11	60.43-60.48	76.31	COMMD1
11	60.68-60.73	85.49	B3GNT2
11	61.18-61.23	82.48	EHBP1
11	61.48-61.53	82.00	OTX1
11	63.53-63.58	69.87	ACTR2
11	75.53-75.58	74.48	KLHL29,***U6***
11	90.83-90.88	75.05	5***S_rRNA***
12	33.83-33.88	76.21	ATP8A2
12	80.78-80.83	101.34	PCCA
12	80.88-80.93	74.15	PCCA
13	11.33-11.38	776.71	ANKRD26
13	17.68-17.73	70.86	FBXO18
13	17.98-18.03	84.92	MASTL,YME1L1
13	47.58-47.63	76.49	***SLC23A2***
13	48.33-48.38	75.47	C20orf196
13	49.53-49.58	127.21	BMP2
13	50.33-50.38	92.49	UBE2D3
13	50.38-50.43	112.50	UBE2D3,UBE2D3
13	52.13-52.18	126.06	ATRN
13	52.33-52.38	87.42	C13H20orf194
13	55.23-55.28	93.84	bta-mir-1-1,bta-mir-133a-1
13	55.73-55.78	71.77	CDH4
13	57.63-57.68	162.48	ZNF831
13	80.08-80.13	83.94	NFATC2
14	0.48-0.53	709.94	***5S_rRNA***
14	0.68-0.73	135.21	***U6***
15	22.28-22.33	80.52	SIK2
15	29.43-29.48	73.00	CD3D,CD3E,CD3G
16	24.33-24.38	89.18	IARS2,RAB3GAP2
16	31.28-31.33	87.50	SNORD112
16	47.58-47.63	76.74	DNAJC11,THAP3
16	47.63-47.68	75.24	KLHL21,PHF13,TAS1R1,THAP3,ZBTB48
16	50.63-50.68	78.78	TP73,WRAP73
16	51.93-51.98	179.53	GABRD,PRKCZ

16	52.18-52.23	82.00	GNB1,NADK,SLC35E2B
16	52.58-52.63	75.97	ACAP3,CPSF3L,GLTPD1,PUSL1,SCNN1D
17	25.08-25.13	902.40	PRAME
17	49.08-49.13	157.56	TMEM132D
17	51.33-51.38	673.85	HSFY2,PRAME
17	51.38-51.43	436.96	PRAME
17	56.48-56.53	84.63	ATP2A2
17	63.53-63.58	74.05	SLC8B1,TPCN1
18	13.03-13.08	95.40	MAP1LC3B,ZCCHC14
18	13.88-13.93	70.93	IL17C,ZC3H18
18	13.93-13.98	73.70	CTU2,CYBA,IL17C,MVD,RNF166,SNAI3
18	13.98-14.03	73.76	CTU2,***PIEZO1,bta-mir-2327***
18	14.03-14.08	107.18	APRT,CDT1,GALNS
18	14.08-14.13	117.71	CBFA2T3,PABPN1L,TRAPPC2L
18	51.53-51.58	78.16	GRIK5
18	57.68-57.73	720.86	SIGLECL1
18	63.18-63.23	157.70	***LILRA4***
19	19.83-19.88	142.29	INOS,ULBP27
19	29.73-29.78	90.87	GAS7
19	44.58-44.63	116.88	ASB16,C17orf53,HDAC5
19	44.63-44.68	76.03	ATXN7L3,TMUB2,UBF
19	45.18-45.23	82.66	EFTUD2,GJC1,HIGD1B,bta-mir-2343
19	45.28-45.33	106.61	C1QL1,KIF18B
19	50.48-50.53	91.60	TBCD,ZNF750
19	51.38-51.43	95.62	DUS1L,FASN,GPS1
20	18.48-18.53	76.70	ELOVL7
21	5.73-5.78	91.98	LRRK1
21	7.08-7.13	83.75	MEF2A
21	7.13-7.18	141.35	MEF2A
21	22.03-22.08	79.63	CIB1,GDPGP1,NGRN,SEMA4B,VPS33B
21	35.03-35.08	70.07	LOXL1,PML,STOML1
21	45.33-45.38	189.58	EAPP
21	45.43-45.48	92.95	SNX6,U3
21	45.83-45.88	112.87	KIAA0391
21	46.03-46.08	68.80	BIKBA

21	69.43-69.48	536.99	EXOC3L4,TNFAIP2
21	70.83-70.88	143.35	***ADSSL1***,INF2,***SIVA1***
22	30.18-30.23	85.23	***bta-mir-1284***
22	36.03-36.08	115.44	***MAGI1***
22	36.08-36.13	82.57	***MAGI1***
22	60.18-60.23	70.29	RUVBL1,SEC61A1
23	8.78-8.83	77.92	UHRF1BP1
23	9.18-9.23	100.24	ZNF76
23	17.53-17.58	100.04	***U6***
23	27.63-27.68	96.77	***BOLA***
23	27.68-27.73	279.11	BT.105339
23	28.48-28.53	110.13	***BOLA***
23	44.53-44.58	76.25	***C23H6ORF105***
23	44.63-44.68	133.08	TMEM170B
23	44.98-45.03	133.99	NEDD9,SMIM13
23	47.93-47.98	77.15	RREB1
24	25.08-25.13	88.60	GAREM
24	40.78-40.83	81.37	PTPRM
24	43.53-43.58	116.37	PTPN2
24	43.63-43.68	138.56	SEH1L
24	43.98-44.03	118.98	MC2R,MC5R
24	55.23-55.28	81.23	TCF7L2
25	37.18-37.23	82.38	CYP3A5
26	13.58-13.63	117.68	BTAF1,CPEB3
26	31.08-31.13	98.18	***7SK***,MXI1
26	42.28-42.33	68.70	TACC2
26	45.93-45.98	75.42	***ADAM12***
26	50.58-50.63	221.47	***5S_rRNA***
27	13.88-13.93	80.04	IRF2
28	32.78-32.83	240.04	KCNMA1
28	41.53-41.58	113.37	***U6,U6***,WAPAL
28	44.33-44.38	71.92	***PARG***
28	44.53-44.58	92.03	ANUBL1
28	44.63-44.68	162.23	8-Mar
29	5.68-5.73	585.06	TRIM43,***U6***

2.4.2.1 GO-BP terms and KEGG pathways related to thermotolerance

The GO-BP terms and KEGG pathways that potentially contribute to the superior thermotolerance ability of African cattle are presented in Figure 2.3 and Table 2.3, respectively. Among the GO-BP terms overrepresented, “regulation of angiogenesis” and “superoxide metabolism” are related to heat tolerance Figure 2.3). Angiogenesis, a mechanism by which new blood vessels grow from pre-existing ones, is essential for healing, growth, development, and maintenance (Birbrair et al. 2014). The body controls angiogenesis by balancing stimulatory and inhibitory factors. Research has shown a decrease in tissue temperature after a period of chronic heat load, which is attributed to adaptation response of the tissue to increase heat dissipation through angiogenesis (Davies et al. 1994; Seese et al. 1998). Angiogenesis influences the flow of blood to the skin and sweating response during heat acclimation (Ely et al. 2014).

Heat stress induces both apoptotic and necrotic cell death (Samali et al. 1999). Regeneration is a process of renewal, restoration, and growth of genomes and cells in response to environmental stress that causes disturbance or damage. It has been previously reported as a stress response in echinoderm species (Patruno et al. 2001). Wnt signaling and calcium signaling pathways were also enriched in the KEGG pathways (Table 2.3) and can be involved in thermotolerance mechanisms in African cattle. These pathways are key for thermal sweating (Cui et al. 2014; Cui and Schlessinger 2015).

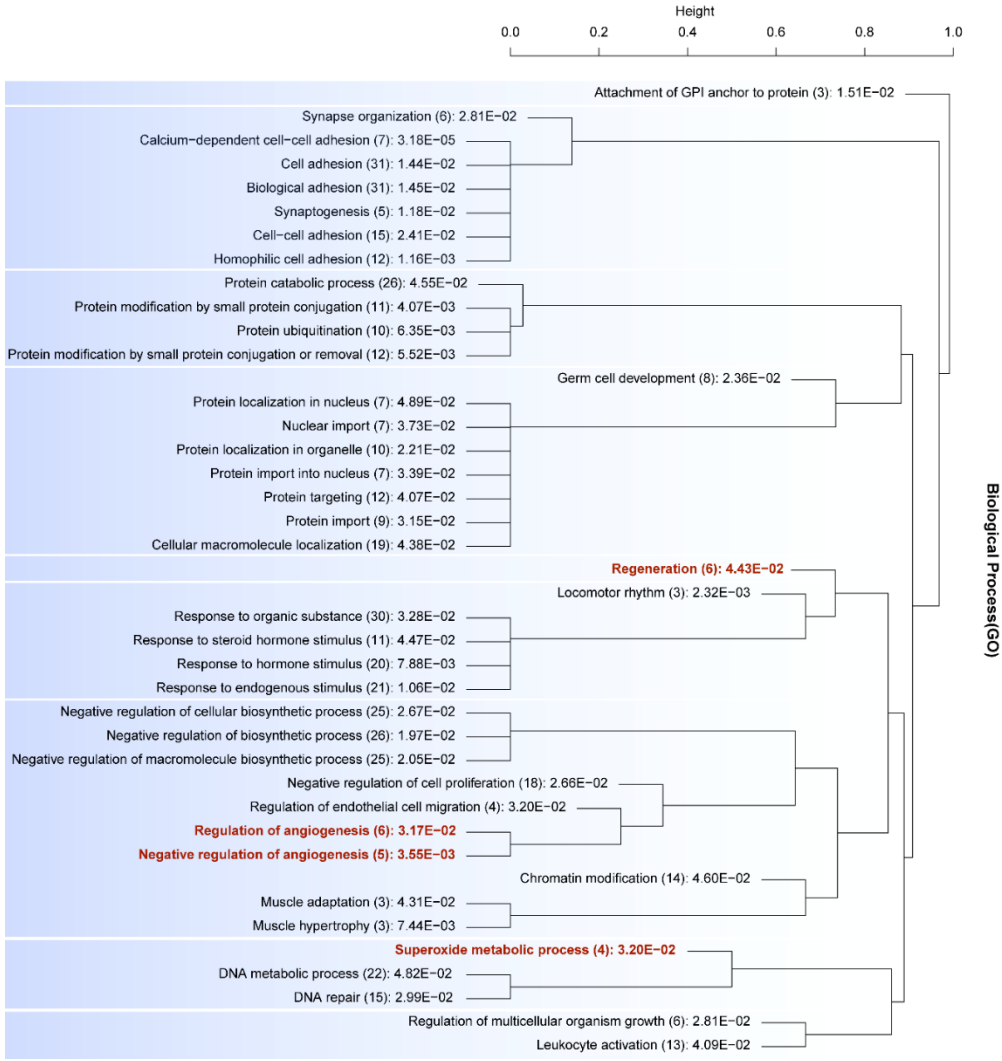


Figure 2.3 Hierarchical clustering of the gene ontology biological process (GO-BP) terms overrepresented from DAVID gene ontology analysis using the genes detected by both XP-CLR and XP-EHH methods. The GO-BP terms related to thermotolerance mechanisms in African cattle are selected manually and are shown in red font color. The numbers of genes in the BP terms are in brackets with the corresponding p-values.

Table 2.3 The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched from DAVID gene ontology analysis using the XP-EHH and XP-CLR combined gene list

KEGG Term	Raw P-value	Genes	Fold Enrichment
hsa04340: Hedgehog signaling pathway	0.0227	CSNK1A1, WNT4, BMP2, CSNK1G3, PRKACB, GLI3	3.6565
hsa04310: Wnt signaling pathway	0.0311	CSNK1A1, TBL1XR1, WNT4, EP300, RUVBL1, PPP3CA, PRKACB, PLCB1, NFATC2, TCF7L2	2.2601
hsa04660: T cell receptor signaling pathway	0.0371	AKT1, PTPN6, LAT, CD3G, CD3D, CD3E, PPP3CA, NFATC2	2.5280
hsa04640: Hematopoietic cell lineage	0.0384	CD3G, CD3D, CD3E, FLT3, CSF1, GP1BA, CD1A	2.7778
hsa04020: Calcium signaling pathway	0.0699	PTGER3, ATP2A2, P2RX3, PPID, TRHR, PPP3CA, PRKACB, PLCB1, CACNA1B, ITPR2	1.9391

2.4.2.2 Genes related to thermotolerance mechanisms

Heat shock and oxidative stress response proteins

Heat shock proteins (HSPs) are among the genes which have been extensively studied for their association with heat stress response in animals. They increase thermotolerance through their functioning as molecular chaperones (Gupta et al. 2013; Belhadj Slimen et al. 2015). Although it is known that heat stress increases HSPs (Belhadj Slimen et al. 2015), HSPs were not identified in this study. Rather, heat shock transcription factor protein (*HSF5*) was detected (Table 2.4). Heat shock factors

(HSFs) are transcription factor proteins that are activated by stress factors and bound to HS sequence element that is transcribed to HSP mRNA and ultimately translated to HSP – this results in protein renaturation. They are essential for the synthesis of HSPs (Gupta et al. 2013). *IGF-I* also increases heat shock proteins in the epidermis (Collier et al. 2008). It is reported to have a thermoprotective role from heat-induced embryo damage during embryonic development (Paula-Lopes et al. 2013). A previous study by Bahbahani et al. (2015) reported the positive selection of genomic regions associated with heat shock and heat stress response in East African Shorthorn Zebu cattle as an adaptation to perform under heat stress conditions.

Heat stress increases lipid peroxidation associated with the production of a large number of free radicals and reactive oxygen species (ROS) which results in cytotoxicity (Sunil et al. 2011; Belhadj Slimen et al. 2015). In response to increased ROS levels, the activities of antioxidant enzymes increase to maintain the steady-state concentrations of generated radicals in the body (Belhadj Slimen et al. 2015). Positively selected genes identified in relation to oxidative stress response include *SOD1*, *GPX7*, *GSTM2*, *GSTM4*, *SLC23A1*, and *SLC23A2* (Table 2.4). *SOD1* is a protein that is able to destroy radicals which are normally produced within the cells and are toxic to biological systems. Overexpression of *SOD1* gene has been previously reported to increase thermotolerance in *Saccharomyces cerevisiae* (Davidson et al. 1996). Analysis of LD and haplotype maps of *SOD1* gene region (Figure 2.4a) showed longer LD patterns and stronger LD block exhibited for the African populations. The haplotype structure was different; for the given region, no blocks and haplotypes were identified for the Commercial cattle populations. *GPX7* is a member of the glutathione peroxidase family which suppresses acidic bile acid-induced ROS (Del Vesco et al. 2015). Similar to *SOD*, *GPX* proteins attenuate the negative effect of environmental stress such as temperature by scavenging both intracellular and extracellular superoxide radicals and preventing lipid peroxidation of the plasma membrane (Sunil et al. 2011; Gupta et al. 2013; Belhadj Slimen et al. 2015). Shortage of *GPX7* in the body

is reported to be associated with ROS accumulations, highly elevated incidence of cancer, auto-immune disorders, and obesity in humans and mouse (Chen et al. 2016). *SLC23A* genes (*SLC23A1*, *SLC23A2*) encodes one of the two sodium-dependent vitamin C transporters which are required for the absorption of Vitamin C into the body and its distribution to organs. Body cells mobilize endogen antioxidants like Vitamin C to detoxify free radicals generated due to heat stress to preserve the steady-state concentrations of ROS. In addition, the beneficial effect of supplementation of vitamins and minerals during heat stress has been reviewed for different species of animals by Belhadj Slimen et al. (2015).

Heat stress induces osmotic stress since water retention is reduced due to the increased electrolyte excretion through urine, feces, and sweat. During heat stress and dehydration, sodium ions (Na+) move from the cellular fluid into the cell increasing the concentration of Na+, thereby inhibiting nutrient uptake by the cell (Pearce et al. 2013). *BHMT2* is involved in the regulation of homocysteine metabolism with beneficial effects in heat stressed animals through its activity against osmotic stress and protection of protein denaturation (Cottrell et al. 2015; Del Vesco et al. 2015). Increased water retention due to the osmolytic effect of betaine increases the volume of the cell, which thereby increases the anabolic activity, the integrity of the cell membrane, and overall performance of the animal. The osmolytic property of betaine permits cellular adaptation to adverse osmotic environments noticed in hot and humid climates (http://www.wattagnet.com/articles/583-betaine-plays-many-roles-in-broiler-diets). Betaine has a modulatory effect on HSPs during heat stress; it declines the effect of HSPs by stabilizing cellular proteins and protecting them from heat stress-induced denaturation (Dangi et al. 2016). Additionally, it inhibits lipid peroxidation and protects the enzymatic antioxidant defense mechanisms ameliorating the tissue and cellular effect of stress (Dangi et al. 2016). The LD and haplotype structures in this gene region were found to be different between African and Commercial cattle populations, that African cattle populations exhibited stronger LD (Figure 2.4b).

PRLH is a hormone that stimulates prolactin release and regulates its expression (Hinuma et al. 1998). Increased plasma prolactin concentration is previously reported to be associated with altered metabolic state of heat-stressed animals (Sunil et al. 2011; Gupta et al. 2013). *PLCB1*, an enzyme that hydrolyzes phospholipids into fatty acids and other lipophilic molecules, is reported to be positively selected in sheep and goats in response to adaptation to the dry arid environments (Kim et al. 2016).

Sweating and sweat gland development

Thermal sweating is one of the mechanisms of heat tolerance (Jian et al. 2014; Lenis Sanin et al. 2016). Thermoregulatory sweating in cattle involves numerous apocrine sweat glands (Akers and Denbow 2008; Collier et al. 2008). The secretion from apocrine glands, in addition to watery sweat, contains fatty acid and some proteins making it less copious, viscous and oily; when mixed with sebum become more liquid and give efficient cooling function (Folk and Semken 1991). Second only to horses, cattle have been shown to lose incredible amounts of heat through sweating (Akers and Denbow 2008).

Several genes and pathways are involved at different stages of sweat gland development and thermal sweating. Wnt signaling is required for sweat gland induction in the progenitor cells of the epidermis and further regulates the successive stages of sweat gland development (Cui et al. 2014; Cui and Schlessinger 2015). The positively selected genes that might be involved in thermal sweating in African cattle include *ITPR2, ITPRIP, CFTR, SCNN1D,* and *SLC9A4. ITPR2* encodes a protein called *InsP3R* that helps calcium ions to move into and out of cells, which in turn is essential for many cell functions. *ITPR2* gene controls a basic cellular process in sweat glands, promoting the release of calcium from extracellular interstitial fluid and release of intracellular Ca^{2+} stores necessary for normal sweat production, and its loss results in impaired sweat secretion called anhidrosis (Klar et al. 2014; Cui and Schlessinger

2015). *ITPRIP* enhances the sensitivity of *ITPR* to intracellular calcium signaling (provided by RefSeq).

CFTR is a multi-functional anion channel that transports chloride (Cl−) and bicarbonate. It regulates a range of transporters including the epithelial sodium channel (McDonagh et al. 2015). *CFTR* contribute to electrochemical driving of sweat secretion and involved in the reabsorption of sodium ions in the sweat duct (Wilke et al. 2007). Partial reabsorption of NaCl in the sweat duct at the subsequent excretion stage results in final hypotonic sweat (Cui and Schlessinger 2015). Mutations in the *CFTR* gene lead to cystic fibrosis, a condition that leads to increased chloride level in sweat (Wilke et al. 2007). Salt concentration in sweat is found to decrease with acclimation to heat (Folk and Semken 1991). *TMEM16* also are calcium-activated chloride channels that play a role in transepithelial anion transport and are expressed in sweat glands (Cui and Schlessinger 2015). *SCNN1D*, also called *ENaC*, are epithelial sodium channels expressed in sweat glands and facilitates sodium absorption (Lee et al. 2008; Cui and Schlessinger 2015). Similarly, *SLC9A4* (also called *NHE1*) is Na+ /H+ exchanger that is expressed in the duct and secretary portion of human sweat glands (Cui and Schlessinger 2015). *SGK1* is involved in the regulation of a wide variety of ion channels, membrane transporters, cellular enzymes, transcription factors, neuronal excitability, cell growth, proliferation, survival, migration, and apoptosis. It stimulates sodium transport into epithelial cells by enhancing the stability and expression of *ENAC* through inhibition of Nedd4-2 (Lee et al. 2008).

Hair coat type and color

Mammalian skin and hair provide physical protection and thermoregulation function (Jian et al. 2014). Some properties of the hair coat and coat color in cattle that enhance conductive and convective heat loss and reduce absorption of solar radiation can be taken as phenotypic markers of heat tolerance (Finch et al. 1984; Hansen 2004; Lenis Sanin et al. 2016). The light-colored, sleek and shiny hair coats of cattle adapted to

the tropical environment (e.g., African cattle) reflect a greater proportion of incident solar radiation reducing heat load (Turner 1964; Finch et al. 1984; Hansen 2004).

Coat color in animals is determined through several processes and the interaction of many color-associated genes (Cieslak et al. 2011). Genes identified in this positive selection scan that are related to coat color include *SLC45A2, MLPH, RAB17, RAB37,* and *ATRN*. *SLC45A2* is a melanosomal membrane channel which plays an important role in melanin synthesis. Melanosomes are specialized organelles where melanin is produced that contain different melanosome-specific components (Cieslak et al. 2011). Mutation in *SLC45A2* affects pheomelanin, producing a cream phenotype in horses and plumage color variation in chicken and Japanese quail (Cieslak et al. 2011). In humans, *SLC45A2* is associated with pale skin, hair or eye color traits (Sturm 2009). *MLPH* is involved in the peripheral transport of melanosomes that can be transferred to the surrounding tissues, and its mutation disrupts the essential melanosome organization resulting diluted coloration in different species of animals (Philipp et al. 2005; Cieslak et al. 2011). Ras oncogene family genes (*RAB17, RAB37*) are also involved in the trafficking, targeting, docking, and fusion of melanosomes (Cieslak et al. 2011). *ATRN* play a role in melanocortin signaling pathway regulating hair color and its level is required for normal pigment production (Gunn et al. 2001). This gene is among those responsible for the change of color of animals as they get older and in relation to their environment (Seo et al. 2007). Regulation of these melanogenesis processes is very important for the differences in pigmentary traits (Sturm 2009).

MC5R, a receptor for αMSH and ACTH, regulates exocrine glands and control the secretion of sebum (Zhang et al. 2006). Sebum is a secretion of sebaceous or oil glands that release their products onto the hair follicles. Sebum lubricates skin and hair with a function of hydrophobic protection against over wetting and for thermal insulation in cattle (Zouboulis 2004; Akers and Denbow 2008). Mice lacking *Mc5r* receptor display reduced production of sebaceous lipids resulting in reduced ability to

repel water from their body effectively and have problems with thermoregulation (Chen et al. 1997).

Homeobox genes (*HOXC12*, *HOXC13*) identified in this study play a role in hair follicle differentiation, growth, and development through regulating keratin differentiation-specific genes. *HOXC13* gene has been found to determine skin thickness (Wu et al. 2009; Wu et al. 2013). Skin thickness and number of hair follicles are reported to have an effect on thermoregulation; thermotolerant cattle have thicker skin than heat susceptible cattle breeds (Alfonzo et al. 2016). In general, *indicus* cattle have thicker skin than *taurus* cattle (Pan 1963; Alfonzo et al. 2016). There was a difference in haplotype and linkage equilibria structure of *HOXC13* gene from African and Commercial cattle breeds that no haplotypes were represented for the Commercial cattle (Figure 2.4c).

Feed intake and energy homeostasis

Heat stress causes a reduction in metabolic rates and alters post-absorptive metabolism in addition to decreasing feed intake. It reorganizes the use of body resources for different biological functions such as growth, reproduction, and health (Belhadj Slimen et al. 2015). Under heat stress conditions, tolerant cattle breeds display a lower reduction in feed intake and production compared to the susceptible ones. The positive selection of genes involved in energy homeostasis, metabolism and feed intake (*FTO*, *ATRN*, *NFATC2*, and *IGF-1*) might help heat tolerant cattle breeds to produce and reproduce under heat stress conditions.

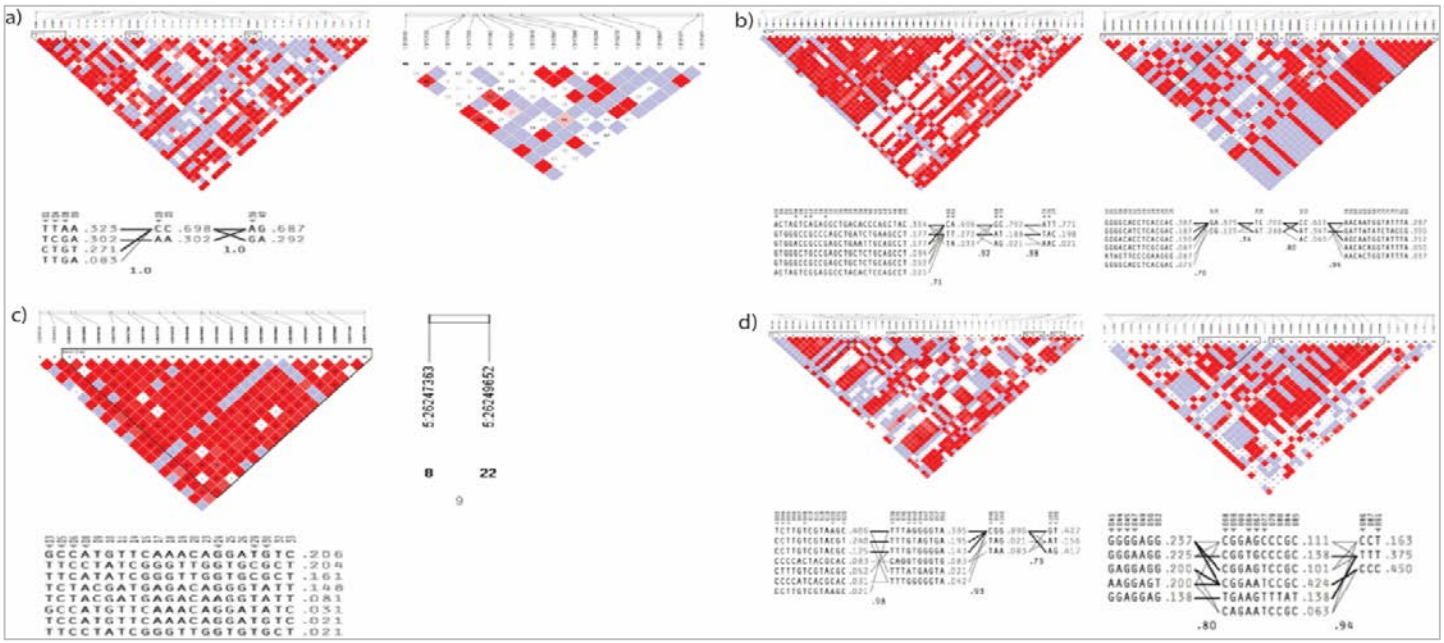


Figure 2.4 Haploview representation of pairwise linkage disequilibria of candidate genes in African and Commercial cattle breeds. a) *SOD1* gene spanning a 8kb region (61 SNPs); b) *BHMT2* gene spanning a 5kb region (82 SNPs); c) *HOXC13* gene spanning a 6kb region (33 SNPs); and d) *SLC9A4* gene spanning a 9kb region (111 SNPs) is shown for African cattle (left, n=48) and Commercial cattle (right, n=41) for each gene. Blocks were defined using the confidence interval method. SNP position is displayed along the top of the diagram. Colors represent D′ values (dark red=high inter-SNP D′; blue=statistically ambiguous D′; white=low inter-SNP D′), and r^2 values are contained within blocks (bold=high r^2).

Table 2.4 Major candidate genes that putatively affect thermotolerance traits in African cattle identified using XP-CLR and XP-EHH population statistics (based on Genecards and literature)

Trait	Gene	Chr.	Gene position (Kbp)	XP-CLR	XP-EHH	XP-EHH p-value
Oxidative stress response	*SOD1*	1	3075.0 - 3125.0	186.33	-	-
	GPX7	3	94325.2 - 94375.2	78.81	-	-
	SLC23A1	7	52225.1 - 52275.1	68.92	-	-
	SLC23A2	13	47575.4 - 47625.4	76.49	-	-
	PLCB1	13	850.0 - 900.0	-	1.1322	0.0044
Osmotic stress response	*PRLH*	3	117600.0 - 117650.0	73.65	1.1692	0.0039
	BHMT2	10	10076.2 - 10126.2	117.23	-	-
Heat shock response	*IGF-I*	5	66525.6 - 66575.6	85.34	-	-
	HSF5	19	9600.0 - 9650.0	-	1.1992	0.0033
Sweating and sweat gland development	*CFTR*	4	51225.0 - 51275.0	77.89	-	-
	ITPR2	5	83950.0 - 84000.0	-	1.1275	0.0045
	SGK1	9	73275.2 - 73325.2	102.4	-	-
	SLC9A4	11	7200.0 - 7250.0	-	1.0014	0.0088
	SCNN1D	16	52575.1 - 52625.1	75.97	-	-
Hair coat type and color properties	*MLPH*	3	117600.0 - 117650.0	73.65	1.1692	0.0039
	RAB17	3	117650.0 - 117700.0	-	1.4019	0.001
	HOXC12, HOXC13	5	26225.6 - 26275.6	74.72	-	-
	RAB37	19	57350.0 - 57400.0	-	1.3376	0.003
	SLC45A2	20	39850.0 - 39900.0	-	1.207	0.0031
	MC5R	24	43975.7 - 44025.7	118.98	-	-
Feed intake and energy homeostasis	*ATRN*	13	52125.4 - 52175.4	126.06	-	-
	NFATC	13	80075.4 - 80125.4	83.94	-	-
	FTO	18	22300.0 - 22350.0	-	1.6631	0.0002
Reproduction function	*CSF1*	3	33575.2 - 33625.2	75.07	-	-
	ESR2	10	76676.2 - 76726.2	126.74	-	-
	RXFP3	20	39850.0 - 39900.0	-	1.207	0.0031
	CIB1	21	22025.0 - 22075.0	79.63	-	-
	MC2R, MC5R	24	43975.7 - 44025.7	118.98	-	-

FTO is a nuclear protein of the AlkB related non-haem iron and 2-oxoglutarate-dependent oxygenase superfamily, which is known to contribute to the regulation of the global metabolic rate, energy expenditure, energy homeostasis and feed intake (Zhang et al. 2011; Zielke et al. 2013). It has also been found to be associated with variations in milk fat yield and total energy content of milk controlling energy homeostasis and energy partitioning (Zielke et al. 2013). *Atrn* is found to be involved in multiple pathways of feeding and metabolism in the rat (Lu et al. 1999; Gunn et al. 2001). *NFATC2* is involved in glucose and insulin homeostasis (Yang et al. 2006). *IGF-1* encodes a protein similar to insulin which is a potent regulator of carbohydrate and lipid metabolism. It mediates post-absorptive nutrient partitioning during heat stress. An increase in insulin level is an adaptation mechanism to heat stress (Collier and Collier 2012).

Reproduction function

The negative effect of heat stress on reproduction function of animals have been reviewed by Hansen (2004, 2009). With this regard, tropically adapted cattle have better reproductive efficiencies under harsh environmental conditions (Hansen 2004, 2009; Makina et al. 2015). Several genes related to reproduction function (*GNRH1, MC2R, MC5R, ESR2, RXFP3, CSF1, CIB1, HOXC12,* and *HOXC13*) were found under selection in this study. *GNRH1* is a gene that is involved in the control of gametogenesis and sex steroid production. The effect of *GNRH1* on reproduction function under heat stress condition might be related to its effect on luteinizing hormone secretion (Chen and Fernald 2008) - heat stress compromises luteinizing hormone secretion (Hansen 2009). Melanocortin receptors (*MC2R, MC5R*) are involved in regulatory mechanisms related to ovarian functions such as ovulation, steroidogenesis, and luteal function (Amweg et al. 2011; Agulleiro et al. 2013). *ESR2* belongs to the nuclear receptor family of transcription factors involved in the regulation of spermatogenesis. It is associated with sperm quality and fertility traits in boars (Gunawan et al. 2012).

SERPINA6, HOXC12, and *HOXC13* are directly or indirectly involved in reproduction pathways and are previously reported to be under selection in an African indigenous cattle (Makina et al. 2015). *RXFP3*, the receptor for relaxin-3, is expressed in the testes of domestic cat and mouse; have a major function during spermatogenesis by an auto-/paracrine mechanism of action (Braun et al. 2015). *CSF1* is known for its function in reproduction with major effects being on neuroendocrine function, ovulation and mammary gland development; it is required for normal male and female fertility (Pollard 1997). *CIB1*, a testicular Secretory Protein Li 9, is an essential protein for male fertility in sheep (Yu et al. 2009).

2.5 Conclusion

Cattle adapted to high-temperature environments develop different adaptation mechanisms. From this study, several putative genes that are related to different thermotolerance mechanisms were identified to be under selection in African cattle breeds. The high environmental temperature of the region might be the possible selective pressures underlying. The identification of the positive selection of these genes in African cattle would help us increase our understanding of the biological mechanisms of heat tolerance in cattle. Breeding and genomic selection programs that will make use of the result need to be designed for African cattle breeds to exploit the genetic resource. This result witnesses the genetic merit of African cattle towards the development of the future heat tolerant commercial cattle breeds that can produce under the current global warming conditions. Genomic selection and genome editing techniques (in the future) could be applied for specialized high producing dairy and beef cattle breeds for thermotolerance traits.

This chapter was published in *Mammalian Genome*

in partial fulfillment of Mengistie Taye Terefe's Ph. D program

Chapter 3. Exploring Evidence of Positive Selection Signatures in Cattle Breeds Selected for Different Traits

3.1 Abstract

Since domestication, the genome landscape of cattle has been changing due to natural and artificial selection forces resulting in several general purpose and specialized cattle breeds of the world. Identifying genomic regions affected due to these forces in livestock gives an insight into the history of selection for economically important traits and genetic adaptation to specific environments of the populations under consideration. This study explores the genes/genomic regions under selection in relation to the phenotypes of Holstein, Hanwoo, and N'Dama cattle breeds using Tajima's D, XP-CLR, and XP-EHH population statistical methods. The whole genomes of 10 Holstein (South Korea), 11 Hanwoo (South Korea), and 10 N'Dama (West Africa - Guinea) cattle breeds re-sequenced to ~11x coverage and retained 37 million SNPs were used for the study. Selection signature analysis revealed 441, 512, and 461 genes under selection from Holstein, Hanwoo, and N'Dama cattle breeds, respectively. Among all these, seven genes including *ARFGAP3, SNORA70,* and other RNA genes were common among the breeds. From each of the gene lists, significant functional annotation cluster terms including milk protein and thyroid hormone signaling pathway (Holstein), histone acetyltransferase activity (Hanwoo), and renin secretion (N'Dama) were enriched. Genes that are related to the phenotypes of the respective breeds were also identified. Moreover, significant breed-specific missense variants were identified in *CSN3, PAPPA2* (Holstein), *C1orf116* (Hanwoo), and *COMMD1* (N'Dama) genes. The genes identified from this study provide an insight into the biological mechanisms and pathways that are important in cattle breeds selected for different traits of economic significance.

3.2 Introduction

Cattle are of significant economic and sociocultural importance all over the world providing meat, milk, hide, and power. Domestication and selection have been changing the genetic landscape of cattle breeds resulting various traits of commercial importance including environmental adaptation, appearance, and production traits (The Bovine HapMap Consortium 2009; Randhawa et al. 2016). Nature selects the genome for adaptation traits (traits like disease resistance, temperature adaptation, and high altitude adaptations) to survive and reproduce under that particular environment. When a mutation that provides a fitness advantage to a particular environment occurs, its frequency increases in the population and creates differences between populations. Similarly, artificial selection by humans, considering various economic and aesthetic traits, also puts its effect on the genome towards the trait of interest and resulted in a number of specialized breeds. For instance, Holstein cattle have been artificially intensively selected for milk and milk-related traits, resulting in animals capable of producing huge amounts of milk (Stella et al. 2010). These days beef breeds (e.g., the Korean Hanwoo cattle) have been intensively selected for meat production and quality traits (Porto-Neto et al. 2014). On the other hand, there are cattle breeds to which no well-designed artificial selection for a particular trait have been done – general breeds now onwards in this manuscript- (e.g., N'Dama, except those simple selection efforts practiced by their owners based on phenotypic characteristics like coat color, horn shape, and the like) (Berthier et al. 2015).

The signature of natural and artificial selection left on the genome can be traced back and help to understand the evolutionary processes shaping the genome (Rothammer et al. 2013; Gouveia et al. 2014). Ascertaining the genes and genomic regions affected due to selection forces is essential in order to understand the biological mechanisms underlying the phenotypic differences observed between livestock

breeds selected for different purposes and those evolved under different environmental conditions (Rothammer et al. 2013). Positive selection signature analysis, by comparing the genomes of phenotypically divergent breeds, have been used previously to identify the genes/gene regions affected due to selection. With this regard, the sequencing of the reference bovine genome facilitated efforts aiming at understanding the genetic basis of phenotypic differences of several cattle breeds through re-sequencing technologies (Zimin et al. 2009). Following this, several methods have been developed and used to assess the positive selection signatures in the genomes of various livestock species and/or breeds. For instance, F_{ST} (Porto-Neto et al. 2014; Bahbahani et al. 2015), XP-EHH (Rothammer et al. 2013; Lee et al. 2014a; Li et al. 2016), ZHp (Rubin et al. 2012), XP-CLR (Lee et al. 2014a), *iHS* and *Rsb* (Bahbahani et al. 2015), and others have been used and revealed genomic regions under selection in relation to domestication, environmental adaptation, and production traits in several livestock species.

In this study, I used Tajima's D, XP-CLR, and XP-EHH population statistical methods to explore genomic selective sweep regions in Holstein, Hanwoo, and N'Dama cattle breeds. Using different methods in signature analysis helps to identify different patterns and regions of selection besides its importance of getting cogent evidence about the sweep regions (Qanbari and Simianer 2014). Tajima's D is a method based on frequency spectrum that compares the number of pairwise differences between individuals with the total number of segregating polymorphisms (Vitti et al. 2013). XP-CLR is a likelihood method for detecting selective sweeps that involve jointly modeling the multilocus allele frequency differentiation between two populations (Chen et al. 2010). XP-EHH is based on linkage disequilibrium which is designed to detect ongoing or nearly fixed selective sweeps by comparing haplotypes from two populations (Sabeti et al. 2007).

3.3 Materials and Methods

3.3.1 Sample preparation and whole genome re-sequencing

Genomic DNA sequences from three cattle breeds (10 Holstein, 11 Hanwoo, and 10 N'Dama) obtained from previously published data was used for this study. Holstein samples were taken from different Korean Holstein heifers inseminated with semen imported from Canada (Lee et al. 2014a). For Hanwoo cattle, blood samples were collected from Hanwoo Improvement Center of the National Agricultural Cooperative Federation (HICNACF), Korea. N'Dama cattle were sampled in the Fouta Djallon area of Guinea, the geographic center of origin of the breed (Kim et al. 2017a). Information on blood sample collection, DNA extraction, and sequencing procedures is described in the manuscript (Kim et al. 2017a). DNA was isolated from whole blood using a G-DEXTMIIb Genomic DNA Extraction Kit (iNtRoN Biotechnology, Seoul, Korea) according to the manufacturer's protocol. Inserts of ~300 bp were generated from 3 μg of genomic DNA randomly sheared using Covaris System. Using the TruSeq DNA Sample Prep. Kit (Illumina, San Diego, CA), library was constructed following the manufacturer's guidelines and whole genome sequencing was performed using the Illumina HiSeq 2000 platform.

I performed a per-base sequence quality check using fastQC software (Andrews 2010). Pair-end sequence reads were mapped to the reference bovine genome (UMD 3.1) using Bowtie2 (Langmead and Salzberg 2012) with default parameters except for the "--no-mixed" option. The overall alignment rate of reads to the reference sequence was 98.50% with an average read depth of 10.8x. On average across the whole samples, the reads covered 98.51% of the genome.

I used open source software packages of Picard tools (http://picard.sourceforge.net), SAMtools (Li et al. 2009), and Genome Analysis ToolKit 1.4 (GATK) (McKenna et al. 2010) for downstream processing and variant

calling. Picard tools were used to filter potential PCR duplicates with options of "REMOVE_ DUPLICATES=true" in "MarkDuplicates". SAMtools was used to create index files for reference and bam files. Genome analysis toolkit 1.4 performed local realignment of reads to correct misalignments due to the presence of indels ("RealignerTargetCreator" and "IndelRealigner" arguments). I used the "UnifiedGenotyper" and "SelectVariants" arguments of GATK to call candidate SNPs. To filter variants and avoid possible false positives, the "VariantFiltration" argument of the same software was adopted with the following options: 1) SNPs with a phred-scaled quality score of less than 30 were filtered; 2) SNPs with MQ0 (mapping quality zero; total count across all samples of mapping quality zero reads) > 4 and quality depth (unfiltered depth of non-reference samples; low scores are indicative of false-positives and artifacts) < 5 were filtered; and 3) SNPs with FS (Phred-scaled P-value using Fisher's exact test) > 200 were filtered since FS represents variation on either the forward or the reverse strand, which is indicative of false positive calls. Using BEAGLE (Browning and Browning 2007), I inferred the haplotype phase and impute missing alleles for the entire set of cattle populations simultaneously. After all the filtering processes, a total of ~37 million SNPs were retained and used for further analysis.

3.3.2 Population stratification

I used STRUCTURE software to identify groups of individuals that are genetically homogeneous (Evanno et al. 2005). STRUCTURE software which implements Bayesian algorithms to detect the true number of clusters (*K)* was used in a sample of individuals of 10 Holstein, 11 Hanwoo, and 10 N'Dama cattle breeds. Beagle was used to generate input files for running STRUCTURE. I used 100,000 iterations with 2,000 burn-in function, and MAF of 0.05.

Principal Component Analysis (PCA) (Jackson 1991) was also performed to examine population variation between the breeds considered. For the analysis, GCTAtool (Browning and Browning 2007) was used to estimate eigenvectors which are equivalent to those estimated by the EIGENSTRAT software tool for PCA. Input data required for GCTA were converted to PLINK format (Purcell et al. 2007), using VCFtools.

3.3.3 Detection of selection signature

The genomes of three cattle breeds (10 Holstein, 11 Hanwoo, and 10 N'Dama) were used to explore the positive selection signatures in each of the cattle populations using Tajima's D, XP-CLR, and XP-EHH statistical methods. Holstein cattle are intensively selected for milk and milk-related traits; Hanwoo cattle are artificially selected for beef traits; and N'Dama breed was used to represent cattle breeds with no or less well-designed intensive artificial selection for a particular trait. Even though whole-genome SNP data from a small number of animals can be used for selective sweep detection (Guo et al. 2016; Kim et al. 2017a), minimum power loss can be expected from this study (Kim et al. 2015b). Three statistical methods were used in order to identify different patterns of selection sweeps and regions of the genome. Detecting the same gene region using different methods provides cogent evidence for selective influences in the region (Qanbari and Simianer 2014).

First, I used Tajima's D statistics to analyze the within-population differentiation of sample populations. Tajima's D compares the number of pairwise differences between individuals with the total number of segregating polymorphisms (Vitti et al. 2013). It detects selective sweep regions going to fixation in the population that makes rare alleles in excess in the population, which results in a negative Tajima's D (Korneliussen et al. 2013). Thus, smaller (i.e., more negative) values of D suggest a surplus of rare alleles, which may be indicative of positive selection or population

expansion (Tajima 1989; Vitti et al. 2013). To calculate the Tajima's D values, I used VCFtools in a window size of 50 kb and interval of 5 kb steps (Danecek et al. 2011). I took the bottom 0.01 (1%) of the empirical Tajima's D values as significant genomic regions under positive selection.

Next, I employed XP-EHH and XP-CLR statistics for cross-population comparisons. To facilitate the analysis, the genome of each of the breeds (as a test population) was compared with the other two breeds combined (e.g., Holstein vs. Hanwoo + N'Dama together) (as a reference population), one after the other. XP-EHH statistics is based on linkage disequilibrium which compares the differences in haplotype frequency and lengths between two populations to control for local variation in recombination rates (Sabeti et al. 2007). XP-EHH values were calculated using software available online (http://hgdp.uchicago.edu/Software/). Using the raw XP-EHH values, I divided the genome into three bins with increments of 500 SNPs combining all windows ≥ 1000 SNPs into one bin. I defined an empirical P-value for each window based on its ranking following previous studies (Lee et al. 2014a; Kim et al. 2017a). The regions with an empirical P-value less than 0.01 (1%) were considered strong signals of selection in the test populations.

XP-CLR is a likelihood method that is based on allele frequency differentiation between populations for detecting selective sweeps in the test population (Chen et al. 2010). The XP-CLR script available online (Chen et al. 2010) was used with non-overlapping sliding windows of 50 kb, and a maximum number of 600 SNPs within each window following previous methods (Kim et al. 2017a). The top 1% (0.01) of the empirical distribution were designated as candidate sweeps and genes that span the window regions were defined as candidate genes. Significant genomic regions identified from the Tajima's D, XP-EHH and XP-CLR tests were annotated to the closest genes (UMD 3.1).

3.3.4 Characterization of genes and candidate association analysis

The Bovine QTL database available online (http://www.animalgenome.org/cgi-bin/QTLdb/BT/search) was searched to identify the overlap of identified genes with previously reported bovine QTL regions. I summarized the overlaps by trait classes defined by the database as (1) Exterior traits, (2) Health traits, (3) Reproduction traits, (4) Production traits, (5) Meat and carcass traits, and (6) Milk traits QTL regions.

Gene enrichment analysis was carried out using the DAVID gene ontology and annotation tool (Huang et al. 2009). I used all the genes detected by three of the methods for each of the breeds considered. Functional annotation clustering of DAVID (version 6.8) with the default settings was used. The DAVID Functional Annotation Clustering report groups/displays similar annotations together, reducing redundancy, which makes the biology clearer and more focused to be read vs. traditional chart report. It is based on the hypothesis that similar annotations should have similar genes and gene members. Accordingly, annotation clusters with an enrichment score of ≥1.3 (equivalent to Fisher's exact test p-value of 0.05, based on the software's recommendations) were taken as significant terms for the identification of enriched clusters. Within an annotation cluster, a representative term was chosen manually considering the number of genes involved in the terms.

I performed genetic variants (amino acid changes) annotation and effect prediction for the candidate genes using SNPEff tool (Cingolani et al. 2012). Missense variants of candidate genes were extracted and carried on for the association analysis. By employing the Chi-squared test and logistic model employed in PLINK V1.07 software, I was able to identify significant SNPs (Chi-squared test) that are specific to each breed. Although no significant results were detected through the logistic model, P-values from the chi-squared test and allele distribution were considered to select

breed-specific significant results. The significance level of Chi-square $P<0.05$ was used and alleles unique for the target breed were selected as significant.

The gene names and descriptions used in this manuscript are based on genecards (http://www.genecards.org/). The Manhattan plot of the $-\log_{10}$ transformed Tajima's D P-values were drawn for the three breeds using R software.

3.4 Result and Discussion

3.4.1 Data description

DNA samples of three cattle breeds (Holstein, Hanwoo, and N'Dama) sequenced to ~11x genome coverage each was used for the study. Following standard data preparation and re-sequencing procedures, an average alignment rate of 98.84 % covering 98.56 % of the taurine reference genome (UMD 3.1) was obtained. Potential PCR duplicates and false-positive calls were filtered using several methods and software, and finally, a total of ~37 million SNPs were obtained and used for further analysis.

3.4.2 Structure and principal component analysis

To understand the admixture level of sample populations, I performed STRUCTURE (Evanno et al. 2005) at two and three population assumptions (Figure 3.1a). When the number of ancestral populations *(K)* was set to 2, Holsteins and Hanwoo breeds showed clear differences from N'Dama. But, at $K = 3$, all the three breeds became different even though Hanwoo showed some level of admixture with Holsteins. This was evidenced by Principal Component Analysis (PCA) (Jackson 1991) that all the

three breeds positioned separately from each other. PC1 explained 20.62% of the variation, separating N'Dama from the other two breeds and PC2 separated Hanwoo and Holsteins explaining 10.05% of the variance (Figure 3.1b). This result is consistent with previous reports that African taurine cattle are more distant from Asian and European taurine breeds (The Bovine HapMap Consortium 2009; Decker et al. 2014); might be because the genome of African taurine cattle is thought to comprise the genomic contribution of African aurochs (Decker et al. 2014).

3.4.3 Positive selection signature

Using Tajima's D, XP-EHH, and XP-CLR statistical methods, I compared the genomes of Holstein, Hanwoo, and N'Dama cattle breeds to explore positive selection signatures in each of the breeds. The Manhattan plot of the $-\log_{10}$ transformed Tajima's D p-values are presented in Figure 3.2. By annotating the 0.01 (1%) outlier regions of the empirical distribution, 441 (209 =Tajima's D, 161 = XP-CLR, and 184 = XP-EHH) genes were detected under selection in Holstein cattle. Among all these genes, 19 were common between three of the methods, and 94 genes were detected by at least two methods (Figure 3.3a). From Hanwoo cattle genome, 512 (202 = Tajima's D, 176 = XP-CLR, and 201 = XP-EHH) genes were identified under selection of which 10 and 58 genes were common for the three and at least two methods, respectively (Figure 3.3a). From N'Dama cattle, 203 (Tajima's D), 159 (XP-CLR), and 190 (XP-EHH) genes were identified under selection. Twelve and 79 of the genes were detected in common by the three methods and at least by two of them, respectively (Figure 3.3a).

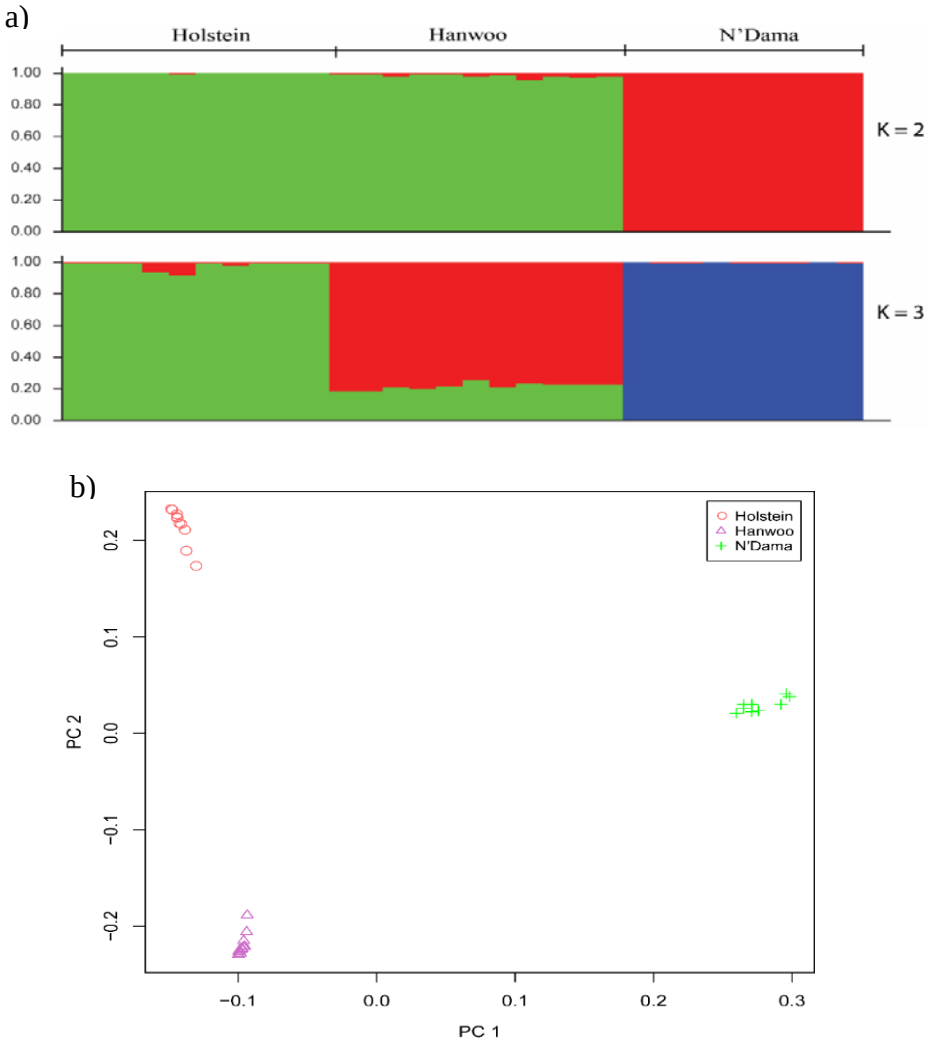


Figure 3.1 Population Stratification of Holstein, Hanwoo, and N'Dama cattle breeds. a) Admixture level of sample populations at different population assumptions (*K*). At *K* = 2, Holstein and Hanwoo breeds showed a clear difference from N'Dama. But, when *K* = 3, all the three breeds became different even though the Hanwoo showed some level of admixture with Holstein cattle breed; b) Principal Component Analysis of three cattle breeds - PC1 (20.62%) separated N'Dama from the other two breeds, and PC2 (10.05%) separated Hanwoo from Holsteins.

I compared the gene list from each of the breeds with previous studies that a substantial number of genes were overlapped (Table 3.2). In Holstein cattle, 151 and 57 of the genes detected by at least one, and two methods, respectively, overlapped with the gene list previously reported by Lee et al. (2014a). Similarly, 34 genes were identified overlapped with those reported by Zhao et al. (2015) to which nine of them were detected by at least two of the methods in this study (Table 3.2). Fortunately, 10 genes (*ACTC1, AQR, GJD2, LYN, POLR1E, PTPN14, RPS20, SV2B, XKR4,* and *ZCCHC7*) were common between these three studies – current study, Lee et al. (2014a), and Zhao et al. (2015). Five genes (*ACVR1C, CD14, BMP7, PAFAH1B3,* and *CHD2*) (Porto-Neto et al. 2014), and two genes (*CTLA4,* and *DBI*) (Lee et al. 2013b) were found overlapping with the gene list from Hanwoo cattle (). Among all genes from N'Dama cattle, 77 of them were previously reported under selection by Kim et al. (2017a); 25 genes were detected by at least two of the methods (Table 3.2). Comparing the gene lists from each of the three breeds, seven genes (*5S_rRNA, 7SK, ARFGAP3, SNORA70, U1, U6,* and *U6atac*) were identified in common. In a genome-wide association study, *ARFGAP3* has been found to affect meat quality traits (Santana et al. 2015) and milk production traits (Mai et al. 2010). RNA genes are transcriptional factors that are required for splicing (Eddy 2001).

Next, I compared the gene list identified with the online bovine QTL previously reported for different traits for overlaps of the gene regions. Figure 3.3b illustrates the overlap of the genes identified to a particular QTL (e.g., how many (%) of the genes identified by a particular method for a particular breed overlaps with a particular QTL). In Holstein cattle, most of the common genes detected by three of the methods overlapped with the bovine QTL regions of milk traits *(ASTN1, HPS1, PAPPA2, PREX2, VPS8,* and *ZBTB20),* reproduction traits *(MNAT1, PAPPA2, PREX2, TRAC,* and *XKR4),* meat and carcass traits *(5S_rRNA, HPS1, MICAL2, PDE11A, PREX2,* and *XKR4),* production traits *(5S_rRNA, PAPPA2, PREX2,* and *VPS8),* and health traits *(XKR4).* Six genes (bta-mir-2291, *LAD1, MAPRE3, SLC28A2, SNORA62,* and *SORD*) identified by three of the methods did not overlap with any of the QTL regions.

Among the Hanwoo genes detected by the three methods, genes overlapping with QTL regions include those affecting meat and carcass traits (*NCOA2, PITPNB, POLR3B, RIC8B,* and *U6)*, production traits *(LRBA, NCOA2, PITPNB,* and *U6)*, milk traits *(LRBA, PHACTR1, RIC8B,* and *U6)*, health *(LRBA, U6)*, and reproduction traits *(LRBA)*. From this list, two genes (*MDFIC,* and *MS4A13*) did not overlap with QTL regions. From N'Dama genes detected by three of the methods, those overlapping with the bovine QTL regions include those affecting meat and carcass traits (*5S_rRNA, 7SK, U1, U6,* and *KCNIP4)*, milk traits (*5S_rRNA, U1, U6,* and *KCNIP4)*, production traits *(U6, CLCA4, COMMD1,* and *KCNIP4)*, reproduction traits (*7SK, U1,* and *KCNIP4)*, and health traits *(U6)*. Among the N'Dama genes detected by three methods, four genes (*CRYGN, POLE4, SNTB1,* and *ZC3H7A*) did not overlap with any QTL region.

The DAVID functional annotation clustering resulted in three (Holstein), two (Hanwoo), and four (N'Dama) significant (enrichment score >1.3) annotation cluster terms (Table 3.3). In this study, I described those enriched functional annotation cluster terms and genes detected by at least two methods and/or those involved in functional annotation clusters that are related to the phenotype of the respective breeds. The gene names and descriptions in this manuscript are based on genecards (http://www.genecards.org/).

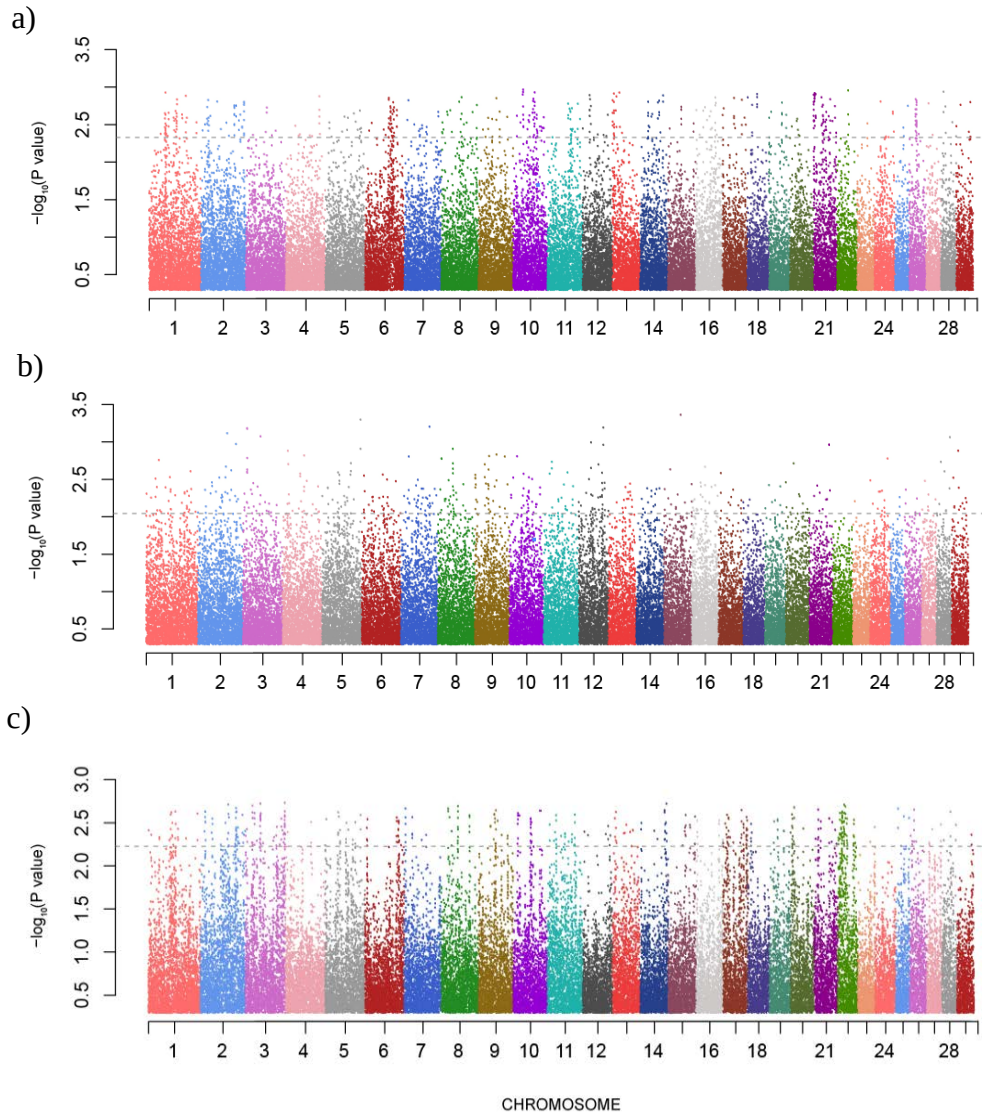


Figure 3.2 Manhattan plot of the $-\log_{10}$ transformed Tajima's D p-values of Holstein (a), Hanwoo (b), and N'Dama (c) cattle breeds. The *y*-axis shows the $-\log_{10}$(*P*-value) of Tajima's D, and the *x*-axis shows chromosomal positions. The horizontal dotted lines represent the 1% outlier Tajima's D regions for all the breeds.

Table 3.1 List of genes identified in common between Tajima's D, XP-EHH and XP-CLR statistical methods in Holstein, Hanwoo, and N'Dama cattle breeds. Genes commonly identified by three of the methods are underlined within the breed row.

Breed	Tajima's D ∩ XP-EHH	Tajima's D ∩ XP-CLR	XP-CLR ∩ XP-EHH
Holstein cattle	5S_rRNA , ADIPOQ, ASTN1, bta-mir-2291, CDC42, CHN1, DPH6, DUOX1, FBXO10, FLRT1, GFPT1, GIPC1, HPS1, ITGA6, LAD1, MACROD1, MAPRE3, MICAL2, MNAT1, PAPPA2, PARVA, PDE11A, POLR1E, PREX2, SHF, SLC28A2, SNORA62, SNORD109A, SNORD116, SORD, SRRM4, STXBP6, TRAC, U6, VPS8, WNT4, XKR4, ZBTB20	5S_rRNA , ACTC1, ANO2, ASTN1, bta-mir-2291, CNTNAP5, CSN3, DNAJC3, FOXO1, GJD2, HPS1, LAD1, MAGI1, MAPRE3, MICAL2, MNAT1, PAPPA2, PDE11A, PREX2, RBAK, SLC28A2, SLC37A1, SNORA56, SNORA62, SORD, TMEM214, TNNI1, TRAC, U2, U6, VPS8, VSTM2B, XKR4, ZBTB20	5S_rRNA, ADAMTS17, AKNA, ASTN1, ATP1A2, ATP1A4, bta-mir-135b, bta-mir-2291, bta-mir-2917, C8orf34, CD300LB, CD48, COL17A1, COL27A1, COL5A2, CPQ, CSRP1, CYTH3, EDIL3, GRIK1, HPS1, KIF12, LAD1, LEMD1, LYN, MAN1A1, MAPRE3, MICAL2, MNAT1, MOS, NFU1, NMNAT2, NUDCD3, OR4S2, PAPPA2, PDE11A, PKNOX1, PLAG1, PLCB1, PLCB4, PREX2, PTPN14, SLC28A2, SMOC2, SNORA19, SNORA62, SNORD112, SORD, THEMIS, TMPRSS11E, TNNT2, TRAC, TTC5, U6, U6atac, UGGT2, VPS8, XKR4, ZBTB20, ZNF175
Hanwoo cattle	5S_rRNA, KCNE2, KIDINS220, LRBA, MDFIC, METTL15, MS4A13, NCOA2, PHACTR1, PITPNB, POLR3B, RIC8B, STRIP1, TXNDC15, U4, U6	5S_rRNA, 7SK, bta-mir-2904-3, C8orf34, COL28A1, ITFG1, LRBA, MBOAT2, MDFIC, MS4A13, NCOA2, PHACTR1, PITPNB, POLR3B, RBL1, RFX4, RIC8B, TSHR, U6	5S_rRNA, ABI2, ACVR1C, ANTXRL, ATP10B, BOK, bta-mir-2413, C1orf116, C21orf62, CCDC91, CIC, DOCK3, DPYD, EVA1C, FAM193A, FIBCD1, FSIP2, GPHN, GRM7, KCTD16, LRBA, MAD1L1, MDFIC, MEGF8, MRPL16, MS4A13, MTERF2, MTMR7, NCOA2, PAFAH1B3, PEBP4, PHACTR1, PITPNB, POLR3B, PRR19, PSTPIP1, RIC8B, SCAPER, SNORA3, TMEM145, U6, YOD1, ZFYVE9
N'Dama cattle	5S_rRNA, 7SK, ANTXR1, ASCC3, bta-mir-2381, CLCA4, COMMD1, CRYGN, HS1BP3, KCNIP4, MYCL, NPAS2, POLE4, PRKCE, SNTB1, TRPA1, U1, U6, U6atac, ZC3H7A	5S_rRNA, 7SK, CLCA4, COMMD1, CRYGN, FAM178B, FAM184B, IGF2BP2, KCNB2, KCNIP4, LCORL, MACF1, METAP2, NOX5, OSR1, PARK2, POLE4, PPP2R5E, SNORA70, SNTB1, SOX6, TRPM8, U1, U6, U6atac, ZC3H7A	5S_rRNA, 7SK, AMZ1, ATF5, ATN1, BSPRY, bta-mir-141, bta-mir-200c, C12orf57, C17orf96, CARD11, CCDC190, CLCA1, CLCA4, COMMD1, COPS7A, CRYGN, CYP46A1, DCDC2C, EPAS1, FHIT, FRMD6, GML, GNA12, GPRC5D, HDHD3, IQGAP2, KCNH5, KCNIP4, LPCAT3, MTUS1, NTM, NUB1, PACSIN2, PHB2, PIANP, PIP5K1B, POLE4, PRKG1, PTPN6, RCBTB1, RPS6KA5, SLC40A1, snoU89, SNTB1, STOM, SVEP1, TRAC, TXNDC11, U1, U6, U6atac, U7, USH2A, WDR86, ZC3H7A

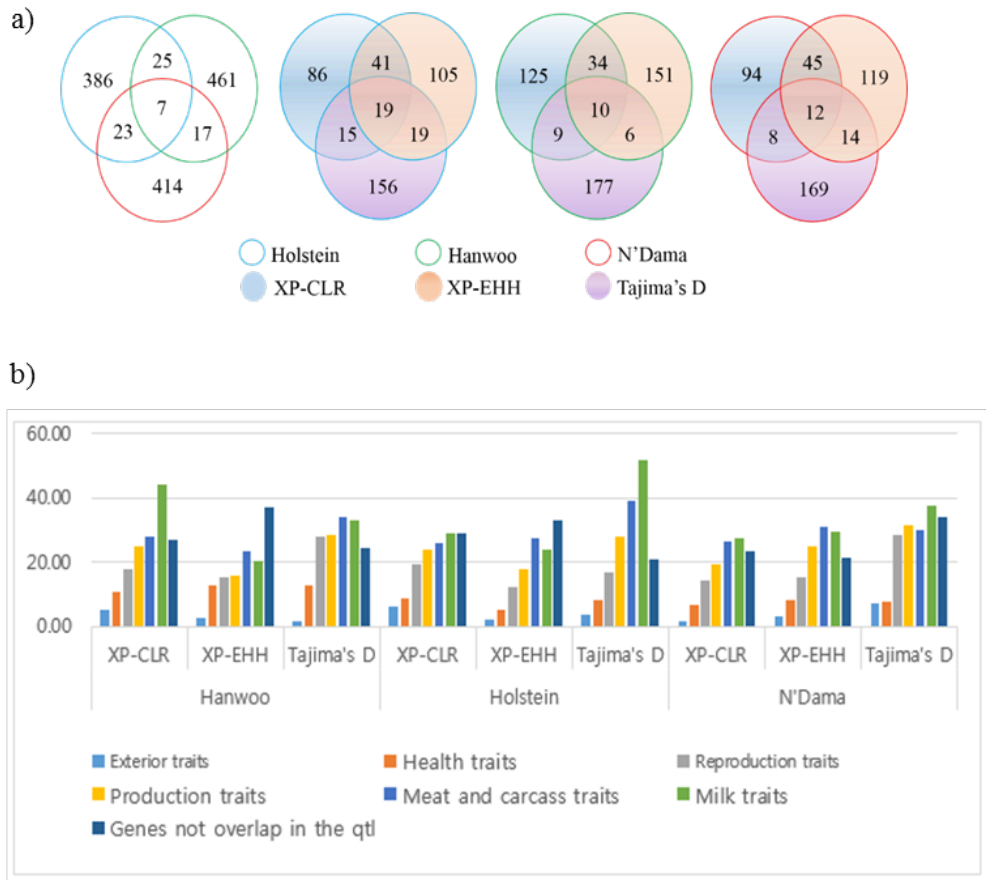


Figure 3.3 a) Summary of the number of genes identified from Tajima's D, XP-CLR, and XP-EHH analysis for Holstein, Hanwoo and N'Dama cattle breeds; b) Illustration of the overlap of genes identified from positive selection analysis to the previously identified bovine QTL regions. The histogram illustrates how many (%) of the genes identified by a specific method (e.g., XP-CLR) for a particular breed (e.g., Hanwoo) overlaps with a particular QTL (e.g., exterior traits). The percentage does not sum up to 100% because of overlapping of QTL regions. Genes that do not overlap with the QTL region are those genes identified under positive selection but not previously reported as QTL regions.

Table 3.2 Genes detected from this study that overlapped with previous studies in the respective breeds. Those genes detected by at ***least two methods*** in this study are ***boldface***

Breeds	Overlapped genes	Reference
Holstein	***5S_rRNA***, *7SK, ACBD6,* ***ACTC1, ADAMTS17, ADIPOQ,*** *AGFG2, AMTN, AQR, ARMC2,* ***ASTN1****, ATF2,* ***ATP1A2, ATP1A4,*** *ATP7B,* ***bta-mir-2291,*** *bta-mir-2457,* ***bta-mir-2917,*** *bta-mir-455, bta-mir-669, CASQ1, CCNB1IP1,* ***CD300LB****, CDKAL1, CHCHD7, CHID1,* ***CHN1****, CILP, CLPX, CLVS1,* ***COL17A1, COL27A1, COL5A2,*** *COQ2, CPNE7,* ***CPQ****, CSN1S1, CSN2,* ***CSN3, CSRP1, DNAJC3, DPH6,*** *DPPA2, DSG3, DTNA,* ***DUOX1****, DUOX2, DUOXA1, DUOXA2, DUSP22, FAM19A1,* ***FBXO10****, FBXW4, GALK2,* ***GJD2****, GLI3, GPR137C, GSKIP, GTF3C1, HECW1, IGHE, IKZF2, IPO9, KCNAB1,* ***KIF12****, KIT, LAMC1, LMOD1, LTBP1,* ***LYN, MAN1A1,*** *MCM6,* ***MICAL2****, MICALCL,* ***MNAT1****, MORC1, MRPS28, NDUFA12,* ***NMNAT2****, NPHP4, NTF3, OR11H4, OR4N2,* ***OR4S2, PAPPA2,*** *PAQR8, PARD3,* ***PARVA, PDE11A****, PDE1B, PGAP3, PHLDB2,* ***PKNOX1****, PLAG1,* ***PLCB4****, PLCD4, PNMT,* ***POLR1E****, PPP1R1A,* ***PREX2****, PSMD12,* ***PTPN14****, RAB1A,* ***RBAK****, RPL13, RPS20, RQCD1, SESN1,* ***SHF****, SHISA4,* ***SLC28A2, SLC37A1, SMOC2,*** *SNAP25, snoMBII-202,* ***SNORA19****, SNORA70,* ***SNORD112****, snoU54,* ***SORD****, SOX5,* ***SRRM4****, STARD3, SULT1B1, SULT1E1, SV2B, SYK, TBC1D16, TCAP,* ***THEMIS****, TIMM17A, TMEM225, TMEM68, TMPRSS11A,* ***TNNI1****,* ***TRAC****, TRMT5, TSPAN4,* ***TTC5****, U1,* ***U2, U6, UGGT2, VSTM2B, XKR4, ZBTB20,*** *ZBTB40, ZCCHC7, ZNF12,* ***ZNF175****, ZNF184*	(Lee et al. 2014a)
	ACTC1*,* ***AQR****, CDC42, CDK10, CHMP1A, COPS5, CPNE7, CSMD3, DNAJB1, DPEP1, GIPC1,* ***GJD2****,* ***LYN****, MC1R, MOS,* ***POLR1E****, PSAT1, PTGER1,* ***PTPN14****, RORA, RPL13,* ***RPS20****, SPATA2L,* ***SV2B****, SYK, TCF25, TECR, TMEM68, TUBB3, UBE3A,* ***XKR4****, ZBTB5,* ***ZCCHC7****, ZNF276*	(Zhao et al. 2015)
Hanwoo	*CTLA4, DBI*	(Lee et al. 2013b)
	ACVR1C, CD14, BMP7, PAFAH1B3, CHD2	(Porto-Neto et al. 2014)
N'Dama	***5S_rRNA, 7SK,*** *ACOT6, ACVR2A, ALLC,* ***AMZ1****, ANKRD34C,* ***ANTXR1****, ARFGAP3, ARL11, BRAT1, C1RL, C2orf82, CAMK2G,* ***CARD11****, CHCHD1,* ***CLCA1****, CLCA2, CLSTN3, COL14A1, COL16A1, COL9A1,* ***COMMD1****,* ***CRYGN****,* ***DCDC2C****, EBPL,* ***FAM184B****, FOXP1, FREM2, GALNT8, GIGYF2,* ***GML****,* ***GNA12****,* ***IQGAP2****,* ***KCNH5****, KCNJ13, KIF2B, KRT80, LY6K, MAP2K5, MGAT4A, NMRK1,* ***NTM****,* ***NUB1****, OPCML, OPRD1, OTUD7A, PNPO, POLD3,* ***PRKCE****,* ***PRKG1****, PRR15L, RBP5,* ***RCBTB1****, RHEB, RSPO2, SBDS, SKAP1, SLC34A2,* ***SLC40A1****, SMARCD3, SNORA31,* ***SNORA70****, SNORD113,* ***STOM****, TANC2, THADA, TMEM156, TRIM62, TSPAN5, TTLL1,* ***U6, U7,*** *VWA3B,* ***WDR86****, YTHDC1, ZNF384, ZSWIM8*	(Kim et al. 2017a)

Holstein breed selected genes

From Holstein cattle genes, the DAVID functional annotation cluster term "milk protein" was enriched to which casein genes (*CSN1S1*, *CSN2,* and *CSN3*) were involved (Table 3.3). Polymorphisms in these milk proteins affect cow milk production parameters and protein quality (Kucerova et al. 2006). Kappa-casein (*CSN3*) is an important protein in stabilizing micelle formation and preventing casein precipitation in milk. It acts together with *CSN2* to form spherical micelles which bind calcium and phosphorous. These proteins are known to affect lactose content in milk (Cecchinato et al. 2014), milk protein (Kucerova et al. 2006; Ogorevc et al. 2009), and previously found under selection in Holstein cattle (Lee et al. 2014a). Searching for non-synonymous mutations, two Holstein-specific significant missense variants – one new (6:87390576; p.Ile157Thr) and one known *(*rs43703016) – were identified in *CSN3* gene region (Table 3.5). Both of the variants are fixed in Holstein cattle as opposed to the other breeds as shown in Figure 3.4a.

Thyroid hormone signaling pathway, enriched in the annotation clusters, is involved in energy homeostasis and modulates energy expenditure (McAninch and Bianco 2014). This pathway might be important for high milk-producing animals in maintaining energy balance during the transition period when the metabolic needs increased dramatically and impact performances during the rest of the lactation period. Thyroid hormones are important for normal growth, development of sexual characteristics, and reproductive function of animals (Fernández et al. 2014). Among the 11 genes involved in this pathway, six of them (*ATP1A4, ATP1A2, FOXO1, PLCB4, PLCB1,* and *WNT4*) were detected by at least two methods. The Pleckstrin homology-like domain cluster, represented by 18 genes (Table 3.3), is involved in intracellular signaling.

Table 3.3 Significant functional annotation clustering terms enriched from DAVID gene enrichment analysis of positively selected genes from Tajima's D, XP-CLR, and XP-EHH for Holstein, Hanwoo, and N'Dama cattle breeds

Representative term[1]	Enrichment score[2]	Genes involved in annotation cluster terms
Holstein breed		
Milk protein	1.755	*CSN1S1, CSN2, CSN3*
IPR011993: Pleckstrin homology-like domain	1.741	*OSBPL6, PREX2, PTPN14, CYTH3, ITSN2, ITPR3, ARHGAP15, SKAP1, MCF2L2, PLCB3, PLCB4, PSD, DCP1B, PLCD4, APBA2, PLCB1, PHLDB2, ARAP2*
bta04919: Thyroid hormone signaling pathway	1.630	*PLCB3, WNT4, PLCB4, PIK3CD, ATP1A4, FOXO1, PLCD4, ATP1A2, BAD, PLCB1, RCAN2*
Hanwoo breed		
GO:0046972~histone acetyltransferase activity (H4-K16 specific)	1.556	*KAT8, WDR5, PHF20*
GO:0060291~long-term synaptic potentiation	1.439	*STX4, STX3, LRRTM2, NTRK2, SNAP25*
N'Dama breed		
bta04924: Renin secretion	2.308	*KCNMA1, CLCA2, PLCB3, CLCA1, CLCA4, PDE3A, PPP3CA*
IPR020683: Ankyrin repeat-containing domain	1.620	*IBTK, ANKS1A, ANKRD33, EHMT1, TANC2, TRPA1, ASB10, ANKRD34C, ACAP2, ANKRD1, FEM1B*
IPR014001: Helicase, superfamily 1/2, ATP-binding domain	1.529	*DDX47, RECQL5, DHX29, ASCC3, INO80, SKIV2L2, SMARCA2, CHD5*
IPR011993: Pleckstrin homology-like domain	1.505	*PPP4R3B, DNM3, ANKS1A, NF2, ROCK2, PREX2, EVL, DGKH, ITSN1, ARHGAP15, SKAP1, PLCB3, FRMD6, ACAP2, SNTB1, SPRED1, KALRN, ARHGAP10*

[1]A representative term was selected manually based on the number of genes involved;

[2]Enrichment score ≥ 1.3 was taken as significant (equivalent to $p<0.05$).

In addition to those involved in functional annotation clusters, genes affecting mammary gland development and health (*WNT4, CDC42, COL5A2, ADIPOQ, MAGI1,* and *PAPPA2*) and milk production traits (*MAGI1*, and *CDC42*) were identified by at least two methods (Table 3.4). *WNT4* is among the genes that play a well-defined essential role in mammary gland development (Ding et al. 2013). Wnt signaling mediates progesterone function during mammary gland morphogenesis (Brisken et al. 2000). The significance of *cdc42* gene in the development of mammary gland prior to pregnancy was shown using transgenic mice (Druso et al. 2016). *CDC42* signaling is important for milk protein synthesis (Akhtar and Streuli 2006). *COL5A2* is an extracellular matrix protein that plays a vital role in mammary gland development (Suárez-Vega et al. 2015). *ADIPOQ* is involved in the process of preventing excessive inflammatory responses in the mammary gland (Lecchi et al. 2015). It also facilitates nutrient partitioning towards the mammary gland which can be taken as an adaptive response to increased energy requirements during milk production (Häussler 2015). A polymorphism in the *MAGI1* gene is found associated with milk yield, fat yield, and protein yield in buffaloes (Venturini et al. 2014). *PAPPA2* have been found associated with milk and protein yield in Holstein cattle (Wickramasinghe et al. 2011). Its expression also increases the availability of *IGF-1* thereby enhancing protein synthesis in the mammary gland (Wickramasinghe et al. 2011). On *PAPPA2* gene region, a significant missense variant specific to Holstein cattle (rs210049354) was identified which is almost (95%) fixed as opposed to Hanwoo to which the reference allele is maintained (Table 3.5; Figure 3.4b). The development and immunity of mammary epithelial cells of lactating mammals are essential for efficient milk production (Rezaei et al. 2016). The positive selection of the genes related to mammary gland development and milk production traits might contribute to the superior milk-producing ability of Holstein cattle (Lee et al. 2014a).

Among the genes detected by at least two methods, those involved in reproduction function include *PAPPA2, CPQ, PLAG1,* and *SMOC2* (Table 3.4). *PAPPA2* affects reproduction and fertility with important roles in pregnancy and postnatal growth (Christians et al. 2013). SNPs in *PAPPA2* gene have been associated with calving ease and productive life in Holstein cattle; influencing the first-calf heifer breeding (i.e., calving interval, days to calving, and pregnancy rate) (Wickramasinghe et al. 2011). *CPQ* plays a role in the liberation of thyroxine hormone from its thyroglobulin precursor that are known for their role in sexual development and spermatogenic functions (Fernández et al. 2014). In bovines, *CPQ* has a contribution to membrane modification of sperm maturation (Kasvandik et al. 2015). *SMOC2* (Höglund et al. 2015) and *PLAG1* (Karim et al. 2011) have been previously reported to affect fertility in cattle.

I also identified genes *(TNNT2, ITGAV,* and *ACTC1)* that are potentially associated with cardiomyopathy, a genetic disorder of the myocardium, that has been previously described in Holstein cattle (Nart et al. 2004). Cardiomyopathy is caused by a genetic mutation that impairs muscle function (Matsson et al. 2008). It follows an autosomal recessive pattern of inheritance to which affected animals display enlarged heart (Owczarek-Lipska et al. 2011). These genes are important components of dilated cardiomyopathy KEGG pathway (http://www.genome.jp/kegg-bin/show_pathway?ko05414+K06487). *TNNT2* encodes the tropomyosin-binding subunit of the troponin complex which regulates striated muscle contraction in response to alterations in intracellular calcium ion concentration (Li et al. 2015). Mutations in *TNNT2* gene were found to cause hypertrophic cardiomyopathy in humans (Li et al. 2015; Gómez et al. 2016). *ACTC1* is an essential protein for cardiac contraction to which its mutation or reduced levels lead to atrial septal defects (Matsson et al. 2008). *ACTC1* and

ITGAV (ITGA6) genes were previously reported associated with cardiomyopathy in Holstein cattle (Lee et al. 2014a).

Genes identified in relation to stature and body size in Holstein cattle include *PLAG1*, *LYN*, and *PAPPA2* (Table 3.4). *PLAG1* is a transcription factor to which its overexpression results in upregulation of *IGF-II*, a gene which in turn possesses growth-promoting activity and plays a role in fetal development (O'Dell and Day 1998). In a previous GWAS analysis, *PLAG1* has been found associated with early life and peripubertal body weight in taurine cattle and taken as a key regulator of mammalian growth (Littlejohn et al. 2012). *PLAG1* (Karim et al. 2011; Zhao et al. 2015) and *LYN* (Utsunomiya et al. 2013) have been previously reported to be under positive selection in relation to stature and body weight in Holstein and Nellore cattle, respectively. *PAPPA2* also affects body size and shape of animals (Christians et al. 2013). Artificial selection for increased body size in Holstein cattle makes them the largest dairy cattle in the world (Hansen et al. 1999).

Table 3.4 Candidate genes putatively affecting the major phenotypes of Holstein, Hanwoo, and N'Dama cattle breeds as detected by at least two statistical methods (Tajima's D, XP-EHH, and XP-CLR)

Genes	Chr.	XP-CLR	XP-EHH	XP-EHH p-value	Tajima's D	Trait and Reference
Holstein breed						
ITGA6	2	-	2.724	0.0018	-2.283	Cardiomyopathy (Lee et al. 2014a)
TNNT2	16	267.48	2.398	0.0063	-	Cardiomyopathy (Li et al. 2015)
ACTC1	10	390.23	-	-	-2.477	Cardiomyopathy (Matsson et al. 2008)
SMOC2	9	346.90	2.527	0.0041	-	Fertility (Höglund et al. 2015)
WNT4	2	-	2.406	0.0060	-2.427	Mammary gland development (Ding et al. 2013)
COL5A2	2	291.89	2.376	0.0072	-	Mammary gland development (Suárez-Vega et al. 2015)
CDC42	2	-	2.371	0.0073	-2.523	Milk protein (Akhtar and Streuli 2006)
CSN3	6	383.08	-	-	-2.344	Milk traits (Kucerova et al. 2006)
ADIPOQ	1	-	2.315	0.0096	-2.349	Milk traits (Venturini et al. 2014)
MAGI1	22	352.51	-		-2.294	Milk traits (Venturini et al. 2014)
PAPPA2	16	263.11	2.372	0.0073	-2.566	Milk traits (Wickramasinghe et al. 2011); Reproduction and Body size (Christians et al. 2013)
CPQ	14	273.03	2.629	0.0025	-	Reproduction (Fernández et al. 2014)
PLAG1	14	287.53	2.628	0.0019	-	Stature (Karim et al. 2011; Zhao et al. 2015)
Hanwoo breed						
PEBP4	8	66.43	1.888	0.0021	-	Meat quality (Li et al. 2016)
NCOA2	14	114.64	1.753	0.0017	-2.314	Meat quality (Wang et al. 2008)
MTMR7	27	84.56	1.584	0.0084	-	Meat traits (Ramayo-Caldas et al. 2012)
ATP10B	7	96.84	1.612	0.0074	-	Meat traits (Esteve-Codina et al. 2013)
ACVR1C	2	66.97	1.792	0.0032	-	Meat traits (Zappaterra et al. 2015)
PITPNB	17	96.75	1.635	0.0065	-2.335	Meat traits (Zheng et al. 2016)
BOK	3	68.92	1.645	0.0062	-	Reproduction (Hsu et al. 1997)
MS4A13	15	69.33	1.626	0.0069	-2.323	Reproduction (Turner et al. 2008)
C1orf116	16	122.58	1.818	0.0030	-	Reproduction (Zhou 2010)
N'Dama breed						
SLC40A1	2	471.63	2.184	0.0097	-	Iron overload (Chen et al. 2015)
CLCA4	3	295.31	2.582	0.0012	-2.389	-
KCNIP4	6	232.72	2.664	0.0008	-2.292	-
LCORL	6	-	2.106	0.0062	-2.226	Body size (Metzger et al. 2013)
STOM	8	247.27	2.293	0.0059	-	Anemia (Yokoyama 2010)
COMMD1	11	296.42	2.644	0.0009	-2.487	Copper metabolism (Sommerhalter et al. 2007)
ZC3H7A	25	227.36	2.225	0.0078	-2.166	Immunity (Liang et al. 2008)

Hanwoo breed selected genes

From Hanwoo selected genes, histone acetyltransferase activity and long-term synaptic potentiation functional annotation clusters were enriched (Table 3.3). Genes involved in histone modification contribute to myogenesis (Zhang 2016), which affects meat quality and quantity (Mozdziak 2006). *NCOA2* and *PITPNB* genes, detected by the three methods (Table 3.4), are involved in fatty acid metabolism. *NCOA2* functions as a transcriptional coactivator for nuclear hormone receptors including steroid, thyroid, retinoid, and vitamin D receptors. It is found to have a concordant effect on lipid metabolism in mammals with a positive association with intramuscular fat in the longissimus dorsi muscle of pigs (Wang et al. 2008). It modulates lipid metabolism and controls energy homeostasis in pigs, found expressed in the liver and adipose tissue of extreme fat-depositing Iberian pigs (Ramayo-Caldas et al. 2014). *PITPNB* is a protein that transfers phospholipids between membranes and is related to body fat and body weight in chicken (Jennen et al. 2004). It is involved in intracellular fatty acid movement and its abundance has been associated with fat deposition in chicken breeds (Zheng et al. 2016).

Other genes detected by two methods that affect meat quality traits in Hanwoo cattle include *PEBP4, MTMR7, ACVR1C,* and *ATP10B. PEBP4* has a pivotal biologic function of lipid binding and inhibition of serine proteases. It is found differentially expressed in the Psoas Major muscle of beef cattle (Moreno-Sánchez et al. 2010), and positively selected in pigs in relation to meat quality traits (Li et al. 2016). *MTMR7* is a phosphatase involved in lipid and carbohydrate metabolism (Sibut et al. 2011) and in the oxidation of lipids and palmitic fatty acid (Ramayo-Caldas et al. 2012). *ACVR1C* is well known for its role in energy and lipid metabolism in porcine muscle and back fat tissue (Zappaterra et al. 2015). It has been previously reported to be positively selected in Hanwoo cattle (Porto-Neto et al. 2014). *ATP10B* is involved in lipid metabolism and fat cell differentiation (Esteve-Codina et al. 2013). *PEBP4, MTMR7,* and *ATP10B* genes overlapped with meat and carcass traits QTL regions. Hanwoo

cattle is known for its high-quality beef with high marbling; it has been intensively artificially selected for beef quality traits since the last four decades (Lee et al. 2014b; Porto-Neto et al. 2014).

Genes involved in reproduction function detected at least by two methods include *MS4A13, BOK,* and *C1orf116. MS4A13*, a gene expressed in testes, is critical to male reproductive success, affecting spermatogenesis, sperm competition, and sperm-egg interaction (Turner et al. 2008). *BOK* is a pro-apoptotic Bcl-2 protein to which its expression is restricted in reproductive organs (Hsu et al. 1997). *C1orf116* is an androgen-specific receptor that plays important roles in male and female reproductive development and function (Zhou 2010). A significant missense variant (rs133059945) was identified in the *C1orf116* gene region (Table 3.5; Figure 3.4c).

Genes involved in the nervous system *(DOCK3, CIC,* and *PAFAH1B3)* and immune system development (*RFX4, ITFG1*) were also detected. *PAFAH1B3,* Platelet-activating factor acetylhydrolase IB subunit gamma, is involved in biological functions of nervous system development and spermatogenesis. *RFX4* is among the regulatory factor X (RFX) gene family transcription factors that are expressed in the testis and brain (Aftab et al. 2008).

N'Dama breed selected genes

From N'Dama cattle positively selected genes, four significant functional annotation clusters were enriched (Table 3.3). Renin secretion is a pathway involved in the regulation of water and electrolyte balance in the body. The renin-angiotensin system has been hypothesized to contribute to arid environmental condition adaptation of animals (Ali et al. 2012). The activation of genes involved in this pathway *(CLCA1, CLCA2,* and *CLCA4)* plays a role in regulating smooth muscle tone, epithelial secretion, and vertebrate olfactory transduction (Piirsoo et al. 2009). In relation to thermotolerance,

protein phosphatase genes *(PPP3CA, PPP2R5E,* and *PPP4R3B),* that regulate phosphorylation were either detected by at least two methods or involved in the enriched functional annotation clusters. Protein phosphatase genes are involved in stress response (Verghese et al. 2012; Bahbahani et al. 2015). *PPP3CA* has a role in the calmodulin activation of calcineurin and dephosphorylation of *HSPB1*, a gene involved in stress resistance and actin organization. *PPP3CA* have been identified associated with sexual precocity in Nellore cattle (Dias et al. 2015). N'Dama cattle are heat tolerant and do well in harsh environments of West Africa (Kahoun 1971).

Zinc is crucial for normal development and function of body cells mediating non-specific immunity such as neutrophils and natural killer cells. *ZC3H7A*, detected by the three methods (Table 3.4), is a family of CCCH zinc finger proteins which are critical regulators of immunity and inflammatory responses (Liang et al. 2008). It has been found enriched in macrophage-related organs such as thymus, spleen, lung, intestine, and adipose tissue (Liang et al. 2008). Serum zinc levels have been proposed to influence susceptibility or resistance of West African cattle to trypanosomiasis that elevated levels of zinc depresses the stimulation of bovine T cells by trypanosomes in vitro and inhibit antigen presentation by macrophages. In comparison to trypanosusceptible animals, resistant animals showed significantly lower serum Zinc levels (Traoré-Leroux et al. 1985). Serum cholesterol and triglyceride levels have also been reported to have a role in trypanosome growth and differentiation in African trypanotolerant N'Dama cattle that resistant N'Dama showed lower cholesterol and triglyceride levels (Ogunsanmi et al. 2000). Zinc finger genes (*ZC3H7A*) are associated with the activation of bone marrow-derived macrophages by lipopolysaccharide (Liang et al. 2008).

Other genes identified by at least two methods in relation to trypanotolerance include *STOM*, *SLC40A1*, and *COMMD1* (Table 3.4). N'Dama cattle, known for its trypanotolerance, resist trypanosomiasis in two main ways: regulating parasite population expansion and resisting anemia (Mattioli et al. 2000). Genes involved in either

of the mechanisms might contribute to the superior trypanotolerance ability of N'Dama cattle. *STOM* is a major integral membrane protein of human erythrocytes, the absence of which is associated with stomatocytosis, a form of hemolytic anemia (Yokoyama 2010). *SLC40A1* is a cell membrane protein that is involved in iron export from duodenal epithelial cells. It encodes ferroportin that expresses macrophages and delivers iron to hepatocytes (Theurl et al. 2016), and its polymorphism is associated with hemochromatosis – a condition caused by the accumulation of iron in the body (Chen et al. 2015). The importance of genetic variants in *STOM* and *SLC40A1* genes for N'Dama trypanotolerance has been previously reported (Kim et al. 2017a). *COMMD1* is a protein that plays a role in NF-kappaB signaling, sodium transport, and copper metabolism (Sommerhalter et al. 2007). NF-kappaB signaling influences innate and adaptive immunity, inflammation, B-cell development, lymphoid organogenesis, and stress response. Malfunctioning of this gene is associated with marked copper accumulation in the hepatocytes (Smedley et al. 2009). Copper is involved in many enzymatic and metabolic process including absorption and transportation of iron (Da Silva et al. 2009). Searching for variants that change protein functions, two N'Dama-specific significant missense variants - rs210601429 (p.Glu4Gln) and rs208532801 (p.Ala52Glu) were identified in *COMMD1* gene region (Table 3.5). Both of the variants are almost (95 %) fixed in N'Dama cattle as opposed to Holstein and Hanwoo breeds to which the reference alleles are maintained (Figure 3.4d). These genes might contribute to the trypanotolerance mechanisms of N'Dama cattle helping them to control the pathogenic effects of trypanosomosis (Berthier et al. 2015).

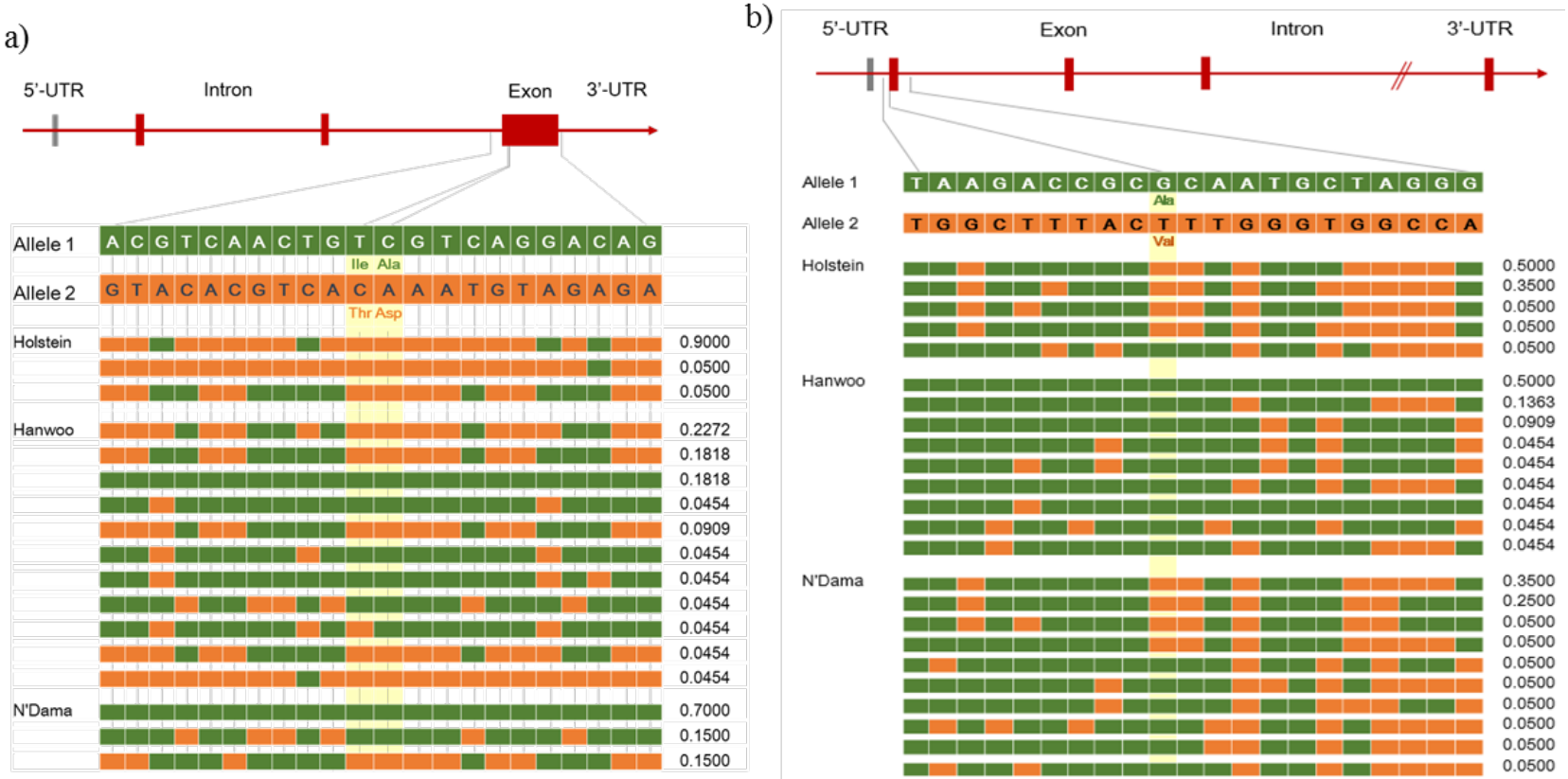
LCORL gene, detected by two methods, has been found to be related to stature and body size in cattle (Pryce et al. 2011; Xu et al. 2015) and height in humans (Horikoshi et al. 2013). Mutation in *LCORL* gene is reported to be responsible for size variation in Horses (Metzger et al. 2013) and pigs (Rubin et al. 2012). N'Dama cattle are compact with short legs of fine bone and small body size and produce a small

amount of milk relative to other breeds (Rege and Tawah 1999). The harsh environmental conditions (high disease, high humidity, and temperature, low quality and quantity of feed etc.) might be the possible selective pressure for the small body size of N'Dama cattle (Hansen 2004). In addition, the smaller body size of N'Dama cattle could be a selective advantage for its superior trypanotolerance that certain tsetse species have a preference for large-sized animals (Leak 1999).

Genes involved in eye development *(CRYGN, SNTB1,* and *U6atac)* were also identified in N'Dama cattle. *CRYGN* is localized to the refractive structure of vertebrate eye lenses (Graw 2009). *SNTB1* is related to myopia (eye problem) (Khor et al. 2013) and *U6atac* is found expressed in the retina of the eye (Baumgartner et al. 2015). The significance of these positively selected genes in relation to the eye in N'Dama cattle needs to be investigated.

Table 3.5 Significant breed-specific missense variants identified in candidate genes that affect the major phenotypes in Holstein, Hanwoo, and N'Dama cattle breeds. Allele 1 is considered as the ancestral allele while allele 2 is considered as the derived allele.

Breed	**Gene**	**Chr.**	**SNP ID**	**Allele 1**	**Allele 2**	**aa change**	χ^2 **P-value**
Holstein	*CSN3*	6	6:87390576	T	C	p.Ile157Thr	7.96E-06
		6	rs43703016	C	A	p.Ala169Asp	3.88E-06
	PAPPA2	16	rs210049354	C	T	p.Ala18Val	5.39E-06
Hanwoo	*C1orf116*	16	rs133059945	C	A	p.Val82Gly	3.47E-06
N'Dama	*COMMD1*	11	rs210601429	C	G	p.Glu4Gln	2.35E-13
		11	rs208532801	A	C	p.Ala52Glu	5.12E-09


a)
5'-UTR
Intron
Exon
3'-UTR
Allele 1
A C G T C A A C T G T C G T C A G G A C A G
Ile Ala
Allele 2
G T A C A C G T C A C A A A T G T A G A G A
Thr Asp
Holstein
0.9000
0.0500
0.0500
Hanwoo
0.2272
0.1818
0.1818
0.0454
0.0909
0.0454
0.0454
0.0454
0.0454
0.0454
0.0454
N'Dama
0.7000
0.1500
0.1500
b)
5'-UTR
Exon
Intron
3'-UTR
Allele 1
T A A G A C C G C G C A A T G C T A G G G
Ala
Allele 2
T G G C T T T A C T T T G G G T G G C C A
Val
Holstein
0.5000
0.3500
0.0500
0.0500
0.0500
Hanwoo
0.5000
0.1363
0.0909
0.0454
0.0454
0.0454
0.0454
0.0454
0.0454
N'Dama
0.3500
0.2500
0.0500
0.0500
0.0500
0.0500
0.0500
0.0500
0.0500
0.0500

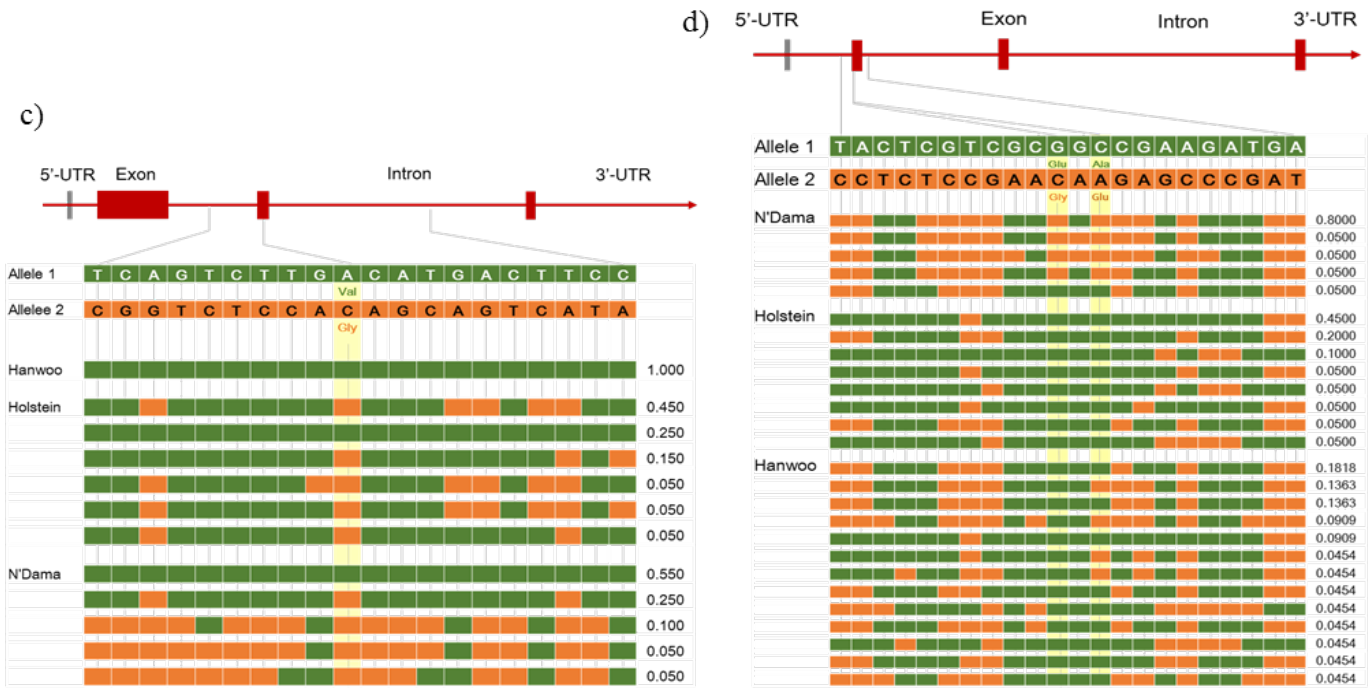


Figure 3.4 The structure of breed-specific non-synonymous variants on candidate gene regions. a) *CSN3* (6:87390576 and rs43703016), b) *PAPPA2* (rs210049354), c) *C1orf116* (rs133059945), d) *COMMD1* (rs210601429 and rs208532801) gene regions. Exons are indicated by vertical brown bars. Alleles are indicated by colored bars, the ancestral allele with green bars and the derived allele with orange bars. Breed specific significant non-synonymous SNPs indicated herein brackets for each of the genes are highlighted in yellow, the amino acid changes are indicated under the respective alleles. The frequency of each haplotype is indicated on the right side of the figure.

3.5 Conclusion

From the analysis of signature of selection, several putative selective sweep gene/genomic regions affecting the phenotypes of cattle breeds were identified. The modest overlap of genes identified between the statistical methods within breeds increases the reliability of the results. In addition, most of the genes identified from all the breeds and statistical methods were overlapped with the previously identified QTL regions which also provide additional evidence for the genes identified. Those genes that did not overlap with any QTL region could be a source of further investigation for experimental QTL analysis.

The genes found under selection are involved in different molecular functions and pathways contributing to the phenotypic differences between cattle breeds considered for the different traits of economic significance. On the other hand, other forces such as genetic drift might be the cause of genetic sweep altering the genetic structure of cattle breeds considered here. Therefore, other validation procedures are required before using the results for application in breeding and selection programs. Moreover, this result should be taken as hypothesis generating, not testing.

This chapter was published in *BMC Genetics*

in partial fulfillment of Mengistie Taye Terefe's Ph. D program

Chapter 4. Whole Genome Scan in African Ankole Cattle Breed Reveals Genetic Signature for Quality Beef

4.1 Abstract

Africa is home to numerous cattle breeds whose diversity has been shaped by subtle combinations of human and natural selection. African Sanga cattle are an intermediate type of cattle resulting from interbreeding between *Bos taurus* and *Bos indicus* subspecies. Recently, research has asserted the potential of Sanga breeds for commercial beef production with better meat quality as compared to *B. indicus* breeds. Here, I identified meat quality-related gene regions that are positively selected in Ankole (Sanga) cattle breeds as compared to *indicus* (Boran, Ogaden, and Kenana) breeds using cross-population (XP-EHH and XP-CLR) statistical methods. I identified 238 (XP-EHH) and 213 (XP-CLR) positively selected genes, of which 97 were detected from both statistics. Among the genes obtained, I primarily reported those involved in different biological process terms and pathways associated with meat quality traits. Those genes (*CAPZB, COL9A2, PDGFRA, MAP3K5, ZNF410*, and *PKM2)* involved in muscle structure and metabolism affect meat tenderness. Genes (*PLA2G2A*, *PARK2, ZNF410*, *MAP2K3*, *PLCD3*, *PLCD1*, and *ROCK1*) related to intramuscular fat (IMF) are involved in adipose metabolism and adipogenesis. *MB* and *SLC48A1* affect meat color. In addition, I identified genes (*TIMP2, PKM2, PRKG1, MAP3K5*, and *ATP8A1*) that are related to feeding efficiency. Among the enriched GO-BP terms, actin cytoskeleton organization, actin filament-based process, and protein ubiquitination are associated with meat tenderness whereas cellular component organization, negative regulation of actin filament depolymerization and negative regulation of protein complex disassembly are involved in adipocyte regulation. The MAPK pathway is responsible for cell proliferation and plays an important role in hyperplastic growth, which has a positive effect on meat tenderness. Results revealed several candidate genes positively selected in Ankole cattle in relation to meat quality characteristics. The genes identified are involved in muscle structure and metabolism, and adipose metabolism and adipogenesis. These genes help in the understanding of the biological mechanisms controlling beef quality characteristics in African Ankole cattle and provide a basis for further research on the genomic characteristics of Ankole and other Sanga cattle breeds for quality beef.

4.2 Introduction

Africa, with its diverse agro-ecological zones, is a home to diverse cattle breeds adapted to their local environments. African cattle breeds are derived from *B. taurus* and *B. indicus* subspecies introduced to the continent at different times, and through interbreeding between them (Rege 1999; Mwai et al. 2015). Since the introduction, their diversity has been shaped by subtle combinations of human and natural selection. Selection in African cattle is mainly for sociocultural concerns and to survive the heterogeneous environment (Hanotte et al. 2010). African cattle have been evolved to adapt to the poor feed availability, high environmental temperature, and high prevalence of internal and external parasite and disease conditions of the continent. These cattle breeds display better heat tolerance, adaptability, tick resistance, reproductive longevity, and maternal characteristics such as fertility, low inter-calf periods and cow efficiency (Strydom et al. 2000; Hansen 2004; Strydom 2008; Mwai et al. 2015).

African Sanga cattle, sometimes referred to as *Bos africanus*, are an intermediate type of cattle believed to be the result of interbreeding between *B. taurus* and *B. indicus*, which dwell in Eastern, Central, and Southern Africa (Rege and Tawah 1999; Mwai et al. 2015). Generally, Sanga cattle can be identified by their long and slender horns, small cervicothoracic hump, and small and unfolded dewlap (Grigson 1991). There are 30 Sanga cattle breeds/strains in Africa that can be subdivided into Sanga of eastern and Sanga of southern Africa based on their geographical distribution (Rege and Tawah 1999). Recently, research outputs are asserting the potential of African Sanga and Sanga-derived breeds to produce carcass and meat quality attributes that favorably compare to British and Continental breeds and are often better than those of the *B. indicus* breeds (Gazzola et al. 1999; Strydom et al. 2000; Strydom et al. 2008; Strydom et al. 2011; Kamatara et al. 2013). Sanga breeds in South Africa (e.g., Bonsmara, Drakensberger and Nguni) were found to produce beef with lower shear

force, shorter myofibrillar fragment length, larger rib fat thickness, and larger soluble collagen when compared with *indicus* (Brahman) cattle (Strydom et al. 2011).

Meat quality is a general term used to describe the attributes of meat which include carcass composition and conformation, the eating quality of meat, health issues associated with meat, and production and environmental issues (Maltin et al. 2003). Meat sensory characteristics such as tenderness, flavor, juiciness, and color are important meat quality parameters which are affected by biological characteristics and proteolytic activities of muscle (Mullen et al. 2006; Bernard et al. 2007). The biological characteristics of muscles such as fiber type, collagen, intramuscular adipose tissue, and protease activities regulate meat tenderness and flavor and are known to be affected by genetic and rearing factors (Bernard et al. 2007). The heritability of beef quality traits is low to moderate which varies between breed groups, methods of estimation, number of records, and other factors (Johnston et al. 2003; Rios Utrera and Van Vleck 2004). The genetic variation within and between breeds is because of the positive selection of gene regions caused by beneficial polymorphisms in the genes affecting the traits. Identification of selection signatures in the genome provides information about the evolutionary processes involved in shaping genomes and functional information about genes/genomic regions (Nielsen 2005).

Studies attempting to detect positive selection signatures in African cattle have reported several genes involved in the immune system, reproduction, energy metabolism, coat coloration, thermoregulation, and tick resistance (Flori et al. 2014; Bahbahani et al. 2015). The detection of immune-related genes might be related to the selective pressure that has been exerted by the long-term presence of pathogens in the continent (Flori et al. 2014), whereas signatures of selection associated with reproduction and thermoregulation is an adaptation to perform under heat stress conditions (Bahbahani et al. 2015). However, there have been no previous studies attempting to

identify genes affecting meat quality traits in African cattle in general and Sanga cattle in particular.

In this study, genes identified as positively selected in Ankole cattle population that are associated with meat quality traits are reported. This was done by scanning the whole genome of four African cattle breeds (African Sanga cattle: Ankole; and three *indicus* breeds: Boran, Ogaden, and Kenana). The XP-EHH and XP-CLR statistical methods were employed in order to detect selection signatures from different data patterns; the two approaches were used as each has its own advantages. XP-EHH compares haplotype lengths of populations to detect selective sweeps when the allele has approached or achieved fixation in one population but remains polymorphic in the other population (Sabeti et al. 2007). XP-CLR is a statistic based on allele frequency differentiation across populations. It is not affected by ascertainment biases and has the advantage of being able to detect older signals and selection on standing variation (Chen et al. 2010).

4.3 Materials and Methods

4.3.1 Ethics statement

Blood samples from African indigenous cattle breeds were collected after consent from the local authorities and owners of the animals. No further specific permissions were required from the Ethics Committee of the International Livestock Research Institute at the time of the sampling.

4.3.2 Sample preparation and whole genome re-sequencing

The data used for this paper was obtained from a previously published paper (Kim et al. 2017a). DNA extracted from whole blood samples (10 ml) taken from four African cattle breeds (10 Ankole, 9 Boran, 9 Ogaden and 10 Kenana) was used for this analysis. G-DEXTMIIb Genomic DNA Extraction Kit (iNtRoN Biotechnology, Seoul, Korea) was used to isolate DNA according to the manufacturer's protocol. To generate inserts of ~300 bp, 3 µg of genomic DNA was randomly sheared using Covaris System. Using the TruSeq DNA Sample Prep. Kit (Illumina, San Diego, CA), the library was constructed following the manufacturer's guidelines and whole genome sequencing was performed using the Illumina HiSeq 2000 platform. To check the quality of the raw sequence data, fastQC software (Andrews 2010) was used. Pair-end sequence reads were mapped to the reference bovine genome (UMD 3.1) using Bowtie2 (Langmead and Salzberg 2012) with default parameters except the "--no-mixed" option. The overall alignment rate of reads to the reference sequence was 98.50% with an average read depth of 10.8×. On average across the whole samples, the reads covered 98.51% of the genome.

Open source software packages of Picard tools (http://picard.sourceforge.net), SAMtools (Li et al. 2009), and Genome Analysis ToolKit 1.4 (GATK) (McKenna et al. 2010) were used for downstream processing and variant calling. Picard tools was used to filter potential PCR duplicates. SAMtools was used to create index files for reference and bam files. Genome analysis toolkit 1.4 performed local realignment of reads to correct misalignments due to the presence of indels ("RealignerTargetCreator" and "IndelRealigner" arguments). The options of "UnifiedGenotyper" and "SelectVariants" arguments of GATK were used to call candidate SNPs. To filter variants and avoid possible false positives, the "VariantFiltration" argument of the same software was adopted with the following options: 1) SNPs with a phred-scaled quality score of less than 30 were filtered; 2) SNPs with MQ0 (mapping quality zero; total count across all samples of mapping quality zero reads) > 4 and quality depth (unfiltered depth of

non-reference samples; low scores are indicative of false positives and artifacts) < 5 were filtered; and 3) SNPs with FS (Phred-scaled P-value using Fisher's exact test) > 200 were filtered since FS represents variation on either the forward or the reverse strand, which is indicative of false positive calls. BEAGLE (Browning and Browning 2007) was used to infer the haplotype phase and impute missing alleles for the entire set of cattle populations simultaneously. After all the filtering processes, a total of ~37 million SNPs were retained and used for further analysis.

4.3.3 Phylogenetic construction

To understand the genetic distance between the breeds considered, a phylogenomic analysis was conducted using neighbor-joining (NJ) and maximum likelihood (ML) methods. A total of 26,427,196 autosomal SNPs from the genomes of 38 individuals of four breeds were used for the phylogenic tree construction.

ML analyses (Felsenstein 1981) were performed using the program TREE-PUZZLE 5.2 (Schmidt et al. 2002) with the GTR model. For the quartet puzzling method (1,000 puzzling steps), nucleotide frequencies and Ts/Tv ratios (3.18) were estimated from the dataset. Quartet puzzling provided reliability values for maximum likelihood analysis (Strimmer and Von Haeseler 1996).

NJ analysis (Saitou and Nei 1987) was performed using the PHYLIP package 3.69 (Felsenstein 1993) based on Kimura's (Kimura 1980) 2-parameter distance. Ts/Tv ratios (3.18) were estimated from the dataset using TREE-PUZZLE 5.2 (Schmidt et al. 2002) and were used as inputs for the SEQBOOT, DNADIST, NEIGHBOUR, and CONSENS programs of the PHYLIP package. A bootstrap test (with 1,000 pseudoreplicates) (Felsenstein 1985) was performed to obtain statistical support for each node of the NJ tree.

4.3.4 Detection of positive selection signals

To detect genome-wide selective sweep regions in the genome, XP-EHH (Sabeti et al. 2007) and XP-CLR (Chen et al. 2010) statistical methods were used. XP-EHH assesses haplotype differences between two populations and is designed to detect alleles that have increased in frequency to the point of fixation or near fixation in one of the two populations being compared (Sabeti et al. 2007; Pickrell et al. 2009).

The genome from Ankole cattle (used as a test population) was compared with *indicus* cattle breeds (Boran, Ogaden, and Kenana grouped into one population), used as a reference population. XP-EHH compares the integrated EHH between two populations for each SNP and the sign of the XP-EHH score determines the direction of selection with extreme values indicating selection in the test population genome. To facilitate comparison of genomic regions across populations, the genome was split into non-overlapping segments of 50 kb and the maximum XP-EHH score was computed for each segment. In order to define the empirical *P*-value, genomic windows were binned in increments of 500 SNPs (combining all windows ≥ 1000 SNPs into one) according to the method used previously (Pickrell et al. 2009). Regions with *P*-values less than 0.01 (1%) were considered strong signals in the Ankole population.

XP-CLR was also performed to identify potential regions differentially selected between the two populations (Chen et al. 2010). XP-CLR is a likelihood method for detecting selective sweeps that involve jointly modeling the multilocus allele frequency differentiation between two populations. XP-CLR scores were calculated using XP-CLR software package (Chen et al. 2010). To calculate XP-CLR scores, non-overlapping sliding windows of 50 kb, a maximum number of 600 SNPs within each window were used and correlation level from which the SNPs contribution to XP-CLR result was down-weighted to 0.95. The regions with the XP-CLR values in

the top 1% of the empirical distribution (XP-CLR > 97.86) were designated as candidate sweeps and the genes that span the window regions were defined as candidate genes (Lee et al. 2014a). Significant genomic regions identified from XP-EHH and XP-CLR were annotated to the closest genes (UMD 3.1).

In order to confirm the positive selection of detected genes using XP-EHH and XP-CLR statistics, Tajima's D and F_{ST} statistics were calculated for the candidate gene regions. Detecting the same gene regions using different methods can provide cogent evidence for selective influences in the region (Qanbari and Simianer 2014). Tajima's D is used to detect selective sweeps going to fixation in the population that makes rare alleles in excess in the population, which results in a negative Tajima's D (Korneliussen et al. 2013). Population differentiation (F_{ST}) is based on the principle that natural selection can change the amount of differentiation between different populations of a species. When populations are differentiated, the amount of genetic differentiation within the region that includes selected locus will increase during when the genetic differentiation in the genomic region is greater than the level expected under neutrality, which can be a consequence of natural selection (Oleksyk et al. 2010). VCFtools was used in a window size of 50 kb at an interval of 5 kb steps to calculate the Tajima's D and F_{ST} values of the candidate gene regions (Danecek et al. 2011).

4.3.5 Characterization of candidate genes under selection

The Database for Annotation, Visualization, and Integrated Discovery (DAVID; version 6.7) gene ontology and annotation tool for gene enrichment analysis was used to further understand the biological functions and pathways of selected genes (Huang et al. 2009). Significant GO terms provide insight into the functional characteristics of annotated genes. The KEGG database was also cross-referenced within DAVID to identify significant pathways. R software (version 3.2.1) was used for hierarchical clustering of GO terms from DAVID. Additionally, Cytoscape software's (version

3.2.0), ClueGO plugin, was used to visualize the integration of GO terms as well as KEGG pathways and create a functionally organized GO/pathway term network (Bindea et al. 2009) with default settings.

4.4 Results and Discussion

4.4.1 Data description

DNA samples extracted from whole blood samples of four African cattle breeds (Boran, Ogaden, Kenana, and Ankole) were sequenced to ~ 11 X genome coverage each. Using a standard sample preparation and whole genome re-sequencing pipeline, an overall alignment rate of 98.84% covering 98.56% of the taurine reference genome was obtained. After filtering false positive calls using several filtering steps, a total of ~37 million SNPs were retained and used for detection of positive selection signature analysis.

4.4.2 Phylogenetic tree

Maximum likelihood (ML) and neighbor-joining (NJ) methods produced consistent features regarding the genetic distance between the breeds considered (Figure 4.1). Ankole cattle are clearly separated from the three *indicus* breeds (100% bootstrap values/quartet puzzling reliability values). Within *indicus* cattle, each of the three breeds was also depicted as a monophyletic group with highly significant values. Ankole cattle, a result of interbreeding between *B. taurus* and *B. indicus* like other Sanga cattle of Africa, are Sanga cattle of east and central Africa (Rege and Tawah 1999; Mwai et al. 2015).

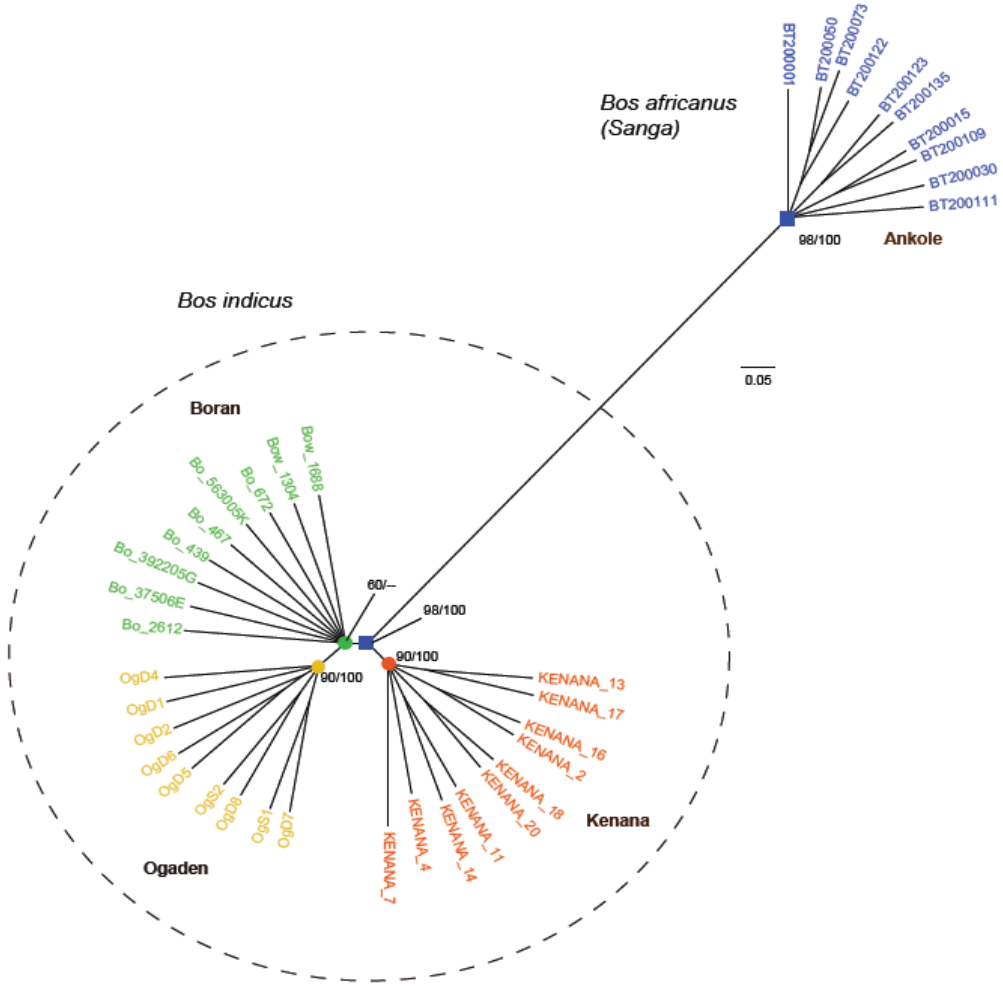


Figure 4.1 Maximum likelihood phylogenomic tree derived from autosomal SNPs of 38 African cattle individuals. The dataset (26,427,196 base pairs) was analyzed with maximum likelihood (ML) and neighbor-joining (NJ) methods which revealed identical topologies. The robustness of the phylogenomic analysis is indicated to the respective nodes: left numbers are bootstrap values for ML tree and right ones are quartet puzzling reliability values for NJ tree.

4.4.3 Positive selective signature in Ankole cattle population

XP-EHH and XP-CLR tests were performed in order to detect positive selection signatures in Sanga (Ankole) cattle. The genome of Ankole population was compared with the genomes of three *indicus* cattle breeds grouped together into one population. Based on the analysis, 238 and 213 putatively advantageous positively selected genes were identified by XP-EHH (Table 4.1) and XP-CLR test statistics (Table 4.2), respectively; of these, 98 genes were detected in both statistics. Gene Ontology Biological Processes (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways within DAVID were used to build on biological modules consisting of clusters of functional terms (Huang et al. 2009). All the 353 genes obtained from both XP-EHH and XP-CLR statistics were included, after removing duplicates, for the analysis. Gene ontology analysis resulted in 44 significantly ($p<0.05$) enriched GO-BP categories (Figure 4.2) and the KEGG-pathway analysis resulted in three significantly enriched pathways ($p<0.05$; Table 4.3). The ClueGO plugin (Bindea et al. 2009) created a functionally organized pathway term networks (Figure 4.3).

The Tajima's D calculated for the candidate gene regions revealed a significant departure from neutrality and indicated the selective maintenance of alleles within the Ankole population as compared to its *indicus* counterparts (Table 4.4). The negative Tajima's D values obtained for the candidate gene regions indicate the presence of an excess of rare alleles in the population. It is known that low-frequency alleles contribute less to the number of pair-wise differences in a sample set than alleles of moderate frequency do; a surplus of rare alleles inflates the latter value disproportionately to the former value (Korneliussen et al. 2013). Similarly, population differentiation analysis supported the positive selection of candidate genes (Table 4.4); candidate gene regions produced higher values of fixation index (Oleksyk et al. 2010). F_{ST} has been widely used to identify selective sweep regions in different livestock species (Qanbari and Simianer 2014). The Tajima's D and F_{ST} plot of candidate gene regions are presented in Figure 4.4.

Table 4.1 Summary of genes identified under selection from the genome of Ankole cattle detected by XP-EHH test statistics

Genes	Chr.	Window (Mbp)	SNP	Max XP-EHH	P-value
GRIP1	5	47.55-47.6	785	3.15	3.00E-05
GSX2	6	71.25-71.3	625	2.99	8.00E-05
U6	20	3.25-3.3	701	2.93	1.00E-04
SORCS1	26	28.35-28.4	778	2.65	2.00E-04
FAU,GRIP1,HELB	5	47.7-47.75	769	2.65	2.00E-04
PDGFRA	6	71.35-71.4	601	2.61	3.00E-04
GRIP1	5	47.65-47.7	924	2.56	3.00E-04
KCNIP1	20	2.25-2.3	956	2.55	3.00E-04
GPR161,IQWD1	3	0.55-0.6	851	2.51	4.00E-04
EVC2,U6	6	105.35-105.4	827	2.48	4.00E-04
SLIT3,bta-mir-218-2	20	0.4-0.45	1032	2.81	5.00E-04
HELB,IRAKM	5	47.75-47.8	664	2.44	5.00E-04
PDGFRA	6	71.4-71.45	858	2.42	6.00E-04
CHIC2	6	71.2-71.25	492	2.38	6.00E-04
SCML4	9	42.85-42.9	410	2.24	7.00E-04
CCDC170	9	89.7-89.75	646	2.36	8.00E-04
SLIT3	20	0.35-0.4	849	2.35	8.00E-04
FGF18	20	3.15-3.2	631	2.33	8.00E-04
CXCL13	6	94-94.05	641	2.33	9.00E-04
SLIT3	20	0.3-0.35	975	2.32	9.00E-04
FUT9	9	54.4-54.45	463	2.24	9.00E-04
RTKN2	28	18.4-18.45	763	2.32	9.00E-04
RTKN2	28	18.45-18.5	890	2.31	1.00E-03
DOCK2	20	1.6-1.65	928	2.3	1.00E-03
PANK3	20	0.2-0.25	700	2.3	1.00E-03
DOCK2,FAM196B	20	1.65-1.7	951	2.27	1.00E-03
CHIC2	6	71.15-71.2	491	2.18	1.00E-03
BAALC	14	63.5-63.55	947	2.26	1.00E-03
ACVR1B,GRASP	5	28-28.05	430	2.18	1.00E-03
GRIP1,U1	5	47.6-47.65	772	2.25	1.00E-03

ZCWPW2	22	2.9-2.95	926	2.24	1.00E-03
SYNE1	9	90.3-90.35	734	2.24	1.00E-03
AHSA1,ISM2, VIPAS39	10	89.7-89.75	637	2.23	2.00E-03
RPL37A	2	105.2-105.25	511	2.22	2.00E-03
MTHFD1L	9	89.2-89.25	855	2.22	2.00E-03
ROCK1,USP14	24	35.5-35.55	520	2.22	2.00E-03
FAIM,PIK3CB	1	131.5-131.55	645	2.21	2.00E-03
ZNF365	28	18.55-18.6	828	2.2	2.00E-03
FOXO3	9	42-42.05	483	2.03	2.00E-03
GRXCR1	6	63.35-63.4	842	2.19	2.00E-03
LSM3	22	58.65-58.7	881	2.18	2.00E-03
TANC1	2	37.4-37.45	837	2.18	2.00E-03
PTGR2,ZNF410	10	85.65-85.7	558	2.18	2.00E-03
PCOLCE2	1	127.15-127.2	869	2.18	2.00E-03
SLC6A6	22	58.45-58.5	723	2.17	2.00E-03
CSMD3	14	52.95-53	628	2.17	2.00E-03
ESR1	9	89.95-90	463	1.99	2.00E-03
RTKN2	28	18.35-18.4	753	2.16	2.00E-03
ATG101,NR4A1	5	27.95-28	404	1.98	2.00E-03
SLIT3	20	0.5-0.55	865	2.15	2.00E-03
MITF	22	31.7-31.75	535	2.15	2.00E-03
PANK3, SPZ1, bta-mir-103-1	20	0.15-0.2	645	2.15	2.00E-03
PARK2	9	98.9-98.95	1175	2.34	2.00E-03
SLIT3	20	0.45-0.5	835	2.15	2.00E-03
PARP6,PKM2	10	18.95-19	358	1.97	2.00E-03
IYD	9	88.55-88.6	770	2.14	2.00E-03
ACVRL1, ANKRD33	5	28.1-28.15	452	1.97	3.00E-03
GLRX3, bta-mir-2397	26	49.7-49.75	781	2.14	3.00E-03
RASGEF1C	7	1-1.05	932	2.13	3.00E-03
NRXN3	10	92.15-92.2	752	2.13	3.00E-03
BAALC	14	63.45-63.5	788	2.12	3.00E-03
KCNIP1	20	2.3-2.35	1014	2.31	3.00E-03
CITED4, KCNQ4	3	105.95-106	709	2.12	3.00E-03
CTDSPL,P LCD1, VILL, bta-mir-26a-1	22	11.45-11.5	638	2.12	3.00E-03
SPTLC2	10	89.8-89.85	567	2.12	3.00E-03
CENPF	16	70.4-70.45	824	2.11	3.00E-03

GLRX3	26	49.75-49.8	687	2.11	3.00E-03
CNGA1	6	68.3-68.35	657	2.11	3.00E-03
RANBP17	20	2.65-2.7	480	1.96	3.00E-03
IL20RB	1	133.25-133.3	879	2.1	3.00E-03
SPTLC2	10	89.75-89.8	783	2.1	3.00E-03
DPYD,U6	3	45.75-45.8	784	2.1	3.00E-03
SORCS1	26	28.3-28.35	795	2.1	3.00E-03
DDX6	15	29.85-29.9	596	2.1	3.00E-03
RANBP17	20	2.7-2.75	410	1.96	3.00E-03
VIT	11	19.2-19.25	854	2.09	3.00E-03
ADAM2	27	34.3-34.35	977	2.09	3.00E-03
MAP3K5	9	75.55-75.6	596	2.09	3.00E-03
GRXCR1	6	63.4-63.45	728	2.08	3.00E-03
FMNL1,HEXIM2	19	45.45-45.5	371	1.95	3.00E-03
PPP1R14C	9	88.5-88.55	916	2.07	3.00E-03
IQWD1	3	0.6-0.65	660	2.07	3.00E-03
LCP2	20	2-2.05	1056	2.19	3.00E-03
OR10A4,OR2D2, OR2D3, ZNF215	15	46.4-46.45	869	2.07	3.00E-03
C7H5ORF15, VDAC1	7	47.2-47.25	603	2.06	4.00E-03
KCNK6	18	48.35-48.4	217	1.94	4.00E-03
BTRC	26	22-22.05	555	2.06	4.00E-03
ADD1,U6	6	107.95-108	692	2.05	4.00E-03
GRAMD2	10	18.9-18.95	487	1.94	4.00E-03
ISPD	4	24.6-24.65	891	2.05	4.00E-03
CCDC93	2	69.9-69.95	923	2.04	4.00E-03
E2F7	5	6.45-6.5	739	2.04	4.00E-03
MAP2K3	19	35.85-35.9	451	1.93	4.00E-03
ZFYVE28	6	108.4-108.45	999	2.03	4.00E-03
PARK2	9	99.25-99.3	1137	2.17	4.00E-03
TOX2	13	73.15-73.2	573	2.02	4.00E-03
ATP8A1	6	62.9-62.95	537	2.02	4.00E-03
RAET1G	9	88.3-88.35	876	2.02	4.00E-03
EIF4E3, PROK2	22	30-30.05	322	1.91	4.00E-03
CTNNA3, LRRTM3	28	23.65-23.7	937	2.01	4.00E-03
NFATC2	13	80-80.05	904	2.01	4.00E-03
PBX1	3	4.45-4.5	715	2.01	4.00E-03

SLC6A6	22	58.4-58.45	927	2.01	4.00E-03
TTLL4	2	107.4-107.45	381	1.9	4.00E-03
HERPUD1, SLC12A3	18	25-25.05	553	2.00	4.00E-03
CCER1	5	20.85-20.9	620	2.00	4.00E-03
APOL6	5	74.2-74.25	627	2.00	5.00E-03
HEXA, TMEM202	10	19.1-19.15	737	2.00	5.00E-03
GDA	8	48.7-48.75	685	1.99	5.00E-03
ESYT3	1	131.7-131.75	714	1.99	5.00E-03
U6	6	71.6-71.65	833	1.99	5.00E-03
GOLPH3	20	41.45-41.5	465	1.9	5.00E-03
FAIM2	5	30.15-30.2	588	1.98	5.00E-03
C7H19orf52, CARM1, YIPF2	7	16.55-16.6	648	1.98	5.00E-03
KCNMA1	28	33.25-33.3	786	1.98	5.00E-03
4-Mar	2	104.8-104.85	829	1.98	5.00E-03
EVC2	6	105.4-105.45	859	1.98	5.00E-03
U7	11	17.4-17.45	991	1.97	5.00E-03
ACBD4, HEXIM1, NMT1, PLCD3	19	45.4-45.45	446	1.89	5.00E-03
MB	5	74.15-74.2	736	1.96	5.00E-03
HNRNPLL,U6	11	20.95-21	494	1.88	5.00E-03
U6	24	51.3-51.35	628	1.96	5.00E-03
KCNIP1	20	2.5-2.55	1019	2.14	5.00E-03
FGF18,NPM1	20	3.1-3.15	607	1.96	5.00E-03
IQWD1, snoZ278	3	0.65-0.7	623	1.96	5.00E-03
MRPL1	6	94.4-94.45	668	1.96	5.00E-03
NFXL1	6	68.2-68.25	457	1.87	5.00E-03
DOK6	24	8-8.05	608	1.96	5.00E-03
COL9A2, SMAP2	3	106.35-106.4	786	1.95	5.00E-03
C1QL4,PRPH, TROAP	5	30.65-30.7	518	1.94	6.00E-03
SMARCAL1	2	105.1-105.15	719	1.94	6.00E-03
DOCK2	20	1.75-1.8	990	1.94	6.00E-03
5S_rRNA	14	0.35-0.4	530	1.94	6.00E-03
CAND1	5	46.7-46.75	334	1.85	6.00E-03
WDR51B	5	19.4-19.45	650	1.94	6.00E-03
GR-A	7	56.35-56.4	427	1.84	6.00E-03
FN3K, SNORA73, TBCD	19	50.5-50.55	825	1.94	6.00E-03
bta-mir-193b, bta-mir-365-1	25	13.3-13.35	622	1.93	6.00E-03

CNGA1,NIPAL1	6	68.35-68.4	436	1.84	6.00E-03
COL9A2, ZMPSTE24	3	106.4-106.45	807	1.93	6.00E-03
U6	6	71.5-71.55	799	1.93	6.00E-03
TOM1L1	19	5.2-5.25	1182	2.07	6.00E-03
FANCA, SPIRE2	18	14.65-14.7	431	1.82	6.00E-03
ATP10D, CORIN	6	67.9-67.95	393	1.82	6.00E-03
NAF1	6	2.65-2.7	658	1.92	6.00E-03
JPH2	13	73.3-73.35	838	1.92	6.00E-03
DENND1A	11	94.6-94.65	357	1.82	6.00E-03
CCDC170	9	89.75-89.8	590	1.92	6.00E-03
SFT2D2	3	0.3-0.35	818	1.92	6.00E-03
IFT57	1	53.3-53.35	503	1.92	7.00E-03
CCDC176, ENTPD5	10	85.75-85.8	463	1.82	7.00E-03
PRSS38	7	3.15-3.2	932	1.91	7.00E-03
EVC	6	105.2-105.25	1008	2.06	7.00E-03
NUDT3,RPS10	23	8.4-8.45	439	1.82	7.00E-03
GDA	8	48.75-48.8	545	1.91	7.00E-03
SLC39A11	19	58.9-58.95	1202	2.05	7.00E-03
VEGFC	27	7-7.05	557	1.9	7.00E-03
RAPGEF4	2	23.8-23.85	547	1.9	7.00E-03
INOS,ULBP27	19	19.8-19.85	953	1.9	7.00E-03
SNORA71	29	26.85-26.9	844	1.9	7.00E-03
HDAC7, RAPGEF3, SLC48A1	5	32.65-32.7	423	1.8	7.00E-03
DSCAM	1	141.95-142	984	1.9	7.00E-03
VIT	11	19.25-19.3	689	1.89	7.00E-03
TM4SF4	1	119.6-119.65	830	1.89	7.00E-03
NRXN3	10	92.2-92.25	851	1.89	7.00E-03
MYB	9	74.25-74.3	477	1.79	7.00E-03
ADCY10	3	0.85-0.9	709	1.89	7.00E-03
HECW1	4	78.3-78.35	788	1.89	7.00E-03
EPHA5	6	82.9-82.95	750	1.89	7.00E-03
RANBP17	20	2.75-2.8	402	1.79	7.00E-03
SMIM23	20	3.5-3.55	642	1.88	7.00E-03
TIMP2,USP36	19	54.1-54.15	864	1.88	7.00E-03
KCNK2	16	70.05-70.1	724	1.88	7.00E-03
CASR	1	67.25-67.3	709	1.88	8.00E-03

PPM1L	1	107.5-107.55	493	1.78	8.00E-03
PIK3CB	1	131.4-131.45	519	1.88	8.00E-03
PARK2	9	99.2-99.25	1100	2.03	8.00E-03
SORCS1	26	28.2-28.25	697	1.88	8.00E-03
CNTN1	5	40.2-40.25	679	1.87	8.00E-03
GRIN2B	5	96.75-96.8	706	1.87	8.00E-03
KCNMA1	28	33.3-33.35	867	1.87	8.00E-03
CHD7	14	28.1-28.15	665	1.87	8.00E-03
RYBP	22	29.3-29.35	393	1.76	8.00E-03
RAPGEF4	2	23.75-23.8	735	1.87	8.00E-03
DNAJC22	5	30.6-30.65	566	1.86	8.00E-03
AGFG1	2	116.4-116.45	698	1.86	8.00E-03
KCNMB1	20	2.15-2.2	833	1.86	8.00E-03
7SK	27	11.9-11.95	535	1.86	8.00E-03
DPH6	10	30.95-31	475	1.76	8.00E-03
ATP8A1	6	63.1-63.15	708	1.86	8.00E-03
MAP3K5, MGC151537	9	75.5-75.55	602	1.86	8.00E-03
NXPH2	2	58.7-58.75	874	1.86	8.00E-03
RPS26	6	16.65-16.7	880	1.86	8.00E-03
CORIN	6	68.05-68.1	381	1.76	8.00E-03
TMEM106B	4	19.9-19.95	820	1.85	8.00E-03
FRMD3	8	77.9-77.95	848	1.85	8.00E-03
MAPKAPK2	16	4.3-4.35	642	1.85	8.00E-03
SCML4	9	42.8-42.85	567	1.85	8.00E-03
EVC,EVC2	6	105.25-105.3	988	1.85	9.00E-03
ITPR2	5	83.45-83.5	799	1.85	9.00E-03
CENPF	16	70.45-70.5	498	1.74	9.00E-03
PLEKHM1	19	45.75-45.8	791	1.85	9.00E-03
MAP4K4	11	6.65-6.7	774	1.85	9.00E-03
COQ6, ENTPD5, FAM161B, ZNF410	10	85.7-85.75	519	1.85	9.00E-03
ARIH1, TMEM202	10	19.15-19.2	475	1.74	9.00E-03
MAP3K5	9	75.6-75.65	796	1.85	9.00E-03
SULT1C4	9	34.5-34.55	852	1.85	9.00E-03
MYO7A	15	57.35-57.4	830	1.84	9.00E-03
TTC39C	24	32.95-33	972	1.84	9.00E-03
LPP,bta-mir-28	1	79.25-79.3	445	1.73	9.00E-03

COLEC12, THOC1, bta-mir-544b-2	24	35.6-35.65	705	1.84	9.00E-03
SLC37A2	29	28.85-28.9	625	1.84	9.00E-03
ZCWPW2	22	2.95-3	929	1.84	9.00E-03
CA10	19	1.15-1.2	750	1.84	9.00E-03
OR10A4	15	46.45-46.5	985	1.84	9.00E-03
SMARCAL1	2	105.15-105.2	508	1.84	9.00E-03
ACVR1B, ACVRL1	5	28.05-28.1	352	1.73	9.00E-03
PIK3CB	1	131.45-131.5	620	1.83	9.00E-03
ASNS	4	15-15.05	844	1.83	9.00E-03
NT5DC1	9	34.9-34.95	863	1.83	9.00E-03
PLEKHG1	9	88.95-89	652	1.83	9.00E-03
CD86	1	67.2-67.25	481	1.73	9.00E-03
PREP	9	45.25-45.3	569	1.83	9.00E-03
ADD1, SH3BP2	6	108-108.05	578	1.83	9.00E-03
KCNMA1	28	33.4-33.45	983	1.83	9.00E-03
C3	7	19-19.05	443	1.72	9.00E-03
HHLA2, MYH15	1	53.5-53.55	604	1.83	1.00E-02
ECHDC3	13	12.55-12.6	789	1.83	1.00E-02
WWOX	18	5.7-5.75	952	1.83	1.00E-02
TTC3	1	151.05-151.1	573	1.83	1.00E-02
EPHA5	6	82.95-83	597	1.83	1.00E-02
CABS1	6	87.45-87.5	682	1.83	1.00E-02
EPHB1	1	135.15-135.2	771	1.82	1.00E-02
AXDND1, NPHS2	16	62.2-62.25	701	1.82	1.00E-02
SPIRE2,TCF25	18	14.7-14.75	429	1.71	1.00E-02
PPP1R14C, RAET1G	9	88.35-88.4	886	1.82	1.00E-02

Table 4.2 Summary of genes identified under selection from the genome of Ankole cattle detected by XP-CLR test statistics

Genes in XP-CLR regions	Chr.	Window (Mbp)	SNPs	XP-CLR
BAALC	14	63.48-63.53	600	628.14
ZCWPW2	22	2.93-2.98	600	578.16
KCNIP1	20	2.28-2.33	600	540.99
FAM136A, XDH	11	14.13-14.18	600	469.61
IYD	9	88.53-88.58	600	449.00
PDGFRA	6	71.33-71.38	516	408.60
U7	11	17.38-17.43	600	377.06
SLIT3, bta-mir-218-2	20	0.43-0.48	600	369.01
GSX2	6	71.28-71.33	540	363.31
FANCL	11	40.73-40.78	600	355.80
U6	6	71.53-71.58	600	353.36
EVC	6	105.23-105.28	600	326.46
BYSL, CCND3, MED20	23	15.68-15.73	559	325.34
GPR161, IQWD1	3	0.58-0.63	600	324.64
SYNE1	9	90.28-90.33	600	291.68
NPM1	20	3.08-3.13	588	287.52
SLIT3	20	0.33-0.38	600	279.92
CHIC2	6	71.23-71.28	308	269.92
ATP8A1	6	63.13-63.18	600	264.27
ATP8A1, SHISA3	6	62.88-62.93	476	258.76
ISM2, SPTLC2	10	89.73-89.78	600	249.22
SLIT3	20	0.48-0.53	600	248.08
FGF18	20	3.18-3.23	463	247.37
U6	15	46.48-46.53	600	244.76
U6	6	71.58-71.63	403	244.14
ACVRL1	5	28.08-28.13	370	243.78
SCML4	9	42.78-42.83	587	243.02
HMGA2	5	48.08-48.13	402	242.19
PPP2R2C	6	104.58-104.63	600	240.81
ANXA3, FRAS1	6	95.03-95.08	600	235.34
ISPD	4	24.53-24.58	600	234.62
DDX6	15	29.88-29.93	350	232.30
RANBP17	20	2.68-2.73	280	231.68
ATP8A1	6	62.98-63.03	441	229.03
COPS3, FLCN, NT5M	19	35.48-35.53	456	227.29

GRIP1	5	47.63-47.68	600	222.50
GRIP1, HELB	5	47.68-47.73	546	221.39
GRIP1	5	47.53-47.58	600	220.03
KHDRBS2	23	0.53-0.58	600	219.09
KCNIP1	20	2.33-2.38	600	216.12
FGF18	20	3.13-3.18	569	215.48
ADAM2	27	34.33-34.38	600	215.07
5S_rRNA, NBAS	11	82.93-82.98	598	214.91
DEF8, MC1-R, TCF25, TUBB3	18	14.73-14.78	329	213.38
GRXCR1	6	63.38-63.43	600	212.55
SORCS1	26	28.38-28.43	600	208.06
LRIT3, RPS26	6	16.68-16.73	600	206.51
PDGFRA	6	71.38-71.43	335	204.49
CNGA1, NIPAL1	6	68.33-68.38	522	201.10
XDH	11	14.18-14.23	374	199.94
NPY1R, NPY5R	6	2.38-2.43	600	199.19
HMGA2, bta-mir-763	5	48.13-48.18	274	198.69
ZNF684	3	106.18-106.23	600	198.00
FAU,HELB	5	47.73-47.78	600	197.79
PLCD1, VILL	22	11.48-11.53	545	195.30
CLIC4	2	128.78-128.83	528	195.05
HMGA2	5	48.03-48.08	307	193.58
SLIT3	20	0.28-0.33	600	192.76
U6	1	133.18-133.23	600	191.24
RTKN2	28	18.38-18.43	600	190.58
EVC2, U6	6	105.38-105.43	600	189.91
GRASP, NR4A1	5	27.98-28.03	292	189.68
CITED4	3	105.93-105.98	478	189.49
RIC3	15	44.98-45.03	600	187.14
RANBP17	20	2.88-2.93	363	185.69
CCER1, EPYC	5	20.88-20.93	502	185.64
PPP1R14C	9	88.48-88.53	569	183.09
KHDRBS2	23	0.83-0.88	577	180.22
DOCK2, FAM196B	20	1.63-1.68	600	179.52
PANK3, bta-mir-103-1	20	0.18-0.23	471	177.17
IRAKM	5	47.78-47.83	399	176.96
SOWAHA	7	45.98-46.03	600	175.21
ZNF746	4	113.28-113.33	588	173.64
CCDC60	17	58.13-58.18	553	173.62

ZNF772	18	64.83-64.88	600	171.76
C10orf67	13	24.53-24.58	506	170.81
C1QL4, DNAJC22, TROAP	5	30.63-30.68	466	168.94
TM4SF4	1	119.63-119.68	600	168.18
SLC25A13	4	13.33-13.38	600	167.85
SLC6A2	18	23.93-23.98	532	166.82
EPHA5	6	82.78-82.83	535	166.75
CSTF2T, PRKG1	26	7.43-7.48	600	166.61
COQ6, FAM161B, ZNF410	10	85.68-85.73	527	165.86
BAALC, SNORA61	14	63.43-63.48	600	165.61
SLIT3	20	0.38-0.43	600	164.20
ATG101	5	27.93-27.98	448	163.88
CHIC2	6	71.18-71.23	154	163.37
C10orf67	13	24.58-24.63	371	161.87
EPHA5	6	82.93-82.98	600	158.54
MT3,MT4	18	24.13-24.18	547	156.28
KCNQ4	3	105.98-106.03	515	156.08
FYN	9	39.08-39.13	338	154.58
TSC22D1	12	14.68-14.73	508	153.94
DYRK1B,FBL	18	49.63-49.68	568	153.77
SORCS1	26	28.23-28.28	600	153.67
SMIM23	20	3.48-3.53	571	153.60
HMGA2	5	48.18-48.23	213	153.52
C19orf38,CARM1, TMED1	7	16.53-16.58	454	152.61
DMTF1, TMEM243	4	33.33-33.38	496	152.44
KHDRBS2	23	0.68-0.73	600	152.39
FAM110B	14	26.08-26.13	543	151.29
RPL37A, SMARCAL1	2	105.18-105.23	525	150.23
TOX2	13	73.18-73.23	600	150.02
RANBP17, TLX3	20	3.03-3.08	493	149.70
M-RIP	19	35.58-35.63	401	148.55
FYN	9	39.13-39.18	459	148.46
MAP3K5	9	75.63-75.68	600	147.37
ATP8A1	6	63.03-63.08	548	146.55
STK32B	6	105.48-105.53	600	145.39
CLDN10	12	76.73-76.78	600	144.40
RIC3	15	44.93-44.98	600	143.94
URB1	1	2.38-2.43	600	143.67
CNTN6	22	25.13-25.18	600	143.44

CAPZB	2	133.78-133.83	600	142.50
ATP8A1	6	62.93-62.98	386	141.94
CXCL13	6	94.03-94.08	487	141.93
CENPF	16	70.43-70.48	543	141.60
RTKN2	28	18.43-18.48	600	141.50
ZPBP	4	5.73-5.78	600	139.76
NUBP1, TVP23A	25	9.48-9.53	600	139.30
PGM5	8	44.68-44.73	531	139.20
GDA	8	48.68-48.73	600	138.65
OTP	10	8.68-8.73	574	138.62
RAB11FIP2	26	38.63-38.68	352	138.49
ZNHIT6	3	58.58-58.63	396	137.00
GDA	8	48.73-48.78	414	136.60
ISPD	4	24.58-24.63	600	136.10
SCN8A	5	28.28-28.33	494	135.73
SCN8A	5	28.23-28.28	585	134.71
GDA	8	48.78-48.83	600	134.69
RLF	3	106.53-106.58	408	134.26
7SK	27	11.88-11.93	498	133.99
WWP1	14	78.63-78.68	600	133.90
LRP11	9	88.18-88.23	600	133.19
CAND1	5	46.68-46.73	545	133.13
PARK2	9	99.13-99.18	600	132.40
CCDC170	9	89.73-89.78	575	131.73
FANCA, SPIRE2	18	14.68-14.73	408	131.67
GRIP1,U1	5	47.58-47.63	576	130.54
ROCK1	24	35.48-35.53	333	130.40
FERMT1	13	48.63-48.68	600	130.36
MB	5	74.18-74.23	579	129.86
ZMPSTE24	3	106.43-106.48	600	129.68
DNM2, TMED1, bta-mir-3604-1	7	16.48-16.53	600	128.71
LNX1	6	70.73-70.78	600	128.67
FLCN, M-RIP, PLD6	19	35.53-35.58	483	128.29
U6	20	36.58-36.63	530	127.91
DPYD	3	45.78-45.83	600	126.60
EFCAB6	5	115.18-115.23	417	126.40
RGS22	14	66.58-66.63	596	126.02
U6	2	87.98-88.03	593	125.97
ALPK2	24	58.23-58.28	600	125.60

WBSCR28	25	33.93-33.98	411	124.42
SCN8A	5	28.18-28.23	287	124.07
DSPP	6	104.23-104.28	600	123.62
PGBD2	7	44.03-44.08	457	123.25
RANBP17	20	2.78-2.83	225	122.94
FAIM, PIK3CB	1	131.53-131.58	538	121.48
C3H3orf22,CHST13, TRXR3	22	61.08-61.13	600	120.96
PTGER4,TTC33	20	33.73-33.78	497	118.72
FRS3,PRICKLE4, TOMM6, USP49	23	15.58-15.63	410	118.11
PLA2G2A	2	133.28-133.33	600	117.48
NUDT3	23	8.33-8.38	240	117.33
LIMA1,U6	5	29.78-29.83	499	117.18
ATP8A1	6	63.08-63.13	535	116.88
CMC1	22	2.68-2.73	580	116.51
ZNHIT6	3	58.53-58.58	418	116.10
MTHFD1L	9	89.23-89.28	600	115.94
ATP2B2, bta-mir-885	22	55.03-55.08	600	115.88
DHX40	19	10.78-10.83	600	115.63
MTHFD1L	9	89.18-89.23	600	115.23
RANBP17	20	2.73-2.78	336	114.04
5S_rRNA, RAB3C	20	20.63-20.68	600	114.03
SORCS1	26	28.33-28.38	600	113.69
U6	20	3.23-3.28	469	113.59
U6	9	46.33-46.38	453	113.51
U6,VEGFC	27	6.98-7.03	410	113.27
MRPL1	6	94.38-94.43	498	112.92
MRAS	1	131.73-131.78	584	112.25
KHDRBS2	23	0.43-0.48	600	112.14
CNTN1	5	40.23-40.28	600	111.96
7SK	13	24.93-24.98	600	111.64
RAET1G	9	88.33-88.38	600	111.28
C1ORF192, SDHC	3	8.18-8.23	516	111.24
5S_rRNA	18	21.18-21.23	569	110.78
5S_rRNA	22	2.53-2.58	600	109.93
VAT1L	18	5.18-5.23	600	108.25
COX6C	14	66.63-66.68	501	108.12
CLYBL	12	80.58-80.63	499	107.76
NIPAL1, TXK	6	68.38-68.43	407	107.49
5S_rRNA	5	47.93-47.98	285	107.40

THOC1, USP14	24	35.53-35.58	557	107.25
COL9A2	3	106.38-106.43	600	107.14
BIN2, DAZAP2, SMAGP	5	28.68-28.73	463	106.30
ALDH1A2,U6	10	52.38-52.43	545	105.65
SNX29	25	11.13-11.18	600	105.62
C8H9ORF85, FAM108B1	8	48.43-48.48	448	105.30
SORCS3	26	26.38-26.43	600	105.29
FMNL3, TMBIM6	5	30.28-30.33	467	104.77
NXPH2	2	58.58-58.63	600	102.95
FAM108B1	8	48.38-48.43	600	102.82
GRHL3, STPG1	2	129.13-129.18	600	102.35
DPYD,U6	3	45.73-45.78	563	102.29
SPTLC2	10	89.78-89.83	491	102.14
FOXD4L1	8	44.63-44.68	324	102.11
SLC25A13	4	13.43-13.48	563	101.59
HDAC7, SLC48A1	5	32.63-32.68	372	101.46
MDM2	5	45.18-45.23	466	100.58
IQWD1, snoZ278	3	0.63-0.68	512	100.21
OTUD1	13	24.63-24.68	600	99.89
EVC	6	105.18-105.23	600	99.69
CD226	24	7.58-7.63	600	98.89
TTC39C	24	32.93-32.98	600	98.63
THOC1, bta-mir-544b-2	24	35.58-35.63	600	97.86

4.4.3.1 Biological process and pathways related to meat quality traits

Meat quality is a multifactorial and complex trait affected by different factors at different levels ranging from molecular to mechanical. Molecularly, genes involved in many cellular mechanisms such as muscle growth, glycolysis, muscle contraction, stress reaction, cell cycle, proteolysis, protein ubiquitination and apoptosis have been reported to be associated with meat quality characteristics (Koohmaraie et al. 2002; Mullen et al. 2006; Guillemin et al. 2011). Previous studies asserted that, as compared with *indicus* breeds, Sanga breeds produce better quality beef (Gazzola et al. 1999; Strydom et al. 2011; Kamatara et al. 2013) with lower shear force, shorter myofibrillar fragment length, larger rib fat thickness, larger soluble collagen, and higher percent drip loss (Strydom et al. 2011). Additionally, Sanga cattle have better feed conversion efficiency, reproductive performances, and tick resistance in the tropics (Schoeman 1989).

From DAVID gene ontology analysis, 44 significant ($p<0.05$) GO-BP terms were enriched (Figure 4.2). The BP terms and gene clusters related to meat quality characteristics were chosen based on their biological function and previous literature. Accordingly, among the enriched GO-BP terms (Figure 4.2), actin cytoskeleton organization (represented by nine genes - *FMNL1, FMNL3, DOCK2, LIMA1, ROCK1, MRAS, PRKG1, CAPZB,* and *ADD1*) and actin filament-based process (additionally contains *MYO7A*) are related to meat tenderness (Gao et al. 2011; Guillemin et al. 2011; Damon et al. 2012). Cellular component organization, a cellular level process which results in the assembly and arrangement of constituent parts or disassembly of a cellular component, is important for beef tenderness (Guillemin et al. 2011). It is also significantly differentially expressed in relation to pork IMF and tenderness (Hamill et al. 2012). Five genes (*WWP1, MDM2, CAND1, PARK2,* and *LNX1*) were involved in protein ubiquitination, which is a key step in protein degradation (Jiang et al. 2010). Ubiquitination pathway affects muscle properties that are relevant for the

quality of meat at postmortem (Ponsuksili et al. 2009), and are expressed in relation to tenderness (Hamill et al. 2012). GO terms of negative regulation of actin filament depolymerization and negative regulation of protein complex disassembly are involved in adipocyte regulation (Gao et al. 2011).

The MAPK KEGG pathway ($p = 0.0215$; Table 4.3), represented by eight genes (*MAP4K4*, *ACVR1B*, *FGF18*, *MAP3K5*, *MAP2K3*, *MRAS*, *PDGFRA*, *PLA2G2A*, *NR4A1*, *MAPKAPK2*, and *NFATC2*) is responsible for cell proliferation and plays an important role in hyperplastic growth (Chang 2007), which has a positive effect on meat tenderness (Koohmaraie et al. 2002). Gap junction, regulation of actin cytoskeleton and MAPK signaling pathways also are important in residual feed intake (Rolf et al. 2012). The ClueGO plugin created a functionally organized pathway term network (Figure 4.3), that the networks actin filament bundle assembly and positive regulation of proteolysis were among enriched networks in relation to meat quality characteristics (Guillemin et al. 2011).

4.4.3.2 Genes affecting meat quality traits in Ankole cattle

In this study, genes identified as positively selected in Ankole Sanga cattle that are potentially associated with meat quality and feed conversion efficiency traits are described based on previous studies and their biological functions (Table 4.4). The gene names and descriptions in this study are based on Genecards (http://www.genecards.org/).

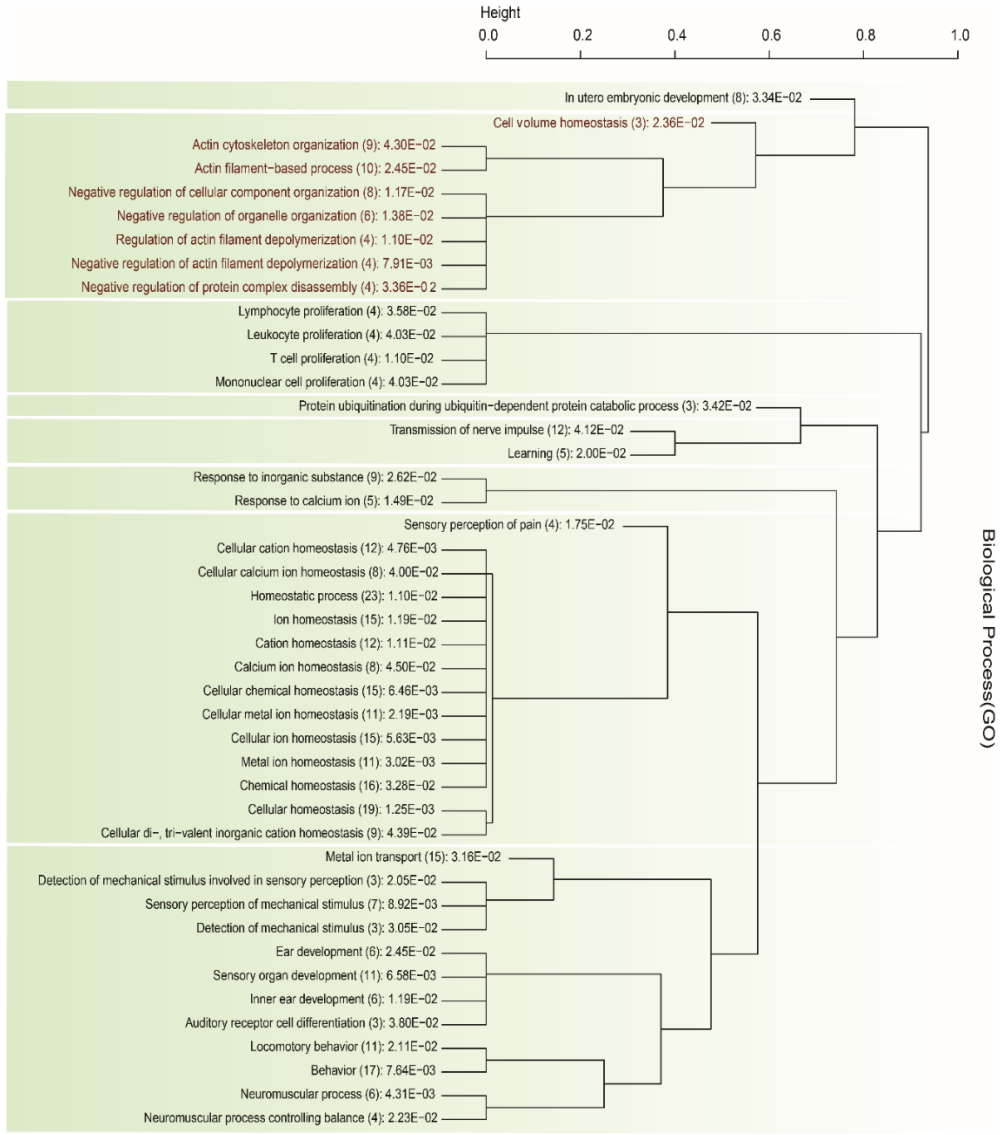


Figure 4.2 Functional clustering of GO-BP terms enriched from DAVID gene ontology analysis. All the 44 significantly ($p<0.05$) enriched BP terms were used for the functional clustering. GO-BP terms in red font color are associated with meat quality traits.

Genes related to meat tenderness

Meat tenderness is an important meat eating quality trait. It is mainly affected by the quantity and solubility of connective tissue, composition and contractile state of muscle fibers, and the extent of proteolysis in rigor muscle (Koohmaraie et al. 2002; Strydom et al. 2011; Joo et al. 2013). Tender meat contains higher levels of soluble collagen, more fat, and lower water content. Myofibril fragmentation index also has a positive correlation with beef loin tenderness (Culler et al. 1978). Sanga breeds have a lower percentage of white muscle fiber and a higher myofibrillar fragmentation index (Strydom et al. 2000; Strydom et al. 2011), which results in lower shear force and more tender beef compared to *indicus* cattle (Strydom et al. 2011). In this study, I have identified genes (*CAPZB, COL9A2, PDGFRA, MAP3K5, ZNF410, LIMA1,* and *PKM2*) that may potentially affect muscle structure and development thereby affecting meat tenderness in Ankole cattle.

Table 4.3 KEGG pathways obtained from DAVID gene enrichment ($p<0.05$) analysis of genes identified under selection in the genome of Ankole cattle. All the genes (354 genes) obtained from both XP-EHH and XP-CLR statistics were used after removing duplicate genes

KEGG pathway term	P-value	Genes	Fold Enrichment
Leukocyte trans-endothelial migration	0.019	*ROCK1, PIK3CB, CLDN10, TXK, RAPGEF4, RAPGEF3, CTNNA3*	3.24
MAPK signaling pathway	0.021	*MAP4K4, ACVR1B, FGF18, MAP3K5, MAP2K3, MRAS, PDGFRA, PLA2G2A, NR4A1, MAPKAPK2, NFATC2*	2.25
Melanoma	0.038	*FGF18, PIK3CB, MITF, PDGFRA, MDM2*	3.85

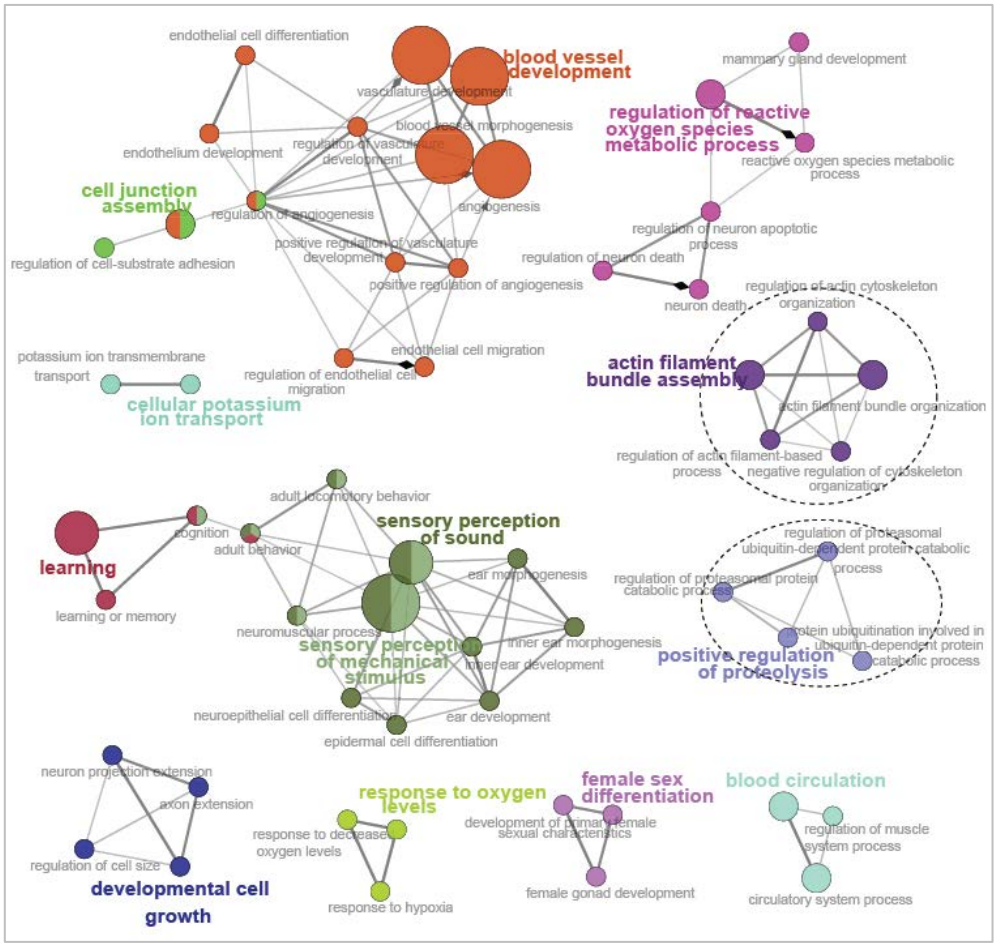


Figure 4.3 ClueGO gene ontology analysis of 354 positively selected genes in Ankole cattle population. ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes. Nodes represent gene ontology terms to which their size reflects the statistical significance of the terms. The most prominent gene ontology term for each group is highlighted in colors, and the circled gene ontology terms are related to meat quality characteristics.

The *CAPZB* (XP-CLR = 142.50) gene encodes the beta subunit of the barbed-end actin binding protein, which belongs to the F-actin capping protein family. It is involved in skeletal muscle development and growth (Xu et al. 2012), and cell signaling and regulation of actin in myofilament contractility (Ponsuksili et al. 2009). When up-regulated, it increases the ability of muscle accretion in pigs (Xu et al. 2012). *CAPZB* contributes to muscle metabolic and structural properties and proteolytic processes providing a link between these functional networks which are important for maturation of muscle to meat (Ponsuksili et al. 2009). A previous functional analysis of meat tenderness revealed a positive correlation between *CAPZB* expression and beef tenderness (Guillemin et al. 2011). In the pig, *CAPZB* is an essential element for protein kinase signaling to the myofilaments and, as a structural protein, it has been shown to influence muscle biochemistry and its postmortem abundance is related to meat quality (Pyle et al. 2002). The Tajima's D and F_{ST} plot of the *CAPZB* gene region (Figure 4.4a) show the presence of an excess of rare alleles in Ankole population and the differentiation of the region between the compared breeds, respectively.

LIMA1 (called *EPLIN)* encodes a cytoskeleton-associated protein that inhibits actin filament depolymerization and cross-links filaments in bundles. It is associated in pigs with functions regarding muscle development and metabolism (Schellander 2010). *ZNF410*, also known as *APA-1*, is an essential component of the stress pathway involved in the meat tenderization process (Guillemin et al. 2011). In previous muscle transcriptome analyses, *ZNF410* has been shown to be highly expressed in the longissimus muscle of Basque pigs that are known to produce pork with higher intramuscular fat and tenderness compared to Large White pigs (Damon et al. 2012). *COL9A2*, a fibrillar collagen, constitutes the largest component of extracellular matrix (ECM) to which its amount, type, and solubility present in muscle tissue have a strong effect on meat tenderness (Chang 2007). This gene was found to be upregulated in the longissimus dorsi muscle of Jeju native piglets (Ghosh et al. 2015), whose meat is known for its preferable taste, tenderness and superior marbling (Cho et al. 2011).

ROCK1, a gene that regulates actin cytoskeleton and cell polarity, is associated with body weight, carcass weight, shank length, shank circumference and other carcass weight traits in chicken (Lu et al. 2012).

Genes involved in MAPK signaling (*MAP3K5, MAP2K3, MAP4K4,* and *MAPKAPK2*) were also identified. MAPK signaling is one of the major intracellular signaling pathways affecting myogenesis (WU et al. 2010) and is relevant to postmortem meat quality (Ponsuksili et al. 2009). *MAP2K3* shows associations with loin muscle area and fat traits in pigs, implying roles in muscle differentiation and growth (WU et al. 2010).

E3 ubiquitin ligase genes (*WWP1,* and *PARK2*) play an important role in the regulation of a wide variety of cellular functions such as protein degradation, transcription, and RNA splicing. These genes catalyze protein ubiquitylation resulting in the targeting of proteins toward various cellular fates, with proteasome-mediated proteolytic degradation (Yin et al. 2010). The ubiquitin-proteasome system is one of the proteolytic systems responsible for the majority of the protein degradation in muscle that is relevant for meat quality postmortem (Clark et al. 2002).

The expression of *PKM2* (XP-EHH = 1.9696; p =2.00.E-03), a gene involved in energy metabolism, is positively correlated with WBSF and has been reported as a functional protein marker for meat tenderness in Thai indigenous chicken (Teltathum and Mekchay 2010) and beef (Guillemin et al. 2011). *PDGFRA* also has an effect on shear force and Loin Eye Area in pig (Wimmers et al. 2007). The Tajima's D and F_{ST} plot of *PKM2* and *PDGFRA* gene regions are shown in Figure 4.4 (b) and (c), respectively. *PRKG1* is reported to be important in the conversion of muscle to meat (Lonergan et al. 2010). *NFATC2* is a calcineurin substrate expressed in skeletal muscle which is responsible for activating new myotubes (Chang 2007). Calcineurin is crucial for myocyte differentiation and determination of the slow oxidative fiber phenotype (da Costa et al. 2007).

Genes related to meat intramuscular fat (IMF)

IMF is a heritable meat quality trait which affects flavor, juiciness, visual characteristics and meat tenderness. It is positively correlated with body fat and red muscle fiber (Joo et al. 2013). Steak from Ankole cattle has been found to be juicy than those Ankole-Boran crossbreds (Kamatara et al. 2013), and Strydom et al. (2008) showed higher levels of rib fat thickness in Sanga as compared to *indicus* cattle.

I identified several genes (*PLA2G2A, PARK2, ZNF410, PKM2, MAP2K3, PLCD3, PLCD1, ROCK1,* and *AHSA1*) which affect the fat content of meat in Ankole cattle. *PLA2G2A* (XP-CLR = 117.48) is a member of the phospholipase A2 family (*PLA2*), which is involved in the hydrolysis of phospholipids into fatty acids and phosphatidylinositol and phospholipid metabolism (Nakamura et al. 2013). Also referred to as Adipose-Specific Phospholipase A2 (*AdPLA*), it is involved in adipocyte metabolism and catalyzes the efficient release of free fatty acids and lysophospholipid from phosphatidylcholine (Duncan et al. 2008). It has been reported in the literature that *PLA2* has a positive effect on porcine fat deposition (IMF) and potentially regulates lipolysis and increases the MUFA deposition rather than the SFA deposition (Wang et al. 2013c). It is also associated with intramuscular fat in beef cattle (Chan and Reverter 2007). *Pla2g2a* has been reported to be a candidate gene in relation to obesity in mice (Sung and Bae 2010).

ROCK1 is involved in pathways relevant to muscle/adipose tissue function in pigs with divergent phenotypes for fatness traits (Cánovas et al. 2010). E3 ubiquitin ligase enzymes have been identified to be involved in the modulation of lipid biology (Yin et al. 2010; Roux et al. 2015). *PARK2* is a strong positional candidate for adiposity in chicken and a positive regulator of fat metabolism (Roux et al. 2015). *PRKG1* is involved in gap junction and is a candidate gene for intramuscular fat in the pigs (Hamill et al. 2012). *MAP2K3* has been shown to be associated with loin muscle and fat traits in pigs (WU et al. 2010). *MAP4K4* is involved in adipogenesis, triglyceride

storage, fatty acid release, fatty acid oxidation and mitochondrial oxidative phosphorylation (Puri et al. 2008). *PKM2* is significantly associated with backfat thickness, an economically important trait in pigs (Cho et al. 2013). *APOL6* is one of the most important known genes involved in lipoprotein metabolism (Corella and Ordovas 2005). Phospholipase C family genes (*PLCD1* and *PLCD3*) generate diacylglycerol and are involved in phosphatidylinositol catabolism and phospholipid synthesis (Nakamura et al. 2013). The transcription of *AHSA1* (*AHA1*) is related to Omega-3 fatty acids in skeletal muscle, which influence meat tenderness, juiciness, and flavor, and are beneficial to human health (Perez et al. 2010). The positive selection of the *AHSA1* gene region is shown in the Tajima's D and F_{ST} plot in Figure 4.4d.

Genes related to meat color, drip loss, and feed conversion efficiency (FCR)

Meat color and water holding capacity of meat are among the quality parameters used as an indicator of freshness and wholesomeness (Mancini and Hunt 2005; Joo et al. 2013). These characteristics are related to variations in the glycolysis rate and muscle temperature decline postmortem. Myoglobin (*MB*; XP-CLR = 129.86; XP-EHH = 1.9640; p=5.00.E-03), a globular single chain protein located in the sarcoplasm, is the principle protein responsible for the red color of meat. *MB* serves as a reserve supply of oxygen and facilitates the movement of oxygen within muscles (Mancini and Hunt 2005; Joo et al. 2013). Figure 4.4e shows the Tajima's D and F_{ST} plot of *MB* gene region in Sanga and *B. indicus* populations. The Solute Carrier Family 48 (Heme Transporter), Member 1 (*SLC48A1*) is responsible for the transport of heme from endosome to the cytosol (Khan and Quigley 2013) and may also have a function in meat color. In general, beef from Sanga cattle breeds showed higher chroma than that of *indicus* cattle breeds (Strydom et al. 2011).

The loss of reddish fluid mainly consisting of water and proteins from meat, called drip loss, is an important meat quality characteristics which is affected by sev-

eral ante- and post-mortem factors (Borchers et al. 2007). A small but significant difference in drip loss is reported between Sanga and *indicus* cattle breeds; meat from Sanga cattle showed higher drip loss (Strydom et al. 2008; Strydom et al. 2011). Higher expression of *PKM2* and *MAP4K4* suppresses the glucose content of muscle cells promoting the onset of anaerobic production of lactate post-mortem, thereby facilitating the decline in pH resulting in higher drip loss (Ponsuksili et al. 2008; Shen et al. 2014).

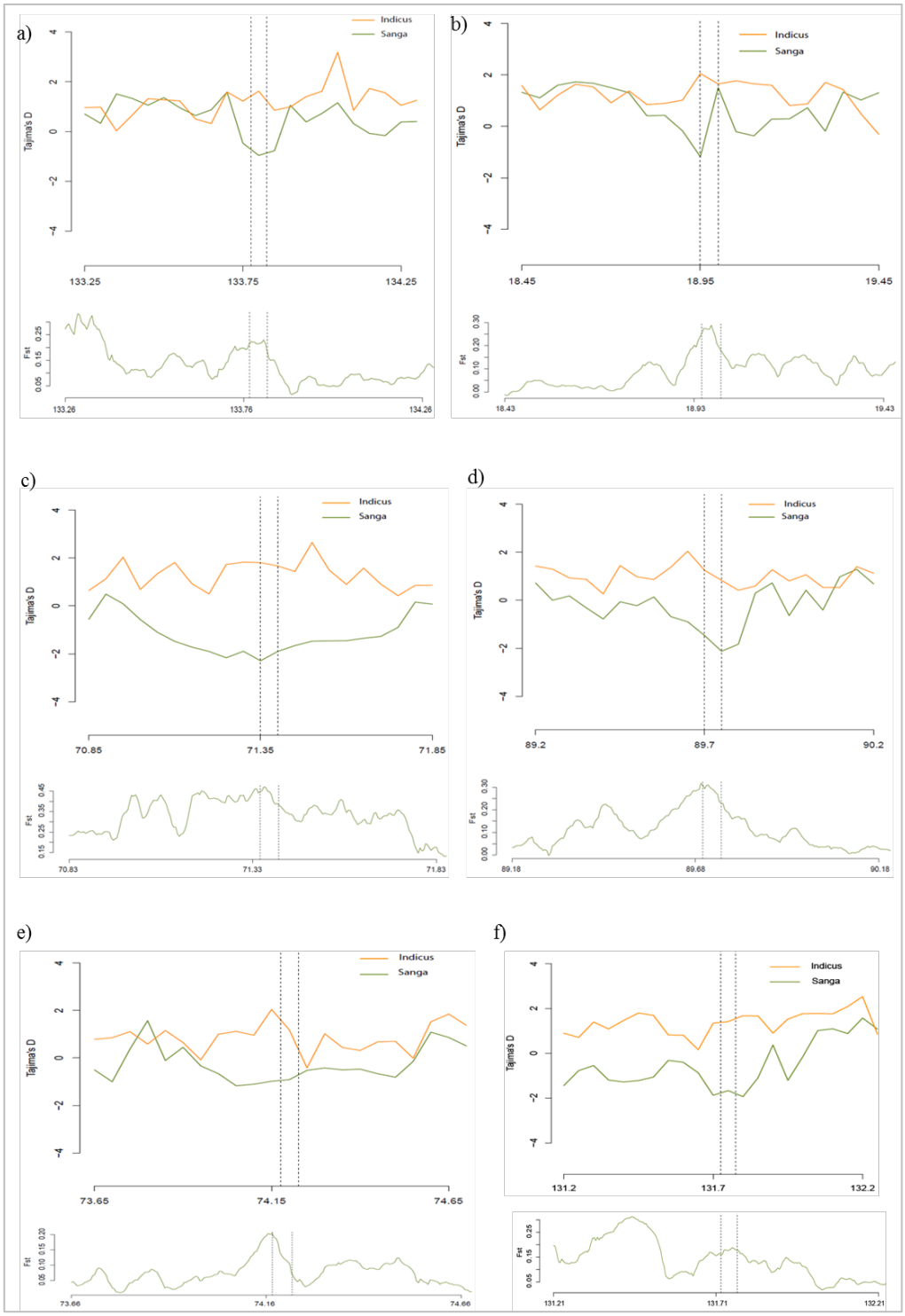
Feed intake and efficiency, measured as residual feed intake (RFI), are economically important traits affecting the cost of beef production (Sherman et al. 2010). Variation in RFI (animals with lower RFI are more efficient) has a genetic component with moderate heritability (Chen et al. 2011). I identified positively selected genes (*TIMP2*, *PKM2*, *PRKG1*, *MAP3K5*, and *ATP8A1*) that are reported in the literature to be related to RFI and feed conversion efficiency. *TIMP2* has been shown to be upregulated in low RFI animals in gene expression profiling studies on genes expressed differentially in cattle with high and low RFI (Chen et al. 2011). *PKM2* was associated with average daily gain, and feed to body weight gain ratio, with a significant additive and/or dominance effects on these traits (Fontanesi et al. 2008). *PRKG1* is involved in gap junction and is also a candidate gene for RFI in cattle (Sherman et al. 2010). *MAP3K5*, also known as apoptosis signal-regulating kinase 1 (*ASK1*), is a candidate gene for residual feed intake in pigs (Do et al. 2014a). *ATP8A1* is also related to feed intake, feed conversion ratio, residual feed intake and weight gain (Santana et al. 2014). Olfactory receptor genes (*OR2D2, OR10A4*, and *OR2D3*) have been shown to affect the perception of taste and smell (Choquette et al. 2012; Do et al. 2014b) and therefore can be related to feed intake and feeding behavior (Do et al. 2014b). *PIK3CB*, and *MRAS* genes involved in the Akt/PI3K and MAPK signaling pathways, respectively, are important for high feed efficiency in chicken (Zhou et al. 2015). The positive selection of these genes may provide clues as to why Ankole cattle are able to use and survive on poor quality feed and withstand severe droughts (Ndumu et al. 2008).

Table 4.4 Summary of major candidate genes related to meat quality characteristics and feed intake in Sanga cattle population detected by XP-EHH and XP-CLR statistics

Candidate genes	Chr.	Window (Mbp)	XP-CLR	XP-EHH	XP-EHH P-value	Tajima's D	Weighted F_{ST}	Species, Trait, and Reference
WWP1	14	78.63 - 78.68	133.90	-	-	1.03	0.18	Tenderness (Yin et al. 2010)[4]
PDGFRA	6	71.35 - 71.40	408.60	2.61	3.00.E-04	-2.29	0.45	Tenderness (Wimmers et al. 2007)[2]
LIMA	5	29.78 - 29.83	117.18	-	-	-0.33	0.39	Tenderness (Schellander 2010)[2]
ROCK1	24	35.48 - 35.53	130.40	2.22	1.61.E-03	0.93	0.29	Tenderness (Lu et al. 2012)[1]; IMF (Cánovas et al. 2010)[2]
PRKG1	26	7.43 - 7.48	166.61	-	-	-1.27	0.13	Tenderness (Lonergan et al. 2010)[4]; IMF (Hamill et al. 2012)[2]; RFI (Sherman et al. 2010)[3]
PKM2	10	18.95 - 19.00	-	1.97	2.00.E-03	-1.18	0.25	Tenderness (Guillemin et al. 2011)[3], (Teltathum and Mekchay 2010)[1]; IMF (Cho et al. 2013)[2]; Drip loss (Ponsuksili et al. 2008; Shen et al. 2014)[2]
ZNF410	10	85.68 - 85.73	165.86	2.18	1.94.E-03	-0.67	0.28	Tenderness (Guillemin et al. 2011)[3], (Damon et al. 2012)[2]
CAPZB	2	133.78 - 133.83	142.50	-	-	-0.47	0.22	Tenderness (Guillemin et al. 2011)[3]
COL9A2	3	106.38 - 106.43	107.14	1.93	5.95.E-03	-1.30	0.15	Tenderness (Ghosh et al. 2015)[2]
NFATC2	13	80.00 - 80.05	-	2.01	4.24.E-03	2.53	0.25	Tenderness (da Costa et al. 2007)[4]
PIK3CB	1	131.50 - 131.55	121.48	2.21	1.69.E-03	-1.06	0.27	RFI (Zhou et al. 2015)[1]
MAP3K5	9	75.55 - 75.60	147.37	2.09	3.00.E-03	-0.23	0.15	RFI (Do et al. 2014a)[2]
TIMP2	19	54.10 - 54.15	-	1.88	7.46.E-03	-0.73	0.18	RFI (Chen et al. 2011)[3]
MB	5	74.18 - 74.23	129.86	1.96	5.00.E-03	-0.98	0.20	Meat color (Mancini and Hunt 2005; Joo et al. 2013)[4]
SLC48A1	5	32.63 - 32.68	101.45	1.80	6.96.E-03	-0.35	0.26	Meat color (Khan and Quigley 2013)[4]

Candidate genes	Chr.	Window (Mbp)	XP-CLR	XP-EHH	XP-EHH P-value	Tajima's D	Weighted F_{ST}	Species, Trait, and Reference
MAP2K3	19	35.85 - 35.90	-	1.93	4.00.E-03	-0.20	0.16	IMF (WU et al. 2010)[2]
PLA2G2A	2	133.28 -133.33	117.48	-	-	0.70	0.27	IMF (Wang et al. 2013c)[2], (Chan and Reverter 2007)[3]
PARK2	9	99.13 - 99.18	132.40	2.34	2.36.E-03	-0.27	0.30	IMF (Roux et al. 2015)[1]
AHSA1	10	89.70 - 89.75	-	2.23	2.00.E-03	-1.47	0.31	IMF (Perez et al. 2010)[3]
PLCD3	19	45.40 - 45.45	-	1.89	5.00.E-03	-0.18	0.03	IMF (Nakamura et al. 2013)[4]
PLCD1	22	11.45 - 11.50	195.30	2.12	3.00.E-03	-1.43	0.14	IMF (Nakamura et al. 2013)[4]
APOL6	5	74.20 - 74.25	-	2.00	5.00.E-03	-0.91	0.13	IMF (Corella and Ordovas 2005)[4]
OR2D2, OR10A4, OR2D3	15	46.40 - 46.45	-	2.07	3.45.E-03	2.61	0.18	Feed intake(Choquette et al. 2012)[4]; (Do et al. 2014b)[2]
MRAS	1	131.73 - 131.78	112.25	-	-	-1.87	0.16	FCE (Zhou et al. 2015)[1]
ATP8A1	6	62.90 - 62.95	264.27	2.02	4.00.E-03	-1.39	0.36	FCE (Santana et al. 2014)[3]
MAP4K4	11	6.65 - 6.70	-	1.85	9.00.E-03	0.72	0.08	Drip loss (Ponsuksili et al. 2008; Shen et al. 2014)[2]

Note: Chr.: Chromosome; Window: start and end positions of the gene region; RFI: residual feed intake; FCE: feed conversion efficiency; IMF: intramuscular fat; Superscripts in the Species, Trait and Reference column indicate the species that the trait has been previously reported for as [1]chicken, [2]pork, [3]beef, [4]general (not for a specific species).


a)
Indicus
Sanga
Tajima's D
133.25
133.75
134.25
Fst
133.26
133.76
134.26
b)
Indicus
Sanga
Tajima's D
18.45
18.95
19.45
Fst
18.43
18.93
19.43
c)
Indicus
Sanga
Tajima's D
70.85
71.35
71.85
Fst
70.83
71.33
71.83
d)
Indicus
Sanga
Tajima's D
89.2
89.7
90.2
Fst
89.18
89.68
90.18
e)
Indicus
Sanga
Tajima's D
73.65
74.15
74.65
Fst
73.66
74.16
74.66
f)
Indicus
Sanga
Tajima's D
131.2
131.7
132.2
Fst
131.21
131.71
132.21

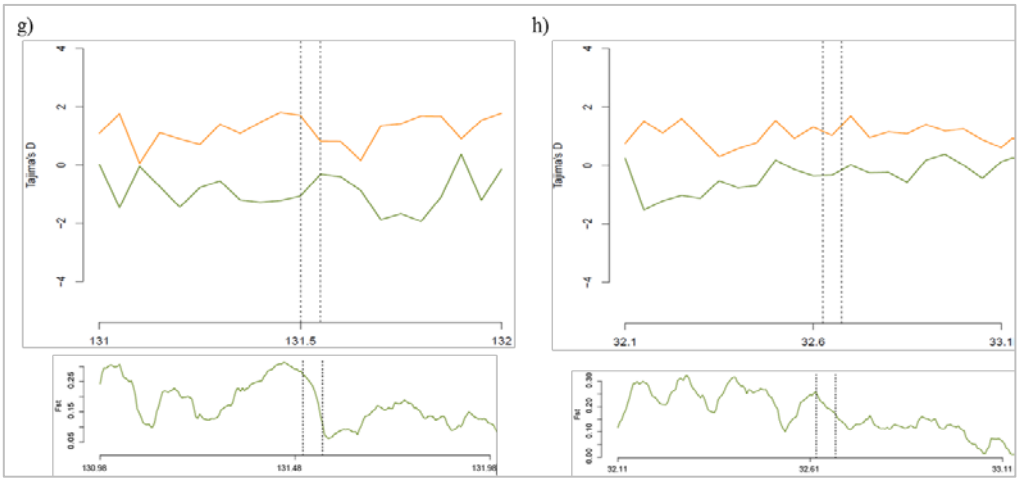


Figure 4.4 Tajima's D and F_{ST} plot of positively selected candidate gene regions in Sanga and *indicus* cattle populations. a) *CAPZB* gene; b) *PKM2* gene; c) *PDGFRA* gene*;* d) *AHSA1* gene; e) *MB* gene; f) *MRAS* gene; g) *PIK3CB* gene; and h) *SLC48A1* gene. The y-axis indicates the Tajima's D and F_{ST} values, and the x-axis is the chromosomal position. The vertical dotted lines indicate the start and end positions of the gene region under consideration. The Tajima's D plot for each gene region (upper plot for each gene) show the Tajima's D value within a 50 kb window plotted for both populations. The smaller (negative) Tajima's D value in the Sanga population shows that the gene region considered is under positive selection. The F_{ST} plot (lower plot for each gene) represents the F_{ST} values calculated within 50 kb window separated by 5 kb window steps.

4.4.4 Implication of the results of this study on Ankole population

The Ankole group is one of the three groups of Sanga cattle representing Sanga cattle in east and central Africa (Rege and Tawah 1999). Ankole breed is a valuable and widely used genetic resource in the region due to its better adaptability. However, there have been no well-designed breed improvement programs for Ankole and other Sanga breeds of eastern Africa (Ndumu et al. 2008; Kugonza et al. 2011). Selective

breeding efforts in other South African Sanga cattle breeds (e.g., Mashona, Tuli, and Afrikander) have resulted in local cattle showing higher beef productivity (Rewe et al. 2009). As cattle genetic resources are being depleted (Hanotte et al. 2010; Mwai et al. 2015) and given the importance of this vital genetic resource, designing breeding programs that would help improve and conserve Ankole cattle is crucial (Kugonza et al. 2011). With this regard, the results provide a basis for further research on the genomic characteristics of Ankole cattle in relation to meat quality traits.

4.4.5 Limitations of the present study

As is typical in this kind of study, there is a possibility of obtaining false positive results. Therefore, validation using other methods such as GWAS, candidate gene approach and gene expression analysis are suggested. In addition, given the multifactorial nature of meat quality traits, limited published literature is available on genes affecting beef quality characteristics.

4.5 Conclusion

Results from the whole genome scan revealed several positively selected genes involved in different biological and cellular functions including those affecting meat quality characteristics. The genes identified in relation to meat quality characteristics are involved in muscle and lipid metabolism that affect tenderness and intramuscular fat content of meat, and help to improve our understanding of the biological mechanisms controlling meat quality traits in beef cattle production. These results provide a basis for further research on the genomic characteristics of Ankole and other Sanga cattle breeds for quality beef production.

This chapter was published in *Genes and Genomics*

in partial fulfillment of Mengistie Taye Terefe's Ph. D program

Chapter 5. Deciphering Signature of Selection Affecting Beef Quality Traits in Angus Cattle

5.1 Abstract

Artificial selection towards a desired phenotype/trait has modified the genomes of livestock dramatically that created breeds that greatly differ in morphology, production, and environmental adaptation traits. Angus cattle are among the famous cattle breeds developed for superior beef quality. This paper aimed at exploring genomic regions under selection in Angus cattle that are associated with meat quality traits and other associated phenotypes. The whole genome of 10 Angus cattle was compared with 11 Hanwoo (A-H) and 9 Jersey (A-J) cattle breeds using a cross-population composite likelihood ratio (XP-CLR) statistical method. The top 1% of the empirical distribution was taken as significant and annotated using UMD3.1. As a result, 255 and 210 genes were revealed under selection from A-H and A-J comparisons, respectively. The WebGestalt gene ontology analysis resulted in sixteen (A-H) and five (A-J) significantly enriched KEGG pathways. Several of the pathways enriched were associated with meat quality traits (insulin signaling, type II diabetes mellitus pathway, focal adhesion pathway, and ECM-receptor interaction), and feeding efficiency (olfactory transduction, tight junction, and metabolic pathways). Genes affecting beef quality traits (e.g., *FABP3, FTO, DGAT2, ACS, ACAA2, CPE, TNNI1),* stature and body size *(*e.g., *PLAG1, LYN, CHCHD7, RPS20),* fertility and dystocia (e.g., *ESR1, RPS20, PPP2R1A, GHRL, PLAG1*), feeding efficiency (e.g., *PIK3CD, DNAJC28, DNAJC3, GHRL, PLAG1*), coat color *(*e.g., *MC1-R)* and genetic disorders (e.g., *ITGB6, PLAG1*) were found to be under positive selection. The findings in this study, after validation using additional or independent dataset, will provide useful information for the study of Angus cattle in particular and beef cattle in general.

5.2 Introduction

Intensive artificial selection within and between breeds of livestock made the development of specialized breeds that are able to produce the intended amount and quality of product a reality. Artificial selection concentrates the genetics of certain individuals that cause differences in the specific patterns of change in allele frequencies, diversity, and haplotype structure that in turn differentiate breeds under selection from others. In beef cattle, such kind of differential selection have resulted in several breeds of high growth rate, superior beef quality, and higher feed efficiency (Albertí et al. 2008).

Apart from its positive impact in improving production and productivity of commercial traits, intensive artificial selection towards a particular trait has been reported to cause several genetic disorders in several beef and dairy cattle breeds (Whitlock et al. 2008). Genetic disorders result in high mortality and reduced reproduction and productivity of herds and greatly impact the profitability of the farm (Ciepłoch et al. 2017). This is because genes causing genetic disorders are linked with those genes affecting economic traits of interest. Ciepłoch et al. (2017), reviewed genetic disorders of beef cattle that are potentially caused by human artificial selection forces. In Angus cattle populations, dwarfism and fawn calf syndrome (Whitlock et al. 2008), arthrogryposis multiplex, neuropathic hydrocephalus, and osteopetrosis (Whitlock 2010) have been reported in Australia and the US. Cardiomyopathy, a genetic disorder that affects the heart muscle, has been reported in Holstein cattle (Guziewicz et al. 2007).

Angus cattle (Aberdeen Angus) are known cattle breeds developed around the early 19th century in Northeast Scotland. During the twentieth century, breeders made enormous changes in the growth, stature and body composition of American Angus cattle through selection (Arthur et al. 2001; McClure et al. 2010). The breed is characterized by its high muscularity, higher growth rate (ADG), wide pelvis and medium height and high level of beef fat (Albertí et al. 2008). Angus cattle are early finishing

with high growth rate, eye muscle and yield (Chambaz et al. 2003). They are naturally polled and predominantly black or red in color (http://www.thecattlesite.com/breeds/beef/7/aberdeen-angus/).

Like that of natural selection, artificial selection towards a particular trait is expected to leave a distinctive signature on the genome that can be traced using genomic and bioinformatics methods. Identification of signature of selection has been used to pinpoint the adaptive events that have generated the enormous phenotypic variation observed between cattle breeds and has a biotechnological relevance (Utsunomiya et al. 2015). Recently, several methods have been developed and applied to scan for footprints of selection in several species and breeds of animals. Using iHS, F_{ST} and CLR methods, The Bovine HapMap Consortium (2009), sought to identify ongoing selection due to domestication, breed formation, and ongoing selection intended to enhance performance and productivity in diverse cattle breeds. From the analysis, genomic regions affecting double muscling (*MSTN*), and those associated with intramuscular fat (*KHDRBS3* and *TG*) were found under positive selection (The Bovine HapMap Consortium 2009). Similarly, Rothammer et al. (2013) explored the genome of 10 beef and dairy cattle breeds for signature of selection and identified genes (*TG*, *ABCG2*, *DGAT1*, *GH1*, *GHR* and the Casein Cluster) that are strongly associated with known QTL for dairy and/or beef traits. In this study, I explored the genome of Angus cattle to identify genes and gene regions under positive selection contributing to the superior meat quality characteristics and associated genetic disorders in Angus cattle. The cross-population composite likelihood ratio (XP-CLR) test (Chen et al. 2010), a population differentiation method, was used for the analysis.

5.3 Materials and Methods

5.3.1 Data preparation and description

In this chapter, I used a whole genome sequencing data of Angus cattle where detailed sample information and resequencing procedures can be found (Kim et al. 2017a). DNA samples of Angus and Jersey cattle breeds were obtained from the Institutional Animal Care and Use Committee of the National Institute of Animal Science, Korea. For Hanwoo cattle, blood samples were collected from Hanwoo Improvement Center of the National Agricultural Cooperative Federation (HICNACF), and DNA was extracted using a G-DEXTMIIb Genomic DNA Extraction Kit (iNtRoN Biotechnology, Seoul, Republic of Korea) according to the manufacturer's protocol. The DNA was checked for its quality and inserts of ~300 bp was generated from a randomly sheared 3 μg of genomic DNA. The fragments of sheared DNA were end-repaired, A-tailed, adaptor-ligated, and amplified using a TruSeq DNA Sample Prep. Kit (Illumina, San Diego, CA, USA). Paired-end sequencing was conducted using the Illumina HiSeq2000 platform with TruSeq SBS Kit v3-HS (Illumina).

Sequence reads were mapped against the reference bovine genome (UMD 3.1) using Bowtie2 (Langmead and Salzberg 2012). The overall alignment rate of reads to the reference sequence was 98.84% with an average read depth of 10.8x, and the reads covered 98.56% of the reference UMD3.1 genome (Kim et al. 2017a).

Open-source software packages were used for downstream processing and variant calling. Picard (https://broadinstitute.github.io/picard/) filtered potential PCR duplicates, and SAMtools (Bindea et al. 2009) created index and bam files. Genome analysis toolkit 3.1 (GATK) (McKenna et al. 2010) was used to perform local realignment of reads. The "UnifiedGenotyper" and "SelectVariants" arguments of GATK was used to call candidate SNPs. In order to filter variants and avoid possible false positives, the "VariantFiltration" argument of the same software was adopted with the

following options: 1) SNPs with a phred-scaled quality score of less than 30 were filtered; 2) SNPs with MQ0 (mapping quality zero; total count across all samples of mapping quality zero reads) > 4 and quality depth (unfiltered depth of non-reference samples; low scores are indicative of false positives and artifacts) < 5 were filtered; and 3) SNPs with FS (Phred-scaled P-value using Fisher's exact test) > 200 were filtered since FS represents variation on either the forward or the reverse strand, which is indicative of false positive calls. For the haplotype information on each chromosome, BEAGLE (Browning and Browning 2007) was used to infer the haplotype phase and impute missing alleles for the entire set of cattle populations simultaneously. Sequences used for this study are available from GenBank with the Bio project accession number of Angus (PRJNA318087), Jersey (PRJNA318089), and Hanwoo (PRJNA210523).

5.3.2 Phylogenetic tree and population structure

I used SNPhylo pipeline (Lee et al. 2014c), using autosomal SNPs, to construct a phylogenetic tree to understand the relationship between breeds. SNPhylo uses the SNPRelate (Zheng et al. 2012) package to check and filter for quality of SNPs applying minor allele frequency (MAF) and missing rate threshold, and make use of linkage disequilibrium; MUSCLE (Edgar 2004) for multiple sequence alignment; and PHYLIP package (Plotree and Plotgram 1989) to determine the phylogenetic tree by a maximum likelihood method. In the SNPhylo pipeline, I used the options of 1000 bootstrapping samples, 29 autosomes, and MAF threshold of 0.05. A total of 14049 SNPs were randomly selected and used for the phylogenetic tree construction. I visualized the phylogenomic tree using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

In addition, analysis of population structure was performed in a Bayesian model-based analysis using STRUCTURE software (Evanno et al. 2005) to identify groups of individuals corresponding to the uppermost hierarchical levels.

STRUCTURE assumes a model in which there are *K* populations (clusters), which contribute to the genotype of each individual characterized by a set of allele frequencies at each marker locus. The software applies a *Markov Chain Monte Carlo* (MCMC) estimation of allele frequencies in each of the *K* populations and the degree of admixture for each animal. The number of clusters were inferred using the options of length of Burnin Period of 2000, the number of MCMC Reps after Burnin period of 200000 and MAF of 0.05. I used PLINK (Purcell et al. 2007) to generate input files used by STRUCTURE using -thin option that retained 4008 loci. I also calculated a population differentiation index (F_{ST}), with 50 kb windows of 5 kb steps, between the populations considered using VCFtools to understand the genetic distance between the breeds considered (Holsinger and Weir 2009).

5.3.3 Signature of positive selection

I performed an XP-CLR statistical test (Chen et al. 2010), which implements composite likelihood methods for detecting selective sweep genomic regions differentially selected between two populations. It is a multi-locus sliding window based test that detects recent selective sweep regions based on allele frequency differentiation between two populations. Here, I carried out pairwise comparisons of the genomes of 10 Angus with 11 Hanwoo (A-H) and 9 Jersey (A-J) cattle breeds using the XP-CLR software package (Chen et al. 2010). Angus cattle are a specialized beef breed native to Aberdeenshire in Scotland (Arthur et al. 2001; McClure et al. 2010). The Hanwoo cattle are a result of interbreeding between taurine and zebu cattle that its history as a draft animal dates back at least 5,000 years, and since recently, it has been intensively selected for high-quality beef (Chung 2014). Jersey cattle are a small dairy cattle known for their better quality milk (The Bovine HapMap Consortium 2009). Since these populations have different demographic and selection histories, comparing Angus cattle with Hanwoo and Jersey breeds could help uncover the

signature of selection related to meat quality traits and other specific phenotypes of Angus cattle.

In order to calculate XP-CLR values, I followed a previously used procedure where non-overlapping sliding windows of 50 kb and a maximum number of 600 SNPs within each window were used (Kim et al. 2017a). According to the software, a weighted composite likelihood ratio (CLR) scheme was adopted in estimating XP-CLR – the pairwise correlation coefficient (r^2) of SNPs from the reference population are used to give weights. When the correlation coefficients are greater than 0.95, CLR scores for the two SNPs are down-weighted. The top 1% (0.01) of the empirical distributions were designated as candidate sweeps and genes that span the window regions were defined as candidate genes (Kim et al. 2017a). Significant genomic regions identified from both comparisons were annotated based on UMD 3.1.

5.3.4 Characterizing genes under selection

WEB-based GEne SeT AnaLysis Toolkit (WebGestalt), which integrates data from centrally and publicly curated databases as well as computational analyses (Wang et al. 2013b), was used for gene enrichment analysis. Gene sets identified from each of the comparisons were submitted separately to identify significantly enriched Gene Ontology Biological Process (GO-BP) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. I employed a hypergeometric statistical method and Bonferroni Multiple Adjustment using a Homo sapiens genome as a reference set.

Additionally, ClueGO a Cytoscape plug-in, was used to integrate GO-BP terms and KEGG pathways in order to functionally organize into GO/pathway term networks (Bindea et al. 2009). Finally, I used the SNPEff variant annotation and effect prediction tool (Cingolani et al. 2012) to annotate and predict the effects of genetic variants (such as amino acid changes) for the genes considered as candidate genes. I

have drawn the Manhattan plot of the -$\log_{10}$ transformed XP-CLR P-values from both comparisons using R software. The gene names and descriptions used are based on genecards (http://www.genecards.org/).

5.4 Result and Discussion

5.4.1 Data description

DNA samples collected from 10 Angus, 11 Hanwoo, and 9 Jersey cattle breeds and sequenced to ~11x genome coverage each was used for the analysis. After standard data preparation and re-sequencing procedures, an average alignment rate of 98.84 % covering 98.56 % of the taurine reference genome (UMD 3.1) was obtained. Using several methods and software, potential PCR duplicates and false positive calls were filtered, and finally, a total of ~37 million SNPs were obtained and used for further analysis.

5.4.2 Phylogenetic tree and structure analysis

I constructed a non-rooted Phylogenetic tree of sample populations using SNPhylo (Lee et al. 2014c). As a result, individual animals within breeds clustered together separately from individuals from other breeds (Figure 5.1a). Similarly, to understand the admixture level of sample populations, STRUCTURE (Evanno et al. 2005) was used to construct structure at 2 and 3 population assumptions (Figure 5.1b). At ancestral populations *(K)* of 2, Angus and Hanwoo breeds clustered together separately from Jersey whereas, at $K = 3$, all the three breeds became different even though Angus showed some sort of admixture with Hanwoo and Jersey. The Weir and

Cockerham weighted F_{ST} estimates between A-H (0.12977), A-J (0.19003), and H-J (0.20374), confirmed the distance of Jersey from both of the other breeds. Historically, three of the breeds are originated separately and developed in different areas for different purposes. The Angus (Arthur et al. 2001; McClure et al. 2010) and Hanwoo (Chung 2014) are beef breeds whereas Jersey (The Bovine HapMap Consortium 2009) are dairy cattle.

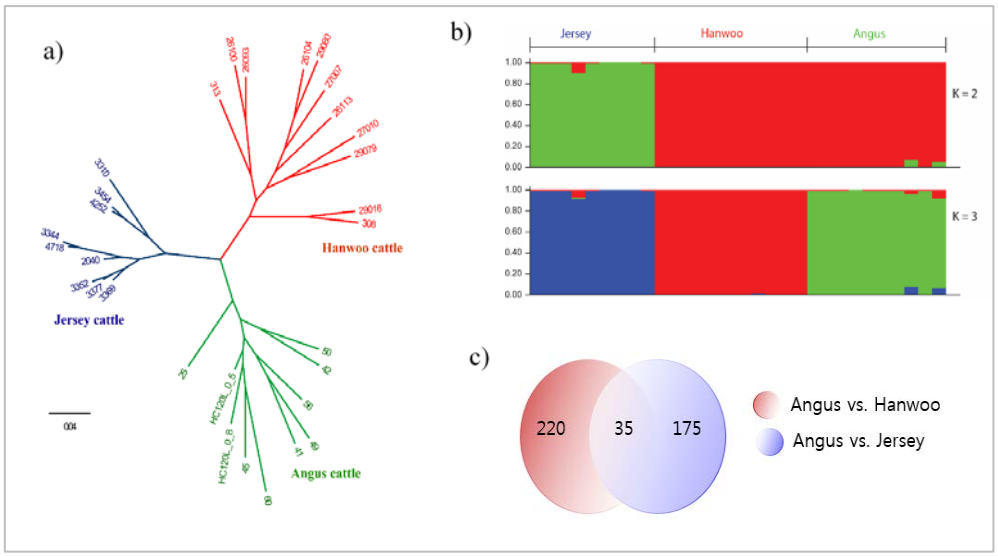


Figure 5.1 Population Stratification of Angus, Hanwoo, and Jersey cattle breeds. a) Phylogenetic tree; b) Population structure; and c) Number of genes identified from Angus vs. Hanwoo and Angus vs. Jersey XP-CLR comparisons.

5.4.3 Signature of positive selection

In order to infer the positive selection signature of gene regions that are related to the phenotypes of Angus cattle, I compared the genomes of Angus cattle with Hanwoo (A-H) and Jersey (A-J) cattle breeds using XP-CLR statistics following previous procedures (Kim et al. 2017a). XP-CLR compares allele frequency differentiation between two populations (Chen et al. 2010). The Manhattan plot of the $-\log_{10}$ transformed XP-CLR score p-values of both comparisons are presented in Figure 5.2. Then, by annotating the top 1% outlier regions of the empirical distribution, 255 and 210 genes were identified from A-H and A-J comparisons, respectively (Figure 5.1c; Table 5.1).

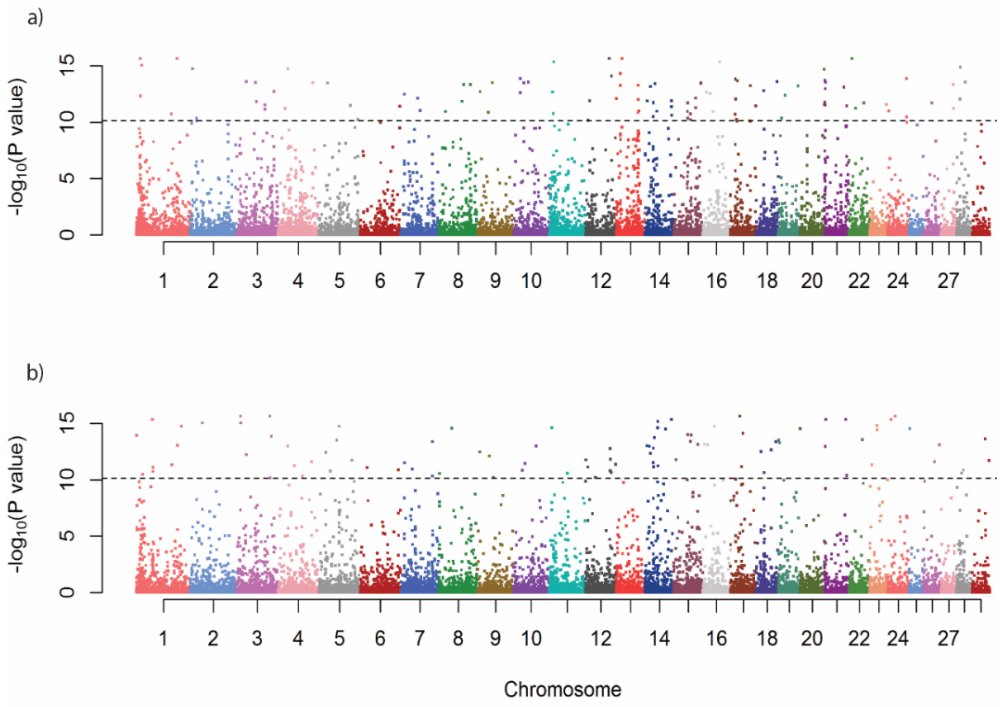


Figure 5.2 Manhattan plot of $-\log_{10}$ transformed XP-CLR score *P*-values of Angus vs. Hanwoo (a), and Angus vs. Jersey (b) comparisons. The *y*-axis shows the $-\log_{10}$ (*P*-value) of XP-CLR p-value, and the *x*-axis shows chromosomal positions. The horizontal dotted lines represent the 1% XP-CLR outlier regions in both of the comparisons.

Table 5.1 Summary of genes identified as positively selected in Angus cattle from a genome-wide comparison of Angus vs. Hanwoo (A-H) and Angus vs. Jersey (A-J) cattle breeds using XP-CLR statistics

Chr.	Start	End	Breed comparison		Gene
			A-H	A-J	
1	1275.0	1325.0	-	55.62	DNAJC28,GART,TMEM50B
1	1325.0	1375.0	-	35.50	SNORA20,TMEM50B
1	15025.0	15075.0	-	35.26	NCAM2
1	18125.0	18175.0	-	36.38	TMPRSS15
1	18175.0	18225.0	-	48.21	CHODL,TMPRSS15
1	50475.0	50525.0	123.92	-	ALCAM
1	108425.0	108475.0	177.39	-	SCHIP1
1	119625.0	119675.0	100.03	-	TM4SF4
1	119775.0	119825.0	170.80	-	TM4SF18
1	144175.0	144225.0	-	59.78	TFF1,TFF2,TMPRSS3
2	***4775.2***	***4825.2***	***135.93***	***32.50***	***GPR17,LIMS2,MYO7B***
2	7125.2	7175.2	166.17	-	COL5A2
2	7325.2	7375.2	-	35.68	COL3A1
2	17675.2	17725.2	106.49	-	SESTD1
2	***18825.2***	***18875.2***	***139.00***	***34.10***	***PDE11A***
2	***18975.2***	***19025.2***	***109.50***	***42.65***	***PDE11A***
2	19125.2	19175.2	137.44	-	PDE11A
2	***36275.2***	***36325.2***	***110.12***	***57.88***	***ITGB6***
2	36475.2	36525.2	-	40.28	PLA2R1,SNORA21
2	72275.2	72325.2	-	38.53	EPB41L5,U4
2	111425.2	111475.2	135.14	-	SGPP2
2	111475.2	111525.2	111.56	-	FARSB,SGPP2
2	112225.2	112275.2	129.11	-	U6
2	122675.2	122725.2	102.16	-	FABP3,TINAGL1
2	134025.2	134075.2	121.78	-	UBR4
3	***7325.2***	***7375.2***	***192.86***	***32.52***	***NOS1AP***
3	8775.2	8825.2	121.44	-	CD244
3	9875.2	9925.2	115.33	-	CCDC19,IGSF9,TAGLN2, bta-mir-1584
3	11176.3	11226.3	-	58.90	OR10Z1,SPTA1
3	25776.3	25826.3	-	33.37	MAN1A2
3	41526.3	41576.3	-	32.93	OLFM3
3	50925.2	50975.2	100.44	-	EVI5

3	51075.2	51125.2	129.38	-	GFI1
3	51225.2	51275.2	118.46	-	RPAP2
3	57725.2	57775.2	148.56	-	CLCA1,CLCA2
3	72825.2	72875.2	104.89	-	U6
3	98476.3	98526.3	-	47.43	SPATA6
3	102126.3	102176.3	-	55.50	RNF220
3	103176.3	103226.3	-	36.33	MED8,SZT2
3	105676.3	105726.3	-	33.02	SCMH1
3	***109525.2***	***109575.2***	***154.19***	***38.01***	***GRIK3***
3	109575.2	109625.2	321.23	-	GRIK3
3	109625.2	109675.2	106.40	-	GRIK3
4	5725.0	5775.0	-	38.69	ZPBP
4	11775.1	11825.1	126.38	-	CASD1
4	32325.1	32375.1	131.71	-	CCDC126
4	32375.1	32425.1	204.82	-	CCDC126
4	32425.1	32475.1	120.78	-	DBF4
4	33825.1	33875.1	176.61	-	GRM3
4	50925.1	50975.1	118.19	-	CTTNBP2,U6
4	52225.1	52275.1	112.17	-	CAV2
4	68425.1	68475.1	103.51	-	JAZF1
4	70075.0	70125.0	-	35.73	SNX10
4	70175.0	70225.0	-	42.98	HNRNPA2B1,NFE2L3
4	92425.0	92475.0	-	42.60	ZNF800
4	99175.0	99225.0	-	33.27	BPGM
4	99225.0	99275.0	-	39.63	BPGM
4	108125.0	108175.0	-	38.12	OR2A2
4	113875.0	113925.0	-	34.71	GIMAP7
5	33175.6	33225.6	-	48.89	PCED1B
5	36925.6	36975.6	-	91.52	PUS7L
5	47675.6	47725.6	-	42.46	GRIP1,HELB
5	58825.6	58875.6	117.47	-	OR6C76
5	59625.6	59675.6	-	40.01	OR10A7
5	78575.6	78625.6	122.90	-	AMN1
5	80175.6	80225.6	-	34.70	TMTC1
5	83025.6	83075.6	-	38.41	STK38L
5	83725.6	83775.6	-	42.72	ITPR2
5	92425.6	92475.6	146.52	-	RERGL
5	95225.6	95275.6	-	60.20	RERG
5	97325.6	97375.6	-	33.20	HEBP1

5	97375.6	97425.6	-	51.00	GPRC5D,HEBP1
5	100775.6	100825.6	-	44.52	CLECL1
5	103375.6	103425.6	-	42.03	CD163L1
5	103425.6	103475.6	-	46.69	BT.105910,CD163L1
5	105525.6	105575.6	-	61.32	KCNA5
5	105625.6	105675.6	133.18	-	KCNA1
5	107525.6	107575.6	-	34.64	IQSEC3
5	114275.6	114325.6	138.14	-	ARFGAP3
5	114625.6	114675.6	113.57	-	SCUBE1,TTLL12
6	20225.5	20275.5	-	49.58	TBCK
6	23125.5	23175.5	-	42.42	SLC9B2
6	59925.2	59975.2	136.43	-	KLHL5
6	60175.2	60225.2	108.87	-	KLB,LIAS,RPL9
6	60225.2	60275.2	191.77	-	LIAS,UGDH
6	68175.5	68225.5	-	35.82	CORIN
6	68225.5	68275.5	-	33.95	NFXL1
6	68375.5	68425.5	-	66.71	NIPAL1,TXK
6	99025.5	99075.5	-	34.32	ENOPH1,TMEM150C,U4
6	116575.2	116625.2	145.87	-	LDB2
6	116925.2	116975.2	132.87	-	LDB2
6	118225.5	118275.5	-	40.09	TBC1D14
7	22525.2	22575.2	108.13	-	C19orf35,LINGO3,LSM7
7	24325.2	24375.2	123.81	-	CDC42SE2
7	25375.2	25425.2	101.50	-	CHSY3
7	39425.2	39475.2	191.54	-	EIF4E1B,GPRIN1,SNCB, TSPAN17
7	39475.2	39525.2	124.40	-	TSPAN17
7	43025.1	43075.1	-	44.74	OR2AJ1
7	43125.1	43175.1	-	33.00	OR2L13
7	43775.2	43825.2	108.04	-	MGC137030
7	53775.1	53825.1	-	40.92	PCDHA13
8	1227.8	1277.8	-	42.17	NEK1
8	1327.8	1377.8	-	48.37	NEK1,U6
8	10777.6	10827.6	119.75	-	PBK,SCARA5,U6
8	18527.6	18577.6	119.00	-	TUSC1
8	31277.6	31327.6	107.08	-	MPDZ
8	47177.8	47227.8	-	38.34	TRPM3
8	60727.8	60777.8	-	121.24	RECK,SNORA40
8	63677.6	63727.6	125.60	-	CORO2A,TRIM14
8	63727.6	63777.6	119.23	-	CORO2A,TBC1D2

8	80627.6	80677.6	106.10	-	GOLM1,NAA35
8	88377.8	88427.8	-	37.99	SYK
8	91577.6	91627.6	101.38	-	CCRK
8	108827.6	108877.6	186.69	-	TLR4
9	34475.2	34525.2	-	51.82	SULT1C4
9	90075.2	90125.2	-	66.94	ESR1
9	103275.2	103325.2	-	34.54	RNASET2
10	7526.2	7576.2	-	34.14	IQGAP2
10	25526.2	25576.2	-	49.06	TRAV3
10	27226.2	27276.2	102.41	-	OR4N2
10	27976.2	28026.2	176.39	-	OR4F15
10	30426.2	30476.2	132.91	-	AQR
10	30476.2	30526.2	124.06	-	AQR,ZNF770
10	30526.2	30576.2	206.78	-	ZNF770
10	47776.2	47826.2	119.09	-	TLN2
10	***55476.2***	***55526.2***	***117.77***	***37.63***	***SNORA25***
10	61326.2	61376.2	106.42	-	COPS2
10	61376.2	61426.2	178.33	-	SECISBP2L
10	65526.2	65576.2	132.79	-	DUOX2,DUOXA1,DUOXA2, SORD
10	70976.2	71026.2	100.11	-	KIAA0586
10	***76526.2***	***76576.2***	***132.97***	***42.50***	***SYNE2***
10	79126.2	79176.2	-	40.74	GPHN
10	82926.2	82976.2	-	37.21	PCNX
11	9180.3	9230.3	100.86	-	GPR45,TGFBRAP1
11	10680.3	10730.3	153.87	-	ACTB,DGUOK
11	10730.3	10780.3	414.54	-	ACTB,STAMBP
11	12880.3	12930.3	207.84	-	DYSF
11	12930.3	12980.3	190.54	-	DYSF
11	12980.3	13030.3	114.28	-	DYSF
11	14130.3	14180.3	169.59	-	FAM136A,XDH
11	15630.3	15680.3	101.50	-	LTBP1
11	16080.3	16130.3	198.34	-	RASGRP3
11	19830.3	19880.3	-	69.18	QPCT
11	31080.3	31130.3	-	33.84	FSHR
11	40530.3	40580.3	-	35.31	VRK2
11	40730.3	40780.3	-	46.52	FANCL
11	44730.3	44780.3	204.47	-	SULT1C2,SULT1C3
11	***45680.3***	***45730.3***	***191.85***	***42.93***	***C11H2orf40***
11	48530.3	48580.3	-	36.16	IMMT,MRPL35

11	***56530.3***	***56580.3***	***100.17***	***38.39***	***REG3G***
11	***56630.3***	***56680.3***	***135.22***	***39.83***	***PTP***
11	68380.3	68430.3	111.85	-	PCBP1
11	75280.3	75330.3	109.39	-	KLHL29
11	97430.3	97480.3	106.39	-	LMX1B
11	100680.3	100730.3	-	43.02	HMCN2
11	104530.3	104580.3	-	41.82	DBH,SARDH
11	106230.3	106280.3	-	37.37	C11H9ORF142,CLIC3,LCN12, LCNL1, PTGDS
12	12726.7	12776.7	182.69	-	TNFSF11
12	***12776.7***	***12826.7***	***149.16***	***42.70***	***TNFSF11***
12	13376.7	13426.7	109.07	-	ENOX1
12	13426.7	13476.7	102.68	-	ENOX1
12	13476.7	13526.7	213.64	-	ENOX1
12	20777.6	20827.6	-	39.27	FAM124A
12	21526.7	21576.7	118.98	-	NEK3,NEK5
12	21576.7	21626.7	100.58	-	CKAP2,NEK3
12	29677.6	29727.6	-	47.51	B3GALTL
12	61027.6	61077.6	-	34.64	SLITRK6
12	69976.7	70026.7	171.02	-	ABCC4
12	76726.7	76776.7	162.39	-	CLDN10
12	76976.7	77026.7	176.01	-	DNAJC3
12	77026.7	77076.7	211.46	-	UGGT2
12	77226.7	77276.7	110.49	-	HS6ST3
13	3526.2	3576.2	-	35.57	MKKS
13	11625.4	11675.4	194.08	-	CCDC3
13	11725.4	11775.4	163.75	-	CAMK1D
13	11775.4	11825.4	129.24	-	CAMK1D
13	12175.4	12225.4	157.45	-	CDC123
13	12225.4	12275.4	128.18	-	CDC123,NUDT5,SEC61A2
13	12475.4	12525.4	117.27	-	PROSER2
13	12525.4	12575.4	125.37	-	PROSER2
13	16725.4	16775.4	133.40	-	U6
13	21226.2	21276.2	-	46.49	MALRD1
13	30325.4	30375.4	113.06	-	ITGA8
13	34576.2	34626.2	-	39.53	ZNF438
13	***51575.4***	***51625.4***	***106.18***	***37.47***	***RNF24***
13	53025.4	53075.4	103.76	-	TMC2
13	57576.2	57626.2	-	67.41	EDN3

13	59975.4	60025.4	104.80	-	FAM209B,GCNT7,RTFDC1
13	61476.2	61526.2	-	38.43	DEFB119
13	63075.4	63125.4	117.91	-	BSP30C
13	63675.4	63725.4	142.39	-	ACTL10,C20orf144,E2F1, NECAB3
13	63775.4	63825.4	127.24	-	CHMP4B,ZNF341
13	63825.4	63875.4	110.95	-	CHMP4B
13	63975.4	64025.4	100.51	-	RALY
13	64325.4	64375.4	157.52	-	ITCH
13	64375.4	64425.4	104.63	-	ITCH
13	64425.4	64475.4	108.54	-	DYNLRB1,ITCH
13	64475.4	64525.4	149.72	-	DYNLRB1,MAP1LC3A,PIGU
13	65125.4	65175.4	107.34	-	MMP24
13	66175.4	66225.4	143.34	-	DLGAP4
13	66275.4	66325.4	129.38	-	DLGAP4,MYL9
13	67525.4	67575.4	118.19	-	RPRD1B,TTI1
14	1175.2	1225.2	-	68.11	TMEM60
14	9775.2	9825.2	110.70	-	KCNQ3
14	9775.2	9825.2	-	33.91	KCNQ3
14	19725.2	19775.2	190.28	-	HAS2
14	24325.2	24375.2	120.62	-	XKR4
14	24375.2	24425.2	140.06	-	XKR4
14	***24425.2***	***24475.2***	***118.63***	***36.80***	***XKR4***
14	24575.2	24625.2	-	50.54	XKR4
14	***24875.2***	***24925.2***	***147.86***	***65.68***	***LYN***
14	24925.2	24975.2	-	34.28	RPS20,U1,snoU54
14	24975.2	25025.2	110.50	-	C-MOS,PLAG1
14	***25025.2***	***25075.2***	***148.19***	***53.29***	***CHCHD7***
14	27875.2	27925.2	-	55.24	RAB2A
14	34425.2	34475.2	112.04	-	C14H8orf34
14	34475.2	34525.2	143.14	-	C14H8orf34
14	37675.2	37725.2	-	58.00	TRPA1
14	39775.2	39825.2	-	43.67	GDAP1
14	45525.2	45575.2	-	37.86	TPD52
14	47975.2	48025.2	-	42.49	SAMD12
14	54175.2	54225.2	-	46.17	CSMD3
14	66375.2	66425.2	-	39.94	FBXO43,SPAG1
14	68675.2	68725.2	132.67	-	LAPTM4B
14	69325.2	69375.2	125.66	-	CPQ
14	75675.2	75725.2	114.73	-	TMEM55A

14	78525.2	78575.2	140.73	-	CPNE3,RMDN1
14	78675.2	78725.2	196.52	-	WWP1
15	6125.6	6175.6	104.97	-	MMP1,MMP1
15	6625.7	6675.7	-	51.12	BIRC3
15	25375.7	25425.7	-	95.32	NXPE4
15	37875.7	37925.7	-	77.46	INSC
15	39675.7	39725.7	-	47.07	SNORA3
15	***40425.6***	***40475.6***	***142.95***	***65.69***	***TEAD1***
15	40875.6	40925.6	208.89	-	MICALCL,PARVA
15	40975.6	41025.6	253.65	-	MICAL2,MICALCL
15	***41025.6***	***41075.6***	***147.95***	***55.80***	***MICAL2***
15	41225.6	41275.6	147.95	-	DKK3
15	41275.6	41325.6	106.42	-	DKK3,USP47
15	***41575.6***	***41625.6***	***112.65***	***36.92***	***GALNTL4***
15	41625.6	41675.6	101.10	-	GALNTL4
15	44225.6	44275.6	113.61	-	AKIP1,ASCL3, C11ORF16, TMEM9B
15	48975.6	49025.6	138.03	-	HBB,HBE4,U1
15	51775.7	51825.7	-	81.93	RRM1
15	***51875.6***	***51925.6***	***119.19***	***55.74***	***STIM1***
15	***54275.6***	***54325.6***	***110.60***	***35.65***	***C2CD3***
15	55925.7	55975.7	-	43.37	DGAT2
15	64125.6	64175.6	111.38	-	CCDC73
15	64175.6	64225.6	150.60	-	CCDC73
15	64225.6	64275.6	108.98	-	CCDC73
15	75025.7	75075.7	-	43.57	ACCSL,ACS,EXT2
16	11325.2	11375.2	153.79	-	U2
16	19575.2	19625.2	102.20	-	USH2A
16	30125.1	30175.1	-	33.76	PARP1
16	30125.2	30175.2	142.83	-	PARP1
16	34525.1	34575.1	-	45.90	SDCCAG8
16	34575.1	34625.1	-	57.25	SDCCAG8
16	***44525.1***	***44575.1***	***105.95***	***33.88***	***CTNNBIP1***
16	44625.2	44675.2	127.22	-	CLSTN1,PIK3CD
16	49275.2	49325.2	122.24	-	LAD1,TNNI1
16	49425.2	49475.2	209.69	-	NAV1
16	49525.2	49575.2	169.63	-	NAV1
16	49575.2	49625.2	101.16	-	IPO9,LMOD1,SHISA4
17	577.0	627.0	-	42.42	CPE
17	10727.0	10777.0	-	43.42	EDNRA

17	14227.0	14277.0	388.94	-	GPA
17	14277.0	14327.0	174.46	-	GYPB
17	14527.0	14577.0	102.32	-	SMARCA5
17	14577.0	14627.0	160.87	-	GAB1,SMARCA5
17	14627.0	14677.0	122.02	-	GAB1
17	14727.0	14777.0	184.64	-	GAB1
17	18177.0	18227.0	-	41.26	MAML3
17	41377.0	41427.0	-	83.06	RXFP1
17	57977.0	58027.0	157.32	-	CIT,PRKAB1
17	58077.0	58127.0	-	36.40	CCDC60
17	***58127.0***	***58177.0***	***137.10***	***44.50***	***CCDC60***
17	68377.0	68427.0	-	40.05	HPS4,SRRD,TFIP11
17	72327.0	72377.0	104.26	-	EIF4ENIF1,SFI1
17	74777.0	74827.0	146.20	-	GP1BB,SEPT5
18	14678.1	14728.1	-	52.66	FANCA,SPIRE2
18	14728.1	14778.1	-	34.92	DEF8,MC1-R,TCF25,TUBB3
18	14928.1	14978.1	-	34.12	SHCBP1
18	22328.1	22378.1	-	42.87	FTO
18	22678.1	22728.1	154.61	-	IRX3
18	24278.1	24328.1	-	45.31	AMFR,NUDT21
18	24328.1	24378.1	-	48.54	AMFR,GNAO1
18	25328.1	25378.1	111.93	-	ARL2BP,PLLP,RSPRY1
18	26528.1	26578.1	116.59	-	GOT2
18	45178.1	45228.1	-	35.39	U6
18	45728.1	45778.1	-	69.72	ZNF599
18	51128.1	51178.1	128.32	-	CEACAM1
18	57428.1	57478.1	116.52	-	KLK10,KLK11,KLK12, KLK13, KLK8, KLK9
18	58428.1	58478.1	-	54.49	5S_rRNA,PPP2R1A
18	61078.1	61128.1	107.46	-	NLRP12
18	62778.1	62828.1	111.98	-	KIR2DL5A,RDH13
19	1025.1	1075.1	-	54.81	CA10
19	5275.1	5325.1	127.86	-	COX11,STXBP4,TOM1L1
19	7575.1	7625.1	-	35.45	NOG
19	8925.1	8975.1	-	38.44	CUEDC1
19	16275.1	16325.1	-	46.97	U6
19	19525.1	19575.1	152.15	-	KSR1
19	24525.1	24575.1	-	70.37	U6
19	24775.1	24825.1	-	102.62	SPATA22

19	46625.1	46675.1	-	43.19	KANSL1,MAPT
19	50625.1	50675.1	-	42.36	FOXK2,WDR45L
19	***50675.1***	***50725.1***	***278.21***	***86.72***	***FOXK2***
19	50925.1	50975.1	-	109.35	bta-mir-2345,bta-mir-2345
19	52575.1	52625.1	-	44.31	RPTOR
19	57575.1	57625.1	157.21	-	CD300LB
20	20475.1	20525.1	174.87	-	RAB3C
20	39075.1	39125.1	-	37.61	PRLR
20	53525.1	53575.1	-	38.76	CDH18
21	725.4	775.4	160.09	-	NDN
21	775.4	825.4	117.49	-	MAGEL2
21	1575.4	1625.4	226.02	-	5S_rRNA,U6
21	1825.4	1875.4	146.75	-	OR5D13
21	2775.4	2825.4	131.25	-	ATP10A
21	2825.4	2875.4	109.29	-	ATP10A
21	***2875.4***	***2925.4***	***159.15***	***45.17***	***ATP10A***
21	33075.4	33125.4	101.20	-	HMG20A
21	45785.2	45835.2	-	33.79	FAM177A1,KIAA0391, PPP2R3C
21	59185.2	59235.2	-	38.95	OTUB2
21	62135.2	62185.2	-	58.24	TCL1A
21	64585.2	64635.2	-	32.92	5S_rRNA
21	70225.4	70275.4	108.85	-	ASPG,TDRD9
22	19276.1	19326.1	144.86	-	GRM7
22	19576.1	19626.1	107.95	-	GRM7
22	33476.1	33526.1	174.96	-	FAM19A1
22	33526.1	33576.1	99.81	-	FAM19A1
22	37376.1	37426.1	111.91	-	PRICKLE2
22	38876.1	38926.1	102.07	-	CADPS
22	39326.1	39376.1	-	54.85	PTPRG
22	43826.1	43876.1	147.84	-	SLMAP
22	54726.1	54776.1	185.18	-	CDCP1
22	54926.1	54976.1	104.08	-	GHRL,SEC13
23	825.7	875.7	-	45.62	KHDRBS2
23	28475.7	28525.7	-	41.23	BOLA
23	28875.7	28925.7	-	37.31	U6,UBD
23	34525.7	34575.7	-	69.11	PRP1,U6
23	37025.7	37075.7	-	42.06	CDKAL1
23	49325.7	49375.7	147.06	-	LYRM4
24	7375.7	7425.7	111.88	-	RTTN

24	21675.7	21725.7	-	59.34	GALNT1
24	49925.7	49975.7	-	34.36	ACAA2,MYO5B,SCARNA17, SCARNA18
24	57175.7	57225.7	-	38.24	ONECUT2,U6
25	2825.0	2875.0	-	56.83	NAA60,ZNF174,ZNF597
25	24925.0	24975.0	134.78	-	C25H16orf82
25	41725.0	41775.0	113.57	-	MAD1L1
25	41775.0	41825.0	186.00	-	ELFN1,MAD1L1
26	24075.6	24125.6	-	33.03	NT5C2
26	24675.6	24725.6	147.91	-	ASMTL,OBFC1
26	47175.6	47225.6	-	53.96	DOCK1
27	30475.6	30525.6	-	43.07	UNC5D
27	36575.6	36625.6	157.84	-	KAT6A,U6
27	***36725.6***	***36775.6***	***291.20***	***68.57***	***AP3M2,PLAT***
27	36775.6	36825.6	144.74	-	IKBKB
28	2125.6	2175.6	112.86	-	RHOU
28	8826.2	8876.2	-	94.90	NID1
28	10826.2	10876.2	-	40.10	5S_rRNA
28	15575.6	15625.6	108.10	-	CCDC6
28	15625.6	15675.6	115.99	-	CCDC6
28	25425.6	25475.6	125.16	-	KBP
28	25525.6	25575.6	312.98	-	U6
28	25625.6	25675.6	129.10	-	SRGN,snoU2-30,snoU2_19
28	***25775.6***	***25825.6***	***190.39***	***75.17***	***HKDC1***
28	25825.6	25875.6	103.52	-	HK1
29	18225.0	18275.0	-	37.31	AAMDC,INTS4
29	***24975.0***	***25025.0***	***112.96***	***69.61***	***DBX1***
29	27525.0	27575.0	135.20	-	TMEM225
29	39425.0	39475.0	-	33.25	MGC157408
29	49225.0	49275.0	-	50.94	CARS,NAP1L4

Common genes between breed comparisons are indicated with **bold** and *italic* font type

To further understand the biological functions of the genes identified from both comparisons, I submitted the gene lists to WebGestalt gene ontology analysis tool (Wang et al. 2013b). Here, I used the KEGG pathway and GO-BP terms analysis. As a result, sixteen and five KEGG pathways were significantly enriched (adj P-value <0.05) from A-H and A-J gene lists, respectively (Figure 5.3). However, no intersecting GO-BP terms were found significant (adj P-value >0.05) from the comparisons. In this study, I described those pathways and genes that affect the phenotypes of Angus cattle based on literature and their biological function (Table 5.2).

5.4.3.1 Pathways and genes related to meat quality traits

Meat quality is a complex and multi-factorial trait affected by genetic and non-genetic factors. Genes that are involved in different biological and cellular functions including muscle growth, glycolysis, adipogenesis, muscle contraction, stress reaction, proteolysis, and apoptosis influence meat quality traits such as intramuscular fat (IMF), tenderness and drip loss (Ladeira et al. 2016). In this study, significantly enriched pathways affecting meat quality traits include insulin signaling (A-H), type II diabetes mellitus pathway (A-H), focal adhesion pathway (both comparisons) and ECM-receptor interaction (A-H) (Figure 5.3). These pathways affect the amount, distribution, and composition of fat in meat which is a determinant factor for its quality (Ladeira et al. 2016).

Insulin signaling and type II diabetes mellitus pathways are related to adipogenesis and IMF deposition (Cui et al. 2012a; Ladeira et al. 2016). Insulin stimulates the expression of genes that encode lipogenic enzymes in the adipose tissue (Ladeira et al. 2016). Focal adhesions are related to cell junction that connects the cytoskeleton of a cell to the ECM (Li et al. 2010). It is related to lipid metabolism and influences IMF deposition (Cui et al. 2012a). Genes in this pathway (*ACTB, COL5A2, MYL9, PARVA, PIK3CD,* and *TLN2*) are involved in muscle development and influences

tenderness and texture of meat (Li et al. 2010; Cui et al. 2012a; Lee et al. 2013a). ECM-receptors have an important role in adipogenesis and meat tenderness (Li et al. 2010). It has been found enriched from genes differentially expressed from fat depots of omental, subcutaneous and intramuscular fat (Hausman et al. 2009; Cui et al. 2012a; Lee et al. 2013a). Genes involved in ECM-receptor interaction were previously found significantly up-regulated in subcutaneous fat and intramuscular fat (Lee et al. 2013a). The positive selection of genes related to meat quality traits in Angus cattle as compared to Hanwoo cattle might be because of the differences in demographic and selection histories that differentiate the allele frequency spectrum between them (Utsunomiya et al. 2015).

Regulation of eicosanoid secretion was significantly enriched ($p<0.05$) in the ClueGO network of A-J gene list (Figure 5.4b). Eicosanoids are signaling molecules derived from enzymatic or non-enzymatic oxidation of essential fatty acids like arachidonic acid (Madsen et al. 2005). Fatty acids influence adipogenesis as pre-cursors for the eicosanoids, as well as regulators of transcription. *PLA2R1*, involved in this network, is a fatty acid transporter that induces cell proliferation and serves as a receptor of a phospholipase for the production of lipid mediators. It has been found associated with fat deposition, body weight and egg production performance in chicken (Gheyas et al. 2015).

Besides pathways, *FABP3* and *TNNI1*genes were identified in A-H comparison in relation to meat quality traits (Table 5.3). *FABP3* is a protein-coding gene that plays a role in intracellular transport of long-chain fatty acids and their acyl-CoA esters. Polymorphisms in *FABP3* gene has been found associated with intramuscular fat and fatty acid composition in porcine meat (Puig-Oliveras et al. 2016), cytosolic fatty acid and lipid binding (Berton et al. 2016), and beef ribeye area and ribeye area to hot carcass weight ratio (Blecha et al. 2015). *TNNI1*, a gene expressed in the slow-twitch skeletal muscle fibers, is a constituent protein of the troponin complex located on the

thin filaments of striated muscle to which its expression affects meat quality traits through its effect on muscle fiber composition (Yang et al. 2010). Polymorphisms in the gene region is associated with intramuscular fat, marbling score and pork color in Large-White Meishan pigs (Yang et al. 2010), drip loss and compression force in Mong Cai pigs (Ngu and Nhan 2012), and pH24 and drip loss of Longissimus Dorsi muscle in pork (Pierzchala et al. 2014).

In the A-J comparison, *ACS*, *ACAA2*, *FTO*, *DGAT2*, and *CPE* genes were identified in relation to meat quality (Table 5.3). Acetyl-CoA (*ACS*, *ACAA2*) genes are essential for de novo fatty acid synthesis (Ladeira et al. 2016). These genes play a role in the activation of long-chain fatty acids for the synthesis of cellular lipids and degradation via beta-oxidation. *ACS/ACSL1*, called Long-Chain-Fatty-Acid-CoA Ligase 1, has been found differentially expressed and upregulated for saturated fatty acids (SFA - palmitic, stearic, oleic fatty acids) in the Longissimus Dorsi muscle of high-fat Nellore cattle (Berton et al. 2016). *ACAA2*, acetyl-CoA acyltransferase 2, is expressed in the subcutaneous fat tissue of beef cattle involved in adipogenesis (Wang et al. 2013a). Screening for non-synonymous mutations representing putative functional variants, four (one known - rs211177037; and three novel – 27:14225708, 27:14245782, 27:14252901) and two new (24:49919351, 24:49933799) missense variants were identified in *ACS/ACSL1* and *ACAA2* gene regions, respectively (Table 5.4).

FTO is a nuclear protein of the AlkB related non-haem iron and 2-oxoglutarate-dependent oxygenase superfamily that has a role in the regulation of global metabolic rate, energy expenditure and homeostasis, body size and fat accumulation, and control of adipocyte differentiation into fat cells. It is found highly expressed in fat tissue contributing to fattier phenotype in pigs (Tempfli et al. 2016), associated with fatness-related traits such as intramuscular fat deposition and backfat thickness in beef cattle (Wei et al. 2011) and associated with marbling score in Hanwoo cattle (Chung 2014).

In this gene region, one novel (14:22243331) and one known (rs381025074) missense variants were identified (Table 5.4).

DGAT2 encodes an enzyme which catalyzes the synthesis of triglycerides. It has been found associated with the accumulation of SFA in the intramuscular tissue of Nellore cattle with extreme values of fatty acid (Berton et al. 2016), and Korean beef cattle (Jeong et al. 2012). A novel missense variant (15:55971894) was identified in this gene. *CPE* is an enzyme involved in the production of neuroendocrine peptide hormones and neuropeptides including insulin, vasopressin, and oxytocin. Mutation in this gene is associated with obesity and infertility in mice (Naggert et al. 1995), marbling score and breeding value of back fat thickness in Hanwoo beef cattle (Shin and Chung 2007), and found upregulated for SFA in the Longissimus Dorsi muscle of high-fat Nellore cattle (Berton et al. 2016). Two novel (17:680132, 17:692409) and one known (rs210567645) missense variants were identified in this gene region.

DNAJ genes (*DNAJC28, DNAJC3*) have an anti-apoptotic role that is important for meat tenderness (O'brien et al. 2014). The expression of a gene family (*DNAJA1* - not identified here) has been reported to characterize 60% of the variations in meat tenderness of Charolaise cattle (Bernard et al. 2007). *MYL9* is a gene related to muscle biology and accretion. O'brien et al. (2014), reported the positive selection of *DNAJA1* gene in Angus cattle. In the gene regions, five (*DNAJC3*), five (*DNAJC28*) missense variants were identified (Table 5.4).

The positive selection of these putative genes and pathways enriched might contribute to the superior beef quality of Angus cattle. Angus cattle are known beef breed developed for its quality, higher growth rate and feed conversion efficiency (Arthur et al. 2001; McClure et al. 2010).

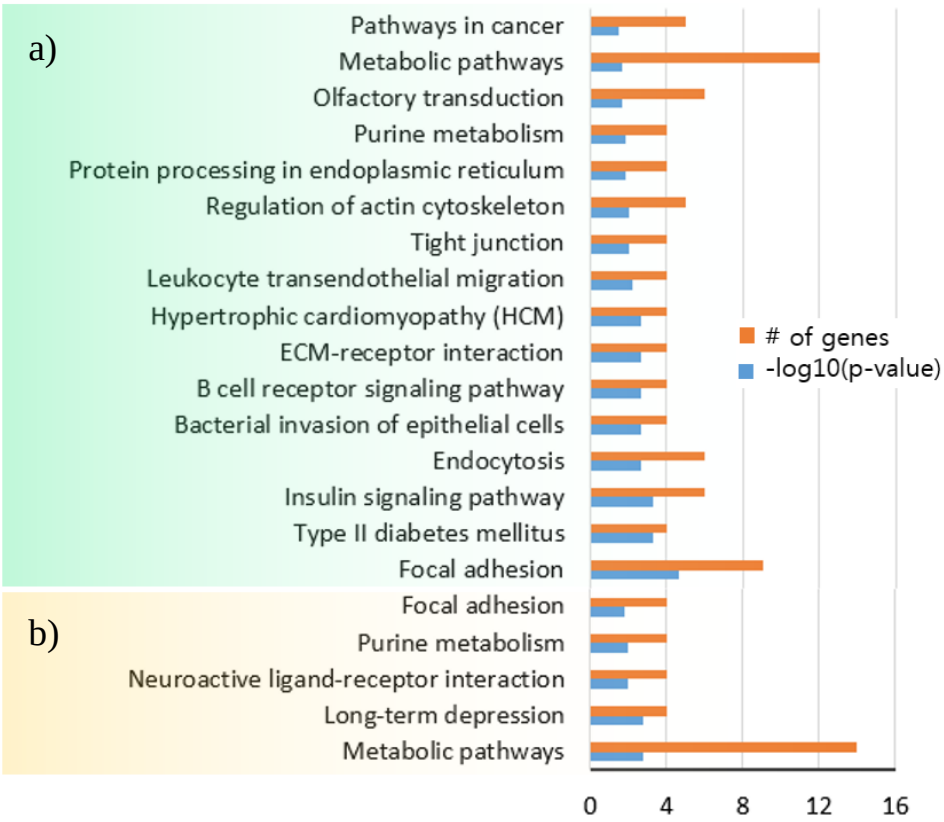


Figure 5.3 Plot of number of genes overrepresented in KEGG pathways enriched from genes identified from Angus-Hanwoo (a), and Angus-Jersey (b) XP-CLR comparisons with the respective *P*-values. Orange color bars are number of genes involved in the pathways and the blue ones are the associated *P*-values.

Table 5.2 KEGG pathways enriched from genes identified under selection in the genome of Angus cattle based on WEB-based GEne SeT AnaLysis Toolkit

KEGG pathways	Genes involved in the pathway	Adj. P-value	Enrich. ratio
Angus-Hanwoo			
Focal adhesion	MYL9, ACTB, PARVA, PIK3CD, COL5A2, TLN2, ITGB6, ITGA8, CAV2	2.5e-05	8.37
Type II diabetes mellitus	HKDC1, HK1, PIK3CD, IKBKB	0.0005	15.49
Insulin signaling pathway	EIF4E1B, HKDC1, PRKAB1, HK1, PIK3CD, IKBKB	0.0005	8.08
Endocytosis	ARFGAP3, WWP1, CHMP4B, STAMBP, CAV2, ITCH	0.0021	5.55
Bacterial invasion of epithelial cells	ACTB, PIK3CD, CAV2, GAB1	0.0021	10.62
B cell receptor signaling pathway	RASGRP3, PIK3CD, LYN, IKBKB	0.0021	9.91
ECM-receptor interaction	GP1BB, COL5A2, ITGB6, ITGA8	0.0024	8.75
Hypertrophic cardiomyopathy (HCM)	ACTB, PRKAB1, ITGB6, ITGA8	0.0024	8.96
Leukocyte transendothelial migration	MYL9, ACTB, PIK3CD, CLDN10	0.0066	6.41
Tight junction	MYL9, MPDZ, ACTB, CLDN10	0.0089	5.63
Regulation of actin cytoskeleton	MYL9, ACTB, PIK3CD, ITGB6, ITGA8	0.0089	4.36
Protein processing in endoplasmic reticulum	UGGT2, SEC13, DNAJC3, SEC61A2	0.0153	4.51
Purine metabolism	NUDT5, DGUOK, XDH, PDE11A	0.0153	4.59
Olfactory transduction	OR4F15, OR6C76, OR5D13, CLCA2, CLCA1, OR4N2	0.0214	2.87
Metabolic pathways	CHSY3, GOT2, SORD, UGDH, COX11, HKDC1, HK1, PIGU, LIAS, GALNTL4, DGUOK, XDH	0.0214	1.97
Pathways in cancer	MMP1, PIK3CD, CCDC6, IKBKB, E2F1	0.0322	2.85
Angus-Jersey			
Metabolic pathways	DBH, SARDH, NT5C2, GALNT1, GART, BPGM, PTGDS, MAN1A2, DGAT2, ACAA2, RRM1, HKDC1, GALNTL4, EXT2	0.0018	2.77
Long-term depression	ITPR2, PPP2R1A, LYN, GNAO1	0.0018	12.77
Neuroactive ligand-receptor interaction	FSHR, RXFP1, PRLR, EDNRA, GRIK3	0.0117	4.11
Purine metabolism	NT5C2, GART, PDE11A, RRM1	0.0117	5.52
Focal adhesion	BIRC3, ITGB6, DOCK1, COL3A1	0.0152	4.47

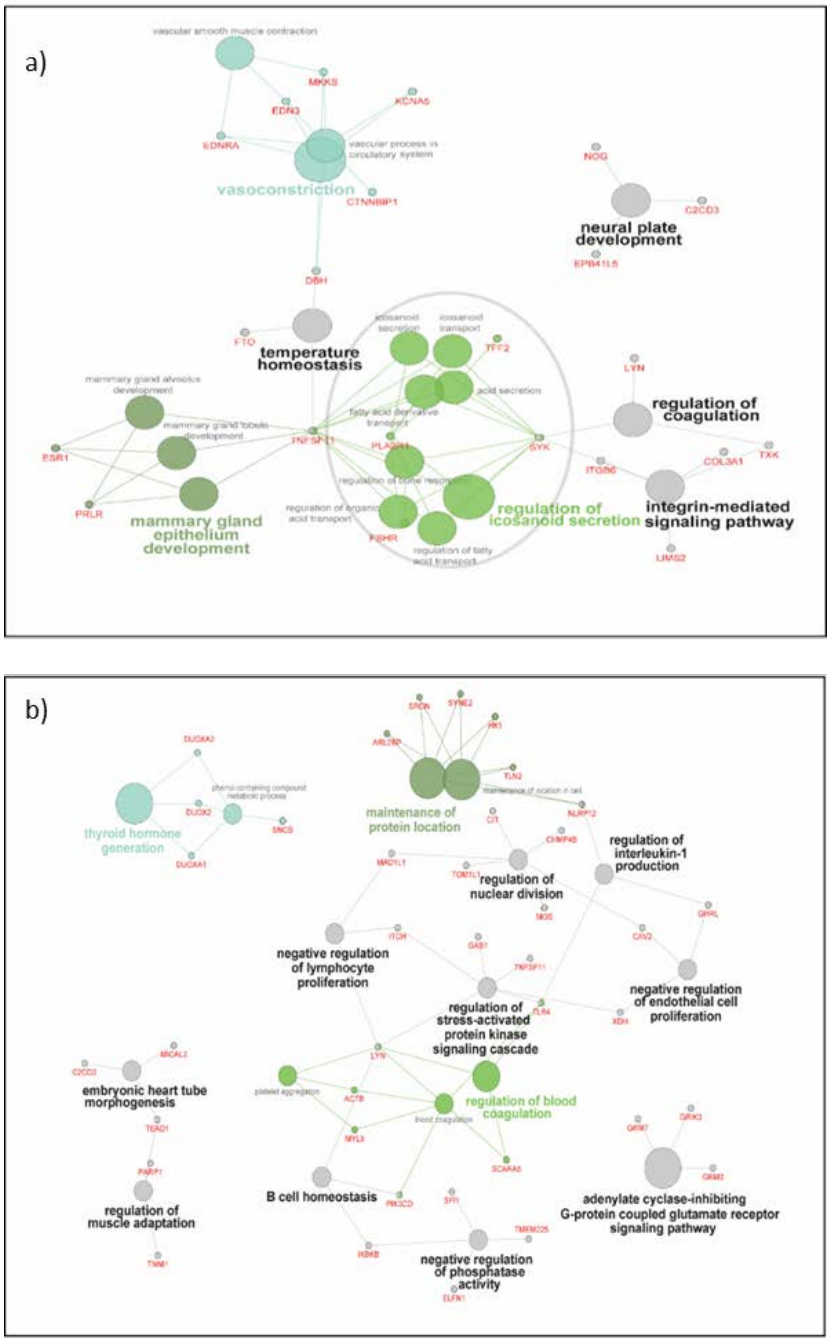


Figure 5.4 Clue-GO network enriched from genes identified under selection from Angus vs. Jersey (a), and Angus vs. Hanwoo (b) XP-CLR comparisons. Nodes represent gene ontology terms to which their size reflects the statistical significance of the terms. The most prominent gene ontology term for each group is highlighted in colors.

5.4.3.2 Genes affecting stature and body size

Body size in cattle, as measured by body weight, is an important trait in meat production. It is influenced by many genes of smaller effect size (Kemper et al. 2012). In relation to this, several genes including *PLAG1* (A-H), *RPS20* (A-J), *LYN*, and *CHCHD7* (both) were identified in this study (Table 5.3). *PLAG1* encodes a zinc finger protein whose activation results in up-regulation of genes that control growth leading to cell proliferation. Its association with height and body weight variation in different cattle breeds (Littlejohn et al. 2012; Utsunomiya et al. 2013; Takasuga 2016; Fink et al. 2017), and carcass weight in Japanese Wagyu cattle (Nishimura et al. 2012) was previously reported. *LYN* is a protein kinase gene that regulates cell proliferation, survival, differentiation, migration, adhesion, degranulation, and cytokine release. *RPS20* also is a ribosomal gene that catalyzes protein synthesis. Both, *LYN* and *RPS20* genes, have been found associated with body weight, stature, and pre-weaning average daily gain in Nellore cattle (Utsunomiya et al. 2013; Fink et al. 2017). *CHCHD7* is a eukaryotic protein consisting of two pairs of cysteines that form two disulfide bonds stabilizing a coiled coil–helix–coiled coil–helix (CHCH) fold (Cavallaro 2010). It affects carcass weight in Japanese Wagyu cattle (Nishimura et al. 2012), and height in Jersey and Holstein cattle (Utsunomiya et al. 2013; Fink et al. 2017). Searching for non-synonymous mutations, I identified one (14:25056041), and two (14:24880380, 14:24881091) novel missense variants in the *CHCHD7* and *LYN* gene regions, respectively (Table 5.4). Domestication and artificial selection for increased meat production have changed the physical and morphological structure of domesticated animals. Moreover, Angus cattle have been intensively artificially selected for higher beef production and quality; this resulted in Angus cattle being among the larger breeds with higher growth rate (Arthur et al. 2001; Chambaz et al. 2003; McClure et al. 2010; Kemper et al. 2012).

5.4.3.3 Genes affecting reproduction – dystocia and fertility

Selection for high growth rate in cattle compromise reproductive performances and cause dystocia. Genes affecting fetal growth, the size of the pelvis of the dam, mature body weight, and other factors have been reviewed to affect dystocia (Zaborski et al. 2016). Therefore, genes controlling these factors cause dystocia. In this study, quite a few genes including *ESR1, RPS20, PPP2R1A* (A-J), *GHRL, PLAG1* (A-H) and zinc finger proteins that directly or indirectly contribute to the factors associated with dystocia were identified under selection in Angus cattle (Table 5.3). *ESR1* is a nuclear receptor family of transcription factors that mediate cellular signaling of estrogens which have effects on reproduction at various stages of development (Lazari et al. 2009). *ESR1* is among the genes which influence calving difficulty (Zaborski et al. 2016). Genes influencing fetal growth including *RPS20* (Zaborski et al. 2016), *PPP2R1A* (Cole et al. 2014), *GHRL* (Maltecca et al. 2011), and *PLAG1* (Utsunomiya et al. 2013; Juma et al. 2016) affect dystocia. *PPP2R1A* is among the major Ser/Thr phosphatases involved in cell growth and signaling (Cole et al. 2014). *GHRL*, a ligand for growth hormone secretagogue receptor type 1 (GHSR), induces the release of growth hormone from the pituitary and reported to be linked to fetal growth (Maltecca et al. 2011). *PLAG1* is associated with birth weight in Nellore cattle (Utsunomiya et al. 2013) and its association with calving ease has been recently reviewed by Takasuga (2016). Zinc finger proteins regulate bone and skeletal development in mammals (Zaborski et al. 2016), and might be associated with birth difficulty. Regardless of higher growth performances in Angus cattle, Archer et al. (1998) reported non-significant differences of reproductive performances and incidence of dystocia between heifers selected for higher growth rate, control lines and those selected for low growth rate. The relatively smaller birth weight but the fast growth rate of Angus cattle might explain this http://extension.psu.edu/animals/beef/reproduction/ articles/regulating-birth-weight-in-beef-cattle.

Genes that are related to fertility (*FSHR*, *CORIN*) were also identified under selection. *FSHR* (A-J), a receptor for the follicle-stimulating hormone, is crucial for follicular development and estradiol production in females, and regulation of Sertoli cell function and spermatogenesis in males (Desai et al. 2013). Loss/gain of function mutation in this gene has been reported previously (Desai et al. 2013). *CORIN* (A-J) plays a role in female pregnancy by promoting trophoblast invasion and spiral artery remodeling in the uterus (Cui et al. 2012b; Soares et al. 2014). It is expressed in the pregnant mouse and human uterus to which its impaired expression is associated with preeclampsia, a major risk factor for placental abruption (Cui et al. 2012b; Nagashima et al. 2013). Sperm and spermatogenesis associated genes (*SPAG1*, *SPATA22*, and *SPATA6*) were also identified and may affect reproduction.

Mammary gland epithelium development was enriched in the ClueGO networks (Figure 5.4b). Genes involved in the network (*ESR1*, *PRLR*) affect reproductive performances in livestock. *PRLR*, a receptor for prolactin, have a role in reproduction function through mammary gland development, lactation, and regulation of maternal behavior. It is associated with embryonic survival rate (Khatib et al. 2009). Prolactin is a pleiotropic hormone that affects many physiological functions including mammary gland development, lactogenesis, and fertility (Donato Jr and Frazão 2016).

5.4.3.4 Genes affecting feeding efficiency

Feeding efficiency in beef cattle production is directly associated with the profitability of the farm (Do et al. 2014b). In relation to this, KEGG pathways of olfactory transduction (A-H), tight junction (A-H), and metabolic pathways (both) were enriched (**Error! Reference source not found.**). Olfactory transduction pathways affect the perception of odor through olfactory receptors and biochemical signaling events which as a result influence food preference and food consumption (Do et al. 2014b; Stafuzza et al. 2017). Olfactory transduction pathway has been found associated with residual feed intake (RFI) in pigs (Do et al. 2014b) and cattle (Abo-Ismail et al. 2014;

Zhao et al. 2015; Stafuzza et al. 2017). In a GWAS analysis of Angus cattle, tight junction and endocytosis pathways were enriched in relation to RFI and average feed intake (AFI), respectively (Rolf et al. 2012), that are measures of feeding efficiency. Olfactory receptor genes involved in olfactory transduction pathway (*OR4F15, OR6C76, OR5D13, OR4N2*) were previously identified under selection in African Sanga cattle for feeding intake (Taye et al. 2017). Additionally, genes associated with feeding efficiency including *GHRL* (Sherman et al. 2008), *PIK3CD, DNAJC28, DNAJC3* (Zhou et al. 2015), and *PLAG1* (Fortes et al. 2013) were identified in this study. *GHRL* is a powerful appetite stimulant and known to play a major role in energy homeostasis. Recently, Sherman et al. (2008) identified a polymorphism in this gene region that is associated with RFI and FCR in beef cattle. Selection efforts towards increased metabolic efficiency of cattle have resulted in decreased feed intake regardless of growth (Stockton 2003; Rolf et al. 2012). The positive selection of the aforementioned genes and pathways might contribute to the feeding efficiency of Angus cattle (Stockton 2003; Rolf et al. 2012).

5.4.3.5 Genes affecting coat color

Angus cattle are characterized by their solid black or red coat color (http://www.the-cattlesite.com/breeds/beef/7/aberdeen-angus/). Genes associated with coat color *(MC1-R, TUBB3* – A-J) were identified in this study. These genes encode a protein that controls melanogenesis and are involved in biological functions of pigmentation and inflammation. Polymorphism in *MC1-R* gene region was previously reported to affect black/red color in Angus cattle (Klungland et al. 1995), Holstein-Frisian cattle (Zhao et al. 2015), and sheep (Våge et al. 1999). Searching for non-synonymous genetic variants in this gene region, I identified three (one known - rs109688013, and two novel - 18:14758030, 18:14758197) missense variants (Table 5.4).

5.4.3.6 Genes affecting genetic disorders

Fawn Calf Syndrome (FCS) is a heritable abnormality of skeletal development reported in Angus cattle. It is a non-lethal developmental genetic defect in calves (Whitlock 2010). FCS is related to Marfan Syndrome in humans, a malformation of connective tissue within the skeletal system (http://calfology.com/library/wiki/contractural-arachnodactyly-fawn-calf-syndrome), that is caused by a mutation in fibrillin-1 gene (*FBN1*) (Jovanović et al. 2007) - not identified here. Instead, an integrin gene (*ITGB6* – A-H) which acts as a receptor for *FBN1* (Jovanović et al. 2007) was found under selection. A mutation in the integrin gene can affect the function of *FBN1* and lead to FCS in Angus cattle. I have identified two known (rs110694377, rs136500299) and one new (2:36325642) missense variants in this gene region (Table 5.4).

In addition, hypertrophic cardiomyopathy (HCM), an inherited cardiac disorder, is significantly ($p<0.05$) enriched in the KEGG pathways of A-H gene list (**Error! Reference source not found.**). Although it is not reported in Angus cattle, HCM is a notorious issue in Holstein cattle (Guziewicz et al. 2007). Genes involved in this pathway include *ACTB, PRKAB1, ITGB6,* and *ITGA8*. Gene families (*ACTC1, ITGAV,* and *ITGA2*) associated with cardiomyopathy were previously reported under selection in Holstein cattle (Lee et al. 2014a).

Dwarfism is a genetic disorder characterized by systemic skeletal disorders, including shortness and deformity of limbs, head, and vertebrae, that is recognized in multiple breeds of cattle including Angus cattle (Whitlock et al. 2008). The *PLAG1* gene that affects stature has been tested for a knockout experiment where the *plag1* K.O. mice showed dwarfism and low fertility (Takasuga 2016).

Table 5.3 Major candidate genes that affect the phenotypes of Angus cattle

Chr.	Gene start – end (Mbp)	XP-CLR score		Gene	Phenotype
		A-H	A-J		
2	122.68 - 122.73	102.15	-	*FABP3*	Meat quality
15	55.93 - 55.98	-	43.36	*DGAT2*	
15	75.03 - 75.08	-	43.57	*ACS/ACLS1*	
16	49.28 - 49.33	122.24	-	*TNNI1*	
17	0.58 - 0.63	-	42.42	*CPE*	
18	22.33 - 22.38	-	42.87	*FTO*	
24	49.93 - 49.98	-	34.36	*ACAA2*	
14	24.88 - 24.93	147.86	65.67	*LYN*	Stature and body size
14	24.93 - 24.98	-	34.28	*RPS20*	
14	24.98 - 25.03	110.49	-	*PLAG1*	
14	25.03 - 25.08	148.18	53.28	*CHCHD7*	
9	90.08 - 90.13	-	66.94	*ESR1*	Birth weight (Dystocia)
18	58.43 - 58.48	-	54.49	*PPP2R1A*	
22	54.93 - 54.98	104.08	-	*GHRL*	
20	39.08 - 39.13	-	37.61	*PRLR*	Fertility
11	31.08 - 31.13	-	33.83	*FSHR*	
6	68.18 - 68.23	-	35.82	*CORIN*	
1	1.28 - 1.33	-	55.62	*DNAJC28*	Feed efficiency
5	58.83 - 58.88	117.46	-	*OR6C76*	
10	27.98 - 28.03	176.38	-	*OR4F15*	
10	27.23 - 27.28	102.41	-	*OR4N2*	
12	76.98 - 77.03	176.00	-	*DNAJC3*	
16	44.63 - 44.68	127.22	-	*PIK3CD*	
21	1.83 - 1.88	146.74	-	*OR5D13*	
18	14.73 - 14.78	-	34.91	*MC1-R*	Coat color
2	36.28 - 36.33	110.12	-	*ITGB6*	Genetic disorder (Fawn syndrome)
13	30.33 - 30.38	113.06	-	*ITGA8*	

Table 5.4 Non-synonymous variants identified in the candidate genes that affect the phenotypes of Angus cattle

Chr.	Gene	Position	SNP ID	Reference allele	Alt allele	aa change
1	*DNAJC28*	1308807	.	G	C	Arg94Thr
		1309122	rs207980987	G	A	Arg199Gln
		1309202	.	A	G	Lys226Glu
		1309583	rs379072367	A	G	Asn353Asp
		1309597	rs381923971	T	A	Asn357Lys
2	*ITGB6*	36316042	rs110694377	G	T	Val400Phe
		36325642	.	C	A	Arg616Ser
		36336237	rs136500299	T	C	Phe667Ser
5	*OR6C76*	58849846	.	C	T	Val137Met
		58850040	rs209013236	A	G	Met72Thr
10	*OR4N2*	27267875	rs381724745	C	G	Ala15Gly
		27267980	.	G	A	Arg50Lys
		27267998	.	C	T	Thr56Met
		27268094	.	A	C	Glu88Ala
		27268115	.	G	C	Arg95Thr
		27268118	.	C	G	Ala96Gly
		27268237	rs210153791	A	C	Met136Leu
		27268349	rs209853697	G	A	Arg173Gln
		27268376	.	C	T	Pro182Leu
		27268502	.	G	A	Arg224His
		27268567	rs380990104	G	A	Val246Ile
		27268601	.	T	A	Ile257Asn
	OR4F15	28022318	.	T	C	Met9Thr
		28022420	.	A	G	Tyr43Cys
		28022542	rs207771141	A	G	Ile84Val
		28022567	rs110551835	G	A	Gly92Asp
		28022597	rs381882370	T	G	Phe102Cys
		28023008	rs380162138	G	T	Arg239Leu
		28023063	.	C	G	Phe257Leu
		28023155	.	C	T	Pro288Leu

Chr.	Gene	Position	SNP ID	Reference allele	Alt allele	aa change
		28023227	rs208795433	G	A	Arg312His
12	*DNAJC3*	77005369	.	C	A	Ser146Tyr
		77005399	rs133195147	G	A	Arg156His
		77005427	.	T	A	Phe165Leu
		77013009	rs381397345	G	A	Asp286Asn
		77013081	.	G	C	Glu310Gln
13	*ITGA8*	30424669	rs384652906	C	G	Ser631Thr
		30427281	.	A	G	Tyr585His
		30427338	rs378771473	T	C	Ile566Val
		30427390	.	T	A	Gln548His
		30434838	.	A	T	Ile508Asn
		30471177	rs137503638	G	T	Leu361Met
		30527780	rs109597731	C	T	Arg68His
14	*LYN*	24880380	.	C	T	Ser106Leu
		24881091	.	G	A	Val175Met
	CHCHD7	25056041	.	G	A	Arg6Gln
15	*DGAT2*	55971894	.	G	A	Arg319Lys
17	*CPE*	680132	.	G	A	Arg380Gln
		692375	rs210567645	G	A	Val414Ile
		692409	.	C	T	Pro425Leu
18	*MC1R*	14757910	rs109688013	T	C	Leu99Pro
		14758030	.	C	T	Ala139Val
		14758197	.	C	T	Leu195Phe
	FTO	22243331	.	G	A	Val57Met
		22354222	rs381025074	C	T	Ser441Leu
	PPP2R1A	58448152	.	G	A	Val560Ile
21	*OR5D13*	1853365	rs134558572	C	T	Ala252Thr
		1853415	rs109158513	C	T	Gly235Glu
		1853470	.	C	A	Val217Phe
		1853901	rs133366938	C	G	Asp74His
		1853915	.	T	A	His69Leu
		1853926	rs137247996	C	A	Leu65Phe

Chr.	Gene	Position	SNP ID	Reference allele	Alt allele	aa change
		1853952	rs134285494	C	G	Ala57Pro
		1853994	.	C	T	Val43Met
		1854083	.	C	T	Gly13Glu
24	*ACAA2*	49919351	.	T	C	Lys191Glu
		49933799	.	C	T	Arg38Lys
27	*ACS/ACLS1*	14225708	.	C	T	Val606Ile
		14236965	rs211177037	C	T	Arg377Gln
		14245782	.	T	C	Ile225Val
		14252901	.	C	T	Val103Met

5.5 Conclusion

In this study, the analysis of positive selection signature identified genes and pathways that contribute to the phenotypes of Angus cattle including meat quality traits, stature and body size, coat color, and genetic disorders to be under positive selection. Intensive artificial selection might be the probable selective pressure for the positive selection of genes affecting meat quality and body size. Genes that are related to genetic disorders might be because of their association with genes affecting other production traits. The findings in this study, if followed by a validation using additional or independent dataset, can provide useful information for the study of beef cattle in general and Angus cattle in particular. These findings will contribute to the detection of functional candidate genes which have undergone positive selection in future studies.

General Discussion

General Discussion

The genome landscape of modern cattle breeds is the results of subtle combinations of domestication, and natural and artificial selection forces in addition to demographic factors including population migration, population expansion, exposure to different diseases and new environment and diet, and admixture. These factors contributed a lot to the diverse and mosaic morphological, production and adaptation characteristics of cattle breeds. Natural and artificial selection are adaptive evolutionary forces that leave a distinct signature in the genome of animals that can be traced back through various statistical methods using SNP data. Genomic regions affected due to natural and artificial selection show lower heterogeneity, skewed allele frequency spectrum and extended haplotype homozygosity as compared to the rest of the genome in the population. Dissecting such an effect from that of other forces like genetic drift that affect the whole genome in the population helps to elucidate signatures related to those adaptation, morphological and production phenotypes. Because of this, statistical methods sought to identify genomic regions affected due to natural and/or artificial selection searches for genomic regions differentially selected as compared to the rest of the genome or as compared to other breeds/populations differentially selected for a particular trait.

In this dissertation, I sought to identify the effect of natural and artificial selection on the genomes of cattle breeds differentially selected for different traits. I used XP-EHH and XP-CLR statistical methods to compare the genomes of cattle breeds to identify signature of selection. By comparing the genomes of African cattle breeds that have evolved in a hot tropical environment with Asian-European commercial cattle breeds, several genes and genomic regions were identified under positive selection including those associated with tropical environment adaptation (e.g., thermotolerance). Thermotolerance, in African cattle, has morphological, physiological and cellular components. The positive selection of genes that are involved in these thermotolerance components (shiny and light coat color, ability to increased sweating, and

cellular and molecular mechanisms) might have been developed as a result of the long-term evolution of these breeds in the high-temperature tropical environment of Africa. This result helps to understand the genetic merit of the breeds in relation to environmental adaptation and will have an implication on the use of African cattle genetic resources for the development of heat-tolerant cattle breeds. Moreover, the genes identified in relation to thermotolerance mechanisms can be useful to use them in high producing commercial dairy and beef cattle breeding. The consistently increasing global warming calls for breeding of cattle for thermotolerance through genomic selection and more specifically genome editing. Currently, genome editing might be too much expensive, however, its use for the future cattle breeding is promising due to the negative correlation between thermotolerance and production traits.

In the studies that explored the genomes of cattle breeds artificially selected for a particular trait (e.g., beef cattle – Hanwoo, dairy cattle – Holstein), and breeds under natural selection without any designed intensive artificial selection for a particular trait (general purpose cattle breeds – N'Dama), several genes including those associated with the major phenotypes of the respective breeds were found under selection. It is intensive artificial selection that is the selective pressure for the genes identified under selection in relation to meat and milk traits in beef and dairy cattle breeds, respectively. In addition to these, genes causing genetic disorders were revealed in intensively selected beef and dairy cattle breeds. The positive selection of genes affecting genetic disorders might be because of their association with those affecting production traits. Genetic association between two or more genes occurs when there is non-random association of their alleles as a result of their proximity on the same chromosome known as genetic linkage or due to linkage disequilibrium.

Several genes revealed from the studies were found to affect polygenic or complex traits – traits affected by two or more genes. In polygenic traits, polymorphisms in two or more genes make a small contribution to the overall outcome of the trait. Almost all of the traits considered in this study including milk traits, meat quality traits, environmental adaptation traits, and coat color are polygenic traits that two or

more genes were found associated with each of the traits. For instance, *CDC42, CSN3, ADIPOQ,* and *PAPPA2* genes were found to affect milk traits in Holstein cattle. The expression of these genes affects milk proteins, milk lactose content, milk yield and other milk-related traits. *CDC42* is a growth regulating protein that contributes to the synthesis of milk proteins. *CSN3* is a casein protein involved in the formation and stabilization of micelles in milk. Together with other casein proteins, it also affects lactose and protein contents in milk. The contribution of *ADIPOQ* for milk traits is through its involvement in the facilitation of nutrient partitioning in the mammary gland. Similarly, *PAPPA2* is associated with milk and protein yield and also affects the expression of other genes affecting protein production. Genes related to oxidative stress response, a mechanism of heat tolerance in African cattle, include *SOD1, GPX7, SLC23A1, SLC23A2,* and *PLCB1* genes. These genes contribute to thermotolerance ability of African cattle through prevention and elimination of reactive oxygen species (ROS) and free radicals produced due to heat stress. ROS and free radicals are determinant to cells. *SOD1* is a homodimer that converts naturally-occurring harmful superoxide radicals to molecular oxygen and hydrogen peroxide. *GPX7* also scavenges both intracellular and extracellular superoxide radicals that help prevent lipid peroxidation of the plasma membrane. *SLC23A* genes are transporter genes which are important for the mobilization of endogen antioxidants like Vitamin C which in turn detoxify free radicals generated due to heat stress to preserve the steady-state concentrations of ROS. In addition to these polygenic effects of genes, the overlap of genes identified with the previously reported QTL regions in relation to the phenotypes of the respective breeds could be an evidence of the polygenicity of the traits.

In pleiotropy, a mutation in a gene may have an effect on multiple characteristics simultaneously due to the gene coding for a product used by a myriad of cells. In this thesis, genes that are associated with more than one trait were identified. *PAPPA2* gene, for example, is associated with milk traits, reproduction, and body size in Holstein cattle; *PKM2* gene is related to tenderness, IMF, and drip loss traits of meat quality in Ankole cattle. Similarly, *PLAG1* gene affects body size and stature, birth weight, and feeding efficiency in Angus cattle. Because of the pleiotropic nature of

genes, selection for one trait may affect other traits simultaneously. In another instance, however, a single gene might be involved in multiple pathways called antagonistic pleiotropy. The expression of a single gene causes competing effects, some of which are beneficial and some of which are detrimental to the fitness of the organism. This is a particular example of genes like *PLAG1* in Angus cattle and *PKM2* in Ankole cattle. *PLAG1* affects traits including body size, stature, birth weight, and feeding efficiency. On the other hand, it causes dwarfism. Similarly, *PKM2* contributes to the IMF and tenderness of beef while causing drip loss.

Finally, studies that identify adaptive signatures in the genome helps to understand the key adaptive events that generated an enormous phenotypic variation observed between cattle breeds today. This in turn help to understand the biological functions of genes affected by selection that contribute to the differences in adaptation and production traits. Moreover, candidate genes and variants that contribute to environmental adaptation traits which are difficult to identify through laboratory experiment can be revealed - and can be used in breeding and selection programs. Augmented with previously reported QTL regions, genomic regions identified under selection in relation to a particular phenotype gives a strong evidence in designing breeding strategies to further improve those breeds, and conservation and use of endangered breeds.

As a limitation in such kind of signature of selection studies, however, other forces (e.g., genetic drift) may cause false positive results. Therefore, validation of these results by employing other methods like qPCR and gene expression analysis, and/or integrating with genome-wide association studies before using in selection and breeding programs is significantly important.

References

Abo-Ismail, M.K., Vander Voort, G., Squires, J.J., Swanson, K.C., Mandell, I.B., Liao, X., Stothard, P. et al. (2014). Single nucleotide polymorphisms for feed efficiency and performance in crossbred beef cattle. *BMC Genetics,* 15 (14): 1-14.

Aftab, S., Semenec, L., Chu, J.S.-C., and Chen, N. (2008). Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evolutionary Biology,* 8 (226): 1-11.

Agulleiro, M.J., Sánchez, E., Leal, E., Cortés, R., Fernández-Durán, B., Guillot, R., Davis, P. et al. (2013). Molecular characterization and functional regulation of melanocortin 2 receptor (MC2R) in the sea bass. A putative role in the adaptation to stress. *PLoS One,* 8 (5): e65450.

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

Akers, R.M., and Denbow, D.M., (2008). Anatomy and physiology of domestic animals. *American Veterinary Medical Association*, Blackwell Publishing, Ames, Iowa.

Akhtar, N., and Streuli, C.H. (2006). Rac1 links integrin-mediated adhesion to the control of lactational differentiation in mammary epithelia. *The Journal of Cell Biology,* 173 (5): 781-793.

Albertí, P., Panea, B., Sañudo, C., Olleta, J., Ripoll, G., Ertbjerg, P., Christensen, M. et al. (2008). Live weight, body size and carcass characteristics of young bulls of fifteen European breeds. *Livestock Science,* 114 (1): 19-30.

Albert, F.W., Carlborg, Ö., Plyusnina, I., Besnier, F., Hedwig, D., Lautenschläger, S., Lorenz, D. et al. (2009). Genetic architecture of tameness in a rat model of animal domestication. *Genetics,* 182 (2): 541-554.

Alfonzo, E.P.M., da Silva, M.V.G.B., dos Santos Daltro, D., Stumpf, M.T., Dalcin, V.C., Kolling, G., Fischer, V. et al. (2016). Relationship between physical attributes and heat stress in dairy cattle from different genetic groups. *International Journal of Biometeorology,* 60 (2): 245-253.

Ali, M.A., Adem, A., Chandranath, I.S., Benedict, S., Pathan, J.Y., Nagelkerke, N., Nyberg, F. et al. (2012). Responses to dehydration in the one-humped camel and effects of blocking the renin-angiotensin system. *PLoS One,* 7 (5): e37299.

Amweg, A., Paredes, A., Salvetti, N., Lara, H., and Ortega, H. (2011). Expression of melanocortin receptors mRNA, and direct effects of ACTH on steroid secretion in the bovine ovary. *Theriogenology,* 75 (4): 628-637.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. *Reference Source*.

Archer, J., Arthur, P., Parnell, P., and Van de Ven, R. (1998). Effect of divergent selection for yearling growth rate on female reproductive performance in Angus cattle. *Livestock Production Science,* 57 (1): 33-40.

Arthur, P., Archer, J., Johnston, D., Herd, R., Richardson, E., and Parnell, P. (2001). Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *Journal of Animal Science,* 79 (11): 2805-2811.

Bahbahani, H., Clifford, H., Wragg, D., Mbole-Kariuki, M.N., Van Tassell, C., Sonstegard, T., Woolhouse, M. et al. (2015). Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. *Scientific Reports,* 5.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* 21 (2): 263-265.

Baumgartner, M., Lemoine, C., Al Seesi, S., Karunakaran, D.K.P., Sturrock, N., Banday, A.R., Kilcollins, A.M. et al. (2015). Minor splicing snRNAs are enriched in the developing mouse CNS and are crucial for survival of differentiating retinal neurons. *Developmental Neurobiology,* 75 (9): 895-907.

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

Bernard, C., Cassar-Malek, I., Le Cunff, M., Dubroeucq, H., Renand, G., and Hocquette, J.-F. (2007). New indicators of beef sensory quality revealed by expression of specific genes. *Journal of Agricultural and Food Chemistry,* 55 (13): 5229-5237.

Berthier, D., Peylhard, M., Dayo, G.-K., Flori, L., Sylla, S., Bolly, S., Sakande, H. et al. (2015). A comparison of phenotypic traits related to trypanotolerance in five West African cattle breeds highlights the value of shorthorn taurine breeds. *PLoS One,* 10 (5): e0126498.

Berton, M.P., Fonseca, L.F., Gimenez, D.F., Utembergue, B.L., Cesar, A.S., Coutinho, L.L., de Lemos, M.V.A. et al. (2016). Gene expression profile of intramuscular muscle in Nellore cattle with extreme values of fatty acid. *BMC Genomics,* 17 (972): 1-16.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H. et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics,* 25 (8): 1091-1093.

Birbrair, A., Zhang, T., Wang, Z.-M., Messi, M.L., Olson, J.D., Mintz, A., and Delbono, O. (2014). Type-2 pericytes participate in normal and tumoral angiogenesis. *American Journal of Physiology: Cell Physiology,* 307 (1): C25-C38.

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

Blecha, I., Siqueira, F., Ferreira, A., Feijó, G., Junior, R.T., Medeiros, S., Sousa, I. et al. (2015). Identification and evaluation of polymorphisms in FABP3 and FABP4 in beef cattle. *Genetics and Molecular Research,* 14 (4): 16353-16363.

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

Borchers, N., Otto, G., and Kalm, E. (2007). Genetic relationship of drip loss to further meat quality traits in purebred Pietrains. *Archiv Fur Tierzucht,* 50 (1): 84-91.

Braun, B., Müller, K., and Jewgenow, K. (2015). Expression profiles of relaxin family peptides and their receptors indicate their influence on spermatogenesis in the domestic cat (Felis catus). *Domestic Animal Endocrinology,* 52 25-34.

Brisken, C., Heineman, A., Chavarria, T., Elenbaas, B., Tan, J., Dey, S.K., McMahon, J.A. et al. (2000). Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes & Development,* 14 (6): 650-654.

Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics,* 81 (5): 1084-1097.

Cánovas, A., Quintanilla, R., Amills, M., and Pena, R.N. (2010). Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. *BMC Genomics,* 11 (372): 1-15.

Cavallaro, G. (2010). Genome-wide analysis of eukaryotic twin CX 9 C proteins. *Molecular BioSystems,* 6 (12): 2459-2470.

Cecchinato, A., Ribeca, C., Chessa, S., Cipolat-Gotet, C., Maretto, F., Casellas, J., and Bittante, G. (2014). Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. *Animal,* 8 (07): 1062-1070.

Chambaz, A., Scheeder, M., Kreuzer, M., and Dufey, P.-A. (2003). Meat quality of Angus, Simmental, Charolais and Limousin steers compared at the same intramuscular fat content. *Meat Science,* 63 (4): 491-500.

Chan, E.K., and Reverter, A. (2007). Integrating Whole-Genome Genetic-Association Studies with Gene Expression Data to Prioritise Candidate Genes Affecting Intramuscular Fat in Beef Cattle Traits. *Association for the Advancement of Animal Breeding and Genetics,* 17 81-84.

Chang, K. (2007). Key signalling factors and pathways in the molecular determination of skeletal muscle phenotype. *Animal,* 1 (5): 681-698.

Chen, C.C., and Fernald, R. (2008). GnRH and GnRH receptors: distribution, function and evolution. *Journal of Fish Biology,* 73 (5): 1099-1120.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Research,* 20 (3): 393-402.

Chen, S.R., Yang, L.Q., Chong, Y.T., Jie, Y.S., Wu, Y.K., Yang, J., Lin, G.L. et al. (2015). Novel gain of function mutation in the SLC40A1 gene associated with hereditary haemochromatosis type 4. *Internal Medicine Journal,* 45 (6): 672-676.

Chen, W., Kelly, M.A., Opitz-Araya, X., Thomas, R.E., Low, M.J., and Cone, R.D. (1997). Exocrine gland dysfunction in MC5-R-deficient mice: evidence for coordinated regulation of exocrine gland function by melanocortin peptides. *Cell,* 91 (6): 789-798.

Chen, Y.-I., Wei, P.-C., Hsu, J.-L., Su, F.-Y., and Lee, W.-H. (2016). NPGPx (GPx7): a novel oxidative stress sensor/transmitter with multiple roles in redox homeostasis. *American Journal of Translational Research,* 8 (4): 1626-1640.

Chen, Y., Gondro, C., Quinn, K., Herd, R., Parnell, P., and Vanselow, B. (2011). Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. *Animal Genetics,* 42 (5): 475-490.

Cho, E.-S., Jeon, H.-J., Lee, S.-W., Park, J.-W., Raveendar, S., Jang, G.-W., Kim, T.-H. et al. (2013). Association of a Pyruvate Kinase M2 (PKM2) Polymorphism with Back Fat Thickness in Berkshire Pigs. *Journal of Animal Science and Technology,* 55 (6): 515-520.

Cho, I., Park, H., Yoo, C., Lee, G., Lim, H., Lee, J., Jung, E. et al. (2011). QTL analysis of white blood cell, platelet and red blood cell-related traits in an F2 intercross between Landrace and Korean native pigs. *Animal Genetics,* 42 (6): 621-626.

Choquette, A.C., Bouchard, L., Drapeau, V., Lemieux, S., Tremblay, A., Bouchard, C., Vohl, M.-C. et al. (2012). Association between olfactory receptor genes, eating behavior traits and adiposity: results from the Quebec Family Study. *Physiology & Behavior,* 105 (3): 772-776.

Christians, J.K., De Zwaan, D.R., and Fung, S.H.Y. (2013). Pregnancy associated plasma protein A2 (PAPP-A2) affects bone size and shape and contributes to natural variation in postnatal growth in mice. *PLoS One,* 8 (2): e56260.

Chung, E.-R. (2014). Novel SNP in the coding region of the FTO gene is associated with marbling score in Hanwoo (Korean cattle). *Journal of Animal Science and Technology,* 56 (27): 1-6.

Ciepłoch, A., Rutkowska, K., Oprządek, J., and Poławska, E. (2017). Genetic disorders in beef cattle: a review. *Genes & Genomics,* 39 (5): 1-11.

Cieslak, M., Reissmann, M., Hofreiter, M., and Ludwig, A. (2011). Colours of domestication. *Biological Reviews,* 86 (4): 885-899.

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

Clark, K.A., McElhinny, A.S., Beckerle, M.C., and Gregorio, C.C. (2002). Striated muscle cytoarchitecture: an intricate web of form and function. *Annual Review of Cell and Developmental Biology,* 18 (1): 637-706.

Cole, J., Waurich, B., Wensch-Dorendorf, M., Bickhart, D., and Swalve, H. (2014). A genome-wide association study of calf birth weight in Holstein cattle using single nucleotide polymorphisms and phenotypes predicted from auxiliary traits. *Journal of Dairy Science,* 97 (5): 3156-3172.

Collier, R., Collier, J., Rhoads, R., and Baumgard, L. (2008). Invited review: genes involved in the bovine heat stress response. *Journal of Dairy Science,* 91 (2): 445-454.

Collier, R.J., and Collier, J., (2012). Environmental physiology of livestock. *Wiley Online Library*.

Corella, D., and Ordovas, J.M. (2005). Single nucleotide polymorphisms that influence lipid metabolism: interaction with dietary factors. *Annual Review of Nutrition,* 25 341-390.

Cottrell, J., Liu, F., Hung, A., DiGiacomo, K., Chauhan, S., Leury, B., Furness, J. et al. (2015). Nutritional strategies to alleviate heat stress in pigs. *Animal Production Science,* 55 (12): 1391-1402.

Cui, C.-Y., Yin, M., Sima, J., Childress, V., Michel, M., Piao, Y., and Schlessinger, D. (2014). Involvement of Wnt, Eda and Shh at defined stages of sweat gland development. *Development,* 141 (19): 3752-3760.

Cui, C.Y., and Schlessinger, D. (2015). Eccrine sweat gland development and sweat secretion. *Experimental Dermatology,* 24 (9): 644-650.

Cui, H.-X., Liu, R.-R., Zhao, G.-P., Zheng, M.-Q., Chen, J.-L., and Wen, J. (2012a). Identification of differentially expressed genes and pathways for intramuscular fat deposition in pectoralis major tissues of fast-and slow-growing chickens. *BMC Genomics,* 13 (213): 1-12.

Cui, Y., Wang, W., Dong, N., Lou, J., Srinivasan, D.K., Cheng, W., Huang, X. et al. (2012b). Role of corin in trophoblast invasion and uterine spiral artery remodelling in pregnancy. *Nature,* 484 (7393): 246-250.

Culler, R., Smith, G., and Cross, H. (1978). Relationship of myofibril fragmentation index to certain chemical, physical and sensory characteristics of bovine longissimus muscle. *Journal of Food Science,* 43 (4): 1177-1180.

da Costa, N., Edgar, J., Ooi, P.-T., Su, Y., Meissner, J.D., and Chang, K.-C. (2007). Calcineurin differentially regulates fast myosin heavy chain genes in oxidative muscle fibre type conversion. *Cell and Tissue Research,* 329 (3): 515-527.

Da Silva, A.S., Hoehne, L., Tonin, A.A., Zanette, R.A., Wolkmer, P., Costa, M.M., Moraes, D.P. et al. (2009). Trypanosoma evansi: Levels of copper, iron and zinc in the bloodstream of infected cats. *Experimental Parasitology,* 123 (1): 35-38.

Damon, M., Wyszynska-Koko, J., Vincent, A., Herault, F., and Lebret, B. (2012). Comparison of muscle transcriptome between pigs with divergent meat quality phenotypes identifies genes related to muscle metabolism and structure. *PLoS One,* 7 (3): e33763.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E. et al. (2011). The variant call format and VCFtools. *Bioinformatics,* 27 (15): 2156-2158.

Dangi, S.S., Dangi, S.K., Chouhan, V., Verma, M., Kumar, P., Singh, G., and Sarkar, M. (2016). Modulatory effect of betaine on expression dynamics of HSPs during heat stress acclimation in goat (Capra hircus). *Gene,* 575 (2): 543-550.

Davidson, J.F., Whyte, B., Bissinger, P.H., and Schiestl, R.H. (1996). Oxidative stress is involved in heat-induced cell death in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences,* 93 (10): 5116-5121.

Davies, C.R., Fukumura, F., Fukamachi, K., Muramoto, K., Himley, S.C., Massiello, A., Chen, J.-F. et al. (1994). Adaptation of tissue to a chronic heat load. *ASAIO Journal,* 40 (3): M514-M517.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Alcalá, A.M., Sonstegard, T.S., Hanotte, O. et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics,* 10 (3): e1004254.

Del Vesco, A.P., Gasparino, E., de Oliveira Grieser, D., Zancanela, V., Soares, M.A.M., and de Oliveira Neto, A.R. (2015). Effects of methionine supplementation on the expression of oxidative stress-related genes in acute heat stress-exposed broilers. *British Journal of Nutrition,* 113 (04): 549-559.

Desai, S.S., Roy, B.S., and Mahale, S.D. (2013). Mutations and polymorphisms in FSH receptor: functional implications in human reproduction. *Reproduction,* 146 (6): R235-R248.

Dias, M., Souza, F., Takada, L., Feitosa, F., Costa, R., Diaz, I., Cardoso, D. et al. (2015). Study of lipid metabolism-related genes as candidate genes of sexual precocity in Nellore cattle. *Genetics and Molecular Research,* 14 (1): 234-243.

Dikmen, S., Cole, J., Null, D., and Hansen, P. (2012). Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. *Journal of Dairy Science,* 95 (6): 3401-3405.

Ding, L., Zhao, Y., Warren, C.L., Sullivan, R., Eliceiri, K.W., and Shull, J.D. (2013). Association of cellular and molecular responses in the rat mammary gland to 17β-estradiol with susceptibility to mammary cancer. *BMC Cancer,* 13 (573): 1-15.

Do, D.N., Ostersen, T., Strathe, A.B., Mark, T., Jensen, J., and Kadarmideen, H.N. (2014a). Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. *BMC Genetics,* 15 (27): 1-15.

Do, D.N., Strathe, A.B., Ostersen, T., Pant, S.D., and Kadarmideen, H.N. (2014b). Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. *Frontiers in Genetics,* 5 (307): 1-10.

Donato Jr, J., and Frazão, R. (2016). Interactions between prolactin and kisspeptin to control reproduction. *Archives of Endocrinology and Metabolism,* 60 (6): 587-595.

Druso, J.E., Endo, M., Lin, M.-c.J., Peng, X., Antonyak, M.A., Meller, S., and Cerione, R.A. (2016). An Essential Role for Cdc42 in the Functioning of the Adult Mammary Gland. *Journal of Biological Chemistry,* 291 (17): 8886-8895.

Duncan, R.E., Sarkadi-Nagy, E., Jaworski, K., Ahmadian, M., and Sul, H.S. (2008). Identification and functional characterization of adipose-specific phospholipase A2 (AdPLA). *Journal of Biological Chemistry,* 283 (37): 25428-25436.

Eddy, S.R. (2001). Non–coding RNA genes and the modern RNA world. *Nature Reviews Genetics,* 2 (12): 919-929.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32 (5): 1792-1797.

Elsik, C.G., Tellam, R.L., and Worley, K.C. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science,* 324 (5926): 522-528.

Ely, B.R., Lovering, A.T., Horowitz, M., and Minson, C.T. (2014). Heat acclimation and cross tolerance to hypoxia: bridging the gap between cellular and systemic responses. *Temperature,* 1 (2): 107-114.

Esteve-Codina, A., Paudel, Y., Ferretti, L., Raineri, E., Megens, H.-J., Silió, L., Rodríguez, M.C. et al. (2013). Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. *BMC Genomics,* 14 (148): 1-14.

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

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution,* 17 (6): 368-376.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* 39 (4): 783-791.

Felsenstein, J. (1993). PHYLIP (phylogeny inference package), 3.5 c ed. *Department of Genetics, University of Washington, Seattle.*

Fernández, M.E., Goszczynski, D.E., Prando, A.J., Peral-García, P., Baldo, A., Giovambattista, G., and Liron, J.P. (2014). Assessing the association of single nucleotide polymorphisms in thyroglobulin gene with age of puberty in bulls. *Journal of Animal Science and Technology,* 56 (17): 1-4.

Finch, V.A., Bennett, I., and Holmes, C. (1984). Coat colour in cattle: effect on thermal balance, behaviour and growth, and relationship with coat type. *Journal of Agricultural Science,* 102 (01): 141-147.

Fink, T., Tiplady, K., Lopdell, T., Johnson, T., Snell, R.G., Spelman, R.J., Davis, S.R. et al. (2017). Functional confirmation of PLAG1 as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. *Scientific Reports,* 7.

Flori, L., Thevenon, S., Dayo, G.K., Senou, M., Sylla, S., Berthier, D., Moazami-Goudarzi, K. et al. (2014). Adaptive admixture in the West African bovine hybrid zone: insight from the Borgou population. *Molecular Ecology,* 23 (13): 3241-3257.

Folk, J.G.E., and Semken, J.A. (1991). The evolution of sweat glands. *International Journal of Biometeorology,* 35 (3): 180-186.

Fontanesi, L., Davoli, R., Costa, L.N., Beretti, F., Scotti, E., Tazzoli, M., Tassone, F. et al. (2008). Investigation of candidate genes for glycolytic potential of porcine skeletal muscle:

Association with meat quality and production traits in Italian Large White pigs. *Meat Science,* 80 (3): 780-787.

Fortes, M., Kemper, K., Sasazaki, S., Reverter, A., Pryce, J., Barendse, W., Bunch, R. et al. (2013). Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. *Animal Genetics,* 44 (6): 636-647.

Gómez, J., Reguero, J.R., and Coto, E. (2016). The Ups and Downs of Genetic Diagnosis of Hypertrophic Cardiomyopathy. *Revista Española de Cardiología,* 69 (1): 61-68.

Gao, Y., Zhang, Y., Jiang, H., Xiao, S., Wang, S., Ma, Q., Sun, G. et al. (2011). Detection of differentially expressed genes in the longissimus dorsi of Northeastern Indigenous and Large White pigs. *Genetics and Molecular Research,* 10 (2): 779-791.

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K. et al. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics,* 10 (550): 1-18.

Gazzola, C., O'Neill, C., and Frisch, J. (1999). Comparative evaluation of the meat quality of beef cattle breeds of Indian, African and European origins. *Animal Science,* 69 (1): 135-142.

Gheyas, A.A., Boschiero, C., Eory, L., Ralph, H., Kuo, R., Woolliams, J.A., and Burt, D.W. (2015). Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Research,* 22 (3): 205-217.

Ghosh, M., Sodhi, S., Song, K.D., Kim, J., Mongre, R., Sharma, N., Singh, N. et al. (2015). Evaluation of body growth and immunity-related differentially expressed genes through deep RNA sequencing in the piglets of Jeju native pig and Berkshire. *Animal Genetics,* 46 (3): 255-264.

Gouveia, J.J.d.S., Silva, M.V.G.B.d., Paiva, S.R., and Oliveira, S.M.P.d. (2014). Identification of selection signatures in livestock species. *Genetics and Molecular Biology,* 37 (2): 330-342.

Graw, J. (2009). Genetics of crystallins: cataract and beyond. *Experimental Eye Research,* 88 (2): 173-189.

Grigson, C. (1991). An African origin for African cattle?—some archaeological evidence. *African Archaeological Review,* 9 (1): 119-144.

Guillemin, N., Bonnet, M., Jurie, C., and Picard, B. (2011). Functional analysis of beef tenderness. *Journal of Proteomics,* 75 (2): 352-365.

Gunawan, A., Cinar, M., Uddin, M., Kaewmala, K., Tesfaye, D., Phatsara, C., Tholen, E. et al. (2012). Investigation on association and expression of ESR2 as a candidate gene for boar sperm quality and fertility. *Reproduction in Domestic Animals,* 47 (5): 782-790.

Gunn, T., Inui, T., Kitada, K., Ito, S., Wakamatsu, K., He, L., Bouley, D. et al. (2001). Molecular and phenotypic analysis of Attractin mutant mice. *Genetics,* 158 (4): 1683-1695.

Guo, X., Fang, Q., Ma, C., Zhou, B., Wan, Y., and Jiang, R. (2016). Whole-genome resequencing of Xishuangbanna fighting chicken to identify signatures of selection. *Genetics Selection Evolution,* 48 (62): 1-6.

Gupta, M., Kumar, S., Dangi, S., and Jangir, B.L. (2013). Physiological, biochemical and molecular responses to thermal stress in goats. *International Journal of Livestock Research,* 3 (2): 27-38.

Guziewicz, K., Owczarek-Lipska, M., Küffer, J., Schelling, C., Tontis, A., Denis, C., Eggen, A. et al. (2007). The locus for bovine dilated cardiomyopathy maps to chromosome 18. *Animal Genetics,* 38 (3): 265-269.

Höglund, J.K., Guldbrandtsen, B., Lund, M.S., and Sahana, G. (2015). Identification of genomic regions associated with female fertility in Danish Jersey using whole genome sequence data. *BMC Genetics,* 16 (60): 1-9.

Häussler, S. (2015). Relevance of bovine adiponectin in the mammary gland of dairy cows. *The Veterinary Journal,* 204 (2): 132-133.

Hamill, R.M., McBryan, J., McGee, C., Mullen, A.M., Sweeney, T., Talbot, A., Cairns, M.T. et al. (2012). Functional analysis of muscle gene expression profiles associated with tenderness and intramuscular fat content in pork. *Meat Science,* 92 (4): 440-450.

Hanotte, O., Dessie, T., and Kemp, S. (2010). Time to tap Africa's livestock genomes. *Science,* 328 (5986): 1640-1641.

Hansen, L., Cole, J., Marx, G., and Seykora, A. (1999). Productive life and reasons for disposal of Holstein cows selected for large versus small body size. *Journal of Dairy Science,* 82 (4): 795-801.

Hansen, P. (2004). Physiological and cellular adaptations of zebu cattle to thermal stress. *Animal Reproduction Science,* 82 349-360.

Hansen, P. (2009). Effects of heat stress on mammalian reproduction. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences,* 364 (1534): 3341-3350.

Hausman, G., Dodson, M., Ajuwon, K., Azain, M., Barnes, K., Guan, L., Jiang, Z. et al. (2009). Board-invited review: the biology and regulation of preadipocytes and adipocytes in meat animals. *Journal of Animal Science,* 87 (4): 1218-1246.

Hinuma, S., Habata, Y., Fujii, R., Kawamata, Y., Hosoya, M., Fukusumi, S., Kitada, C. et al. (1998). A prolactin-releasing peptide in the brain. *Nature,* 393 (6682): 272-276.

Holsinger, K.E., and Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature reviews - Genetics,* 10 (9): 639.

Horikoshi, M., Yaghootkar, H., Mook-Kanamori, D.O., Sovio, U., Taal, H.R., Hennig, B.J., Bradfield, J.P. et al. (2013). New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nature Genetics,* 45 (1): 76-82.

Howard, J.T., Kachman, S.D., Snelling, W.M., Pollak, E.J., Ciobanu, D.C., Kuehn, L.A., and Spangler, M.L. (2014). Beef cattle body temperature during climatic stress: a genome-wide association study. *International Journal of Biometeorology,* 58 (7): 1665-1672.

Hsu, S.Y., Kaipia, A., McGee, E., Lomeli, M., and Hsueh, A.J. (1997). Bok is a pro-apoptotic Bcl-2 protein with restricted expression in reproductive tissues and heterodimerizes with selective anti-apoptotic Bcl-2 family members. *Proceedings of the National Academy of Sciences,* 94 (23): 12401-12406.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research,* 37 (1): 1-13.

Jackson, J.E. 1991. A User's Guide to Principal Components. JSTOR, USA.

Jennen, D., Vereijken, A., Bovenhuis, H., Crooijmans, R., Veenendaal, A., Van der Poel, J., and Groenen, M. (2004). Detection and localization of quantitative trait loci affecting fatness in broilers. *Poultry Science,* 83 (3): 295-301.

Jensen, J.D., Foll, M., and Bernatchez, L. (2016). The past, present and future of genomic scans for selection. *Molecular Ecology,* 25 (1): 1-4.

Jeong, J., Kwon, E., Im, S., Seo, K., and Baik, M. (2012). Expression of fat deposition and fat removal genes is associated with intramuscular fat content in longissimus dorsi muscle of Korean cattle steers. *Journal of Animal Science,* 90 (6): 2044-2053.

Jian, W., Duangjinda, M., Vajrabukka, C., and Katawatin, S. (2014). Differences of skin morphology in Bos indicus, Bos taurus, and their crossbreds. *International Journal of Biometeorology,* 58 (6): 1087-1094.

Jiang, C., Shi, P., Li, S., Dong, R., Tian, J., Wei, J., and Luo, S. (2010). Gene expression profiling of skeletal muscle of nursing piglets. *International Journal of Biological Sciences,* 6 (7): 627.

Johnston, D., Reverter, A., Ferguson, D., Thompson, J., and Burrow, H. (2003). Genetic and phenotypic characterisation of animal, carcass, and meat quality traits from temperate and tropically adapted beef breeds. 3. Meat quality traits. *Australian Journal of Agricultural Research,* 54 (2): 135-147.

Joo, S., Kim, G., Hwang, Y., and Ryu, Y. (2013). Control of fresh meat quality through manipulation of muscle fiber characteristics. *Meat Science,* 95 (4): 828-836.

Jovanović, J., Takagi, J., Choulier, L., Abrescia, N.G., Stuart, D.I., Van Der Merwe, P.A., Mardon, H.J. et al. (2007). αVβ6 Is a Novel Receptor for Human Fibrillin-1 Comparative Studies of Molecular Determinants Underlying Integrin-Rgd Affinity and Specificity. *Journal of Biological Chemistry,* 282 (9): 6743-6751.

Juma, A.R., Damdimopoulou, P.E., Grommen, S.V., Van de Ven, W.J., and De Groef, B. (2016). Emerging role of PLAG1 as a regulator of growth and reproduction. *Journal of Endocrinology,* 228 (2): R45-R56.

Kahoun, J. (1971). Heat tolerance in West African cattle. *Ghana Journal of Science,* 11 (1): 19-26.

Kamatara, K., Mpairwe, D., Christensen, M., Mutetikka, D., and Madsen, J. (2013). Sensory characteristics and tenderness of meat from Ankole bulls: Influence of crossbreeding and feeding system. *South African Journal of Animal Science,* 43 (5): 107-110.

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J.A., Baurain, D., Cambisano, N. et al. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics,* 43 (5): 405-413.

Kasvandik, S., Sillaste, G., Velthut-Meikas, A., Mikelsaar, A.V., Hallap, T., Padrik, P., Tenson, T. et al. (2015). Bovine sperm plasma membrane proteomics through biotinylation and subcellular enrichment. *Proteomics,* 15 (11): 1906-1920.

Kemper, K.E., Visscher, P.M., and Goddard, M.E. (2012). Genetic architecture of body size in mammals. *Genome Biology,* 13 (4): 244.

Khan, A.A., and Quigley, J.G. (2013). Heme and FLVCR-related transporter families SLC48 and SLC49. *Molecular Aspects of Medicine,* 34 (2): 669-682.

Khatib, H., Huang, W., Wang, X., Tran, A., Bindrim, A., Schutzkus, V., Monson, R. et al. (2009). Single gene and gene interaction effects on fertilization and embryonic survival rates in cattle. *Journal of Dairy Science,* 92 (5): 2238-2247.

Khor, C.C., Miyake, M., Chen, L.J., Shi, Y., Barathi, V.A., Qiao, F., Nakata, I. et al. (2013). Genome-wide association study identifies ZFHX1B as a susceptibility locus for severe myopia. *Human Molecular Genetics,* 22 (25): 5288-5294.

Kim, E.-S., Sonstegard, T.S., and Rothschild, M.F. (2015a). Recent artificial selection in US Jersey cattle impacts autozygosity levels of specific genomic regions. *BMC Genomics,* 16 (302): 1-10.

Kim, E., Elbeltagy, A., Aboul-Naga, A., Rischkowsky, B., Sayre, B., Mwacharo, J., and Rothschild, M. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity,* 116 (3): 255-264.

Kim, J., Cho, S., Caetano-Anolles, K., Kim, H., and Ryu, Y.-C. (2015b). Genome-wide detection and characterization of positive selection in Korean Native Black Pig from Jeju Island. *BMC Genetics,* 16 (3): 1-9.

Kim, J., Hanotte, O., Mwai, O.A., Dessie, T., Bashir, S., Diallo, B., Agaba, M. et al. (2017a). The genome landscape of indigenous African cattle. *Genome Biology,* 18 (34): 1-14.

Kim, S.-J., Ka, S., Ha, J.-W., Kim, J., Yoo, D., Kim, K., Lee, H.-K. et al. (2017b). Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama. *BMC Genomics,* 18 (371): 1-18.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequence. *Molecular Evolution,* 16 (2): 111-120.

Klar, J., Hisatsune, C., Baig, S.M., Tariq, M., Johansson, A., Rasool, M., Malik, N.A. et al. (2014). Abolished InsP3R2 function inhibits sweat secretion in both humans and mice. *The Journal of Clinical Investigation,* 124 (11): 4773-4780.

Klungland, H., Vage, D., Gomez-Raya, L., Adalsteinsson, S., and Lien, S. (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mammalian Genome,* 6 (9): 636-639.

Koohmaraie, M., Kent, M.P., Shackelford, S.D., Veiseth, E., and Wheeler, T.L. (2002). Meat tenderness and muscle growth: is there any relationship? *Meat Science,* 62 (3): 345-352.

Korneliussen, T.S., Moltke, I., Albrechtsen, A., and Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics,* 14 (289): 1-14.

Kucerova, J., Matejicek, A., Jandurová, O., Sorensen, P., Nemcova, E., Stipkova, M., Kott, T. et al. (2006). Milk protein genes CSN1S1, CSN2, CSN3, LGB and their relation to genetic values of milk production parameters in Czech Fleckvieh. *Czech Journal of Animal Science,* 51 (6): 241.

Kugonza, D., Nabasirye, M., Mpairwe, D., Hanotte, O., and Okeyo, A. (2011). Productivity and morphology of Ankole cattle in three livestock production systems in Uganda. *Animal Genetic Resources,* 48 13-22.

Ladeira, M.M., Schoonmaker, J.P., Gionbelli, M.P., Dias, J.C., Gionbelli, T.R., Carvalho, J.R.R., and Teixeira, P.D. (2016). Nutrigenomics and Beef Quality: A Review about Lipogenesis. *International Journal of Molecular Sciences,* 17 (6): 918.

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

Lazari, M.F.M., Lucas, T.F.G., Yasuhara, F., Gomes, G.R.O., Siu, E.R., Royer, C., Fernandes, S.A.F. et al. (2009). Estrogen receptors and function in the male reproductive system. *Arquivos Brasileiros de Endocrinologia & Metabologia,* 53 (8): 923-933.

Leak, S.G., (1999). Tsetse biology and ecology: their role in the epidemiology and control of trypanosomosis. *CABI publishing Wallingford, UK.*

Lecchi, C., Giudice, C., Uggè, M., Scarafoni, A., Baldi, A., and Sartorelli, P. (2015). Characterisation of adiponectin and its receptors in the bovine mammary gland and in milk. *The Veterinary Journal,* 203 (3): 296-301.

Lee, H.-J., Jang, M., Kim, H., Kwak, W., Park, W., Hwang, J.Y., Lee, C.-K. et al. (2013a). Comparative transcriptome analysis of adipose tissues reveals that ECM-receptor interaction is involved in the depot-specific adipogenesis in cattle. *PLoS One,* 8 (6): e66267.

Lee, H.-J., Kim, J., Lee, T., Son, J.K., Yoon, H.-B., Baek, K.-S., Jeong, J.Y. et al. (2014a). Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genome Biology and Evolution,* 6 (6): 1366-1374.

Lee, I.H., Campbell, C.R., Cook, D.I., and Dinudom, A. (2008). Regulation of epithelial Na+ channels by aldosterone: role of Sgk1. *Clinical and Experimental Pharmacology and Physiology,* 35 (2): 235-241.

Lee, K.-T., Chung, W.-H., Lee, S.-Y., Choi, J.-W., Kim, J., Lim, D., Lee, S. et al. (2013b). Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. *BMC Genomics*, 14 (519): 1-14.

Lee, S.-H., Park, B.-H., Sharma, A., Dang, C.-G., Lee, S.-S., Choi, T.-J., Choy, Y.-H. et al. (2014b). Hanwoo cattle: origin, domestication, breeding strategies and genomic selection. *Journal of Animal Science and Technology*, 56 (2): 1-8.

Lee, T.-H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014c). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, 15 (162): 1-6.

Lee, W., Taye, M., Kwon, T., Yoon, J., Jang, D., Suzuki, S., and Kim, H. (2017). Identifying candidate positive selection genes in Korean imported pig breeds. *Genes & Genomics*, 39 (5): 557-565.

Lenis Sanin, Y., Zuluaga Cabrera, A.M., and Tarazona Morales, A.M. (2016). Adaptive Responses to Thermal Stress in Mammals. *Revista de Medicina Veterinaria*, 31 121-135.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25 (16): 2078-2079.

Li, Y.-D., Ji, Y.-T., Zhou, X.-H., Li, H.-l., Zhang, H.-t., Xing, Q., Hong, Y.-f. et al. (2015). TNNT2 Gene Polymorphisms are Associated with Susceptibility to Idiopathic Dilated Cardiomyopathy in Kazak and Han Chinese. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 21 3343-3347.

Li, Y., Xu, Z., Li, H., Xiong, Y., and Zuo, B. (2010). Differential transcriptional analysis between red and white skeletal muscle of Chinese Meishan pigs. *International Journal of Biological Sciences*, 6 (4): 350-360.

Li, Z., Chen, J., Wang, Z., Pan, Y., Wang, Q., Xu, N., and Wang, Z. (2016). Detection of selection signatures of population-specific genomic regions selected during domestication process in Jinhua pigs. *Animal Genetics*, 47 (6): 672-681.

Liang, J., Song, W., Tromp, G., Kolattukudy, P.E., and Fu, M. (2008). Genome-wide survey and expression profiling of CCCH-zinc finger family reveals a functional module in macrophage activation. *PLoS One*, 3 (8): e2880.

Littlejohn, M., Grala, T., Sanders, K., Walker, C., Waghorn, G., Macdonald, K., Coppieters, W. et al. (2012). Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus. *Animal Genetics*, 43 (5): 591-594.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M., and Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences*, 91 (7): 2757-2761.

Lonergan, E.H., Zhang, W., and Lonergan, S.M. (2010). Biochemistry of postmortem muscle—Lessons on mechanisms of meat tenderization. *Meat Science*, 86 (1): 184-195.

Lu, X.-Y., Gunn, T.M., Shieh, K.-R., Barsh, G.S., Akil, H., and Watson, S.J. (1999). Distribution of Mahogany/Attractin mRNA in the rat central nervous system. *FEBS letters*, 462 (1-2): 101-107.

Lu, Y., Chen, S., Liu, W., Hou, Z., Xu, G., and Yang, N. (2012). Polymorphisms in Wnt signaling pathway genes are significantly associated with chicken carcass traits. *Poultry Science,* 91 (6): 1299-1307.

Ma, T., Xu, L., Wang, H., Chen, J., Liu, L., Chang, G., and Chen, G. (2015a). Mining the key regulatory genes of chicken inosine 5′-monophosphate metabolism based on time series microarray data. *Journal of Animal Science and Biotechnology,* 6 (21): 1-11.

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

Madsen, L., Petersen, R.K., and Kristiansen, K. (2005). Regulation of adipocyte differentiation and function by polyunsaturated fatty acids. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1740 (2): 266-286.

Mai, M., Sahana, G., Christiansen, F., and Guldbrandtsen, B. (2010). A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *Journal of Animal Science,* 88 (11): 3522-3528.

Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E., Taylor, J.F., Makgahlela, M.L., and Maiwashe, A. (2015). Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetics Selection Evolution,* 47 (1): 1-14.

Maltecca, C., Gray, K., Weigel, K., Cassady, J., and Ashwell, M. (2011). A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. *Animal Genetics,* 42 (6): 585-591.

Maltin, C., Balcerzak, D., Tilley, R., and Delday, M. (2003). Determinants of meat quality: tenderness. *Proceedings of the Nutrition Society,* 62 (02): 337-347.

Mancini, R., and Hunt, M. (2005). Current research in meat color. *Meat Science,* 71 (1): 100-121.

Matsson, H., Eason, J., Bookwalter, C.S., Klar, J., Gustavsson, P., Sunnegårdh, J., Enell, H. et al. (2008). Alpha-cardiac actin mutations produce atrial septal defects. *Human Molecular Genetics,* 17 (2): 256-265.

Mattioli, R.C., Pandey, V.S., Murray, M., and Fitzpatrick, J.L. (2000). Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (Bos taurus) and trypanosusceptible Gobra zebu (Bos indicus) cattle. *Acta Tropica,* 75 (3): 263-277.

McAninch, E.A., and Bianco, A.C. (2014). Thyroid hormone signaling in energy homeostasis and energy metabolism. *Annals of the New York Academy of Sciences,* 1311 (1): 77-87.

McClure, M., Morsci, N., Schnabel, R., Kim, J., Yao, P., Rolf, M., McKay, S. et al. (2010). A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Animal Genetics,* 41 (6): 597-607.

McDonagh, E.M., Clancy, J.P., Altman, R.B., and Klein, T.E. (2015). PharmGKB summary: very important pharmacogene information for CFTR. *Pharmacogenetics and Genomics,* 25 (3): 149-156.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20 (9): 1297-1303.

Metzger, J., Schrimpf, R., Philipp, U., and Distl, O. (2013). Expression levels of LCORL are associated with body size in horses. *PLoS One,* 8 (2): e56497.

Meuwissen, T., Hayes, B., and Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers,* 6 (1): 6-14.

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

Moreno-Sánchez, N., Rueda, J., Carabaño, M.J., Reverter, A., McWilliam, S., González, C., and Díaz, C. (2010). Skeletal muscle specific genes networks in cattle. *Functional & Integrative Genomics,* 10 (4): 609-618.

Mozdziak, P. (2006). Linking embryonic myogenesis to meat quantity and quality. *Polish Journal of Food and Nutrition Sciences,* 15/56 (2): 117-121.

Mullen, A., Stapleton, P., Corcoran, D., Hamill, R., and White, A. (2006). Understanding meat quality through the application of genomic and proteomic approaches. *Meat Science,* 74 (1): 3-16.

Mwai, O., Hanotte, O., Kwon, Y.-J., and Cho, S. (2015). African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. *Asian-Australasian Journal of Animal Sciences,* 28 (7): 911-921.

Nagashima, T., Li, Q., Clementi, C., Lydon, J.P., DeMayo, F.J., and Matzuk, M.M. (2013). BMPR2 is required for postimplantation uterine function and pregnancy maintenance. *The Journal of Clinical Investigation,* 123 (6): 2539-2550.

Naggert, J.K., Fricker, L.D., Varlamov, O., Nishina, P.M., Rouille, Y., Steiner, D.F., Carroll, R.J. et al. (1995). Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity. *Nature Genetics,* 10 (2): 135-142.

Nakamura, Y., Kanemarum, K., and Fukami, K. (2013). Physiological functions of phospholipase Cδ1 and phospholipase Cδ3. *Advances in Biological Regulation,* 53 (3): 356-362.

Nart, P., Williams, A., Thompson, H., and Innocent, G. (2004). Morphometry of bovine dilated cardiomyopathy. *Journal of Comparative Pathology,* 130 (4): 235-245.

Ndumu, D., Baumung, R., Hanotte, O., Wurzinger, M., Okeyo, M., Jianlin, H., Kibogo, H. et al. (2008). Genetic and morphological characterisation of the Ankole Longhorn cattle in the African Great Lakes region. *Genetics Selection Evolution,* 40 (5): 467-490.

Ngu, N.T., and Nhan, N.T.H. (2012). Analysis of troponin I gene polymorphisms and meat quality in Mongcai pigs. *South African Journal of Animal Science,* 42 (3): 288-295.

Nguyen, T.T., Bowman, P.J., Haile-Mariam, M., Pryce, J.E., and Hayes, B.J. (2016). Genomic selection for tolerance to heat stress in Australian dairy cattle. *Journal of Dairy Science,* 99 (4): 2849-2862.

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

Nishimura, S., Watanabe, T., Mizoshita, K., Tatsuda, K., Fujita, T., Watanabe, N., Sugimoto, Y. et al. (2012). Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. *BMC Genetics,* 13 (40): 1-11.

O'Dell, S.D., and Day, I.N. (1998). Molecules in focus Insulin-like growth factor II (IGF-II). *The International Journal of Biochemistry & Cell Biology,* 30 (7): 767-771.

O'brien, A.M.P., Utsunomiya, Y.T., Mészáros, G., Bickhart, D.M., Liu, G.E., Van Tassell, C.P., Sonstegard, T.S. et al. (2014). Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genetics Selection Evolution,* 46 (19): 1-14.

Ogorevc, J., Kunej, T., Razpet, A., and Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics,* 40 (6): 832-851.

Ogunsanmi, A., Taiwo, V., Onawumi, B., Mbagwu, H., and Okoronkwo, C. (2000). Correlation of physiological plasma lipid levels with resistance of cattle to trypanosomosis. *Veterinarski Arhiv,* 70 (5): 251-257.

Okeyo, A., Hanotte, O., Kwon, Y.-J., and Cho, S. (2015). African indigenous cattle: Unique genetic resources in a rapidly changing world. *Asian-Australasian Journal of Animal Sciences,* 28 (7): 911-921.

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

Owczarek-Lipska, M., Plattet, P., Zipperle, L., Drögemüller, C., Posthaus, H., Dolf, G., and Braunschweig, M.H. (2011). A nonsense mutation in the optic atrophy 3 gene (OPA3) causes dilated cardiomyopathy in Red Holstein cattle. *Genomics,* 97 (1): 51-57.

Pan, Y. (1963). Quantitative and morphological variation of sweat glands, skin thickness, and skin shrinkage over various body regions of Sahiwal Zebu and Jersey cattle. *Australian Journal of Agricultural Research,* 14 (3): 424-437.

Patruno, M., Thorndyke, M.C., Carnevali, M.C., Bonasoro, F., and Beesley, P.W. (2001). Growth factors, heat-shock proteins and regeneration in echinoderms. *Journal of Experimental Biology,* 204 (5): 843-848.

Paula-Lopes, F., Lima, R., Satrapa, R., and Barros, C. (2013). Physiology and Endocrinology Symposium: Influence of cattle genotype (vs.) on oocyte and preimplantation embryo resistance to increased temperature. *Journal of Animal Science,* 91 (3): 1143-1153.

Pearce, S.C., Mani, V., Boddicker, R.L., Johnson, J.S., Weber, T.E., Ross, J.W., Rhoads, R.P. et al. (2013). Heat stress reduces intestinal barrier integrity and favors intestinal glucose transport in growing pigs. *PLoS One,* 8 (8): e70215.

Perez, R., Cañón, J., and Dunner, S. (2010). Genes associated with long-chain omega-3 fatty acids in bovine skeletal muscle. *Journal of Applied Genetics,* 51 (4): 479-487.

Philipp, U., Hamann, H., Mecklenburg, L., Nishino, S., Mignot, E., Günzel-Apel, A.-R., Schmutz, S.M. et al. (2005). Polymorphisms within the canine MLPH gene are associated with dilute coat color in dogs. *BMC Genetics,* 6 (34): 1-15.

Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S. et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Research,* 19 (5): 826-837.

Pierzchala, M., Hoekman, A., Urbanski, P., Kruijt, L., Kristensen, L., Young, J.F., Oksbjerg, N. et al. (2014). Validation of biomarkers for loin meat quality (M. longissimus) of pigs. *Journal of Animal Breeding and Genetics,* 131 (4): 258-270.

Piirsoo, M., Meijer, D., and Timmusk, T. (2009). Expression analysis of the CLCA gene family in mouse and human with emphasis on the nervous system. *BMC Developmental Biology,* 9 (10): 1-11.

Plotree, D., and Plotgram, D. (1989). PHYLIP-phylogeny inference package (version 3.2). *Cladistics,* 5 (163): 6.

Pollard, J.W. (1997). Role of colony-stimulating factor-1 in reproduction and development. *Molecular Reproduction and Development,* 46 (1): 54-61.

Ponsuksili, S., Jonas, E., Murani, E., Phatsara, C., Srikanchai, T., Walz, C., Schwerin, M. et al. (2008). Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. *BMC Genomics,* 9 (367): 1-14.

Ponsuksili, S., Murani, E., Phatsara, C., Schwerin, M., Schellander, K., and Wimmers, K. (2009). Porcine muscle sensory attributes associate with major changes in gene networks involving CAPZB, ANKRD1, and CTBP2. *Functional & Integrative Genomics,* 9 (4): 455-471.

Porto-Neto, L., Lee, S.-H., Sonstegard, T., Van Tassell, C., Lee, H., Gibson, J., and Gondro, C. (2014). Genome-wide detection of signatures of selection in Korean Hanwoo cattle. *Animal Genetics,* 45 (2): 180-190.

Pryce, J.E., Hayes, B.J., Bolormaa, S., and Goddard, M.E. (2011). Polymorphic regions affecting human height also control stature in cattle. *Genetics,* 187 (3): 981-984.

Puig-Oliveras, A., Revilla, M., Castelló, A., Fernández, A.I., Folch, J.M., and Ballester, M. (2016). Expression-based GWAS identifies variants, gene interactions and key regulators

affecting intramuscular fatty acid content and composition in porcine meat. *Scientific Reports,* 6.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J. et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics,* 81 (3): 559-575.

Puri, V., Virbasius, J., Guilherme, A., and Czech, M. (2008). RNAi screens reveal novel metabolic regulators: RIP140, MAP4k4 and the lipid droplet associated fat specific protein (FSP) 27. *Acta Physiologica,* 192 (1): 103-115.

Pyle, W.G., Hart, M.C., Cooper, J.A., Sumandea, M.P., de Tombe, P.P., and Solaro, R.J. (2002). Actin capping protein an essential element in protein kinase signaling to the myofilaments. *Circulation Research,* 90 (12): 1299-1306.

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

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

Ramayo-Caldas, Y., Ballester, M., Fortes, M.R., Esteve-Codina, A., Castelló, A., Noguera, J.L., Fernández, A.I. et al. (2014). From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine. *BMC Genomics,* 15 (232): 1-15.

Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A., Corominas, J., Castelló, A., Ballester, M., Estellé, J. et al. (2012). Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. *BMC Genomics,* 13 (1): 1.

Randhawa, I.A., Khatkar, M.S., Thomson, P.C., and Raadsma, H.W. (2016). A Meta-Assembly of Selection Signatures in Cattle. *PLoS One,* 11 (4): e0153013.

Raymond, M., and Rousset, F. (1995). An exact test for population differentiation. *Evolution,* 49 (6): 1280-1283.

Rege, J. (1999). The state of African cattle genetic resources I. Classification framework and identification of threatened and extinct breeds. *Animal Genetic Resources Information,* 25 1-25.

Rege, J., and Tawah, C. (1999). The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains. *Animal Genetic Resources Information,* 26 1-25.

Rewe, T., Herold, P., Kahi, A., and Valle Zárate, A. (2009). Breeding indigenous cattle genetic resources for beef production in Sub-Saharan Africa. *Outlook on Agriculture,* 38 (4): 317-326.

Rezaei, R., Wu, Z., Hou, Y., Bazer, F.W., and Wu, G. (2016). Amino acids and mammary gland development: nutritional implications for milk production and neonatal growth. *Journal of Animal Science and Biotechnology,* 7 (20): 1-22.

Rios Utrera, A., and Van Vleck, L.D. (2004). Heritability estimates for carcass traits of cattle: a review. *Genetics and Molecular Research,* 3 (3): 380-394.

Rolf, M., Taylor, J., Schnabel, R., McKay, S., McClure, M., Northcutt, S., Kerley, M. et al. (2012). Genome-wide association analysis for feed efficiency in Angus cattle. *Animal Genetics,* 43 (4): 367-374.

Rothammer, S., Seichter, D., Förster, M., and Medugorac, I. (2013). A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics,* 14 (908): 1-14.

Roux, P.-F., Boitard, S., Blum, Y., Parks, B., Montagner, A., Mouisel, E., Djari, A. et al. (2015). Combined QTL and Selective Sweep Mappings with Coding SNP Annotation and cis-eQTL Analysis Revealed PARK2 and JAG2 as New Candidate Genes for Adiposity Regulation. *G3: Genes| Genomes| Genetics,* 5 (4): 517-529.

Rubin, C.-J., Megens, H.-J., Barrio, A.M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C. et al. (2012). Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences,* 109 (48): 19529-19536.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X. et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature,* 449 (7164): 913-918.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution,* 4 (4): 406-425.

Samali, A., Holmberg, C.I., Sistonen, L., and Orrenius, S. (1999). Thermotolerance and cell death are distinct cellular responses to stress: dependence on heat shock proteins. *FEBS letters,* 461 (3): 306-310.

Santana, M., Kadarmideen, H., Pant, S., Alexandre, P., Junior, G.O., Gomes, R., Utsunomiya, Y. et al., 2014. Systems Genetics Investigations for Feed Intake, Feed Efficiency and Performance in Nellore (Bos indicus) Cattle. In: 10th World Congress on Genetics Applied to Livestock Production, August 17-22, 2014, Vancouver, BC, Canada

Santana, M., Ventura, R., Utsunomiya, Y., Neves, H., Alexandre, P., Oliveira Junior, G., Gomes, R. et al. (2015). A genomewide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nellore cattle. *Journal of Animal Breeding and Genetics,* 132 (6): 420-427.

Schellander, K. (2010). Identifying genes associated with quantitative traits in pigs: integrating quantitative and molecular approaches for meat quality. *Italian Journal of Animal Science,* 8 (2 Suppl): 19-25.

Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics,* 18 (3): 502-504.

Schoeman, S. (1989). Recent research into the production potential of indigenous cattle with special reference to the Sanga. *South African Journal Animal Sciences,* 19 (2): 55-61.

Seese, T.M., Harasaki, H., Saidel, G.M., and Davies, C.R. (1998). Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of muscle tissue to chronic heating. *Laboratory Investigation,* 78 (12): 1553-1562.

Seo, K., Mohanty, T.R., Choi, T., and Hwang, I. (2007). Biology of epidermal and hair pigmentation in cattle: a mini-review. *Veterinary Dermatology,* 18 (6): 392-400.

Shen, L., Lei, H., Zhang, S., Li, X., Li, M., Jiang, X., Zhu, K. et al. (2014). The comparison of energy metabolism and meat quality among three pig breeds. *Animal Science Journal,* 85 (7): 770-779.

Sherman, E., Nkrumah, J., and Moore, S. (2010). Whole genome single nucleotide polymorphism associations with feed intake and feed efficiency in beef cattle. *Journal of Animal Science,* 88 (1): 16-22.

Sherman, E., Nkrumah, J., Murdoch, B., Li, C., Wang, Z., Fu, A., and Moore, S. (2008). Polymorphisms and haplotypes in the bovine NPY, GHR, GHRL, IGF2, UCP2, and UCP3 genes and their associations with measures of growth, performance, feed efficiency and carcass merit in beef cattle. *Journal of Animal Sciences,* 86 (1): 1-16.

Shin, S., and Chung, E. (2007). SNP detection of carboxypeptidase E gene and its association with meat quality and carcass traits in Korean cattle. *Asian Australasian Journal of Animal Sciences,* 20 (3): 328-333.

Sibut, V., Hennequet-Antier, C., Le Bihan-Duval, E., Marthey, S., Duclos, M.J., and Berri, C. (2011). Identification of differentially expressed genes in chickens differing in muscle glycogen content and meat quality. *BMC Genomics,* 12 (112): 1-13.

Smedley, R., Mullaney, T., and Rumbeiha, W. (2009). Copper-associated hepatitis in Labrador Retrievers. *Veterinary Pathology,* 46 (3): 484-490.

Soares, M.J., Chakraborty, D., Kubota, K., Renaud, S.J., and Rumi, M.K. (2014). Adaptive mechanisms controlling uterine spiral artery remodeling during the establishment of pregnancy. *The International Journal of Developmental Biology,* 58 (2-4): 247-259.

Sommerhalter, M., Zhang, Y., and Rosenzweig, A.C. (2007). Solution structure of the COMMD1 N-terminal domain. *Journal of Molecular Biology,* 365 (3): 715-721.

Stafuzza, N.B., Zerlotini, A., Lobo, F.P., Yamagishi, M.E.B., Chud, T.C.S., Caetano, A.R., Munari, D.P. et al. (2017). Single nucleotide variants and InDels identified from whole-genome re-sequencing of Guzerat, Gyr, Girolando and Holstein cattle breeds. *PLoS One,* 12 (3): e0173954.

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

Stockton, M. (2003). Genetics of Feed Efficiency in Angus Cattle, Florida Beef Report.

Strimmer, K., and Von Haeseler, A. (1996). Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution,* 13 (7): 964-969.

Strydom, P. (2008). Do indigenous Southern African cattle breeds have the right genetics for commercial production of quality meat? *Meat Science,* 80 (1): 86-93.

Strydom, P., Frylinck, L., and Smith, M. (2011). Variation in meat quality characteristics between Sanga (Bos taurus africanus) and Sanga-derived cattle breeds and between Sanga and Brahman (Bos indicus). *Animal,* 5 (03): 483-491.

Strydom, P., Frylinck, L., Van der Westhuizen, J., and Burrow, H. (2008). Growth performance, feed efficiency and carcass and meat quality of tropically adapted breed types from different farming systems in South Africa. *Animal Production Science,* 48 (5): 599-607.

Strydom, P., Naude, R., Smith, M., Scholtz, M., and Van Wyk, J. (2000). Characterization of indigenous African cattle breeds in relation to carcass characteristics. *Animal Science,* 70 (2): 241-252.

Sturm, R.A. (2009). Molecular genetics of human pigmentation diversity. *Human Molecular Genetics,* 18 (R1): R9-R17.

Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Robert-Granie, C., Tosser-Klopp, G., and Arranz, J.J. (2015). Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing. *Scientific Reports,* 5.

Sung, M.K., and Bae, Y.J. (2010). Linking obesity to colorectal cancer: application of nutrigenomics. *Biotechnology Journal,* 5 (9): 930-941.

Sunil, K., Kumar, A., and Kataria, M. (2011). Effect of heat stress in tropical livestock and different strategies for its amelioration. *Journal of Stress Physiology & Biochemistry,* 7 (1): 45-54.

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

Takasuga, A. (2016). PLAG1 and NCAPG-LCORL in livestock. *Animal Science Journal,* 87 (2): 159-167.

Taye, M., Kim, J., Yoon, S.H., Lee, W., Hanotte, O., Dessie, T., Kemp, S. et al. (2017). Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genetics,* 18 (11): 1-14.

Teltathum, T., and Mekchay, S. (2010). Relationships between Pectoralis muscle proteomes and shear force in Thai indigenous chicken meat. *Kasetsart Journal - Natural Sciences,* 44 (1): 53-60.

Tempfli, K., Kiss, B., Szalai, K., Simon, Z., Pongrácz, L., and Bali Papp, Á. (2016). Differential expression of six genes in fat-type Hungarian Mangalica and other pigs. *Archives Animal Breeding,* 59 (2): 259-265.

The Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science,* 324 (5926): 528-532.

Theurl, I., Hilgendorf, I., Nairz, M., Tymoszuk, P., Haschka, D., Asshoff, M., He, S. et al. (2016). On-demand erythrocyte disposal and iron recycling requires transient macrophages in the liver. *Nature Medicine,* 22 (8): 945-951.

Traoré-Leroux, T., Fumoux, F., and Roelants, G. (1985). Correlation of serum zinc levels with resistance of cattle to trypanosomiasis. *Acta Tropica,* 42 (1): 39-44.

Turner, H. (1964). Coat characters of cattle in relation to adaptation. *Proceedings of Australian Society of Animal Production,* 5 181-187.

Turner, L.M., Chuong, E.B., and Hoekstra, H.E. (2008). Comparative analysis of testis protein evolution in rodents. *Genetics,* 179 (4): 2075-2089.

Utsunomiya, Y.T., Do Carmo, A.S., Carvalheiro, R., Neves, H.H., Matos, M.C., Zavarez, L.B., O'Brien, A.M.P. et al. (2013). Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC Genetics,* 14 (52): 1-12.

Utsunomiya, Y.T., O'Brien, A.M.P., Sonstegard, T.S., Sölkner, J., and Garcia, J.F. (2015). Genomic data as the" hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. *Frontiers in Genetics,* 6 (36): 1-13.

Våge, D.I., Klungland, H., Lu, D., and Cone, R.D. (1999). Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome,* 10 (1): 39-43.

Vatsiou, A.I., Bazin, E., and Gaggiotti, O.E. (2016). Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular Ecology,* 25 (1): 89-103.

Venturini, G., Cardoso, D., Baldi, F., Freitas, A., Aspilcueta-Borquis, R., Santos, D., Camargo, G. et al. (2014). Association between single-nucleotide polymorphisms and milk production traits in buffalo. *Genetics Molecular Research,* 13 (4): 10256-10268.

Verghese, J., Abrams, J., Wang, Y., and Morano, K.A. (2012). Biology of the heat shock response and protein chaperones: budding yeast (Saccharomyces cerevisiae) as a model system. *Microbiology and Molecular Biology Reviews,* 76 (2): 115-158.

Vitti, J.J., Grossman, S.R., and Sabeti, P.C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics,* 47 97-120.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *PLoS biology,* 4 (3): e72.

Wang, H., Zheng, Y., Wang, G., and Li, H. (2013a). Identification of microRNA and bioinformatics target gene analysis in beef cattle intramuscular fat and subcutaneous fat. *Molecular BioSystems,* 9 (8): 2154-2162.

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013b). WEB-based gene set analysis toolkit (WebGestalt): update 2013. *Nucleic Acids Research,* 41 (W1): W77-W83.

Wang, W., Xue, W., Jin, B., Zhang, X., Ma, F., and Xu, X. (2013c). Candidate gene expression affects intramuscular fat content and fatty acid composition in pigs. *Journal of Applied Genetics,* 54 (1): 113-118.

Wang, X., Chen, J., Liu, H., Xu, Y., Wang, X., Xue, C., Yu, D. et al. (2008). The pig p160 co-activator family: Full length cDNA cloning, expression and effects on intramuscular fat content in Longissimus Dorsi muscle. *Domestic Animal Endocrinology,* 35 (2): 208-216.

Wei, S., Zan, L., Ujan, J.A., Wang, H., Yang, Y., and Adoligbe, C. (2011). Novel polymorphism of the bovine fat mass and obesity-associated (FTO) gene are related to backfat thickness and longissimus muscle area in five Chinese native cattle breeds. *African Journal of Biotechnology,* 10 (15): 2820-2824.

Whitlock, B.K. (2010). Heritable birth defects in cattle. *Appl Reprod Strat Proc,* 146-153.

Whitlock, B.K., Kaiser, L., and Maxwell, H.S. (2008). Heritable bovine fetal abnormalities. *Theriogenology,* 70 (3): 535-549.

Wickramasinghe, S., Rincon, G., and Medrano, J. (2011). Variants in the pregnancy-associated plasma protein-A2 gene on Bos taurus autosome 16 are associated with daughter calving ease and productive life in Holstein cattle. *Journal of Dairy Science,* 94 (3): 1552-1558.

Wilke, K., Martin, A., Terstegen, L., and Biel, S. (2007). A short history of sweat gland biology. *International Journal of Cosmetic Science,* 29 (3): 169-179.

Wimmers, K., Murani, E., Ngu, N., Schellander, K., and Ponsuksili, S. (2007). Structural and functional genomics to elucidate the genetic background of microstructural and biophysical muscle properties in the pig. *Journal of Animal Breeding and Genetics,* 124 (s1): 27-34.

Wright, D. (2015). The genetic architecture of domestication in animals. *Bioinformatics and Biology Insights,* 9 (Suppl 4): 11.

WU, H., ZHAO, S., and FAN, B. (2010). Investigation of effects of the MKK3 and MKK6 genes on meat production traits in the pig (Brief Report). *Archiv Tierzucht,* 53 (2): 242-245.

Wu, J., Sun, H., Wang, F., Li, Y., Zhao, C., and Zhang, W. (2013). Adaptive evolution of Hoxc13 genes in the origin and diversification of the vertebrate integument. *Journal of Experimental Zoology, Part B: Molecular and Developmental Evolution,* 320 (7): 412-419.

Wu, J., Zhang, W.-g., Li, J.-q., Jun, Y., and Zhang, Y.-j. (2009). Hoxc13 expression pattern in cashmere goat skin during hair follicle development. *Agricultural Sciences in China,* 8 (4): 491-496.

Xu, L., Bickhart, D.M., Cole, J.B., Schroeder, S.G., Song, J., Van Tassell, C.P., Sonstegard, T.S. et al. (2015). Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Molecular Biology and Evolution,* 32 (3): 711-725.

Xu, Y., Qian, H., Feng, X., Xiong, Y., Lei, M., Ren, Z., Zuo, B. et al. (2012). Differential proteome and transcriptome analysis of porcine skeletal muscle during development. *Journal of Proteomics,* 75 (7): 2093-2108.

Yang, H., Xu, Z., Lei, M., Li, F., Deng, C., Xiong, Y., and Zuo, B. (2010). Association of 3 polymorphisms in porcine troponin I genes (TNNI1 andTNNI2) with meat quality traits. *Journal of Applied Genetics,* 51 (1): 51-57.

Yang, T.T., Suk, H.Y., Yang, X., Olabisi, O., Raymond, Y., Durand, J., Jelicks, L.A. et al. (2006). Role of transcription factor NFAT in glucose and insulin homeostasis. *Molecular and Cellular Biology,* 26 (20): 7372-7387.

Yaro, M., Munyard, K., Stear, M., and Groth, D. (2016). Combatting African Animal Trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance. *Veterinary Parasitology,* 225 43-52.

Yin, H., Gui, Y., Du, G., Frohman, M.A., and Zheng, X.-L. (2010). Dependence of phospholipase D1 multi-monoubiquitination on its enzymatic activity and palmitoylation. *Journal of Biological Chemistry,* 285 (18): 13580-13588.

Yokoyama, H. (2010). Three-dimensional structure of membrane protein stomatin and function of stomatin-specific protease. *Yakugaku Zasshi,* 130 (10): 1289-1293.

Yu, Y., Song, X., Du, L., and Wang, C. (2009). Molecular characterization of the sheep CIB1 gene. *Molecular Biology Reports,* 36 (7): 1799-1809.

Zaborski, D., Grzesiak, W., and Pilarczyk, R. (2016). Detection of difficult calvings in the Polish Holstein-Friesian Black-and-White heifers. *Journal of Applied Animal Research,* 44 (1): 42-53.

Zappaterra, M., Braglia, S., Bigi, M., Zambonelli, P., and Davoli, R. (2015). Comparison of expression levels of fourteen genes involved in the lipid and energy metabolism in two pig breeds. *Livestock Science,* 181 156-162.

Zhang, B., Zhang, Y., Zhang, L., Wang, J., Li, Z., and Chen, H. (2011). Allelic polymorphism detected in the bovine FTO gene. *Molecular Biotechnology,* 49 (3): 257-262.

Zhang, L., Li, W.-H., Anthonavage, M., and Eisinger, M. (2006). Melanocortin-5 receptor: a marker of human sebocyte differentiation. *Peptides,* 27 (2): 413-420.

Zhang, R. (2016). Identification of candidate genes for porcine meat quality and investigation of effects of sulforaphane on porcine satellite cell, Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn, Diss., 2016.

Zhao, F., McParland, S., Kearney, F., Du, L., and Berry, D.P. (2015). Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics Selection Evolution,* 47 (49): 1-12.

Zheng, A., Chang, W., Liu, G., Yue, Y., Li, J., Zhang, S., Cai, H. et al. (2016). Molecular Differences in Hepatic Metabolism between AA Broiler and Big Bone Chickens: A Proteomic Study. *PLoS One,* 11 (10): e0164702.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics,* 28 (24): 3326-3328.

Zhou, N., Lee, W.R., and Abasht, B. (2015). Messenger RNA sequencing and pathway analysis provide novel insights into the biological basis of chickens' feed efficiency. *BMC Genomics,* 16 (195): 1-20.

Zhou, X. (2010). Roles of Androgen Receptor in Male and Female Reproduction: Lessons From Global and Cell-Specific Androgen Receptor Knockout (ARKO) Mice. *Journal of Andrology,* 31 (3): 235-243.

Zielke, L.G., Bortfeldt, R.H., Reissmann, M., Tetens, J., Thaller, G., and Brockmann, G.A. (2013). Impact of variation at the FTO locus on milk fat yield in Holstein dairy cattle. *PLoS One,* 8 (5): e63406.

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F. et al. (2009). A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology,* 10 (4): R42.41-R42.10.

Zouboulis, C.C. (2004). Acne and sebaceous gland function. *Clinics in Dermatology,* 22 (5): 360-366.

국문초록

소의 전장유전체에서 적응적 흔적 발굴

멩기스티 타예 테레페

농생명공학부 동물생명공학전공

서울대학교 대학원 농업생명과학대학

소는 가장 흔하고 그 수가 많은 가축화 된 우제류 중 하나이다. 가축화된 소 품종들의 유전체에는 자연 선택과 인위 선택의 복합적 효과에 기인한 가축화와 품종 형성의 역사가 남아 있다. 진화생물학은 현대에 집단 내 및 집단 간에서 관찰되는 다양한 형태적 및 생산적 표현형질의 변화들 (Variations)을 일으켜온 핵심적인 적응적 특성들 (Adaptive Features)을 이해하는 것을 추구한다. 진화생물학적 관점에서, 소 품종들의 전장 유전체 내에서 이러한 선택의 힘 (Selection Forces)의 종적들 (Footprints)을 해독하는 것은 매우 흥미롭다. 최근에, 분자 집단 및 진화 유전학 전문가들은 소 품종들을

포함한 여러 생물들의 유전체에서 '선택 (Selection), 특히 양성 선택 (Positive Selection),'의 대상이 되는 분자 변이들과 중립적인 분자 변이들을 구별하는데 관심을 쏟고 있다. 저비용 및 고효율 차세대 염기서열해독 (NGS, Next Generation Sequencing) 기술의 출현 이후에 이어진 소 참조유전체의 구축과 지리적 및 생물학적으로 분화된 소 품종들의 단일 염기 다형성 (SNP, Single Nucleotide Polymorphism) 데이터의 축적은, 이런 변이들을 발견하고 이해하기 위한 전례 없는 기회를 창출하고 연구를 가능하게 하고 있다.

이 박사 논문에서는, 소 품종들 - 아프리카, 엔다마 (N'Dama), 앙콜 (Ankole), 홀스타인 (Holstein), 한우, 그리고 앵거스 (Angus) - 의 전장 유전체 NGS SNP 데이터를 활용하여, 각각의 소 품종들의 주요 표현형질에 기여하는 자연 선택의 힘 (Natural Selection Forces)과 인위 선택의 힘 (Artificial Selection Forces)의 종적을 설명하고자 한다. 선택에 영향을 받은 유전자/유전자지역을 탐색하기 위해서, 집단 간 확장된 단상형 동형접합성 (XP-EHH, Cross-population Extended Haplotype Homozygosity)과 집단 간 합성의 우도비 (XP-CLR, Cross-population Composite Likelihood Ratio)에 대한 통계적 분석을 활용했다. 소의 참조유전체 (UMD3.1)는 이러한 분석에서 선택을 받은 것으로 밝혀진 이상치 (Outlier) 지역에서 유전자를 주석 (Annotate)하는데 활용했다. DAVID (Database for Annotation, Visualization, and

Integrated Discovery) 데이터베이스의 분석도구인 유전자 온톨로지 (Gene Ontology)와 주석 (Annotation) 도구는 선택 받은 유전자들의 생물학적인 기능들과 경로들을 이해하기 위한 유전자군 강화 (Gene Enrichment) 분석에 활용하였다.

제 1장에서는, 아프리카 소 품종들에 대해 특별한 강조하며 소 품종들 내의 변화들 (Variations), 양성 선택 흔적의 배경적 원리들, 그리고 양성 선택 흔적을 규명하는 목적과 방법들을 서술하였다. 추가적으로, 유전적으로 분화된 소 품종들로부터 선택 흔적에 대한 선행 연구 결과들에 대해 검토하였다.

제 2장에서는, 열대 환경에 적응한 형질에 연관되어서 아프리카 소 품종들 내에서 선택을 받은 유전체 지역을 찾기 위해, 아프리카 소 품종들의 유전체와 상업용 아시아-유럽의 타우린 (taurine) 소 품종들을 비교하였다. 아프리카 소 품종들은, 그들의 내재적인 우수한 열저항적 능력이 발달하도록 도와준 더운 열대기후에서, 수 천년간 진화해왔다. 이 연구에서 유전자군 강화 분석을 통해서 몇몇 선택 받은 유전자/유전자지역이 다른 생물학적 과정 (BP, Biological Process) 용어들과 경로들에 과대표적인 것으로 밝혀졌다. 열 스트레스 반응에 연관해서 “혈관생성 (Angiogenesis)”와 “재생 (Regeneration)” BP 용어들이 강화 (Enriched)되었다. 게다가, 몇몇 선택을 받은 유전자들은 열 저항성 기작들

(mechanisms)에 연관된 해부학적 구조, 그리고 생리적이고/이거나 분자적 기능들에 관여했다. 이러한 유전자들은 산화적 스트레스 반응, 삼투압적 스트레스 반응, 열 충격 반응, 모발과 피부 성질, 땀샘 발달 및 땀의 분비, 먹이 섭취와 대사, 그리고 생식 기능들에 연관되었다. 그러므로, 이 연구에서 밝혀진 유전자들과 BP 용어들은 직·간접적으로 아프리카 소 집단의 우수한 열 저항성 기작들에 기여한다. 이런 소 품종들이 수 천년 간 진화해온 높은 열대 기온은 열 저항성 기작들의 발달에 대한 선택 압(Selective Pressure)이 되었을 것이다.

제 3장에서는, 우유 형질들, 육류 생산과 육질 형질들, 그리고 환경적 적응 형질들 각각에 대한 분기적인 선택으로 영향을 받은 유전체 지역을 판독하기 위해, 홀스타인, 한우, 그리고 엔다마의 유전체를 탐색하였다. 소의 특정 형질에 대한 인위 선택과 자연 선택은 소의 유전체를 유의하게 변화시켰다. 이 때문에, 몇몇 소 품종들은 형태적, 생산적, 그리고 적응 특성의 모자이크를 가지게 되었다. 인위 선택 하에 홀스타인과 한우는 각각 유용 품종과 육용 품종으로, 반면, 자연 선택 하에 엔다마는 일반적 목적의 품종 (특정한 목적으로 인위 선택 되지 않은 품종)으로 진화되었다. 이러한 인위 선택 힘들과 자연 선택 힘들에 의해서 영향을 받은 유전체 지역을 규명하는 것은, 집단 특이적 환경에 대한 유전적 적응과 경제적으로 중요한 형질에 대해 선택 받은 역사에 대한 통찰을 제공해줄 것으로 기대된다. 이

연구로, 홀스타인, 한우, 엔다마 소 품종 유전체 각각에서, 우유 형질 (예, *CSN3*, *PAPPA2*, 그리고 *ADIPOQ*), 육류 생산 및 육질 형질 (예, *NCOA2*와 *PITPN3*), 그리고 환경적 적응 형질 (예, *SLC40A1*, *STOM*, 그리고 *COMMD1*)에 연관된 유전자/유전자지역들이 양성 선택을 받은 것으로 나타났다. 게다가, 홀스타인, 한우, 그리고 엔다마 각각 소 품종에서 선택 받은 것으로 알려진 유전자들로부터, 유의한 기능적 주석 군집 용어들은 홀스타인에서 유단백질과 갑상선 호르몬의 신호전달경로를, 한우에서 히스톤 아세틸화효소 (Acetyltransferase) 활동을, 엔다마에서 레닌 분비를 포함하였다.

제 4장에서는, 육질 형질에 관련된 양성 선택 하의 유전자와 유전체 지역을 규명하기 위해 앙콜 소 (African Sanga cattle)의 유전체를 연구하였다. 아프리카 상가 소는 타우르스 (*Bos taurus*)와 인도원우 (*Bos indicus*) 아종들 (sub-species) 간 교배로 생산된 중간 유형의 품종이다. 최근에 아프리카 상가 소와 인도원우와의 육질 비교의 실험 결과 아프리카 상가 소의 육질이 뛰어남이 보고되었다. 이 장에서, 아프리카 상가 소의 전장 유전체 SNP 데이터와 인도원우의 유전체를 XP-EHH와 XP-CLR 통계 방법을 활용하여 비교하였다. 그 결과, 아프리카 상가 소에서 연함 (Tenderness), 근내지방함량 (IMF, Intermuscular Fat), 육류 색상과 같은 육질 형질에 영향을 미친 몇몇 유전자들이 양성 선택을 받은 것으로 밝혀졌다. 이 유전자들은

BP 용어들 및 KEGG 대사경로들에서 육질에 영향을 주는 근 구조와 대사, 지방대사 및 생성과 관련이 있음이 밝혀졌다. 이 연구 통해, 앙콜 소는 열대환경조건 하에서 우수한 육류 생산 및 품질 형질을 가질 수 있음을 밝혀냈다. 이러한 결과들은 열대 아프리카 지역에서 더 나은 품질의 소고기를 생산하기 위해서 앙콜 소와 다른 상가 소 품종들의 유전체적 특성을 연구하는데 기초를 제공할 것이다.

제 5장에서는, 앵거스 소 품종의 우수한 육질 특성과 다른 연관 표현형질들에 대한 유전적 청사진(Genetic Blueprint)을 규명하였다. 앵거스 소 품종은 지난 수 십년 간 우수한 육질 특성들에 대해서 집중적으로 선발되어 왔다. 앵거스 소 품종의 유전체에서 선택을 받은 유전체 지역을 규명한 결과로, 몇몇 유전자들이 육질 형질과 모색에 관련되어 있음을 확인하였다. 추가적으로, 앵거스 소 품종에서 유전적 결함을 일으킬 수 있는 추정유전자를 확인하였다. 이 연구 결과는 앵거스 소 품종의 육질을 더욱 향상 시키고, 생산과 생산성을 감소시키는 유전적 장애를 궁극적으로 예방 조치를 수행하는데 도움이 될 수 있다.

결론적으로, 상기 연구들을 통해, 각각 소 품종들에서 선택 되어온 주요 경제 형질 및 적응 형질에 연관된 아프리카, 엔다마, 앵콜, 홀스타인, 한우, 그리고 앵거스 소 품종들의 양성 선택 받은 유전자들의 목록을 규명했다. 이런 발견들은 오늘날에 우세한 소 품종들 간에 관측되는 다양한

표현형질적 변화를 발생시켜온 적응적 현상의 이해를 높일 것이다. 지역 환경 적응에 기여한 분자 표지 (Molecular Markers, 예시: 다른 실험실의 실험방법을 통해서 발굴하기 어려운 열 내성 기작 관련 표지)는 유생산과 우유 품질, 육류 생산과 육질, 생식과 그 외 연관 형질 같은 생산 형질들에 영향을 미치는 것이 밝혀졌다. 이 연구를 통해 발굴된 표지들은 소 품종들의 유전적 가치를 이해하고 다른 생산 시스템에서 적절하게 유전체 선발과 육종 프로그램을 개발하는데 사용될 수 있을 것이다.

키워드: 아프리카 소, 생물학적 기전, KEGG 경로, 양성선택 흔적, 바이오마커, XP-CLR, XP-EHH

학번: 2014-31453