University of Nottingham
UK | CHINA | MALAYSIA

GENOME DIVERSITY AND ADAPTATION OF AFRICAN TAURINE AND ZEBU CATTLE

Submitted by

Abdulfatai Tijjani, DVM, M.Sc.

A thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2018

Declaration

I hereby declare that this thesis has not been previously presented or submitted for examination to this University or any other one. The work described herein is entirely by me and references to other people's work has been duly acknowledged.

Abdulfatai Tijjani

September 2018

Abstract

There is an overwhelming phenotypic diversity observed among domesticated animals compared with what is observed in wild species. It follows domestication, migrations, human and natural selection of animals including adaptation to different climatic conditions, local environmental and production conditions. Indigenous African cattle represent a unique set of cattle population as of their adaptation to the conditions of their local environments, which is characterized by a diversity of climatic conditions (hot and dry, hot and humid, cold etc.), prevalence of livestock diseases, including parasitic and tick-borne diseases, and periods of food shortages as results of semi-intensive to extensive management system. However, the productivity of African cattle is generally low compared to the exotic cattle breeds in other parts of the world (Andersson 2001, Groeneveld *et al*., 2010).

We are faced with the challenges of improving livestock productivity on the African continent in order to meet the ever increased demands for livestock product due to population growth. This may be achieved by breeds replacement (importation of exotic breed, but it means changing the production system), within breed improvement (but takes time) or crossbreeding of exotic with local breeds (work but often only for the F1 generation not beyond). In such a context, a possible way forward would be to do 'informed' crossbreeding by combining the adaptive traits of the indigenous breeds with the productivity traits of the exotic ones. However, a prerequisite for this approach is the understanding of the genetic mechanism and identification of the genome regions that have been subjected to both natural and artificial selection and underlying economically important traits including environmental adaptation (Ellegren and Sheldon 2008, Cosart *et al*., 2011, Lee *et al*., 2013). This may be typically done using

genome-wide association in pedigree resource populations. Alternatively, recent approaches have been suggested to identify key genome regions with remarkable footprints due to selection pressures; the detection of signatures of positive selection. The latter is the focus of this work.

Thus, the primary goal of the research presented herein is to unravel at genome-wide level the genome diversity and particularly the genetic basis of economically important phenotypes including morphology (growth traits), productivity (milk production traits) and local environmental adaptation traits such as disease resistance, heat tolerance and tick resistance traits of indigenous African taurine and zebu cattle through signatures of selection analyses. The full genome sequences of an average of ten representative samples of ten different African cattle breeds were recruited in our attempt to achieve this goal. The African breeds involved are two West African taurine; Muturu from Nigeria and N'Dama from Guinea, and eight East African zebu breeds; Aryashai, Baggara, Kenyan Boran, Butana, Fulani, Gash, Kenana and Ogaden. In addition, samples of five reference cattle breeds including Ankole (African Sanga), Holstein and Jersey (European taurine), Hanwoo (Asian taurine) and Gir (Asian zebu) were also included for ease of comparison and interpretation of results.

In the course of this study, 70 new genome cattle sequences were generated, these were complemented with cattle genome sequences obtained from previous studies with sequences data publicly available. We adopted several selection scan approaches including two within population tests; iHS and ZHp, and four population comparison tests; Rsb, XPEHH, XPCLR and F_{ST}.

Prior to signature of positive selection analyses, in chapter 2, we performed variants (SNPs and indels) discovery by comparing the sequences of our cattle samples to the UMD3.1 bovine genome assembly. We subsequently carried out an assessment of genetic diversity, population structure and phylogenetic relationships among the different cattle breeds. The aim was to reveal the genome variations and population structure among the cattle breeds, which in turn will guide the basis for the selection breeds for the comparative genomic tests. As expected, we observed higher mean variant numbers (SNPs and indels) and nucleotide diversity in the zebu breeds than in the taurine breeds, with the ratio as high as 2:1 in most zebu breeds. A significant difference of population structure and differentiation was revealed between West African and European taurine populations, and also within the West African taurine breeds. All African zebu were shown to be of at least two ancestry background (indicine and taurine).

In subsequent chapters, we aim to identify signatures of positive selection. In order to increase the reliability of our results, our strategy entailed the combination of within-population tests and between population tests in order to identify a comprehensive list of candidate genes in the different cattle populations studied. The between-population approach involved the comparison of groups of cattle breeds to identify candidate genomes regions and genes related to contrasting phenotypes and environmental challenge. We then performed the functional annotation of the candidate genes to elucidate the function of the candidate genes putatively linked to the phenotypes under study.

Like most other West African *Bos taurus*, the shorthorn Muturu is under imminent threat of replacement or crossbreeding with zebu population and their populations have been reduced to a few hundred breeding individuals only. Hence, they have been classified as an endangered breed. In chapter 3, for the first time, we present the genome-wide selection signatures of the endangered trypanotolerant West African shorthorn Muturu based on two complimentary selection scan tests; iHS and Rsb. The results were also compared to N'Dama, a West African longhorn trypanotolerant taurine, and two European taurine breeds (Holstein and Jersey). Among the most remarkable selection signatures regions found in the Muturu cattle, are regions which overlap with members of the major histocompatibility complex (MHC) class I and class II genes and other genes with functions related to both innate and adaptive immunity. These genes particularly the MHC class II genes and genes linked to heat tolerance, such as *INTS6*, are shared with the N'Dama. Considering the production environment of both West African *Bos taurus* (WAT) breed studied, which is characterized by high disease prevalence and harsh environmental climatic pressures such as heat (high temperature) and UV radiation, the signature of selection signals detected here may be the consequence of the innate adaptation mechanisms contributing to the survival of these two taurine cattle living in tropical areas.

In chapter 4, the genomic signatures of the two trypanotolerant West African taurine breeds; shorthorn Muturu and longhorn N'Dama, were further investigated. Our analysis was based on candidate genes identified by six selection scan tests; iHS, ZHp, Rsb, XPEHH, XPCLR and F_{ST}. Each of the trypanotolerant cattle was compared to two groups of trypanosusceptible cattle populations (African zebu and European taurine) in order to detect candidate genes that may be related to their common trypanotolerance

phenotype. Following the annotation of the detected candidate gene sets in each breed, the candidate genes were revealed to be involved in pathways relevant to trypanosomiasis disease progression. The list of common genes between the breeds in these pathways was then investigated further. A major finding of this chapter is the major role that the *PTPN6* gene may play in the genetic control of trypanotolerance in West African cattle, with several lines of supportive evidence. These are – a unique WAT haplotype in the region of the gene and a protein-protein interaction network indicating the most likely central role of *PTPN6*, through its interactions with bovine MHC class II genes and other genes to initiate a cascade of biological processes that may confer protective immunity for the survival of cattle following trypanosome infection.

Finally, in chapter 5, we investigated the genetic control of tropical adaptation and production traits in eight indigenous African zebu breeds. Among the common selective sweeps identified in a majority of the African zebu breeds are genome regions which overlapped candidate genes such as *HMGA2* and *PTPRG* that are related to growth and conformation traits. Candidate genes related to immune response, feeding behaviour, coat colour were also identified across breeds. In addition, heat stress response and tick resistance response gene were investigated by comparing African cattle to non-African taurine breeds. Last but not least we identified candidate genes related to milk yield, protein yield and fat yield in the African zebu dairy breeds (Kenana and Butana) following the comparison of each of this breed to African non-dairy zebu.

Put together, the findings of this thesis buttress the facts that African cattle are a group of unique cattle whose genetic resources could be exploited for the genetic

improvement of livestock productivity in sub-Saharan African. The new sequences generated within the course of this project serve as a valuable addition to the limited publicly available genetic resources of livestock from Africa. We envisaged that our results will encourage further investigation into the genome dynamics of the African cattle.

Publication

Bahbahani, H., **Tijjani, A**., Mukasa, C., Wragg, D., Almathen, F., Nash, O., Sonstegard, T. & Hanotte, O.(2017). Signatures of Selection for Environmental Adaptation and Zebu × Taurine Hybrid Fitness in East African Shorthorn Zebu. *Frontiers in Genetics*, *8*, 68

Selected Conferences

Symposium on African Livestock Trypanosomes 2018 (#SALTed2018). The Roslin Institute, Easter Bush Campus Midlothian, United Kingdom. (Attended)

36th International Society for Animal Genetics (ISAG) Conference, UCD, Dublin. 16-22, July 2017 – Poster and oral presentations on "Towards the unravelling of the genomic basis of milk production traits in African dairy zebu cattle"

School of Life Sciences Postgraduate Research Symposium, University of Nottingham, UK. 13-14, July 2017. – Oral presentation on "Genome diversity and adaptation of African trypanotolerant taurine and zebu dairy cattle"

ISCB NGS-Barcelona - Structural Variation and Population Genomics Conference, Barcelona, Spain. 3-4 April 2017 – Poster presentation on "Whole genome copy number variation (CNV) detection in West Africa trypanotolerant taurine cattle population"

35th International Society for Animal Genetics (ISAG) Conference, Salt Lake City, Utah, USA. 23-27, July 2016 – Poster and oral presentations on "Comparative genomics reveal common diversity and signature of selection in West African taurine cattle population"

49th UK Population Genetics Meeting, Edinburgh, UK. 15-18, December 2015 – poster presentation on "Comparative genomics reveal common diversity and signature of selection in West African taurine cattle population"

Awards

The International Research Collaboration Award (IRCA), Graduate School, University of Nottingham, UK	2017
2016 ISAG Yoshi Suzuki Award	2016
ISAG Travel bursary	2016
School of Life Sciences Travel Bursary	2015
Vice-Chancellor's Scholarship for Research Excellence (International)	2014

Acknowledgment

Firstly, I give all glory and honor to Almighty Allah for His countless blessings and for the opportunity to have come this far in my academic pursuit, Alhamdulillah Robil al-Amin.

My special thanks goes to my Supervisor, Professor Olivier Hanotte for his guidance and for sharing with me from his wealth of knowledge and research experience. I especially appreciate your scientific insight and encouragement which has helped me to come this far in my research study as well as career development. I thank the University of Nottingham for making my PhD dream a reality through the award of the Vice Chancellor Scholarship. Thanks to the management and staff of National Biotechnology Development Agency (NABDA), Abuja for the opportunity to further my studies. I particularly thank my Director, Professor Nash for his encouragement and support, this means a lot to me. Many thanks to Dr Jummai Adamu-Tutuwa for her advice and support too.

I am indebted to a number of people for whom their advice and assistance contributed to the success of this thesis. I appreciate our various collaborators including Professor Hassan Musa of the University of Khartoum for providing the new sequences of the Sudan zebu breeds studied. Worthy of mentioning are Tad Sonstegard for hosting me in his lab at Recombinetics, USA, Dr. Yuri Utsunomiya and Dr. Bashir Salim for their regular support and fruitful discussions. I wish to acknowledge the support of the staff of the School of Life Sciences, UoN and my colleagues in Professor Olivier Hanotte' s lab for their support.

My deepest appreciation goes to my family for their enormous support in various ways. To my dear wife, Ganiyat and our adorable children, Faizaan, Fatima and Zahra, I appreciate you all for your patience and for accommodating my numerous excuses during the periods of my absence from home, I will always adore you all. Special thanks to my mother, Alhaja Sherifat for her unflinching support, prayers and encouragements. To my Big sisters, Mrs Fatima B. Bamidele, Hon, Mulikat Adeaola Akande and Mrs Idayat Raji, I thank you very much for your love and support always. Also, I appreciate all my other siblings for supporting me in different ways, you are all part of my success story, Thank you!

Finally, this thesis is dedicated to the loving memory of my beloved Father, Alhaji Tijjani Lawal Akande (ATC).

Table of Contents

List of Figures

List of Tables

Abbreviations

AFDZ	African dairy zebu
AFNDZ	African non-dairy zebu
AFZ	African zebu
ARY	Aryashai
BGR	Baggara
BOR	Kenyan Boran
bp	base pairs
BTA	*Bos taurus* Autosome
BTN	Butana
CNV	copy number variation
FLN	Fulani
F_{ST}	fixation index
GATK	Genome analysis toolkits
GDR	Gene desert region
GIR	Gir
GSH	Gash
HNW	Hanwoo
HOL	Holstein
Hp	Pooled Heterozygosity
iHS	Integrated haplotype score
InDel	insertion/deletion
JER	Jersey
KNN	Kenana
LD	Linkage disequilibrium

MAF	minor allele frequency
MHC	major histocompatibility complex
MUT	Muturu
NDM	N'Dama
QC	quality control
QTL	Quantitative trait loci
SD	standard déviation
SNP	single nucleotide polymorphism
WAT	West African taurine

CHAPTER 1:

General Introduction

1.1 Background

Cattle are one of the most economically and scientifically important livestock species. Since domestication, they have been widely kept by humans worldwide as a means of livelihood, especially across Africa where they continue to serve as a major source of draft power, milk and meat besides fulfilling socio-cultural roles. As a model organism, they have already contributed to our knowledge of the mammalian endocrine function, fertilization, and growth physiologies (Gibbs *et al.,* 2002). They have adapted to intense selective pressures since domestication while keeping and/or developing unique environmental adaptation, for example, their ability to utilize low quality forage for their energy requirements and for meat and milk production (Bradley and Magee 2006, Snelling *et al.,* 2007, Nijman *et al.,* 2008). In this respect, they represent an important model to discover and analyses the genetic control of adaptation to environmental pressures in mammals (Porto-Neto *et al.,* 2013).

The worldwide cattle population have been classified into approximately 1,000 recognized breeds based on their origin, unique characteristics and adaptation to different environments (Mason 1969). Over 145 breeds comprising over 192 million cattle are indigenous to Africa (Rege 1999). Despite being poor producers in comparison to commercial exotic breeds, African cattle are well adapted to their local environment. They represent a unique genetic resource that may be characterized so as to provide us with information for the improvement of livestock productivity (Loftus *et al.,* 1999, Troy *et al.,* 2001, Hanotte *et al.,* 2002). However, our knowledge of African cattle in terms of their genetic diversity and genome variation is yet inadequate compared to commercial breeds calling for the investigation of the genetic basis of

fitness, phenotypic differentiation, adaptation, including disease resistance (Cosart *et al.,* 2011).

Advances in genomic technology combined with the availability of reference genomes provide us with the opportunity to carry-out studies aimed at unravelling the basis of the genetic and genomic variation of African cattle population and the characterization of their genomes (Achilli *et al.,* 2009, Hanotte *et al.,* 2010). Studies such as genetic characterization, genome-wide association studies, detection of signature of selection are some of the important studies which have been performed extensively on commercial cattle breeds and now require more attention in indigenous African cattle populations (Mbole-Kariuki *et al.,* 2014).

These studies generally require the identification of a substantial number of genome-wide SNPs that could be used as genetic markers. Generating these markers by next generation DNA sequencing technology is still costly even though costs are rapidly declining (Shendure and Ji 2008). DNA sequencing of the protein coding region of the genome (exome) in some studies can be an alternative because of its lower cost (Cosart *et al.,* 2011). In any case, application of NGS platform for the discovery and analysis of high quality markers from the complex genomes of plants and animals is now the focus of many genetic studies (Ramos *et al.,* 2009, Kilian and Graner 2012).

Among the different types of genetic marker types, the Single Nucleotide Polymorphisms (SNPs) are increasingly gaining popularity in functional and neutral genetic diversity studies. They are the most common source of genome-wide genetic variation associated with heritable differences between individuals (Suh and Vijg

2005). With low mutation rate, they are thought to be good genetic markers of potential disease phenotypes as well as for other complex traits. Moreover, SNP markers are amenable to high throughput genotyping platforms and are valuable for a variety of genetic and genomic applications such as the construction of genetic and physical maps as well as the analysis of genetic diversity (Gray *et al.,* 2000)**.**

Investigation of the genomic variation and ultimately the identification of causative mutation(s) responsible for the adaption of African cattle to their environment at the genome wide level is a multi-step process. The most powerful approach is the analysis of full genome data. African cattle genome sequence may be mapped to the cattle reference genome, currently either Btau_4.6.1 (Liu *et al.,* 2009) and UMD 3.1 (Zimin *et al.,* 2009), in order to detect SNVs (SNPs and indels), structural variants (large insertions, deletions, inversion) and Copy Number Variation (CNV). Then, typically a genome wide search for signature of selection through the statistical analysis of these polymorphisms will be performed, this may be done with or without phenotypic information. Subsequently, candidate polymorphisms within these regions may be further investigated through the analysis of more cattle populations and association studies with inferred phenotypes.

Selection signature analyses have been performed in both humans and domestic animal species (Qanbari and Simianer 2014)**.** Positive, purifying and balancing selection are different selection types that can be detected. However, positive selection is easier to detect in the population as a result of increase in the prevalence of advantageous alleles which manifest as a unique pattern of genetic variation in the genome. Depending on the approach, whether intra- or inter-population studies, several analysis methods have

been designed and one or a combination of these methods may be used to identify positive selection in domestic species. These methods have been developed on the basis of haplotype structure, local variability and allele frequency spectrum analysis which are all informative to detect selective sweeps (Biswas and Akey 2006, Gouveia *et al.,* 2014, Qanbari and Simianer 2014).

1.2 The domestication of indigenous African cattle

All modern cattle are classified into two main subspecies or species namely; the humpless taurine (*Bos taurus taurus* or *Bos taurus*) found in Europe, northern Asia, and West Africa and the humped zebu cattle (*Bos taurus indicus or Bos indicus*) primarily found in southern Asia and Africa. They belong to the Bovini tribe within the Bovidae family, which also include other members such as the Saola, Buffalos, Yak and Bison. They are believed to have originated from the extinct wild aurochs, *Bos primigenius* (Grigson 1978, Grigson 1980, Epstein 1984, Loftus *et al.,* 1994).

The domestication of the two cattle sub-species is believed to have occurred in different centres and dates as far back as 10,000 years (Bradley *et al.,* 1996). While *B. indicus* is believed to have originated from southern Asian aurochs (*Bos primigenius nomadicus*) and domesticated in South Asia (Indus valley), *B. taurus* is thought to have originated from the northern Eurasian aurochs (*Bos primigenius primigenius*) and it was domesticated in the Near East region (Fertile Crescent). These two separate domestication events are supported by both archaeological evidences (Meadow 1984) as well as genetic data including evidence from *Y*-chromosome (Nijman *et al.,* 2008) and mitochondrial DNA (Achilli *et al.,* 2009) studies (Epstein 1984, Grigson 1991, Loftus *et al.,* 1994, Bradley *et al.,* 1996). Also, genome wide analysis of the bovine

SNP polymorphism 50K SNP chips (Bovine Hapmap Consortium, 2009) and bovine high density 770,000 SNP markers support the distinction between zebu and taurine cattle and as well a further distinction between the European taurine and African taurine breeds (Porto-Neto *et al.,* 2013).

There have been several debates on the origin of African cattle. With the proposition of, in addition to the two centres of domestication in the Fertile Crescent and Indus Valley, a third domestication event in north-east Africa. The propositions of both two and three domestication events seems to be supported by controversial archaeological and genetic data (Grigson 1991, Loftus *et al.,* 1994, Hanotte *et al.,* 2002, Stock and Gifford-Gonzalez 2013). However, recent full mitochondrial DNA analysis of cattle favours only two separate domestication centres, with the African cattle mtDNA diversity embedded within the European and Asian cattle mtDNA (Achilli *et al.*, 2009). Also, whole-genome SNP arrays supported two domestication events for taurine and indicine lineages followed by introgression from wild aurochs in Africa, East Asia and Europe (Decker *et al.,* 2014).

The introduction of cattle into the African continent occurred in different phases with the taurine cattle considered to have been introduced first approximately 7000 years ago and subsequently zebu approximately 4,500 years ago (Lott and Hart 1979). Among the African taurine cattle, the longhorn taurine are thought to be the first on the African continent. They are believed to have dispersed with Hamitic people southwards through present-day Sudan, to the west along the northern coastal region, southwest into West Africa and also centrally through a much-reduced Saharan region. The cattle which entered West Africa are also believed to have encountered

disease challenges forest areas infested by the vector of trypanosomiasis. They subsequently evolved tolerance to trypanosomiasis and were able to thrive in the southern forested regions. These African longhorn taurine cattle are represented today as the trypanotolerant N'Dama breed predominantly found in West African countries, a center of origin in the Fouta Djallon region of Guinea (Murray 1988).

Around 3000 BC .the shorthorn taurine cattle were brought into Africa through the North Africa region, specifically Egypt. They gradually out-numbered the initial longhorn taurine population and subsequently, spread throughout North, East and West Africa. They also encountered trypanosomiasis in the West Africa region and eventually developed tolerance trait like the longhorn taurine (Payne 1970). These African shorthorn taurine cattle are represented today, for example, by the trypanotolerant Muturu, Baoule, Somba (Murray 1988).

The cervico-thoracic zebu cattle were believed to be the first humped zebu that were brought into in Africa through the Horn of Africa (Somalia and Ethiopia regions), from the Arabian Peninsula around 2000 BC (Payne 1970). The population of taurine in the regions began to be impacted when pastoralists began to crossbreed them with zebu due to the superior adaptation to the arid environment. These ancient crossbred taurine x zebu cattle, often referred as the Sanga cattle population, subsequently dispersed into the southern and central African regions by the migrating Khosian and Bantu speaking people (Epstein 1984). It is assumed that African taurine x zebu crossbreeding was predominantly a zebu male mediated process as no zebu mitochondrial DNA haplotypes have been found so far on the African continent (MacHugh *et al.,* 1997, Hanotte *et al.,* 2002). This

is further supported by combining findings from different studies using mitochondrial DNA markers (Bradley *et al.,* 1996), *Y*-chromosome marker (Hanotte *et al.,* 2000) and autosomal microsatellite markers (Rege *et al*, 2001).

During the Islamic Arab conquests and expansion of influence from 7th century AD, the thoracic-humped zebu were introduced into Africa through Horn of Africa and its coastal area. They were mostly preferred to the Sanga cattle due to their higher milk production capability in addition to their adaptive traits. Subsequently, they spread to other parts of Africa including West Africa where they were crossbred with the trypanotolerant taurine living in the tsetse infested areas. It is also believed that the rinderpest epizootics which occurred in 1890's in east and southern Africa, wiped out the majority of the taurine cattle, which were more susceptible to the disease than animals with predominant zebu background, thereby resulting in the presence predominantly of more resistant zebu cattle types in the regions affected by the rinderpest epidemics (Denbow and Wilmsen 1986).

1.3 African cattle types and their adaptation to tropical environment and diseases

As previously mentioned, three main types of indigenous African cattle are nowadays recognisable in Africa, these are: Zebu (humped cattle), widely distributed in East and West Africa; taurine (humpless cattle), predominantly found in Central and some parts of West Africa; Sanga populations, found mainly in Eastern and Southern Africa (Grigson 1991). Figure 1.1 shows the different cattle types in Africa.

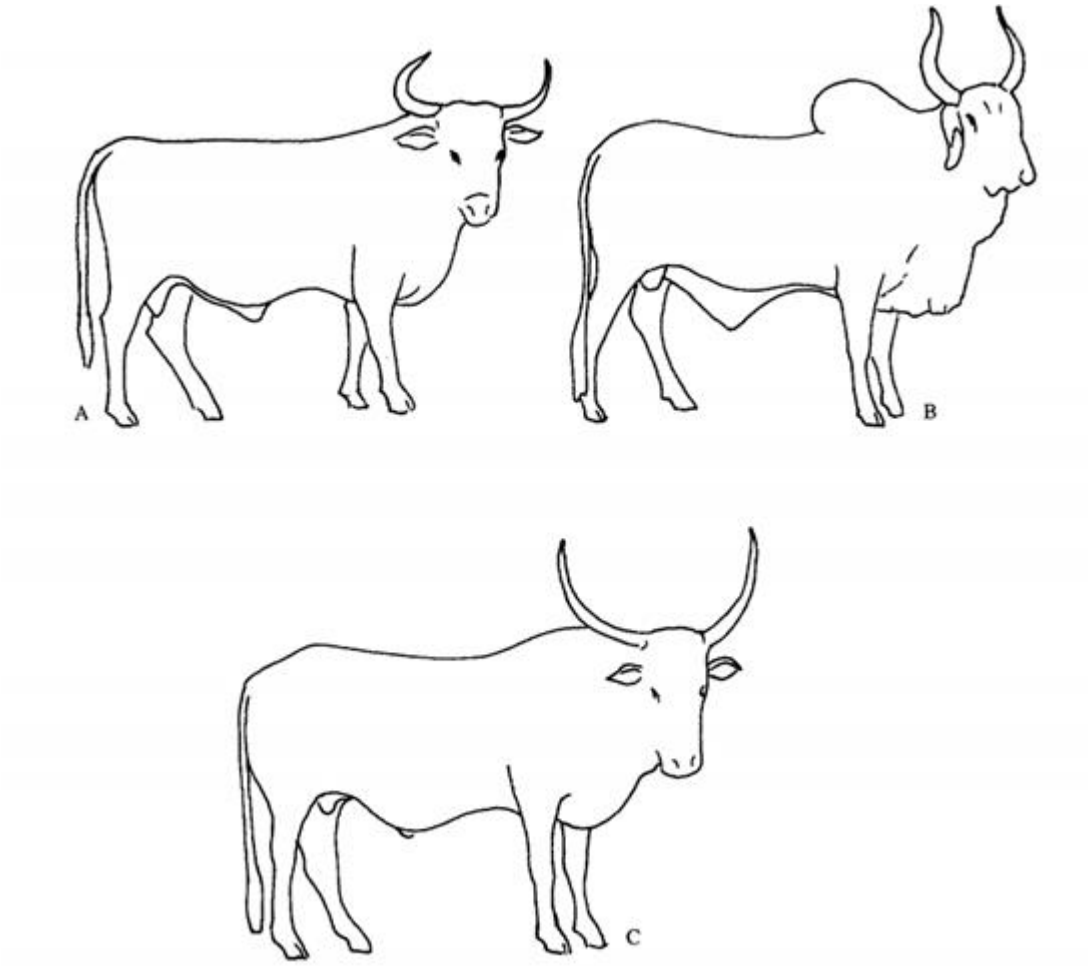


Figure 1.1 | The three main cattle types in Africa. A is typical taurine, B is typical zebu and C is a typical Sanga cattle. Picture adapted from (Grigson 1991)

The unique feature of the indigenous cattle in Africa is their particular adaptation to conditions of their local environments, which is characterized by hot and humid weather, prevalence of livestock diseases, including parasitic infectious diseases, and a semi-intensive to extensive management system. These animals have been able to adapt to these local conditions by being tolerant to thermal stress and endemic diseases while surviving on less nutritious grasses. However, indigenous African cattle breeds whether zebu or taurine or admixed types have shown varying degree of adaptation due to the different environmental conditions (Latif *et al.,* 1991, Mattioli *et al.,* 1993, Carvalho *et al.,* 1995, Wambura *et al.,* 1998, Hansen 2004, Mbole-Kariuki *et al.,* 2014).

1.3.1 Trypanotolerance in West African taurine

Tolerance to certain diseases is among the unique feature of the African cattle. It is an established fact that certain indigenous Africa cattle show tolerance and/or to some extent resistance to diseases such as trypanosomiasis (sleeping sickness) and cowdriosis (heart water) compared to their exotic animal and crossbred counterpart (Mirkena *et al.,* 2010). Trypanosomiasis is a common disease in sub-Saharan African caused by the parasitic protozoa of the genus *Trypanosoma* that affects many species including cattle and humans. In Africa, it is mainly transmitted by the tsetse fly of the *Glossina* genus (Courtin *et al.,* 2008, Lamy *et al.,* 2012). Trypanosomes affecting livestock in Africa include *T. congolense, T. vivax, T. evansi*, and *T. brucei*, all of which infect blood and tissues. Major symptoms in cattle include hyperthermia, anaemia, rapid weight loss, mucous pallor, miscarriage, 'petering out', pica, splenomegaly, cachexia, and death (Naessens 2006, Courtin *et al.,* 2008).

Connor (1994) has observed that the vector fly appears to prefer cattle hosts over other domestic animals even though they feed on a range of hosts. Africa indigenous cattle, particularly the West African taurine breeds living in areas infested with tsetse fly have developed tolerance to trypanosomiasis whereas other cattle imported from non-endemic areas are highly susceptible (Lamy *et al.,* 2012). The West African longhorn N'Dama cattle and shorthorn breeds including Muturu, Lagune, Bauole, Somba are renowned for their ability to withstand trypanosome infection and often remained productive under the disease condition, hence they are referred to as trypanotolerant breeds (Roberts and Gray 1973). Their tolerance ability is due to less severe anaemia developed during trypanosomiasis infection and the control of parasitemia level (Murray 1988). However, the physiological mechanisms involved in dampening the

effects of the protozoan infection as well as the genetic mutation associated to trypanotolerance remain unclear (Spickett 1994)

1.3.2 Heat tolerance in African cattle

According to Lamy *et al*. (2012)**,** animals generally adopt various strategies to achieve balance between heat gain and heat loss from the body in order to avoid an increase in body temperature and maximize heat loss when subjected to high ambient temperatures. These strategies include: (1) increasing surface area per unit of body weight; (2) increasing the temperature gradient between animal and air; (3) increasing conduction of heat from the body core to the skin; (4) decreasing solar radiation reflection; (5) increasing metabolic rate and feed intake, and (6) adjusting cellular mechanisms. *B. indicus* show higher thermo-tolerance to hot and dry environment than both the African and European taurine because of their physiological abilities to combine the above mentioned strategies (Lamy *et al.*, 2012).

For instance, zebu cattle have ability to dissipate heat more than the *B. taurus* through their sweats glands, skin appendages and thoracic hump. The sweat glands are numerous and large, producing greater sweat, the skin appendages and hump help to increase the skin surface area thereby reducing the amount of heat produced by the body and increasing heat dissipation. The genetic difference in heat tolerance between zebu and taurine also extends to the cellular level as the effect of high temperature on cellular functions such as feed intake, growth rate, milk yield as well as reproductive function appears more deleterious in the taurine breeds (Malayer *et al.*, 1990, Hernández-Cerón *et al.*, 2004). Another advantage of thermal stress tolerance in zebu is due to their characteristic lighter colour, sleeker and shinier coat cover as opposed to the darker,

denser and typically woolly coat of the taurine species especially in the European taurine (Lamy *et al.*, 2012).

However, adaptation to heat stress is also expected in West African taurine living in hot and humid environments. Of relevance here is the recent finding of the genetic control of the slick phenotype in tropically adapted taurine Senepol cattle in which mutations within prolactin (*PRL*) and prolactin receptor (*PRLR*) genes been associated with the slick coat phenotype in heat tolerant animals (Littlejohn *et al.*, 2014, Porto-Neto *et al.*, 2018)

1.4 Sequencing of the bovine genome

The development and advances in genomics technology have led to the possibility of generating genomic sequence of any organism. Sequencing of the genome of a species and a good quality reference genome(s) assembly are no doubt an invaluable requirement for ongoing and future mammalian genomic research. Landmark genome sequencing started with the Human Genome project (1999 - 2003) (Lander *et al.*, 2001, Collins *et al.*, 2003, Wheeler *et al.*, 2008). Genome sequencing of other species only started a few years after the successful completion of the Human genome project in 2003. Sequencing is now being carried out almost routinely due to development of much faster sequencing platforms and the rapidly declining associated cost (Cosart *et al.*, 2011).

The domestic cattle genome was one of a few species initially sequenced after the human genome. The Human Genome Sequencing Centre at Baylor College of Medicine released the initial bovine genome assembly, BosTau 4.0 (BCM4) in 2007 (Liu *et al.*,

2009), an alternative assembly, University of Maryland assembly of *B. taurus*, release 2 (UMD2) was also developed in 2009 by the Centre for Bioinformatics and Computational Biology (CBCB) at University of Maryland (Zimin *et al.,* 2009). Both assemblies were developed with the sequences of a single inbred female Hereford cow, L1 Dominette 01449, and her sire, which were sequenced using a combination of hierarchical sequencing and whole-genome shotgun approaches. While BCM4 was built using a BAC-based Atlas assembler, UMD2 was built using a modified Celera Assembler algorithm. Contrasting features between these two initial assemblies include 53 Mb of sequences assigned to the X chromosome in UMD2 not found in the BCM4 assembly. Also, the UMD2 contain some sequences assigned to the *Y* chromosome whereas the BCM4 does not assign any sequence to the *Y* chromosome. Overall, the UMD2 is the larger assembly with an estimate of approximately 2 percent of assigned sequences missing from the BCM4 assembly (Zimin *et al.,* 2009). Although both assemblies have had several improved versions, they are still characterized with many issues. The current versions of the two assemblies are Btau5.0.1 and UMD3.1.1 respectively (https://www.ncbi.nlm.nih.gov/assembly), as of 25 July 2018. The latest version of Btau5.0.1 takes the UMD 3.1 assembly contigs and scaffolds and incorporates ~19x Pacific Biosciences long-read data using the PBJelly method (Richards *et al*. 2012). Until very recently, they were known in the research community as the reference genome assemblies for the *B. taurus taurus* subspecies.

Nevertheless, improvements in long read technologies and assembly methods have motivated the construction of a new *de novo* bovine reference assembly referred to as ARS-UCD1.2. This new assembly was constructed from approximately 80X PacBio and 80X Illumina coverage of Dominette sequences. The new assembly statistics

include a N50 contig (a set of overlapping segments of DNA) size of 25.9 Mb and an N50 scaffold size of 103 Mb representing 267- and 16-fold improvements over UMD3.1.1 (contig N50=0.097Mb, scaffold N50=6.4Mb) and 94- and 15-fold improvements over Btau5.0.1 (contig N50=0.276Mb, scaffold N50=6.8Mb) (https://www.ncbi.nlm.nih.gov/assembly). The full-lengths of the transcripts from 30 Dominette tissues were additional sequenced in order to support improved annotation. Comparison of the old (UMD3.1.1 and BTau5.0.1) and new (ARS-UCD) *B. taurus* assemblies is presented in Table 1.1

Table 1.1 | Comparison of the current three bovine reference genome assemblies.

Assembly statistics	Btau.5.0.1	UMD3.1.1	ARS-UCD1.2
Total sequence length	2,724,980,740	2,670,123,310	2,715,853,792
Total assembly gap length	11,947,874	20,737,887	28,162
Gaps between scaffolds	2,856	3,193	0
Number of scaffolds	5,998	6,336	2,211
Scaffold N50	6,806,220	6,380,747	103,308,737
Scaffold L50	104	107	12
Number of contigs	42,267	75,617	2,597
*Contig N50	276,285	96,955	25,896,116
*Contig L50	2,849	7,930	32
Total number of chromosomes and plasmids	31	30	31
Number of component sequences (WGS or clone)	42,574	75,597	2,211

* Contig refers to a set of overlapping DNA segments

1.5 Tools for genome characterization

Molecular characterization aiming at assessing the genetic variation within and between different indigenous African cattle populations is a very crucial step towards achieving the objective of conserving the genetic resources of Africa indigenous cattle breeds and improving their productivity through genetic improvement. Protein, as well as DNA markers, maybe used but the latter are the markers of choice since they can provide

information on every region of the genome, regardless of their level of gene expression. DNA polymorphic markers typically include microsatellites, Single Nucleotide Polymorphism (SNP), insertions/deletion (indels), Copy Number Variations (CNV) etc. They may be studied using different methodological approaches, the most powerful being the analysis of genome sequences (Meghen *et al.,* 1994, Sunnucks 2000, Metta *et al.,* 2004).

DNA based markers have been used in animals for various population genetic studies including diversity studies, marker assisted selection (MAS) and for mapping quantitative trait loci (QTL) (Erhardt and Weimann 2007). They have also been used to assess population conservation genetic parameters such as effective population size, the identification of past bottlenecks, inbreeding status and gene flow (Goodnight and Queller 1999, Lynch and Ritland 1999, Hedrick 2001, Garrigan and Hedrick 2003).

A number of studies on the genetic characterization of African cattle such as genetic diversity studies, introgression, relationships among different breeds as well as evolution of Africa cattle have utilized DNA markers, including microsatellites (MacHugh *et al.,* 1997, Moazami-Goudarzi *et al.,* 1997, Hanotte *et al.,* 2002, Ibeagha-Awemu *et al.,* 2004, Ibeagha-Awemu and Erhardt 2005, Freeman *et al.,* 2006, Erhardt and Weimann 2007), mitochondrial (mtDNA) DNA sequence (Loftus *et al.,* 1994, Loftus *et al.,* 1994, Bradley *et al.,* 1996), *Y* chromosome polymorphism (Bradley *et al.,* 1994, Hanotte *et al.,* 2000). Single Nucleotide polymorphisms (SNP) have now gained popularity as the molecular marker of choice for animal diversity studies with the possibility of SNP genotyping and full genome sequencing of an animal species in a single, simple and reliable experiment.

In cattle, genome wide characterization of SNP polymorphism became possible following the design of SNP genotyping arrays such as the low density BovineSNP50 Beadchip and the High-Density Bovine BeadChip (BovineHD) designed by Illumina (Matukumalli *et al.,* 2009, Matukumalli *et al.,* 2011), or the high density Axiom Genome-Wide BOS 1 Array (BOS 1) and ultra-high-density Affymetrix BOS 1 pre-screening assay (AFFXBIP) released by Affymetrix Inc. (Santa Clara, CA) (Rincon *et al.*, 2011, Ramey *et al.,* 2013). The low and high density arrays can genotype up to 54,000 and 770,000 SNPs respectively. Although these tools have been criticized due to their ascertainment bias to European taurine breeds upon which they are validated, they have been used for a number of genomics studies in different cattle populations worldwide. These studies include genetic diversity and structure analysis (Bovine Hapmap Consortium, 2009, Lin *et al.,* 2010, Mbole-Kariuki *et al.,* 2014) and the detection of signatures of selection (Gautier *et al.,* 2009, Gautier and Naves 2011, Utsunomiya *et al.,* 2013, Xu *et al.,* 2014)**.**

Compared to SNP chip genome information, genome wide sequencing is obviously more informative, providing a complete window of the gene polymorphism present in a species or population, It has been applied to several population genomics studies (Davey *et al.,* 2011). It started with the Sanger sequence technique that was used for the Human Genome Project (1990 – 2003) followed by the development of Next Generation Sequencing (NGS) technologies such as SOLiD, Illumina, PacBio. Several reviews of the various NGS technology platforms, bioinformatics tools as well as file formats are available (Shendure and Ji 2008, Metzker 2010, Pareek *et al.,* 2011, Pavlopoulos *et al.,* 2013). Although the associated costs are still very high, it is expected

that the cost and turnaround time for accomplishing whole genome sequencing will continually be reduced. However, a major concern requiring urgent attention is how to efficiently handle the enormous data being generated after genome sequencing.

The genomes of humans and animals are now increasingly been investigated for the detection of structural variants including copy number variation (CNV), this is as a result of their association with important phenotypes including complex traits (Ionita-Laza *et al.,* 2009, Pinto *et al.,* 2010, Almal and Padh 2012, Hou *et al.,* 2012, Zhang *et al.,* 2014). To date, several studies on the detection of CNV have been accomplished in different species such as cattle (Liu *et al.,* 2008, Bickhart *et al.,* 2012, Hou *et al.,* 2012, Jiang *et al.,* 2012, Liu and Bickhart 2012, Cicconardi *et al.,* 2013, Xu *et al.,* 2014, Zhang *et al.,* 2014, da Silva *et al.,* 2016), sheep and goat (Fontanesi *et al.,* 2010, Fontanesi *et al.,* 2011, Liu *et al.,* 2013), pig (Fadista *et al.,* 2008, Ramayo-Caldas *et al.,* 2010, Chen *et al.,* 2012), dogs (Berglund *et al.,* 2012), chicken (Volker *et al.,* 2010, Wang *et al.,* 2010) and humans (Sebat *et al.,* 2004, Conrad *et al.,* 2010, Sudmant *et al.,* 2010, Zarrei *et al.,* 2015).

1.6 Signatures of selection

The diversity which exists among various species is notably observed as variation in their phenotypes such as physical appearance (morphology), production traits, environmental adaptations, and disease resistance/susceptibility. The ultimate goal of any evolutionary scientist is to understand the underlying genetic variation contributing to phenotypic diversity in different species (Biswas and Akey 2006). This can be achieved by accurate identification of loci under selection and possibly the associated causal mutation(s) through signature of selection study.

By definition, ‘selection signatures’ are defined as regions of the genome that harbour functionally important sequence variants and therefore have been under either natural or artificial selection (Qanbari and Simianer 2014). In molecular evolutionary terms, three different forms of selection are recognized: Positive selection, negative or purifying or background selection and balancing selection. Each will impact the diversity of the genome in specific ways (typically changes in frequency of alleles at the selected loci and surrounding loci as well as genotype frequencies changes). While balancing selection leads to increase variability in a population, negative selection acts upon a new deleterious mutation thereby removing it and maintaining the functional integrity of DNA sequence (Oleksyk *et al.,* 2010, Gouveia *et al.,* 2014).

Positive selection arises when an existing allele or a new mutation is advantageous, leading to its increases in frequency in a population, they may be associated with genes related to adaptation (e.g. genes related to tropical adaptation) or the evolution of new form or function (Kreitman 2000, Nielsen 2005, Gouveia *et al.,* 2014). Even if only a single allele is being selected it may result in a local selective sweep following a hitchhiking effect affecting surrounding polymorphisms or alleles (Kim and Stephan 2002, Qanbari and Simianer 2014). The model of selective sweep in which a newly arisen allele with a strong selective advantage increases quickly in frequency until reaching fixation is known as “hard sweep”. In contrast, when the selected allele is part of existing genetic variation, it causes a “soft sweep” in which the footprint left by selection tends to be less pronounced and the frequency of the selected allele at the beginning of the selected phase is the crucial factor influencing the selective sweep (Prezeworski *et al.,* 2005, Pritchard *et al.,* 2010).

The detection of genomic signatures is now possible as a result of the availability of genomic data which are now commonly generated following the advancements in high throughput next generation sequencing technologies and associated bioinformatics tools (Prezeworski *et al.,* 2005, Qanbari and Simianer 2014). The detection of signatures of selection at a genome wide level is particularly suitable for population genomics studies in the absence of pedigree information and phenotypic data. It, therefore, by-passes the limitations of genome wide association studies (GWAS) which require pedigrees and/or phenotype information and the candidate gene approach which rely on prior information from other studies (Biswas and Akey 2006, Qanbari and Simianer 2014)

1.7 Approaches for detecting signature of positive selection

Several approaches have been designed for the detection of selection. Most of these methods were validated with human genomic data but have been implemented in other species including cattle. The majority of the tests are designed to detect loci under selection either by within population analysis or by comparison of the genome of two populations (between population approaches). As the span of the regions under selection is generally wide, selection tests are based on the detection of outlier regions often referred to as candidate regions (Randhawa *et al.*, 2016).

1.7.1 Tests based on linkage disequilibrium pattern (haplotype structure)

The determination of the pattern of linkage disequilibrium (LD) is the underlying principle for the design of some statistical tests to detect selection signatures. Strong positive selection is expected to accelerate the frequency of an advantageous allele

faster than recombination breaks down LD at the selected haplotype. Some of the methods based on this principle include; Long Range Haplotype (LRH), integrated Haplotype Score (iHS), extended haplotype homozygosity (EHH) e.g. cross population (XP) EHH and Rsb tests, relative extended Haplotype homozygosity (REHH), Whole Genome Long Range Haplotype (WGLRH) and LD decay (LDD) (Sabeti *et al.,* 2002, Biswas and Akey 2006, Voight *et al.,* 2006, Sabeti *et al.,* 2007, Tang *et al.,* 2007, Gouveia *et al.,* 2014, Qanbari and Simianer 2014).

The *iHS* statistical test is one of the approaches used in this thesis. This method is based on the estimation of extended haplotype homozygosity (EHH) for each core bi-allelic SNP, which is defined as either ancestral or derived. The sum of EHH computed over both directions away from the core SNP is referred to as integrated EHH. It is denoted iHHA for the ancestral allele, iHHD for the derived allele and iES for the SNP site (Qanbari *et al.,* 2011, Gouveia *et al.,* 2014). *iHS* is based on the extent of decay of linkage disequilibrium surrounding a variant subjected to natural selection, it is thus described as a within population score as the ratio between iHHA and iHHD:

$$iHS = \ln(\frac{iHHA}{iHHD}) \qquad \textbf{\textit{Equation 1.1}}$$ (Voight *et al.* 2006)

The log ratio of iHH for the ancestral and derived alleles is then standardized such that it has a mean of 0 and variance of 1 irrespective of allele frequency at the core SNP. Large positive and negative values of *iHS* indicate unusually long haplotypes carrying the ancestral and derived allele, respectively (Biswas and Akey 2006, Voight *et al.,* 2006).

Rsb and *XPEHH* statistics are other methods also based on EHH, they are also utilized in this thesis. In contrast to the *iHS* test where EHH is compared between alleles within

a population, both methods involves the comparison of the haplotype patterns of the same allele (denoted as "iES") between two populations, A and B. Thus, the statistics "ln(iESA/iESB)" is then calculated. The extreme positive values suggest positive selection in population A whereas negative values indicate selection in population B. The main difference between them is that Rsb calculates the EHH based on the status of each core SNP allele while XP-EHH calculates the EHH based on the core SNP site (Sabeti *et al.,* 2007, Tang *et al.,* 2007).

While the iHS test has power to detect selected alleles that are at low to intermediate frequencies, both tests have power to detect selected alleles that have risen to near fixation or in complete fixation in one population but remain polymorphic in the other population (Pickrell *et al.*, 2009).

1.7.2 Tests based on reduced local variability

Runs of Homozygosity (ROH) and pooled heterozygosity (H_p) tests have been proposed with the aim to detect regions of the genome where there are reduced variations compared to the rest of the genome. The ROH test examines persistent regions of the genome with homozygosity whereas the H_p statistic is designed to estimate variability by looking for regions of the genome with low heterozygosity. The *Hp* test specifically involves the count of reads with the most and least abundantly observed alleles at every detected SNP position across sliding windows of adjacent loci. The distribution of *Hp* values are then normalized by transforming *Hp* to Z-scores (*ZHp*) using (ZHp= (Hp - μHp)/σHp) (Rubin *et al.*, 2010, Qanbari and Simianer 2014). The *Hp* approach was initially applied to scan the chicken genome (Rubin *et al.,* 2010, Qanbari *et al.,* 2012) but has also been applied in other domestic species including pig (Amaral *et al.,* 2011,

Bosse *et al.,* 2012, Rubin *et al.,* 2012), dogs (Boyko *et al.,* 2010, Quilez *et al.,* 2011) and cattle (Liao *et al.,* 2013, Choi *et al.,* 2015, Bahbahani *et al.,* 2017). The *Hp* approach is also adopted in this thesis

1.7.3 Tests based on deviation from allele frequency spectrum

Examples of statistical tests designed to identify selected loci based on deviation from allele frequency spectrum include Composite likelihood ratio (CLR), Cross population (XP) CLR, Tajima's D (TD) and Fay and Wu H tests among others. Tajima's D test calculates the difference between two different estimators of population mutation rate, these are Watterson's estimator (θ) and mean pairwise difference between sequences in a sample (π). A negative value of Tajima's D is a deviation from allele frequency spectrum and it is caused by excess of low frequency alleles which is indicative of positive selection whereas a value of zero is expected in a standard neutral model. The Fay and Wu H test is similar to Tajima's D test but the former can identify ancestral allele at each polymorphic site, a limitation of the latter test (Biswas and Akey 2006, Qanbari and Simianer 2014)

The XP-CLR test, also used in this thesis, is a likelihood method that is based on multi-locus allele frequency differentiation between two populations. It increases resolution by enhancing the signals and localization of the position of a selected allele (Chen *et al.,* 2010). This method is sensitive to SNP ascertainment bias and provides a confidence interval for the position of an advantageous allele with surprisingly good resolution. Unlike the haplotype-based tests, this method does not require phase information but it requires the estimates of recombination rates at selective sweep loci. More so, it is able to detect older signals (since a strong XP-EHH signal is expected to

disappear within several hundred generations due to LD breakdown), and also selection on standing variation (O'Reilly *et al.,* 2008, Chen *et al.,* 2010).

1.7.4 Tests based on population differentiation

Genetic differentiation in a population can be determined from the estimation of Wright's fixation index F_{ST} (Wright 1943). Signatures of positive selection especially between populations can also be detected by comparing the multi-locus values of this metric with the values in a neutral model, as it is expected that the values of F_{ST} estimated for regions of the genome under selection to exceed the background level. An elevated F_{ST} values is expected for alleles under selection advantage (Lewontin and Krakauer 1973, Gouveia *et al.,* 2014, Qanbari and Simianer 2014). A major concern surrounding this approach is the reliability of the estimates of F_{ST} values under neutrality. A number of efforts such as (Beaumont and Nichols 1996, Vitalis *et al.,* 2001, Vitalis *et al.,* 2003, Beaumont and Balding 2004, Foll and Gaggiotti 2008, Excoffier *et al.,* 2009) have been made to address this concern but the various propositions are not without their strengths and limitations (Gouveia *et al.,* 2014). This approach has been employed to identify selection signature and population structure in several cattle populations (Barendse *et al.,* 2009, Gautier *et al.,* 2009, Hayes *et al.,* 2009, Ramey *et al.,* 2013) including West African cattle breeds (Gautier *et al.,* 2009). The F_{ST} test is also used in this thesis.

1.7.5 Tests based on synonymous and non-synonymous substitutions rates

Estimates of the ratios of non-synonymous (d_n) to synonymous (d_s) substitutions (d_n/d_s test) has also been used to identify signature of selections where values of $d_n/d_s = 1$, < 1 and > 1 is an indication of neutral, negative and positive selections respectively

(Biswas and Akey 2006, Gouveia *et al.,* 2014). The principle of this selection test is the basis upon which computation tools such as *PAML* (Yang 2007) and *MEGA* (Kumar *et al.,* 2008) are designed.

1.8 Signatures of selection in worldwide cattle

Since domestication in parallel to their dispersion, cattle have been subjected to both human selection (selective breeding) and natural selection (environmental adaptation) leading to the development of breeds across the world. In the developed world, we find mainly specialized cattle breeds that have been selected by humans for specific traits such as production (meat and milk yield) and morphological traits (coat colour, presence or absence of horns). On the other hand, the majority of cattle breeds or populations in the developing nations are diversified mainly on the basis of their morphology (coat colour) and adaptive traits (e.g. disease resistance and heat tolerance) following natural selection. The over 1000 cattle breeds are the reflection of their genetic diversity following domestication, adaptation, production traits and morphology (Rischkowsky and Pilling 2007, Randhawa *et al.,* 2016). Thus, cattle represent a good model for the study of the genetic control and its molecular basis of different traits, however, most economically important traits in livestock are expected to be quantitative (Andersson and Georges 2004, Gouveia *et al.,* 2014).

Today, several selection signatures in worldwide cattle breeds have been identified using the different approaches mentioned above, (Barendse *et al.,* 2009, Bovine HapMap Consortium, 2009, Flori *et al.,* 2009, Hayes *et al.,* 2009, Larkin *et al.,* 2012, Utsunomiya *et al.,* 2013, Porto-Neto *et al.,* 2014, Qanbari and Simianer 2014, Randhawa *et al.,* 2014, Choi *et al.,* 2015, Xu *et al.,* 2015, Mei *et al.,* 2017). These

studies have identified a large number of candidate selection signatures especially in the group of European breeds (Randhawa *et al.,* 2016). Reviews of these studies and genomic regions of the candidate bovine signatures of selection are presented by Gouveia *et al* (2014), Gutiérrez-Gil *et al.* (2015) and Randhawa *et al.* (2016), amongst others.

Randhawa *et al.* (2016) has implemented a novel approach referred to as "meta-assembly" to infer hotspots of positive selection in the bovine genome by using several published selection signatures studies across worldwide cattle breeds. Considering four cattle architypes and 16 breeds, several genomic regions of variable gene density where positive selection has been replicated within and across cattle breed-populations were highlighted. These shared signatures, in four or more breeds, were located on all *Bos taurus* autosomes (BTA) except BTA-23, BTA-24, BTA-25, and BTA-28. The common genomic regions across various populations reveal historical selection shared between cattle breeds, most likely due to ancestral, geographical and/or commercial similarities, while the unique genomic regions under selection, which are private to a single breed, can be related to particular characteristics of that population (Randhawa *et al.,* 2016).

The hotspots of genomic selection signatures are mostly related to adaptation (disease, climate, feed resources), appearance (polledness, coat colors) and production (milk, meat, fertility) traits, each of which have economic importance in various environments and production systems. Some of the reported genes associated with these traits are *NCAPG, LCORL, PLAG1*, related to body size and stature, *ABCG2, DGAT1, GHR* genes, related to milk production while *MCIR, KIT* are related to coat colour. In

addition, strong selection signature regions for which QTLs and functional variants have been localized include chromosomes BTA-6 and BTA-20 reported to harbor QTLs that affect milk traits in dairy cattle breeds (Boettcher *et al.,* 2004, Flori *et al.,* 2009, Ogorevc *et al.,* 2009, Qanbari *et al.,* 2010, Weikard *et al.,* 2012). Regions on BTA-2 have been associated with feed efficiency and intramuscular fat in beef cattle (Barendse *et al.,* 2009, Qanbari *et al.,* 2011), while selection signatures on BTA-14 have been reported to be involved in the control of marbling and fatness trait in cattle (Casas *et al.,* 2005, Hayes *et al.,* 2009, Veneroni *et al.,* 2010).

1.9 Signatures of selection in African cattle

The majority of the genome-wide studies of detection of signature of selection have focused on European *Bos taurus* breeds with much fewer studies for zebu and African cattle breeds. The investigation of selection signatures in indigenous African cattle may provide insight into regions targeted by selection during domestication and breed formation in addition to adaptation to the tropics. A few genome-wide studies have targeted African cattle for the investigation of their selection signatures (Dayo *et al.,* 2009, Gautier *et al.,* 2009, Flori *et al.,* 2014, Xu *et al.,* 2014, Makina *et al.,* 2015, Bahbahani *et al.,* 2017, Kim *et al.,* 2017, Kim *et al.,* 2017, Taye *et al.,* 2017, Bahbahani *et al.,* 2018). These studies have also applied different selection scan approaches and have utilized genome-wide SNP data ranging from low density SNP data (50K chip) to full genome sequences. The focus of these studies has largely been to identify selection signatures related to tropical environmental adaptation (disease resistance, tick resistance, and heat tolerance), production traits (milk and meat production) and morphological traits (coat colour and growth traits). Studies using SNP data could not be totally relied upon because of false positive results as a result of the ascertainment

bias of the available bovine SNP genotyping assay (Matukumalli *et al.,* 2009) with results following the use of full genome sequences predicted to be more reliable.

Several studies have reported results in the West African longhorn N'Dama cattle from Guinea and Gambia using low to high density SNPs data (Barendse *et al.,* 2009, Bovine HapMap Consortium, 2009, Dayo *et al.,* 2009, Gautier *et al.,* 2009, Xu *et al.,* 2014) and full genome sequence data (Kim *et al.,* 2017, Kim *et al.,* 2017, Taye *et al.,* 2017). The reason being that N'Dama cattle is the most studied among the trypanotolerant West African breeds and it is often referred to as the trypanotolerant breed of reference. Unravelling the genetic basis of trypanotolerance has been the major drive for the genome wide studies in N'Dama cattle. Candidate genes such as *ARHGAP15, TICAM1, INHBA, SLC40A25* and *HCRTR1,* which are relevant to trypanotolerance traits including the control of parasitemia, anaemia and feeding behaviour, have been reported. The validation of these findings in other trypanotolerant cattle breeds will be a worthwhile pursuit and the results will no doubt receive the interest of members of the scientific community.

A limited number of selection signature studies involving African zebu have only recently been accomplished, they involved zebu breeds from South Africa (Makina *et al.,* 2015) and East Africa (Kim *et al.,* 2017, Msalya *et al.,* 2017, Taye *et al.,* 2017, Bahbahani *et al.,* 2018). Only two of these studies (Kim *et al.,* 2017, Taye *et al.,* 2017) have utilized full genome sequencing. These two studies include three zebu breeds (Kenana, Ogaden, Boran) and one Sanga breed (Ankole).

Kim *et al*. (2017) compared the genome of African cattle to cattle living in temperate regions (European taurine and Asian taurine) with the aim of detecting genes involved in adaptation to the tropics. In regards to heat tolerance, they identified a long-range haplotype across African cattle within one of the heat tolerance QTLs on chromosome 22 (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index) and also identified one heat shock protein of 70 kDa (heat shock protein 4, *HSPA4*). The degree of haplotype sharing at these two regions was observed to be more extensive in the African zebu than African taurine, this is in concordance with a previous report regarding zebu's superior ability to regulate body temperature during exposure to heat stress (Hansen 2004). They also reported positive selection signals in genome regions including two genes, *SOD2* and *PRLH,* and they suggest their possible roles in heat stress response in Africa zebu. In contrast to the non-African cattle breeds, non-synonymous mutations near fixation in the zebu cattle were identified in these genes. In particular, the suggestive role of *PRLH* gene in thermotolerance is based on its regulatory function on the expression of prolactin (*PRL*), a gene which has been reported to impact on hair morphology and thermoregulation in cattle with the slick coat phenotype (Littlejohn *et al.*, 2014, Kim *et al.,* 2017). Other genes in African cattle within candidate genome regions proposed to be under selection are the *BOLA* gene, reported to be involved in tick resistance, and the coat colour genes, *KIT* and *MC1R* identified when Ankole cattle was compared to other African cattle (Kim *et al.*, 2017).

1.10 African cattle breeds included in the study

1.10.1 Muturu

The Muturu cattle (*Bos taurus*) (Figure 1.2) is one of the humpless shorthorn breeds in West Africa. They are believed to descent from the shorthorn (brachyceros) humpless

cattle which appeared in ancient Egypt around 2000 BC, likely migrating from the Centre of cattle domestication in the Near East and were first recorded in West Africa during the first half of the first millennium AD (Gwaza and Momoh 2016).

Their distribution in the past was across West and Central Africa (Rege *et al.,* 1994). They were initially most widespread in Nigeria until the Fulani invasion in 1820 when zebu cattle were gradually introduced (Felius 1995). With an unofficial census, a population size of 115,000 cattle heads was reported in Nigeria in 1992 (Bourn *et al.,* 1994). The existence of the Muturu cattle has been consistently threatened since the introduction of *Bos indicus* (white Fulani) in the region. Farmers also prefer the latter subspecies to Muturu and they were indiscriminately crossbred with the larger white Fulani to form the breed known as Keteku (Muturu x White Fulani) (Jabbar *et al.,* 1998).

Nowadays, they are only sparsely distributed in the humid forest zone and in a few savannah areas of Nigeria and are now classified as an endangered cattle breed (Adebambo 2001, Gwaza and Momoh 2016). The breed is also found in most coastal areas of West Africa countries including south eastern coastal areas of Ghana and in the southern part of Upper Volta (Burkina Faso). The distinction between the different populations is not clear (Gwaza and Momoh 2016). However, the breed has been classified into four separate populations, namely Ghana Dwarf Muturu (also called Muturu or Forest Shorthorn), Liberia Dwarf Muturu (also called Liberian Muturu), Nigerian Forest Muturu (also called Kirdi), and Nigerian Savannah Muturu (also termed Muturu) (DAGRIS; http://dagris.ilri.cgiar.org).

The typical coat colour of Muturu varies from black for the forest Muturu to black and white colour in the Savannah type. Specifically, various shades of ebony black, fawn, black and white, black with white patches, white with brown or black spots have been observed among the Nigerian Muturu cattle (Rege 1999, Adebambo 2001). Muturu has been described as the smallest African cattle breed, possessing a compact body with fine-boned limbs, the reported height at withers is 95 cm and 88 cm for the male and female, respectively (Maule 1990, Rege 1999, Adebambo 2001). In the past, the breed was seldom kept for milk or meat production but rather for socio-cultural purposes especially in the south eastern part of Nigeria where the breed is considered as property of the local deities or is dedicated to a shrine or sacrificed during funeral ceremonies (Adebambo 2001).

Figure 1.2 | Muturu cattle from Nigeria (photo credit: Abdulfatai Tijjani)

Like the N'Dama and other trypanotolerant West African taurine, Muturu is also known for their tolerance to trypanosomiasis. Thus, they represent valuable genetic resources for the study of livestock disease resistance. The animals included in this study were sampled in Nsukka and Igbo-Etiti towns in Enugu state, south eastern Nigeria.

1.10.2 N'Dama

N'Dama cattle (Figure 1.3) are one of the two main humpless longhorn taurine (*Bos taurus*) in West Africa, the other being the Kuri, the two breeds are quite distinct in their morphology (Rege and Tawah 1999). As mentioned earlier in section 1.2, they are believed to be the descendants of the first cattle introduced into the African continent (Payne 1970). The formation of the N'Dama as a distinct breed is thought to have taken place on the Fouta Djallon plateau in Guinea from where they dispersed across of coastal West and Central Africa, in the regions infested by tsetse fly. N'Dama have an inherited tolerance to trypanosomiasis disease (Murray *et al.*, 1982). They are often referred to as the West African trypanotolerant reference breed and they are recognised as a valuable genetic resource especially for introducing animal husbandry into other tsetse-infested areas of Africa (MacHugh 1996). They are also resistant to tick and tick-borne infections (Mattioli *et al.*, 1995), and are well adapted to stressful humid and dry tropical climates (DAGRIS; http://dagris.ilri.cgiar.org).

Figure 1.3 | N'Dama cattle (adapted from https://www.roysfarm.com/ndama-cattle/)

The typical N'Dama is medium sized, weighing approximately 230 kg to 370 kg, the height at withers is approximately 100 cm and 120 cm for the cow and bull respectively. Their coat colour is typically fawn or brown (MacHugh 1996). They are a multipurpose breed with relatively low milk production (DAGRIS; http://dagris.ilri.cgiar.org). The N'Dama samples included in this study are samples from the Guinea (Fouta Djallon area) with the genome data made available in the public domain following the recent publication of Kim *et al*. (2017).

1.10.3 Kenana

Kenana cattle (Figure 1.4) are one of the two main types of zebu cattle in northern Sudan. They originate from an area between the White and Blue Nile rivers known as the Fung which is composed primarily of acacia scrubland. They are largely kept under the custody of semi-nomadic pastoralists in northern Sudan. They are medium to large size with average body weights of between 300 and 500 kg in males and 250-350 kg in females (MacHugh 1996). Coat colour is white or steel grey with darker shading on the neck and shoulders and a black switch to the tail. A typical Kenana cow of northern Sudan is characterised for its superior milk production capacity compared to other African zebu. The documented average level of milk production per lactation is 538.26 kg under local management, while levels of between 1200 – 2000kg have been recorded in animals kept within an improved production facility. They are also used for beef and traction (DAGRIS; http://dagris.ilri.cgiar.org). The Kenana samples included in the present study were sampled from different herds around the city of Rabak located approximately 260 kilometres south of Khartoum in Sudan.

Figure 1.4 | A typical Kenana cattle, female (left) and male (right) (photo credit: Professor Hassan Musa, University of Khartoum, Sudan)

1.10.4 Butana

Butana cattle are the other main type of zebu cattle found in northern Sudan, aside Kenana. The Butana cattle has similar body size and good dairy production characteristics, similar to the Kenana breed. Averages of 598.73 kg and up to 1662 kg of milk production per lactation have been recoded for Butana animals kept under local management and within an improved facility, respectively. Butana is also used for meat and drought power. Butana breed are found in east of Khartoum and in the Butana plain, including the Atbara River in the east and the River Nile in the south. They are also found in Gazira in central Sudan and along the river Nile in the northern region. They are owned mainly by the Batahin and Shukria tribes. They form approximately 9-15% of the total population of cattle in Sudan. The most common coat color of this breed is red or dark red, however, calves are grey, similar in color to Butana, they turn red when they reach adulhood. Butana breeds together with the Kenana breeds have been used widely for cross-breeding purposes with European cattle in Sudan (DAGRIS; http://dagris.ilri.cgiar.org).

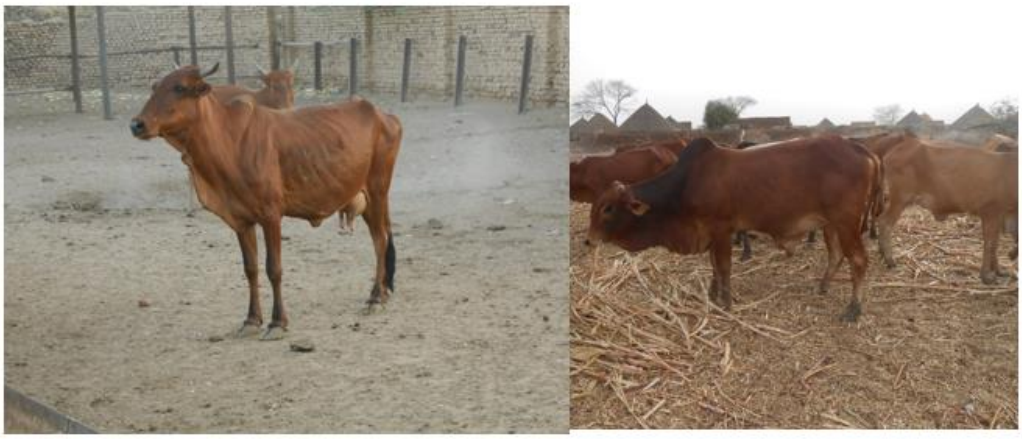


Figure 1.5 | A typical Butana cattle, female (left) and male (right) (photo credit: Professor Hassan Musa, University of Khartoum, Sudan)

1.10.5 Aryashia

Aryashai zebu cattle (Figure 1.6) are indigenous to eastern Sudan, found mainly in the Algash delta stretching in an area including Aroma in the South to Dordeib in the North and to Atbara River. They look like Butana in size and body conformation but similar to Kenana in colour. The dominant colour of the Aryashai cattle is white, which is from the word "Erashai", meaning white cow in Beja slang, other colours include red brown or mixed colours. They are characterized by low milk productivity (Rahman 2007).

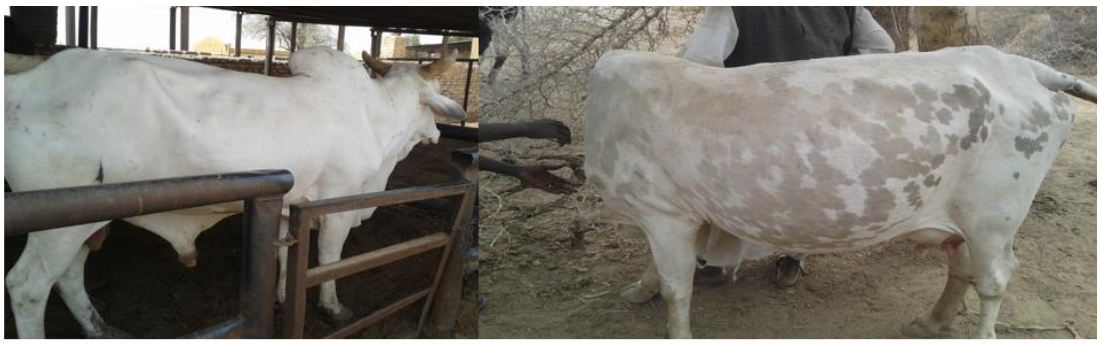


Figure 1.6 | A typical Aryashia cattle, male (left) and female (right) (photo credit: Professor Hassan Musa, University of Khartoum, Sudan)

1.10.6 Gash

Like Aryashai, Gash cattle (Figure 1.7) are also indigenous to eastern Sudan. They are a medium sized, strong and aggressive cattle breed that are owned by the Hadandawa

and Beni Amir tribes. Their common colours are a mixed red and white, red and black or black.

Figure 1.7 | A typical Gash cattle, female (left) and male (right) (photo credit: Professor Hassan Musa, University of Khartoum, Sudan)

1.10.7 Baggara

Baggara cattle (figure 1.8) is a short horned large East African zebu breed that are owned by the nomadic tribes in southern Darfur, central and southern Kordofan and Nuba Mountains, and in areas west of the White Nile. They are given different names in different areas (Rahman 2007). They have variable size which is attributed to admixture with small Nilotic and large size Fulani cattle, which share their seasonal grazing lands and migratory routes. They are also influenced by a number of factors such as seasonal availability of water and feed, high market prices and infectious diseases (Bashir and El Zubeir 2013). They also display variable coat colour: white or white with some black spots or markings, while brown or red cattle are also found (Alsiddig *et al.*, 2010). Baggara cattle constitute approximately 33% of the Sudanese national herd and are the main producer of beef for local consumption and export (DAGRIS; http://dagris.ilri.cgiar.org).

Figure 1.8 | A typical Baggara cattle (photo credit: Professor Hassan Musa, University of Khartoum, Sudan).

1.10.8 Sudanese Fulani

The Sudanese Fulani are included amongst the West African zebu and are known by several other names including Wodabe, Red Fulani, Red Bororo, Abori, Bodadi and Brahaza. They are kept by the Fulani herders Wodabe tribes (Kuri) in Sudan. They have a common origin with Gobra cattle of Senegal, but they are mainly found all over southern and south-western Mali, along the flood plains of the Niger river from Segon to Timbuktu and also in the Ivory Coast and Burkina Faso where interbreeding with taurine cattle takes place. A special characteristic of this breed is their hardiness with the ability to walk up to 300 km in search for food and water. They are considered a good milk producer, with daily milk production as high of 24.4 litres reported (DAGRIS; http://dagris.info.org).

Figure 1.9 | A typical Sudanese Fulani cattle (photo credit: Professor Hassan Musa, University of Khartoum, Sudan)

1.10.9 Boran

Boran cattle belong to the large East African Zebu breeds of cattle. The Boran cattle likely originated in Ethiopia and today they are found in this country as well as in Kenya (Kenyan Boran, Orma Boran) and Somalia. The Kenyan Boran is a meat type of breed. They are usually white, fawn, or grey coats, but colours can vary from brown to red, the hooves and muzzle are always black. Boran cattle have the reputation to be very versatile and to adapt well to various environments. They have the ability to thrive well in dry and low rainfall areas. The breed can walk long distances in hot climate in search of feed and water (DAGRIS; http://dagris.ilri.cgiar.org). The animal studied here are Kenya Boran from the ILRI Kapiti Ranch, they have been included in a previous study and their sequence data are publicly available (Kim *et al.*, 2017).

Figure 1.10 | A typical male Boran cattle (picture adapted from DAGRIS Ethiopia; http://mpalalive.org/field_guide/boran_cattle)

1.10.10 Ogaden

Ogaden zebu, also known as a lowland zebu, are classified amongst the small rather than the large East African zebu (Figure 1.11). They originated from the Ogaden rangelands of south-eastern Ethiopia and are nowadays found in the Ogaden area of the Somali region of eastern Ethiopia. Their characteristics is similar to the Ethiopian Boran with a well-developed hump and large dewlap and they have short horns. The coat colour pattern is uniform plain white but some animals have shades of black around the face, hump, and rump (Figure 1.11). They can be dual purpose (milk and meat) but are mainly used for milk (Rege and Tawah 1999) (DAGRIS Ethiopia; http://eth.dagris.info/node/2334). The Ogaden samples included here were previously studied by (Kim *et al.*, 2017).

Figure 1.11 | A typical male Ogaden cattle (picture adapted from DAGRIS Ethiopia; http://eth.dagris.info/node/2334).

CHAPTER 2:

Genetic diversity and population structure of African indigenous zebu and taurine breeds

2.1 Introduction

The goal of cattle genomic research is the identification of genetic variants underlying phenotypic traits (Das *et al.,* 2015). The construction of the currently available SNP chips for cattle genotyping were based on now publicly available SNP databases, which make them biased toward variants useful for characterizing of taurine breeds, in particular, European taurine. Genomic studies of zebu breeds and indigenous African cattle breeds may therefore be affected by these biases (Matukumalli *et al.,* 2009, Helyar *et al.*, 2011). Also, linkage disequilibrium with the causal SNP variants may be incomplete, reducing the power of the approach. Whole genome resequencing technology remove these limitations and may allow direct identification of putative functional causative SNPs as well as causative structural variants (Daetwyler *et al.,* 2014, Das *et al.*, 2015).

Genomic variants including single-nucleotide polymorphisms (SNPs), insertions/ deletion (indels) can be identified by comparing whole-genome sequence for selected taurine or indicine breeds to the reference genome assembly. In cattle, a large number of variants have so far been identified. As at October 2017, dbSNP database *version* 150 (http://www.ncbi.nlm.nih.gov/snp/) included approximately 104.35 million single nucleotide variants (SNV) and 8.46 million insertions/deletions (Eck *et al.,* 2009, Zhan *et al.,* 2011, Stafuzza *et al.*, 2017). Despite a large number of SNPs identified in bovine genome-sequencing projects, few have been validated in indigenous African cattle populations (Edea *et al.,* 2013). The dbSNP database includes mainly SNPs identified in non-African cattle with only a couple of studies having generated whole genome sequence of indigenous African cattle breeds and deposited the identified variants in the public database (Bahbahani *et al.,* 2017, Kim *et al.,* 2017).

The presence of genetic diversity in domestic livestock species is of great importance for sustainable genetic improvement of breeds in various environments, as well as to facilitate rapid adaptation to potential changes in breeding goals (Notter 1999, Bijma and Woolliams 2002). The history of this diversity involves the spread of original livestock populations from their centres of domestication. Subsequently, different local populations are developed following the application of the original populations to various selection pressures in their new environments and the effect of genetic drift (FAO 2013). Local breeds contribute to global animal genetic resources (AnGR), even though many local breeds have small population size which puts them at risk of extinction (Bittante 2011). Genetic diversity provides the raw material for farmers and pastoralists to improve their breeds and adapt livestock populations to changing environments and changing demands (Rischkowsky and Pilling 2007).

Several studies have undertaken genetic characterization such as genetic diversity, introgression, relationships between worldwide cattle populations (Loftus *et al.,* 1994, Bradley *et al.,* 1996, Martín-Burriel *et al.,* 1999, Kantanen *et al.,* 2000, Cañón *et al.,* 2001, Mukesh *et al.,* 2004, Lai *et al.,* 2006, Lin *et al.,* 2010, Ai *et al.,* 2013, Edea *et al.,* 2013) as well as in African populations (Gwakisa *et al.,* 1994, MacHugh *et al.,* 1997, Moazami-Goudarzi *et al.,* 1997, Hanotte *et al.,* 2002, Ibeagha-Awemu *et al.,* 2004, Ibeagha-Awemu and Erhardt 2005, Freeman *et al.,* 2006, Erhardt and Weimann 2007, Makina *et al.,* 2014, Mbole-Kariuki *et al.,* 2014). These studies have utilized molecular markers such as microsatellites and single nuclear polymorphisms (SNPs), *Y* chromosome polymorphism, mitochondrial DNA (mtDNA) and genome-wide SNP markers. The latter are now the markers of choice for livestock diversity studies.

The results of previous studies have shown that diversity among modern cattle population declines with distance from primary centres of domestication, also that genetic diversity and relatedness of breeds are linked to their areas of origin, with breeds that have diverged recently generally geographically closer. These studies have also showed larger genetic differences between taurine and indicine groups than across breeds within each of the cattle subspecies (Loftus *et al.,* 1999, Hanotte *et al.,* 2002, McKay *et al.,* 2008, Edea *et al.,* 2013). In particular, low diversity is reported within and across European taurine breeds, possibly a reflection of the low diversity in the founding population and/or strong bottlenecks at breed formation level. Small effective population sizes and epistasis can lead to reduced individual fitness and reduced response to selection.

In this chapter, we report the discovery of genomic variants (SNVs and InDels) in indigenous African taurine and zebu cattle breeds based on full genome sequence analysis with the aim to contribute to the bovine SNP database and to provide new resources for future genomic studies on indigenous African cattle breeds. The variants information on variants will help to improve applications of genomic selection as well as increase our understanding of the genetic basis of local adaptation and production traits of African cattle. Furthermore, the discovered SNPs were used to assess population structure and genetic diversity among the African cattle breeds' studies and in comparison with non-African breeds (Holstein, Jersey, Hanwoo and Gir)

.

2.2 Materials and methods

2.2.1 Cattle breeds

A total of 151 cattle samples representing fifteen cattle breeds were studied. These breeds comprise of 11 indigenous African and four non-Africa breeds. The African breeds include two West African trypanotolerant taurine (WAT) (Muturu (MUT) *n* = 10 and N'Dama (NDM) *n* = 10), eight African zebu (AFZ) (Aryshai (ARY) *n* = 10, Butana (BTN) *n* = 10, Kenana (KNN) *n* = 10), Baggara (BGR) *n* = 10), Gash (GSH) *n* = 10, Fulani (FLN) *n* = 10, Ogaden (OGD) *n* = 9 and Kenyan Boran (BOR) *n* = 10), and one African admixed (Sanga) breed, Ankole (ANK) *n* =10). The four non-Africa breeds include Asian zebu (Gir cattle (GIR) *n* = 12), Asian taurine (Hanwoo (HWN), *n* = 10) and European taurine (ET) (Holstein-Friesian (HOL) *n* = 10 and Jersey (JER) *n* = 10).

Kenana and Butana breeds are grouped as African zebu dairy (AFZD) population, while Baggara, Boran, Gash and Ogaden breeds were grouped as African zebu non-dairy (AFZND) population. The African cattle breeds are the focus of the present study while other samples are included as reference breeds/populations for ease of comparison and interpretation of results. Figure 2.1 shows the sampling location of the African cattle breeds while the details of all cattle breeds included in this study are presented in Table 2.1

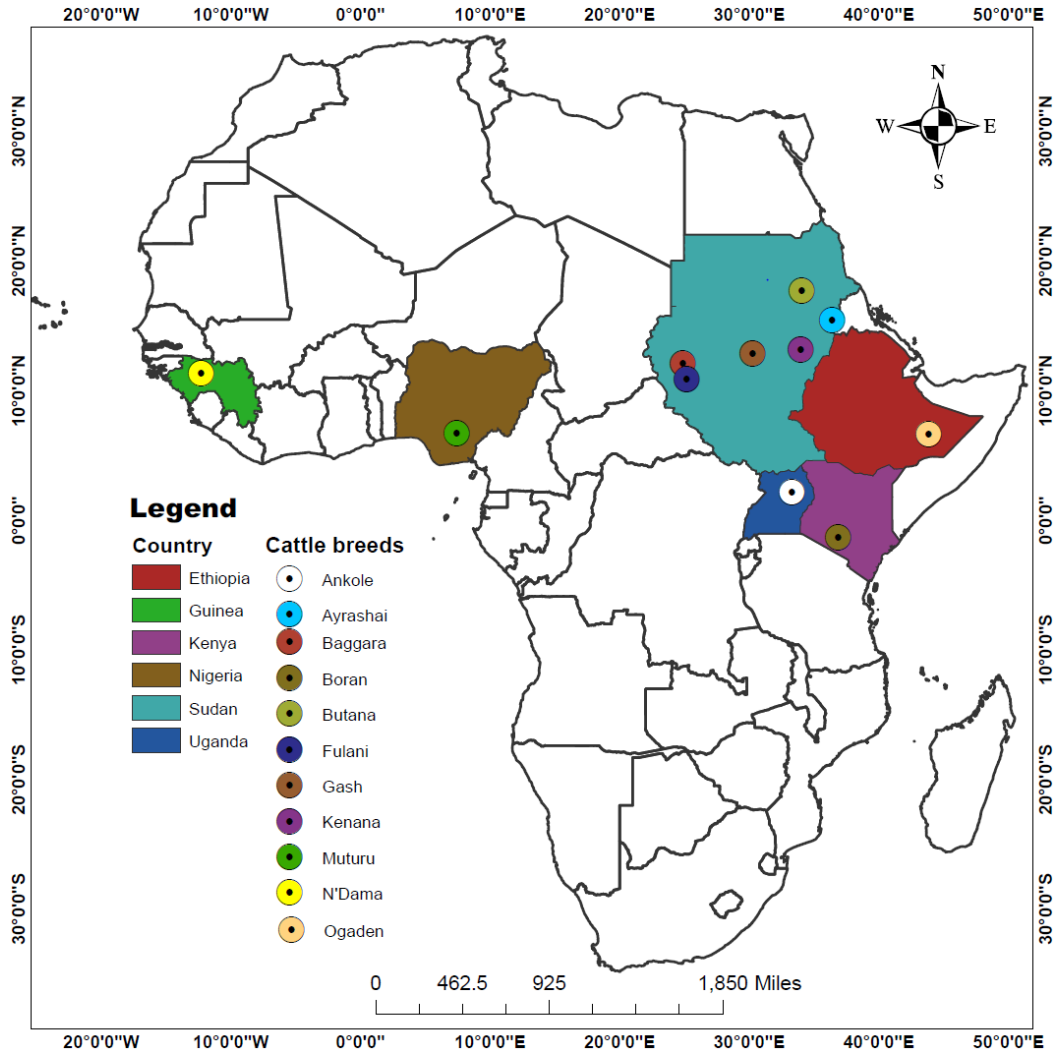


Figure 2.1 | Map of Africa showing the sampling locations of Indigenous African cattle studied

Table 2.1 | Details of cattle breeds included in the present study

S/No	Breed	code	Type	Origin	Main use	Sample (n)	Data source	Bio-project accession numbers
1	Baggara	BGR	Zebu	Sudan	meat	10	New sequences	NA
2	Fulani	FLN	Zebu	Sudan	meat	10	New sequences	NA
3	Gash	GSH	Zebu	Sudan	meat	10	New sequences	NA
4	Aryshai	ARY	Zebu	Sudan	milk	10	New sequences	NA
5	Butana	BTN	Zebu	Sudan	milk, meat	10	New sequences	NA
6	Kenana	KNN	Zebu	Sudan	milk, meat	10	New sequences	NA
7	Gir	GIR	Zebu	Brazil	milk, meat	12	New sequences	NA
8	Muturu	MUT	Taurine	Nigeria	meat	10	(Bahbahani *et al.,* 2017)	PRJNA386202
9	Kenyan Boran	BOR	Zebu	Kenya	meat	10	(Kim *et al.,* 2017)	PRJNA312138
10	Ogaden	OGD	Zebu	Ethiopia, Somalia	meat, milk	9	(Kim *et al.,* 2017)	PRJNA312138
11	Ankole	ANK	Sanga	Uganda	meat	10	(Kim *et al.,* 2017)	PRJNA312138
12	N'Dama	NDM	Taurine	Guinea	meat	10	(Kim *et al.,* 2017)	PRJNA312138
13	Holstein	HOL	Taurine	Europe	milk	10	(Stothard *et al.,* 2015)	PRJNA176557
14	Jersey	JER	Taurine	Europe	milk	10	(Daetwyler *et al.,* 2014)	PRJNA238491
15	Hanwoo	HNW	Taurine	Korea	meat	10	(Kim *et al.,* 2017)	PRJNA312138

2.2.2 Whole genome resequencing of cattle samples

Eighty-two new genomes sequences were generated and 69 cattle sequences were from previous studies (Table 2.1). For the latter, raw sequences stored as fastq format were retrieved from the NCBI SRA database except for Jersey cattle for which binary alignment (Bam) formats were obtained (see (Daetwyler *et al.*, 2014, Stothard *et al.*, 2015, Kim *et al.*, 2017) as well as Table S2.1 for further details). Details for the new sequences are provided below.

2.2.2.1 - Muturu samples

Although the raw sequence data for the 10 Muturu samples have been included in a recent publication (Bahbahani *et al.*, 2017), requiring the data to be made available in the public domain, the whole genome re-sequencing of the samples were carried out in the course of this study. Specifically, tissue samples were collected from the ear lobe of 10 Muturu cattle samples from the south-eastern part of Nigeria. Total genomic DNA was extracted using the Macherey-Nagel NucleoSpin® Tissue DNA extraction kit according to the manufacturer's protocol. Samples were sent to a commercial company (Novogene) for sequencing where paired-end libraries with an insert size of 150 bp were constructed from at least 5-µg genomic DNA according to Illumina's library preparation protocol. Paired-end sequencing of libraries were performed using the Illumina Hiseq2500 platform (Illumina Inc., San Diego, CA, USA).

2.2.2.2 - Sudanese zebu cattle samples

Sixty Sudanese zebu samples representing six zebu breeds were sampled in various locations in Sudan as shown in Figure 2.1. Approximately 10 ml of whole blood sample was collected from each of Sudanese zebu animals into EDTA VACUETTE® tubes

following standard procedures, under the supervision of a qualified veterinarian. Total genomic DNA was extracted from the whole blood samples using the Qiagen DNeasy extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. The quality and quantity of the DNA were evaluated using a Nano-Drop spectrophotometer (NanoDrop Technologies, USA) and gel electrophoresis. Whole genome resequencing was performed on the Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA).

2.2.2.3 - Gir Samples

The Gir samples included in this study were part of the progeny test program from the National Program for Improvement of Dairy Gir (PNMGL), headed by Embrapa Dairy Cattle (Juiz de Fora, Minas Gerais, Brazil) in cooperation with the Brazilian Association of Dairy Gir Breeders (ABCGIL) and the Brazilian Association of Zebu Breeders (ABCZ). Semen samples were collected from the animals and genomic DNA was extracted using a modified phenol/chloroform method as described briefly below. Samples were washed with lysis buffer and incubated for 2 hours with extraction buffer containing dithiothreitol 10% and RNase. Pellets were incubated overnight with saline-proteinase K buffer and protein was removed by phenol-chloroform extraction. Quality and quantity of DNA for all samples were evaluated by NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Mate-paired and paired-end libraries (200 bp and 2 x 100 bp, respectively) were prepared according to Illumina protocol and subsequently sequenced on the Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA).

2.2.3 Sequence mapping

The re-sequenced data for all 151 samples were aligned to the *Bos taurus* reference genome (UMD3.1) (Zimin *et al.,* 2009) using the Burrows-Wheeler Alignment tool (BWA) version 0.7.5a with option "bwa-mem" (Li and Durbin 2009). Following the mapping, the alignment files in SAM format files were converted to BAM files using SAMtools *ver* 1.19 (Li *et al.,* 2009). We then followed the Broad institute recommended Genome Analysis Toolkit (GATK) Best Practices pre-processing workflow which preceded variant discovery in High-Throughput Sequence Data (https://www.broadinstitute.org/gatk/guide/bp_step.php). Picard tools *ver* 1.119 (http://sourceforge.net/projects/picard) was used to sort the alignment files by coordinate, index, mark duplicates reads and to generate quality matrices for mapping. We then performed Local re-alignment around InDels and recalibration of the base quality scores with the use of Genome Analysis Toolkit (GATK *ver* 3.4) (McKenna *et al.,* 2010) using default settings, the dbSNP Build 148 (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/cow_9913/chr_rpts/) was also used for known sites.

2.2.4 Single nucleotide variants (SNVs) detection and annotation

Discovery of SNVs (single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels)) was performed using the HaplotypeCaller tool included in the Genome Analysis Toolkit (GATK; *ver* 3.4) (McKenna *et al.,* 2010). The joint genotyping approach (GenotypeGVCF mode) was adopted to identify variants simultaneously in all 151 cattle samples. We then used the GATK "Select Variants" mode to separate SNPs and InDels into different files before subjecting each to different hard filtering steps in order to reduce false positive variants. For the SNPs the filtering criteria were

“DP < 4” (SNPs covered by less than 4 reads), "MQ < 40.0" (SNPs with low mapping quality) and “QUAL < 30.0" (SNPs with a Phred-scaled quality score of less than 30). The InDels were filtered as follows QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0). Bcftools *v1.2* (Danecek *et al.*, 2011) was used to obtain the estimates of the number of total, bi-allelic, multi-allelic, singleton SNPs (found only within one individual) and the transition versus transversion ratio for individual cattle breeds based on autosomal bi-allelic SNPs. Only bi-allelic SNPs, which passed the SNP filtering criteria stated above, were considered and included in further analyses.

SnpEff *ver4.3t* (Cingolani *et al.*, 2012) was used to ascertain genome locations and effects of all discovered variants based on *Ensembl* gene annotation database (UMD3.1) and dbSNP Build 148 (Sherry *et al.*, 2001). Variants were further classified as either known if the non-reference allele is found in the dbSNP database or novel. ‘Upstream gene variants’ refers to a variant that overlaps within the 5 kb region upstream of the gene codon start site. ‘Stop gain’ means that a nonsynonymous SNP leads to the creation of a stop codon at the variant site. ‘Stop loss’ means that a nonsynonymous SNP leads to the elimination of a stop codon at the variant site. ‘Splicing’ means that a variant is within 2 bp of a splice junction. ‘Downstream’ means that a variant overlaps with the 5 kb region downstream of the gene end site. ‘Upstream/Downstream’ means that a variant is located in downstream and upstream regions for two different genes.

2.2.5 Data quality control

In addition to the stringent filtering criteria applied on the whole SNP dataset to reduce false positive variant calls, we performed further quality control by estimating the proportion of missing genotypes in individual samples and the relatedness among all

individuals in order to exclude closely related individuals and those with a high proportion of missing genotypes (> 50%). The missing genotype and relatedness statistics for all individuals were calculated using the options "*--missing-indv*" and "*--relatedness*", respectively, of Vcftools *v0.1.14* (Danecek *et al.,* 2011). Pairs of animals belonging to different breeds with high relatedness (> 0.8) but belonging to different breeds were excluded from the dataset, while within breed of animals with high relatedness (> 0.8) was removed from further analyses. Following these criteria, seven samples were removed (Butana (BTN015 & BTN020), Gash (GSH007 & GSH021) and Gir (GIR894 & GIR861) and Baggara (BGR002). No animal were excluded due to excessive missing data.

2.2.6 Estimation of genetic diversity and population structure

For the genetic diversity study, VCFtools *ver 0.1.14* (Danecek *et al.,* 2011) was used to calculate inbreeding coefficients (F) for individual cattle samples, nucleotide diversity (π) for individual cattle breeds and pairwise population-differentiation (F_{ST}) between breeds. Nucleotide diversity (π) and F_{ST} estimates were based on 50 kb non overlapping sliding windows along the cattle autosomes, while the inbreeding coefficient was estimated using the *"--het*" option in VCFtools. By this, F equals (O-E)/(N-E), Where O is the observed number of homozygotes, E is the expected number and N is the number of genotyped loci.

Population structure among the 15 cattle breeds was inferred with principal component analysis (PCA) and estimation of proportion of ancestry (admixture). VCFtools *ver 0.1.14* was used to select all autosomal bi-allelic SNP with a minimum minor allele frequency of 0.1 (MAF > 0.1) and to convert the vcf file of selected genotype into the

Plink format (Purcell *et al.,* 2007). Principal component analysis was performed using Plink software v1.19; the first two eigenvectors were plotted. PCA analysis was performed in four different categories using SNP datasets consisting of: i) all 15 cattle samples, ii) all taurine breeds, iii) all African zebu samples only and iv) all Sudanese samples only. The block relaxation algorithm implemented in the ADMIXTURE *ver 1.3.0* program (Alexander *et al.,* 2009) was used to further infer population structure within the cattle populations with Kinship (K) set from 2 to 10. However, the admixture analysis was preceded by subjecting the autosomal bi-allelic SNPs to additional filtering criteria including removal of genotypes with more than 10% missing data (*--geno 0.1*) and linkage disequilibrium (LD) pruning using default option (50 SNPs step 5 SNPs, r2 0.5) of PLINK software (Purcell *et al.,* 2007). GENESIS (Buchmann and Hazelhurst 2014) was used to visualize PCA and Admixture plots.

2.2.7 Phylogenetic relationships

To investigate the phylogenetic relationships among the individual cattle samples as well as at the breed level, a maximum-likelihood (ML) tree based on all cattle samples was constructed using SNPhylo (Lee *et al.,* 2014) while the MEGA 7 software (Kumar *et al.,* 2016) was used to visualize the tree output from SNPhylo. SNPs having a minimum MAF of 0.1 and not missing in any cattle sample were selected and SNPhylo was run with the options to include the removal of SNPs in LD of r^2 of less than 0.5 and a bootstrap value of 100. We further modelled the pattern of splits and gene flow among the different cattle breeds using the Treemix software (Pickrell and Pritchard 2012). Treemix was run a number of times to include up to six migration events, the models were visualized using the R script provided in Treemix.

2.2.8 Linkage disequilibrium (LD) and effective population size (*Ne*)

Plink software *v1.19* (Purcell et al. 2007) was used to estimate the levels of LD in the cattle breeds using genome-wide pairwise r^2 metrics for SNPs that were not more than 1000 SNPs apart and separated by a maximum distance of 1,000,000 base pairs. The decay of LD was plotted in R. Effective population sizes (*Ne*) at different times were estimated based on the genome-wide LD using the SNeP tool option for unphased SNP data (Barbato *et al.,* 2015). This tool estimates *Ne* following the equation:

$$N_{T(t)} = (4f\,(ct))^{-1}\left(E\left[r^2_{adj}|ct\right]^{-1} - \alpha\right) \quad \textbf{\textit{Equation 1}} \quad \text{(Corbin } \textit{et al.,} \text{ 2012)}$$

Where N_T is the effective population size and where the age of *Ne* (t) for any distance is calculated as $t = (2f(c_t))^{-1}$ (Hayes *et al.,* 2003). r^2_{adj} is the LD value adjusted for sample size, the recommended default α value = 1 was used to correct for the occurrence of mutation while ct is the recombination rate for a specific distance between SNP markers (Barbato *et al.,* 2015).

2.3 Results

2.3.1 Variant discovery and annotation

Following the alignment of all cattle samples sequence data to the UMD 3.1 bovine reference genome (Zimin *et al.,* 2009) we achieved an average 98% genome coverage and a range of 9.64 – 11 X mean depth coverage across samples. Variant calling was performed on all cattle samples simultaneous and detected variants were subjected to strict filtration criteria (see Materials and Methods section) in order to reduce false positive variant calls. Consequently, the total number of SNPs detected within the zebu breeds ranged from 23.4 million to 25.7 million SNPs, with the lowest number detected in Gash and the highest number in Fulani. In the taurine breeds, the number of SNPs ranged from 8.7 million in Muturu to approximately 14 million in N'Dama (Table 2.2).

Table 2.2 | SNP statistics in each cattle breed

Breed	Total number of SNPs	Average per sample	Het/Hom	TS/TV	No of Singletons (%)	dbSNP (%)
Fulani	25,729,904	11,289,375	1.89	2.32	19.6	92.7
Baggara	25,369,598	11,161,034	1.76	2.32	21.2	92.8
Aryshai	25,306,348	11,497,656	1.62	2.32	18.2	92.9
Kenana	25,028,266	11,505,880	1.55	2.32	17.5	93.0
Ogaden	24,862,705	11,520,736	1.56	2.32	19.3	94.3
Gir	24,170,381	11,982,657	0.94	2.33	15.0	92.6
Butana	24,103,512	11,491,354	1.60	2.32	20.4	93.4
Boran	23,864,277	11,341,483	1.71	2.32	16.6	94.8
Gash	23,473,765	11,440,971	1.52	2.33	19.9	93.7
Ankole	22,047,326	9,636,223	1.95	2.32	18.9	96.0
N'Dama	14,107,603	7,116,855	1.53	2.31	12.7	95.0
Hanwoo	13,935,436	5,943,980	1.72	2.28	25.7	95.8
Holstein	10,277,798	5,180,989	1.56	2.28	16.3	99.6
Jersey	9,838,049	4,650,501	1.16	2.28	16.9	99.5
Muturu	8,794,483	4,841,893	0.69	2.29	13.9	94.7

Table 2.3 | InDel statistics in each cattle breed

Breed	Total number of InDel	No of Insertions	No of Deletions	dbSNP (%)	Average per sample
Fulani	5,002,728	2,525,059	2,477,669	64.2	1,542,093
Baggara	4,926,820	2,487,782	2,439,038	64.4	1,518,017
Aryshai	4,893,156	2,470,389	2,422,767	64.3	1,532,091
Kenana	4,888,734	2,475,814	2,412,920	64.2	1,556,799
Ogaden	4,870,321	2,459,035	2,411,286	65.0	1,569,648
Boran	4,757,449	2,410,450	2,346,999	65.1	1,544,602
Butana	4,687,797	2,372,646	2,315,151	64.8	1,545,287
Gash	4,639,715	2,357,344	2,282,371	64.6	1,538,248
Ankole	4,503,750	2,297,000	2,206,750	65.9	1,330,077
Gir	4,429,961	2,210,734	2,219,227	65.3	1,483,275
N'Dama	3,433,613	1,776,432	1,657,181	63.8	1,058,827
Hanwoo	3,327,163	1,694,323	1,632,840	64.8	901,355
Holstein	2,887,371	1,494,939	1,392,432	66.3	813,498
Jersey	2,557,876	1,308,297	1,249,579	67.7	642,114
Muturu	2,328,066	1,197,435	1,130,631	63.3	695,439

A ratio of 2.32 Ti/Tv ratios were obtained in the zebu breeds and approximately 2.28 in the taurine breeds. Among the different cattle breeds, Hanwoo breeds has the highest number of singletons (25.7%), while the two West African taurine, N'Dama and Muturu have the lowest numbers; 12.7% and 13.9%, respectively. Singletons in the African zebu breeds comprised approximately 20% of the total number of SNPs in each of the breeds (Table 2.2).

The number of InDel follow similar trend as the SNPs with Muturu having the lowest number (approximately 2.3 million) and highest number of approximately 5 million identified in the Fulani (Table 2.3).

Annotation of the detected variants in each cattle breed reveal that approximately 65% and 26% variants are located in the intergenic regions and introns, respectively, whereas variants within exons are comprised of less than 1% of the detected variants in each cattle breed. In addition, following comparison with cattle variant database (dbSNP Build 148; (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/cow_9913/chr_rpts/), approximately 4 - 8% of the SNPs in each of the cattle breeds are not present in the SNP database, with the exception of the European cattle breeds (Holstein and Jersey) (Table 2.2). With regards to indels, approximately 30% of indels were not reported before (Table 2.3). These novel variants are a valuable addition to the database of known bovine variants.

2.3.2 Genetic diversity and population differentiation

Genetic parameters including inbreeding coefficient (F), nucleotide diversity and F_{ST} values were estimated in order to examine the genetic diversity and population

divergence among the different cattle breeds. The inbreeding level was estimated as the probability that any two alleles at any given locus are identical by descent (IBD) (Leutenegger *et al.*, 2003). The level of inbreeding among the 15 individual cattle breeds were estimated as the average measure of homozygosity on a per-individual basis within the same breed (Danecek *et al.*, 2011). Average inbreeding coefficients range from 0.9% in Holstein to 17.1% in Gir cattle (Table 2.4). Average inbreeding coefficients are generally low (less than 10%) among the cattle breeds with the exceptions for Muturu, Jersey and Gir cattle breeds for which the estimates were 15.5%, 15.7% and 17.1%, respectively.

The average nucleotide diversity was highest in zebu breeds (> 3.0×10^{-3}) followed by the Gir (2.75×10^{-3}) and Ankole (2.65×10^{-3}) (Table 2.4). A range of values between 1.1×10^{-3} to 1.7×10^{-3} were observed in the taurine breeds with the exception of Muturu where an exceptionally low average value of 9.2×10^{-4} was observed (Table 2.4). Genetic relationships among the different cattle breeds was also estimated through pairwise population differentiation (F_{ST}) which is a widely used statistic in population genetics with values ranging from 0 to 1 indicating no differentiation and complete divergence between populations, respectively (Weir and Cockerham 1984, Hartl *et al.*, 1997). F_{ST} values were the lowest population between Sudanese Gash and Aryashai (F_{ST} = 0.005) and between Gash and Butana (F_{ST} = 0.005), while the highest population differentiation observed was between Gir and Muturu (F_{ST} = 0.476) (Table 2.4). Within the Sudanese zebu breeds, genetic differentiation is generally low with the highest F_{ST} value of 0.030 observed between Gash and Baggara and between Gash and Fulani. Within the African zebu breeds, population differentiation is higher between Boran and the Sudanese breeds than between Ogaden and Sudanese breeds. A somewhat similar

level of population differentiation (F_{ST} = 0.1) is observed between Gir cattle and all African zebu breeds. Between African zebu and taurine breeds, the highest population divergence is observed between Gash and the taurine breeds with F_{ST} values of 0.360, 0.306, 0.252, 0.241 and 0.235 in Muturu, Hanwoo, N'Dama, Jersey and Holstein, respectively. The Fulani cattle is the closest to the taurine breeds. Within the taurine breeds, genetic diversity is lower between the European taurine breeds (Holstein *versus* Jersey, F_{ST} = 0.124) compared to between the African taurine breeds (N'Dama and Muturu, F_{ST} =0.237) (Table 2.4).

2.3.4 Population structure as revealed by PCA and Admixture analyses

The population structure among the different breeds was determined by principal component analysis using genome-wide autosomal SNP data. Following the removal of SNPs with minor allele frequency of less than 0.1 approximately 17 million SNPs were available for PCA analysis. The first and second principal components accounted for 36.26% and 5.92% of the proportion of variation among the 15 cattle breeds. PC 1 clearly distinguished the taurine from the zebu breeds, while PC 2 separates the European taurine from African taurine as well as Sudanese zebu breeds from Asian zebu. The Ankole breed which is an African admixed breed is located at an intermediate position between the African zebu and African taurine breeds **(**Figure 2.2**)**. A plot of PC 3 and PC 4 (Figure 2.3) shows the separation of West African taurine from European taurine while the Asian taurine Hanwoo is positioned at an intermediate position. Also, Asian zebu is clearly separated from African zebu. PC 4 separates Muturu from N'Dama and Sudanese zebu from other African zebu (Figure 2.3).

Table 2.4 | Estimates of genetic diversity parameters among the 15 cattle breeds

Breeds	Inbreeding coefficient (F)	Nucleotide diversity (π)	Het/Hom	Proportion of indicine	Proportion of African taurine	Proportion of European taurine
Fulani	0.059	3.22E-03	1.89	0.71	0.27	0.02
Butana	0.092	3.20E-03	1.60	0.79	0.19	0.02
Aryashai	0.084	3.18E-03	1.62	0.81	0.19	0.00
Kenana	0.095	3.18E-03	1.55	0.80	0.20	0.00
Ogaden	0.097	3.12E-03	1.56	0.73	0.26	0.01
Baggara	0.079	3.09E-03	1.76	0.72	0.26	0.02
Gash	0.088	3.03E-03	1.52	0.85	0.15	0.00
Boran	0.044	3.00E-03	1.71	0.76	0.17	0.07
Gir	0.171	2.75E-03	0.94	1.00	0.00	0.00
Ankole	0.038	2.64E-03	1.95	0.53	0.39	0.09
N'Dama	-0.013	1.71E-03	1.53	0.05	0.95	0.00
Hanwoo	**0.038**	**1.54E-03**	**1.72**	**0.00**	**0.16**	**0.84**
Holstein	0.009	1.23E-03	1.56	0.00	0.00	1.00
Jersey	0.155	1.17E-03	1.16	0.00	0.00	1.00
Muturu	0.157	9.21E-04	0.69	0.00	1.00	0.00

In order to infer regional population structure among the African breeds, we performed separate PCA analysis using only the African breeds. Examining first the African zebu, the Sudanese zebu are separated from the other African zebu (Boran and Ogaden) (PC 1 of Figure 2.4). PCA analysis using only the Sudanese zebu breeds (Figure 2.5) shows the separation of Baggara, Fulani and Kenana from the remaining three Sudanese breeds (PC 1 = 2.06%), while PC 2 of Figure 2.5 shows a further separation of Kenana and Butana (zebu dairy) from the other Sudanese breeds.

Table 2.5 | Pairwise Population fixation (F_{ST}) values among 15 cattle breeds

Breeds	F_{ST}														
	KNN	**BTN**	**ARY**	**BGR**	**GSH**	**FLN**	**OGD**	**ANK**	**BOR**	**MUT**	**NDM**	**HOL**	**JER**	**GIR**	**HNW**
Kenana	0.000														
Butana	0.018	0.000													
Aryashai	0.019	0.010	0.000												
Baggara	0.024	0.018	0.021	0.000											
Gash	0.028	0.005	0.005	0.030	0.000										
Fulani	0.024	0.019	0.021	0.009	0.030	0.000									
Ogaden	0.029	0.024	0.025	0.028	0.034	0.026	0.000								
Ankole	0.092	0.085	0.087	0.070	0.097	0.067	0.073	0.000							
Boran	0.053	0.049	0.049	0.048	0.059	0.047	0.029	0.086	0.000						
Muturu	0.353	0.345	0.348	0.310	0.360	0.305	0.356	0.272	0.354	0.000					
N'Dama	0.244	0.238	0.241	0.206	0.252	0.201	0.243	0.181	0.249	0.237	0.000				
Holstein	0.229	0.225	0.228	0.206	0.235	0.202	0.228	0.199	0.227	0.287	0.241	0.000			
Jersey	0.235	0.231	0.234	0.211	0.241	0.207	0.232	0.203	0.232	0.294	0.246	0.124	0.000		
Gir	0.097	0.095	0.096	0.110	0.104	0.110	0.089	0.198	0.117	0.476	0.371	0.314	0.317	0.000	
Hanwoo	0.284	0.289	0.279	0.245	0.306	0.241	0.284	0.210	0.280	0.254	0.211	0.137	0.147	0.399	0.000

A separate PCA plot of the taurine breeds reveals a clear separation of the African taurine from the European and Asian taurine (PC 1 = 7.01%) and interestingly, at PC 2 (3.95%) separation within the African taurine breeds, while European cattle (Jersey and Holstein) are not distinguishable (Figure 2.6).

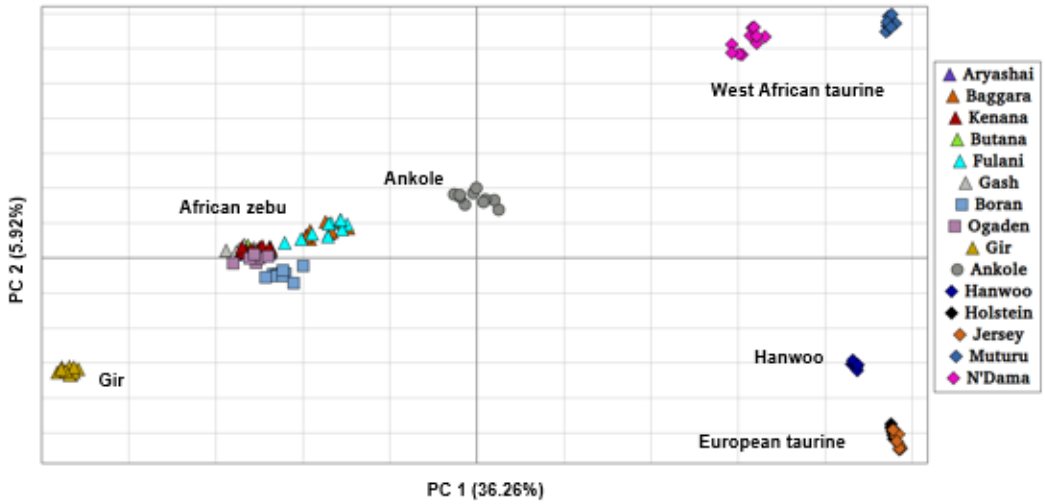


Figure 2.2 | Principal component (PC 1 versus PC 2) analysis for African and reference breeds

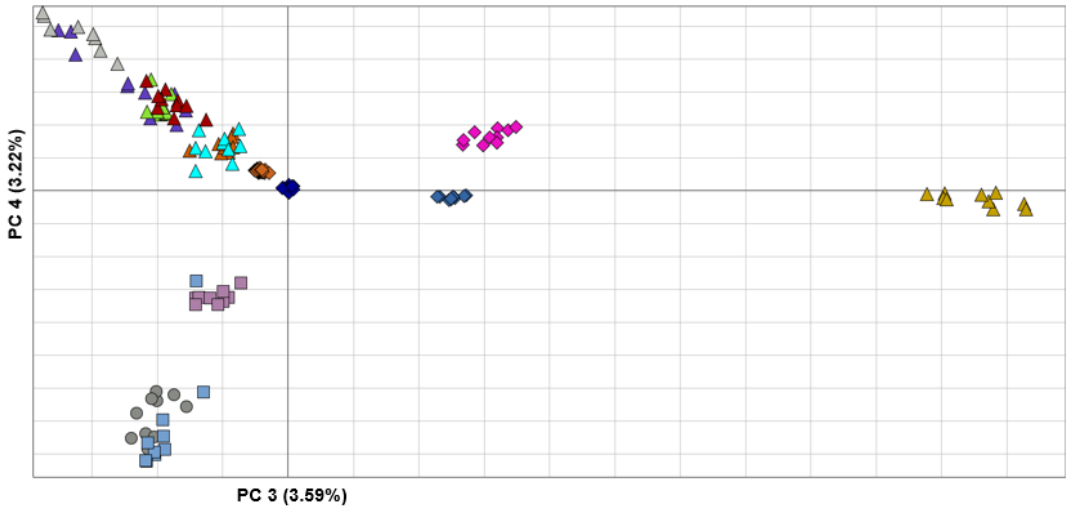


Figure 2.3 | Principal component (PC 3 versus PC 4) analysis for African and reference breeds.

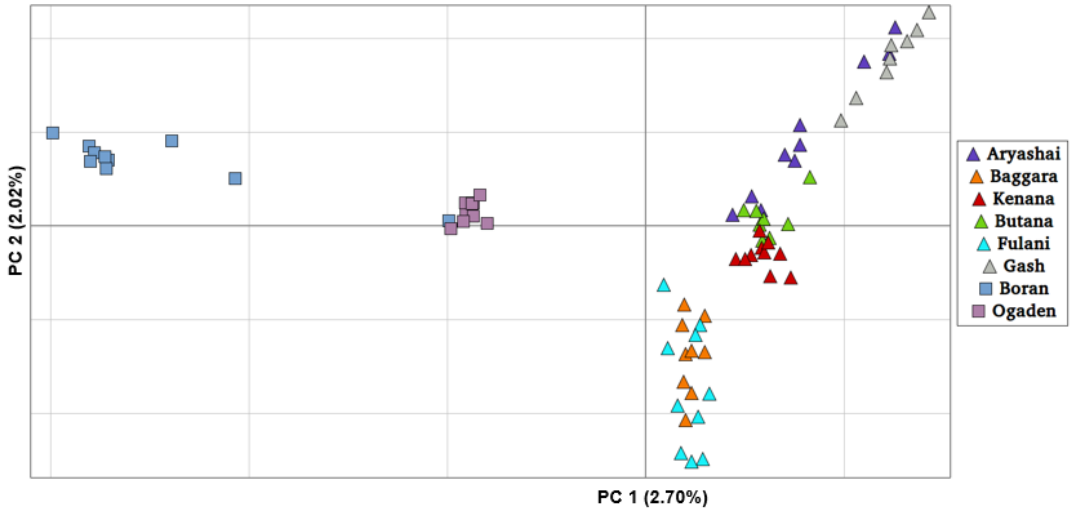


Figure 2.4 | Principal component (PC 1 *versus* PC 2) analysis for the African zebu breeds.

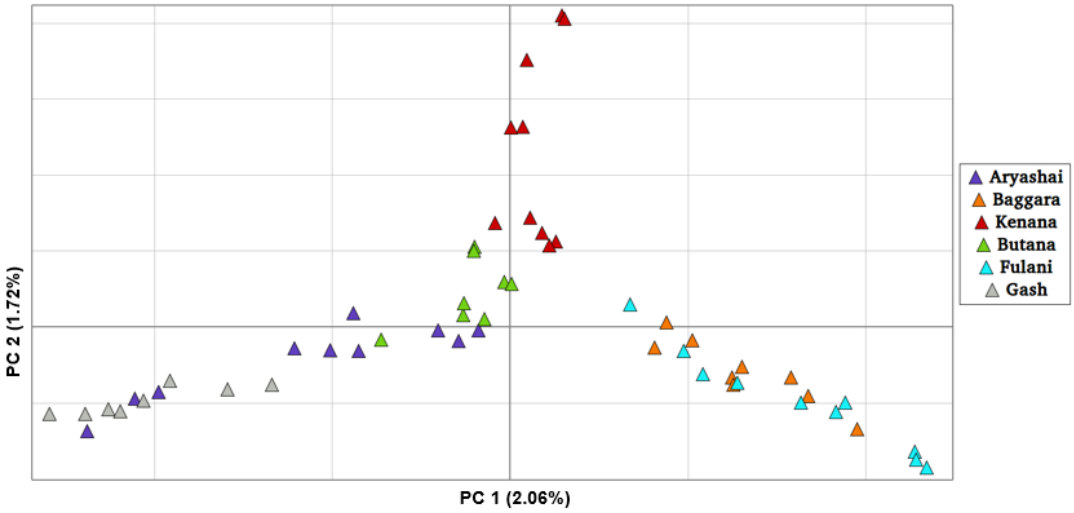


Figure 2.5 | Principal component (PC 1 *versus* PC 2) analysis of Sudanese zebu breeds.

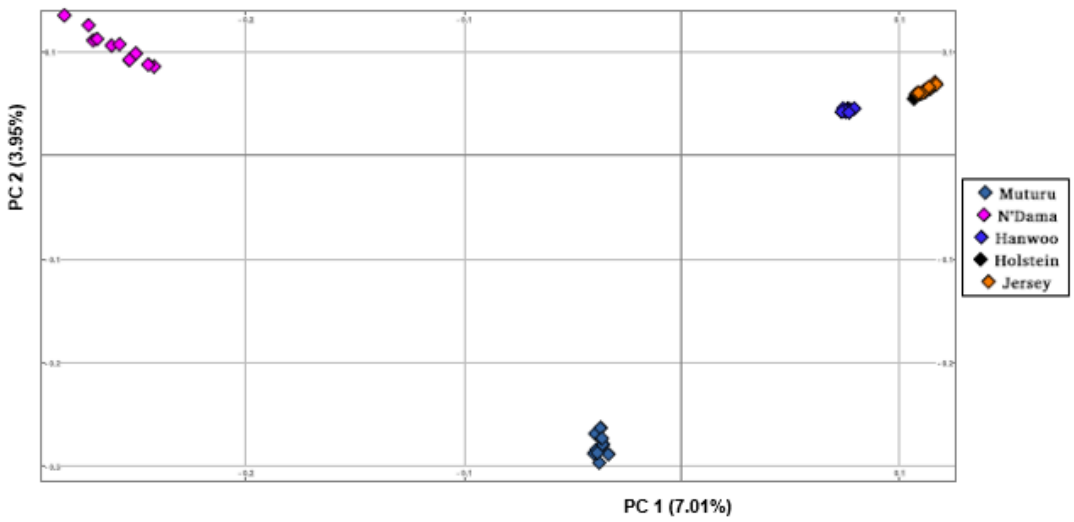


Figure 2.6 | Principal component (PC 1 *versus* PC 2) analysis for the taurine breeds.

Using approximately 1.9 million SNPs in high linkage equilibrium (obtained after LD pruning of the 17 million SNPs used for PCA), we explored the genetic ancestry of the 15 cattle breeds for the possible number of ancestral population, K, ranging from 2-10 (Figure 2.7). Results of admixture analyses revealed that at K = 2, all African zebu breeds share some common ancestry with the taurine breeds, while the N'Dama and to a small extent, Hanwoo taurine cattle show a common ancestry with zebu cattle.

Using the cross validation (CV) procedure to determine the optimum K with less CV error (Alexander *et al.*, 2009), the result of our analyses supports K = 3 as the most likely number of ancestral populations for our study breeds. The three ancestral populations inferred here are zebu, African taurine and European taurine. At this level, the African zebu show similar zebu admixed genetic background as in K = 2 with an indication of the taurine background to be of African taurine origin. However, Baggara, Fulani and Boran breeds showed but very small level additional genetic background from European taurine, the proportion of European taurine genetic background is

highest in the Boran cattle. Among the Sudanese zebu, Baggara and Fulani also show higher taurine background than the others (Figure 2.7a).

As expected, Ankole cattle which is an African admixed breed is shown to be of both zebu and African taurine genetic makeup with additionally a slight proportion of European background. A majority of the N'Dama samples (7 out of 10) are shown to contain varying levels of zebu genetic background. The Hanwoo samples are shown to be of two taurine backgrounds but mainly of the European genetic background. Interestingly, Muturu cattle is shown to be of a single taurine genetic background and it also differentiated from the N'Dama cattle from K = 5 up to K = 10. The Gir cattle is shown to be of single zebu ancestry while the commercial breeds (Holstein and Jersey) are also of a single genetic background.

Based on the assumption of three ancestral populations (K = 3), the estimated average ancestry proportions of the different cattle breeds are provided in Table 2.4 while the proportions of ancestry for individual cattle samples are provided in Supplementary Table S2.1. The admixture plots for only the African and European *B. taurus* populations (Figure 2.7b) reveal clustering at a continental level when K = 2. When K = 3, the European populations are separated while the African populations are still clustered together, however, when K = 4, all four *B. taurus* breeds have separate ancestry (Figure 2.7b).

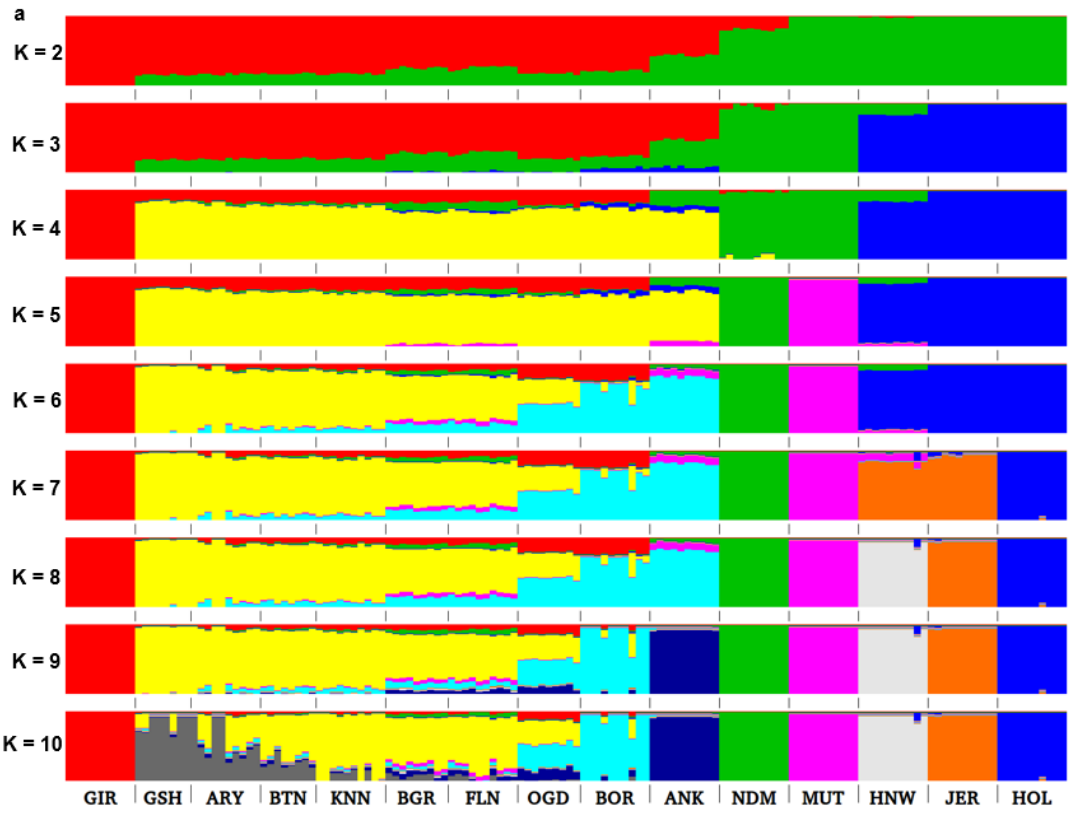


Figure 2.7a | Admixture plot showing ancestry proportions for the 15 cattle breeds. Population structure assessed using ADMIXTURE *ver*.1.3.0, bar plot, generated by GENESIS. Each individual/population is represented by a vertical bar and partitioned into coloured segments with the length of each segment representing the proportion of the individual's genome from K = 2 to K = 10 ancestral populations. GIR – Gir, GSH – Gash, ARY – Aryashai, BTN – Butana, KNN – Kenana, BGR – Baggara, FLN – Fulani, OGD – Ogaden, BOR – Boran, ANK – Ankole, NDM – N'Dama, MUT – Muturu, HNW – Hanwoo, JER – Jersey and HOL – Holstein

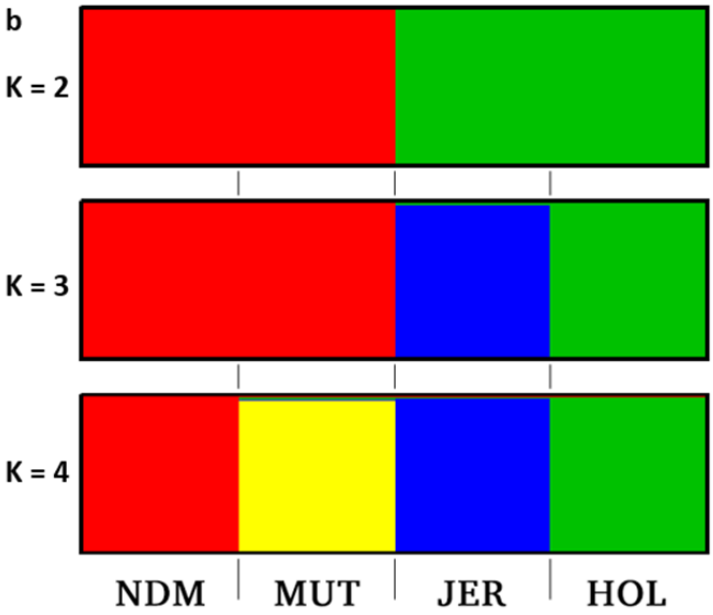


Figure 2.7b | Admixture plot for African and European taurine breeds only. Population structure assessed using ADMIXTURE ver1.3.0 program, bar plot, generated by GENESIS. Each individual/population is represented by a vertical bar, often partitioned into coloured segments with the length of each segment representing the proportion of the individual's genome from K = 2 to K = 4 ancestral populations. NDM – N'Dama, MUT – Muturu, HNW – Hanwoo, JER – Jersey and HOL – Holstein.

2.3.5 Phylogenetic relationships

The clustering of all 144 cattle samples based on the maximum likelihood tree generated by SNPhylo software agreed with the results of PCA and admixture analysis with clustering of the samples according to their respective breeds and their geographic regions of origin (Figure 2.8). Therefore, African zebu are clustered together and separately from the taurine breeds. Within the African zebu, the Sudanese zebu are clustered separately from Boran and Ogaden while Baggara and Fulani samples are placed at intermediate positions between the zebu and taurine breeds.

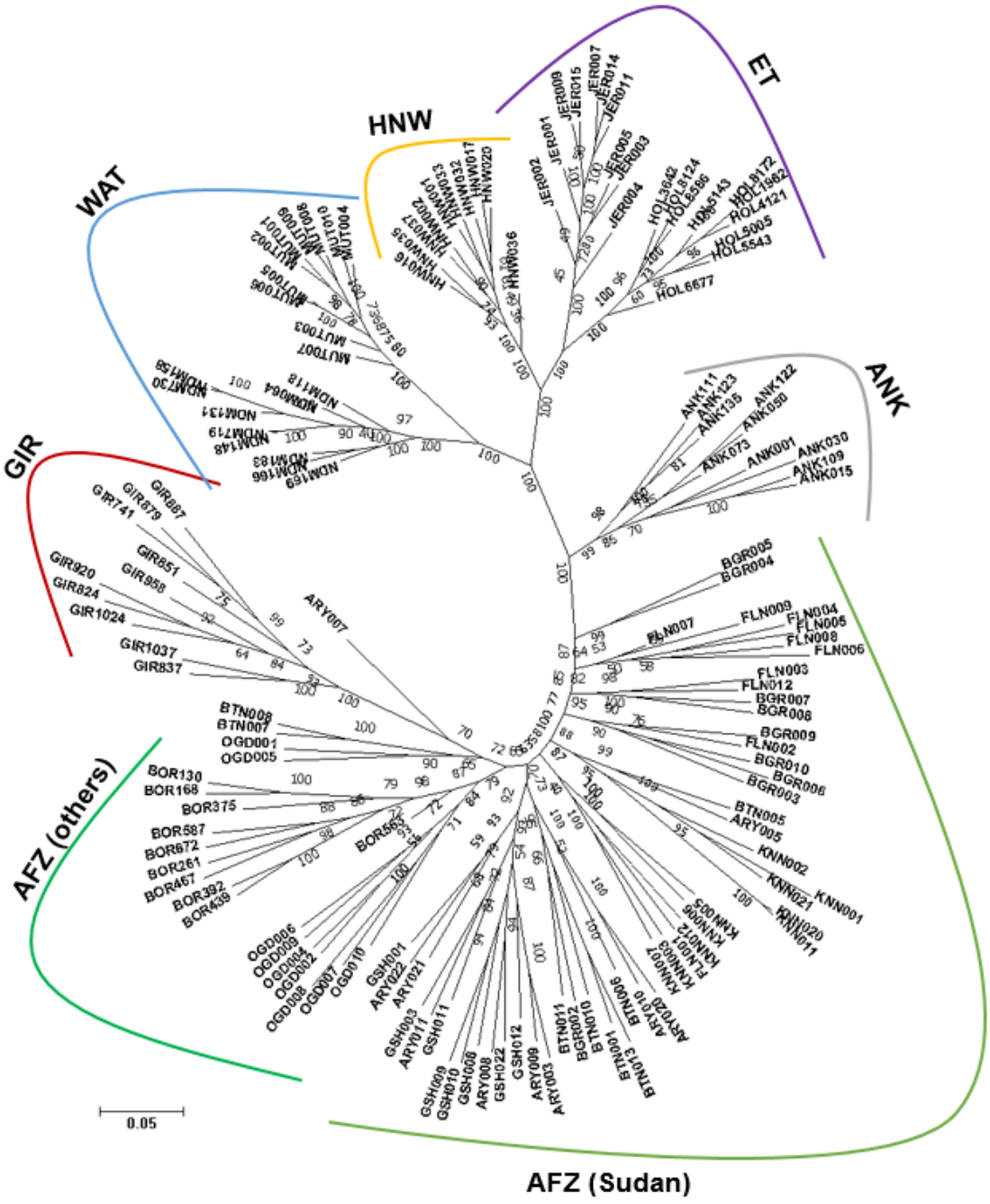


Figure 2.8 | Phylogenetic relationships among the 144 cattle samples based on unrooted maximum likelihood tree constructed with a bootstrap value of 100. WAT – West African taurine, HWN – Hanwoo, ET – European taurine, ANK – Ankole, AFZ – African zebu, GIR - Gir

We further investigated the population structure and possible gene flow among the 15 cattle breeds using the Treemix software. The initial plot of evolutionary relationships without inferring any migration events reveals a population structure similar to the admixture results whereby the zebu breeds are separated from taurine breeds and the African zebu breeds placed at an intermediate position due to their admixed nature. Therefore, the African zebu breeds were positioned according to their proportions of the two ancestry's genetic backgrounds in the following order, starting from the closest to the taurine: Fulani – Baggara – Boran - Ogaden – Kenana – Butana – Aryashai – Gash (Figure 2.9a). We sequentially added up to 6 migratory events to the tree that reveal high levels (migration weight) of gene flow between breeds (Figure 2.9b). This tree shows possible introgression from N'Dama ancestry into Fulani and Baggara, and from Ankole into Ogaden, Boran and Kenana. Within the Sudanese zebu, there is an evidence of gene flow between Kenana and Butana. Interestingly there is also a strong indication of introgression between the two West African taurine breeds with direction of gene flow from Muturu ancestry to the N'Dama cattle. However, we did not identify introgression of European taurine breeds into any of the African cattle breeds.

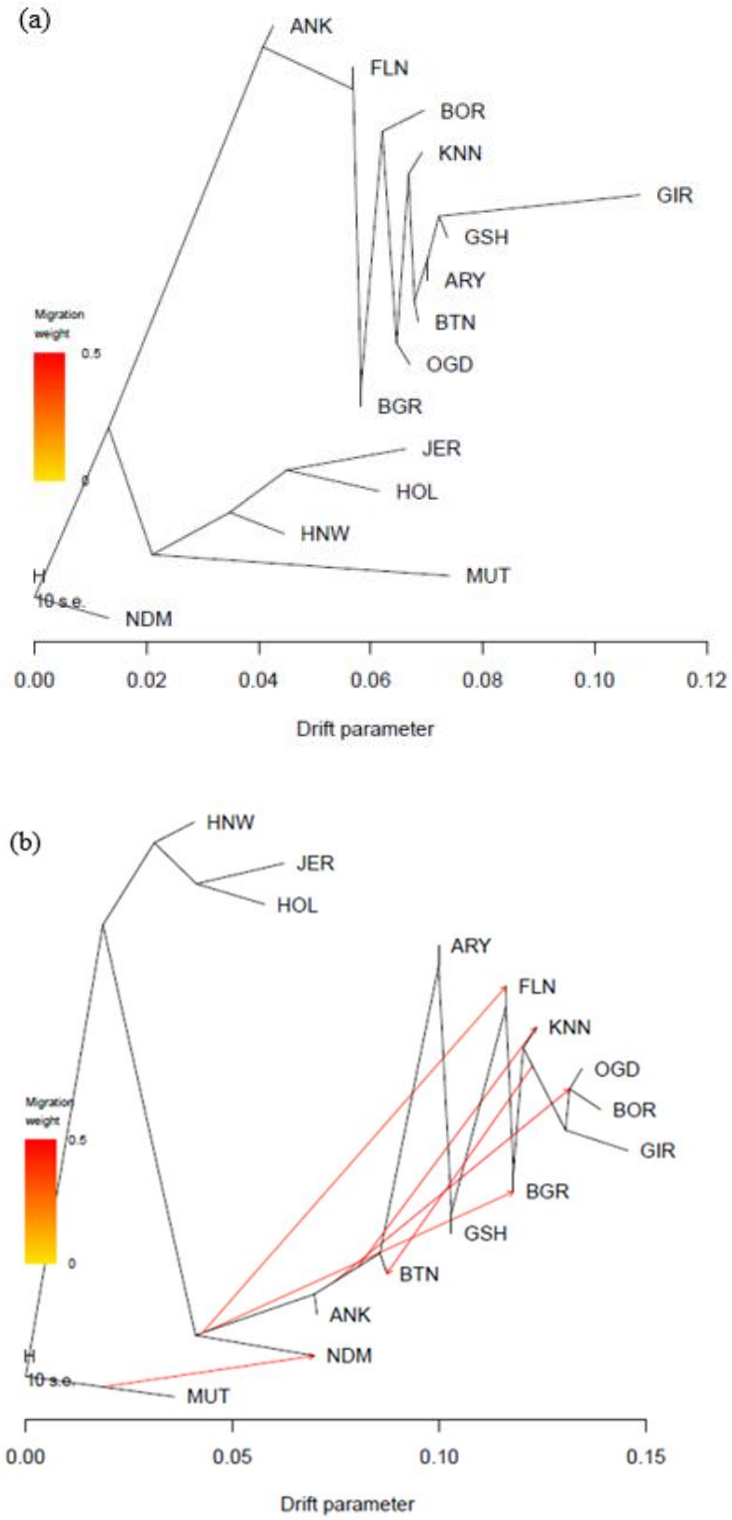


Figure 2.9 | Pattern of splits and gene flow among 15 cattle breeds, a) without migration events, b) with 6 migration events.

2.3.6 Linkage disequilibrium (LD) decay and effective population size Ne

The plot of LD decay against genome distance revealed a faster decay LD in the majority of zebu breeds (Figure 2.10b) compared to most of the taurine breeds (Figure 2.10a). Among the taurine breeds, the N'Dama breed shows the highest average LD level and Hanwoo the lowest. In European populations, the LD average r^2 reached 0.2 at a distance of approximately 200 kb while at the same distance, the value is approximately 0.25 for Muturu and reaching above 0.3 in N'Dama and less than 0.2 in the majority of the zebu breeds. Among the zebu, the highest average genome LD is observed in Gash and the lowest in Fulani cattle.

The average r^2 across genomic distances was used to model historical effective population size (N_e). At around 1000 generations ago, the effective population sizes of all cattle breeds were less than 2000 except for Hanwoo breed that is slightly above 2000 (Figure 2.11). A similar trend in declining *Ne* is observed in Muturu and the European populations from approximately 200 generations ago down until recent (Figure 2.11a). At all times the effective population sizes of zebu breeds were higher than those of taurine populations. Both zebu and taurine populations experience a declining trend from historical N_e peaks more than 900 generations ago and the effective population sizes in the generation before the current one was estimated to be in the interval of 27 to 50. This result is suggestive that all the cattle breeds experienced an effective population size bottleneck in the past.

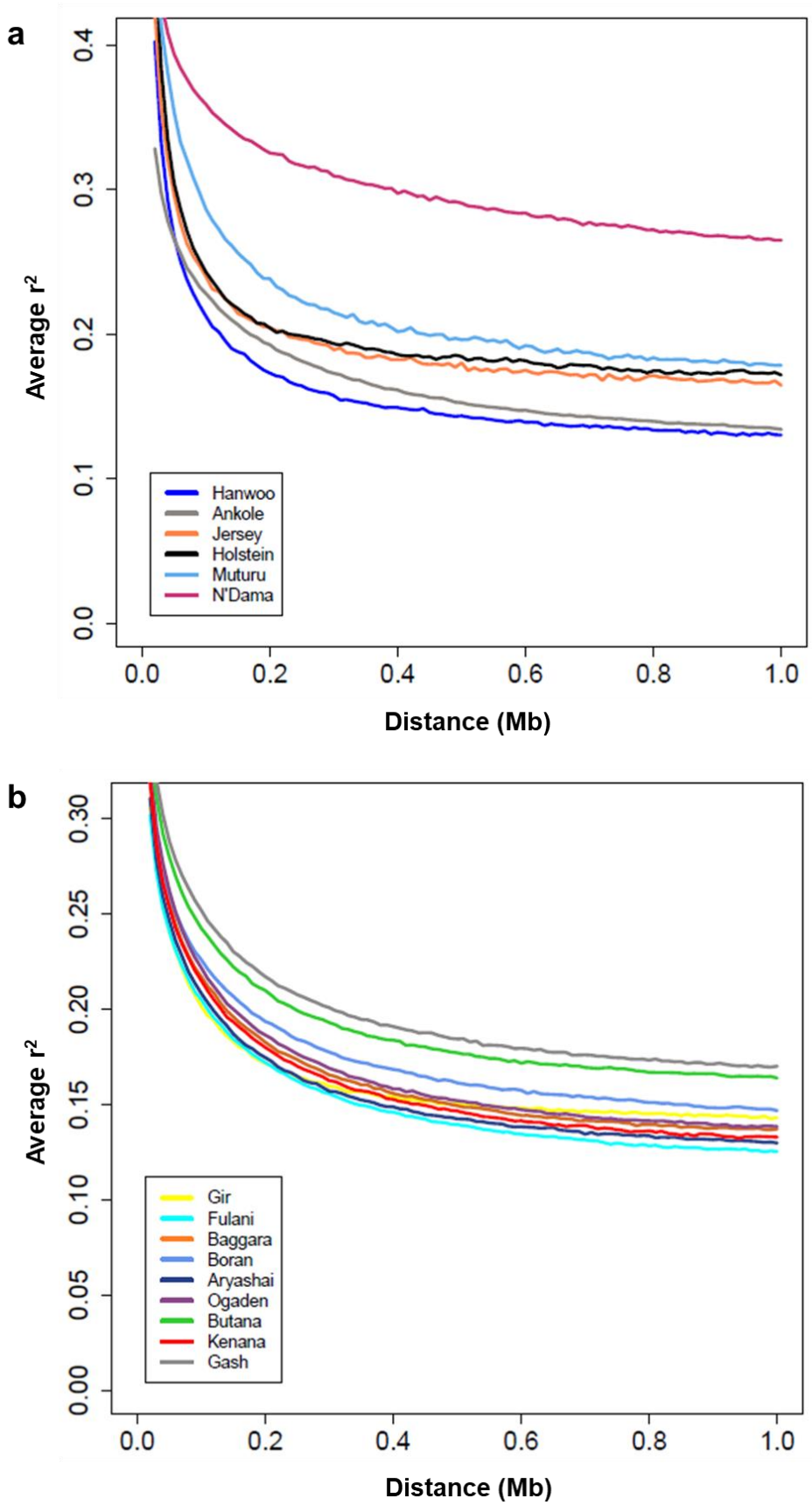


Figure 2.10 | Decay of average r^2 from markers 0 to 1Mb apart across all autosomes in (a) taurine breeds and (b) zebu breeds.

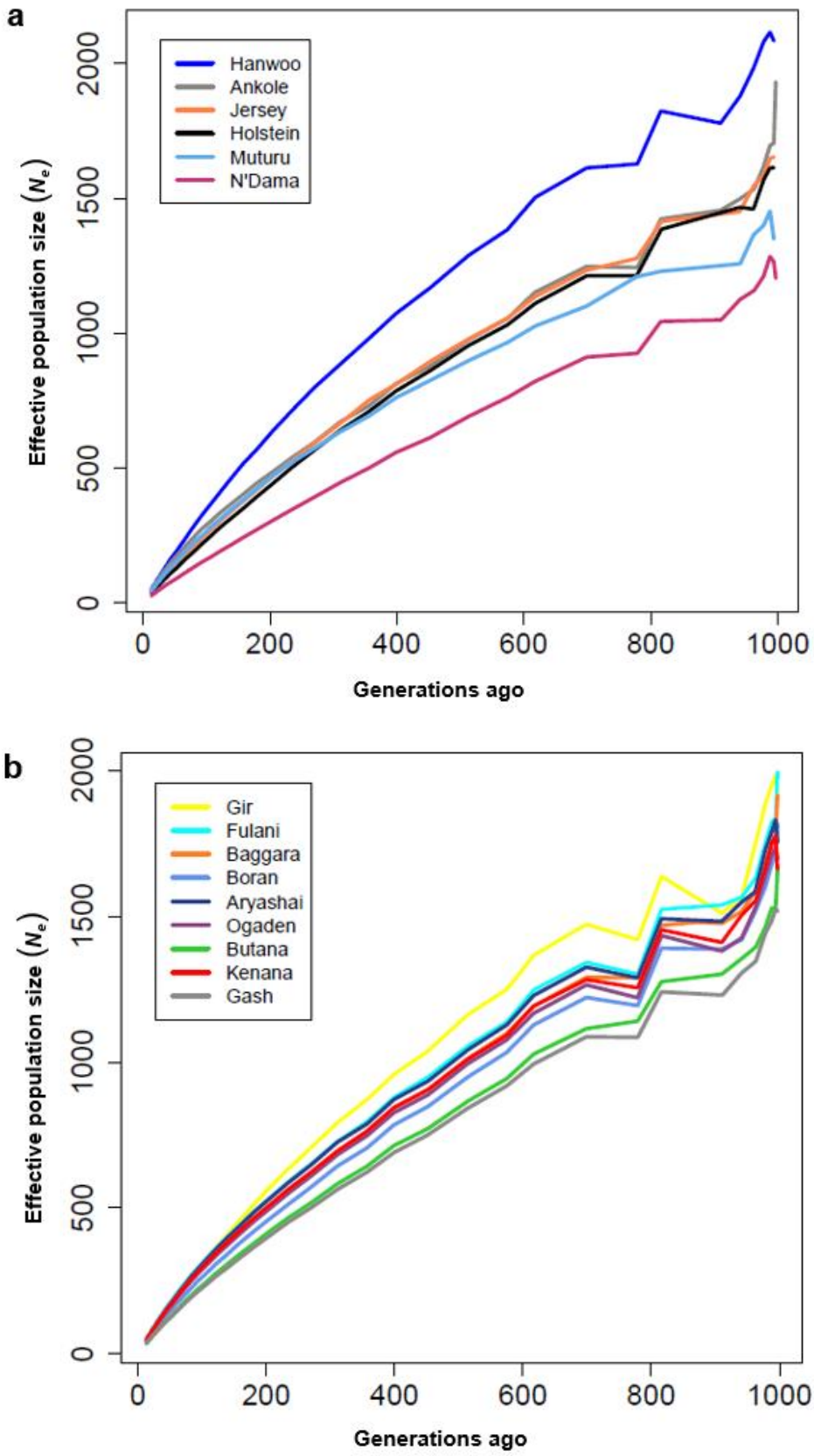


Figure 2.11 | Effective population size (Ne) estimated based on genome-wide LD in the (0 to 1000 generations ago), (a) taurine and (b) zebu breeds.

2.4 Discussion

In this chapter, we report the discovery of nucleotide sequence (SNPs and InDels) polymorphism in 15 cattle breeds following their comparisons with the bovine reference assembly UMD3.1. The numbers of variants identified in the zebu breeds are approximately 2 – 2.5 times higher than the numbers of detected variants in the taurine breeds. The observed higher numbers of detected variants in the zebu breeds is expected as the cattle reference assembly was constructed with sequences from a taurine breed, the Hereford (Zimin *et al.*, 2009). However, within the African taurine breeds, the difference in the number of SNPs detected in Muturu and N'Dama is quite high and unexpected as both populations are from Africa in contrast to the European populations which are expected to be genetically closer to the reference Hereford cow. A possible reason for this could be as a result of the high level of inbreeding in the Muturu, also supported by the low ratio of heterozygous SNPs to non-reference homozygous SNPs compared to N'Dama. Another factor could be the zebu admixture (approximately 0.05%) in N'Dama as revealed in the admixture plot (Figure 2.6).

Genetic diversity among the different cattle breeds was assessed using genetic parameters such as inbreeding coefficient (*F*), nucleotide diversity and population differentiation (F_{ST}). Inbreeding level is generally low among the cattle breeds with the exception of Muturu, Jersey and Gir. The high inbreeding levels in these breeds are also apparent by their low levels of the observed ratios of heterozygous SNPs to homozygous non-reference SNPs (Table 2.4). Higher inbreeding in a population will result in the loss of genetic variation and could also lead to potential loss of favourable alleles within the population (Santana *et al.*, 2010)**.** It is noteworthy to mention, however, that although the highest average inbreeding coefficients among the cattle

breeds studied here is approximately 17%, some individual samples within these breeds have inbreeding coefficient less than 10% and others have inbreeding levels of 17% (Supplementary Table S2.1).

The average nucleotide diversities observed in the taurine breeds are approximately 2 times less than those of the zebu breeds. Apart from Muturu, average nucleotide diversity values observed for all breeds are similar to values for taurine and zebu cattle breeds supporting the possibility that most modern cattle including African cattle breeds experienced bottlenecks during domestication (Kim *et al.,* 2017, Mei *et al.,* 2017). The much lower level of variation within the Muturu breed as reflected by the low nucleotide diversity, low ratio of heterozygous versus homozygous SNPs and higher inbreeding coefficient may suggest that this breed experienced an even stronger bottleneck than the other breeds (Wiener and Wilkinson 2011). The higher nucleotide diversity observed in the zebu breeds may indicate a large ancestral effective population size for *Bos indicus* while the lower genetic diversity within taurine breeds could reflect a lower diversity within the pre-domestication ancestral population and/or post-domestication effects of stronger bottlenecks at breed formation and stronger selection (Bovine HapMap Consortium, 2009).

The admixture results, particularly the relative ancestry proportion of the African zebu breeds, is in concordance with previous reports of the admixed nature of the majority of indigenous African zebu breeds. In addition, the observed zebu introgression in N'Dama is consistent with the report of previous studies that have utilized both high density SNP data and full genome sequence data (McTavish *et al.,* 2013, Decker *et al.,* 2014, Mbole-Kariuki *et al.,* 2014, Kim *et al.,* 2017, Bahbahani *et al.,* 2018). However,

an observation from these previous reports, involving HD SNP data, and this study is the suggestion of N'Dama share ancestry at K = 2 with zebu. This admixture vanishes making N'Dama a pure clade, when K = 3. The inherent ascertainment bias of the bovine HD SNP chip (Matukumalli *et al.,* 2009) may be the explanation here. On the other hand, our admixture results in respect to the Muturu samples support the existence of pure taurine ancestry in Africa (Decker *et al.,* 2014). Surprisingly, the population divergence between the African taurine breeds (0.237) is as high as the level of divergence between N'Dama and African zebu and in some cases even higher. For instance, F_{ST} values of 0.201, 0.206 and 0.238 were observed between N'Dama and zebu breeds such as Fulani, Baggara and Butana, respectively.

Put together, the results presented here are clearly supporting significant population differentiation among the different cattle populations, particularly between the West African and European *B. taurus*. The different population structure within the West African *B. taurus* is an indication of clear population divergence between Muturu and N'Dama, a possible legacy of different demographic histories for the West African shorthorn and longhorn *B. taurus* cattle populations (Payne and Hodges 1997).

CHAPTER 3:

Genome-wide analysis reveals candidate signatures of positive selection in the endangered trypanotolerant Muturu cattle

3.1 Introduction

The Muturu cattle indigenous to the West African sub-region is among the least characterized shorthorn (brachyceros) humpless *Bos taurus* cattle, which include also the Somba, Baoulé, and Lagune breeds. The word Muturu, the Hausa word for humpless, is used for all taurine shorthorns in English-speaking West African countries. The past distribution of Muturu ranged across West and Central Africa, particularly in Cameroon, Liberia, Ghana and Nigeria. Nowadays, Muturu is sparsely distributed in the humid forest zone and in a few savannah areas (Gwaza and Momoh 2016).

It is presumed that Muturu cattle populated most of Nigeria before the arrival of zebus in the semi-arid regions in the 17th century. The late 19th century cattle rinderpest plague seems to have caused the extinction of many Muturu herds (Blench 1998). Two Muturu ecotypes may be recognised in Nigeria, both sparsely distributed, the Savannah Muturu in pockets of North Central of Nigeria (Benue Plateau and surroundings) and the Forest Muturu found in South - East and South West Nigeria (DAGRIS 2005).

Muturu are the smallest cattle of Nigeria with a withers height of 83-93 cm and a weight of 150 - 225 kg (DAGRIS, 2005). Muturu cattle are important in traditional culture and there is a strong spiritual attachment to the animals, they are frequently used for prestige or dowry purposes. The animals play an important role in all kind of ceremonies and they are used for ritual sacrifices, particularly at funerals when their hides are used as wrapping for the deceased. Many bulls are therefore killed for ceremonial use in the traditional Muturu keeping areas, which may lead to shortages of breeding bulls (Blench 1998). Muturu cattle are known for their resistance/tolerance to the parasitic protozoal disease trypanosomiasis (Murray 1988). Muturu cattle are also tolerant to tick and tick-borne diseases but they are susceptible to rinderpest (DAGRIS, 2005).

In this chapter, we report a genome-wide assessment of candidate positive selection signatures in a Muturu population from Nigeria using whole genome re-sequencing. These analyses involved the detection of signatures of positive selection using two complimentary extended haplotype homozygosity based selection tests (integrated haplotype score (*iHS*) and between population (*Rsb*) tests). Muturu results were compared to N'Dama, a West African longhorn trypanotolerant taurine, and two non – trypanotolerant European taurine breeds (Holstein and Jersey).

3.2 Materials and Methods

3.2.1 Study populations and genome-wide SNP data

A total of 40 cattle samples representing four different cattle populations were included in the analysis, namely Muturu (MUT, $n = 10$ from the forest ecotype sampled in South-East Nigeria), N'Dama (NDA, $n = 10$), Holstein (HOL, $n = 10$) and Jersey (JER, $n = 10$) (see Chapters 1 and 2 for further details). After quality control steps (Chapter 2), 8,794,483, 14,107,603, 10,277,798 and 9,838,049 SNPs were available for analysis in Muturu, N'Dama, Holstein and Jersey, respectively (see Chapter 2, Table 2.2).

3.2.2 Signature of selection using integrated haplotype score (*iHS*) test

The *REHH* package (Gautier and Vitalis 2012) in R was used to compute the *iHS* score for all SNPs with MAF $\geqslant 0.05$ within each of the four *Bos taurus* populations separately and using the option "*freqbin =0.1"*. The *freqbin* option enables the calculation of empirical *P*-values for every score based on the number of SNPs within a bin of similar allele frequency of size 0.1 (Voight *et al.* 2006). *P*-values were derived as

−log10(1−2|Φ(iHS)−0.5|), where Φ(iHS) represents the cumulative distribution function of the Gaussian density (Gautier and Vitalis 2012)

To infer candidate regions under selection, we utilized an in-house R script to summarize the *iHS* scores for each SNP within a sliding window of 100 kb with an overlap step size of 50 kb, selecting SNPs with the highest test score to represent a particular window. We applied a threshold of $P < 0.0001$, equivalent to $iHS > 4$, for any window to be considered as candidate selection signature window**.** Bedtools *v2.25.0* (Quinlan and Hall 2010) was used to merge the overlapping selected windows to define a selection region.

As this test require phased haplotypes (Sabeti *et a*l. 2002, Voight *et al*. 2006), phasing was performed separately for each cattle population and for each chromosome using *SHAPEIT* (Delaneau *et al*. 2013). We adopted the "read aware" phasing approach which takes advantage of the phase information contained in sequencing reads to improve the quality of the resulting haplotypes. Following this approach, we extracted Phase Informative Reads (PIRs) from the sequence alignment files (Bam files) for all samples and then performed the actual phasing of the genotypes contained in the variant file using the extracted PIRs as an additional input to the *SHAPEIT* software (Delaneau *et al.* 2013).

3.2.3 Signature of selection using extended haplotype homozygosity (*Rsb*)

Evidence of positive selection was further investigated by comparing Muturu to each of the other three *B. taurus* breeds using the *Rsb* test described in Chapter One. To obtain *Rsb* scores, we utilized the different iES statistics estimated using the REHH package during the *iHS* analyses in each of the Muturu, N'Dama, Holstein and Jersey

populations as described above. The iES values were summarized using the same overlapping window size of 100 kb and step size of 50 kb, as described above. *Rsb* scores were then computed by comparing Muturu with N'Dama, Holstein and Jersey using Equation 3.2. *Rsb* scores were standardized (normal deviate) in R environment. As the *Rsb* score is directional, a positive *Rsb* score suggests selection in population 1 while the negative score suggests selection in population 2. In our case, selection signatures in Muturu are represented by the positive values. To select candidate selection signature windows, the extreme 0.5% window values were considered. Bedtools *v2.25.0* (Quinlan and Hall 2010) was used to merge the overlapping selected windows to form a selection region.

3.2.4 Functional annotation of candidate selected genes

Cattle genes which overlap genomic windows passing the significant selective sweep thresholds were retrieved based on *Ensembl Genes 92 database* using the *Ensembl BioMart* online tool (http://www.ensembl.org/biomart) (last accessed on June 10, 2018) (Smedley *et al.*, 2009). The list of candidate genes was processed in the web-based Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (http://david.abcc.ncifcrf.gov/) (Huang *et al.*, 2009, Huang *et al.*, 2009) for functional annotation and identification of over-represented genes involved in biological processes (GO) and KEGG pathways. Further characterization of candidate selected genes was performed by comparing the missense variants identified within the regions among the four taurine breeds.

3.3 Results

3.3.1 Comparison of SNPs detected in four *B. taurus* breeds

A total of 19,947,406 autosomal bi-allelic SNPs was identified in the *B. taurus* breeds when compared to the cattle genome of reference UMD 3.1. Amongst these, 4,876,129 SNPs were common to all four populations, a proportion equivalent to 55.4%, 49.6%, 47.4% and 34.6% of autosomal SNPs in Muturu, Jersey, Holstein and N'Dama respectively. 1,022,193 (11.6%) were commonly identified in Muturu and N'Dama, while 999,226 (5%) and 5,435,777 (27%) are uniquely identified in Muturu and N'Dama, respectively. So more than 37% of the SNPs were uniquely present in African cattle compared to the two European breeds.

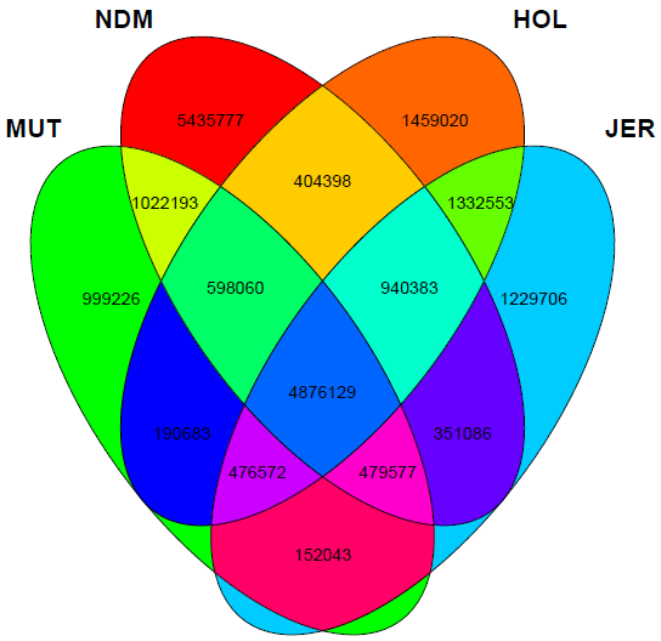


Figure 3.1 | Comparison of the autosomal bi-allelic SNPs between Muturu and other *Bos taurus* breeds. MUT - Muturu, NDM - N'Dama, HOL - Holstein, JER – Jersey

3.3.2 Genomic signatures as revealed by *iHS*

To investigate ongoing within breed recent selection in Muturu, we used the EHH-based iHS test. Genome-wide within population *iHS* scores was calculated using all

SNPs with minimum allele frequencies of 5% in the breed. The distributions of the standardized *iHS P*-values along the 29 bovine autosomes is shown at Figure 3.1. Using a threshold of $iHS \geq 4$ (*P*-value < 0.0001), we identified 266 significant genomic regions to be under positive selection in Muturu (Supplementary Table S3.1). Of these, 169 regions overlap with or are in nearby positions to 430 candidate genes while the remaining 97 selected regions are gene desert regions (GDR) according to the *Ensembl* cow genes database 92 (Supplementary Table S3.1).

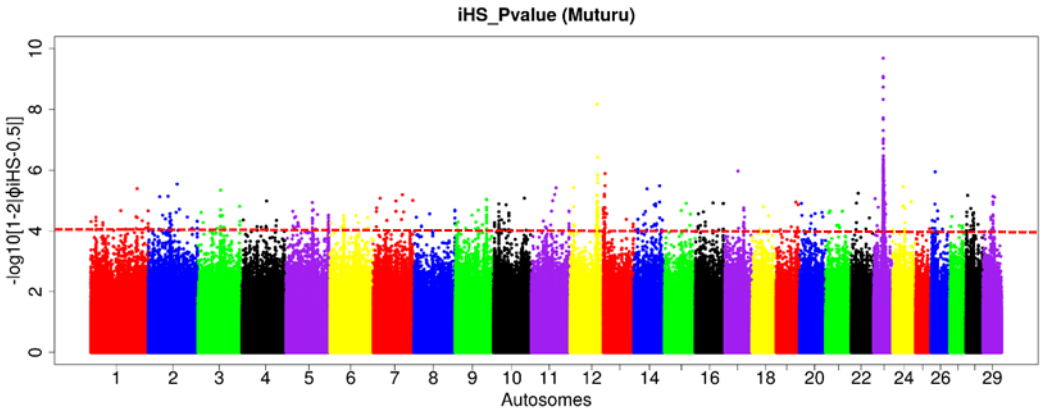


Figure 3.2 | Manhattan plot of the genome-wide distribution of iHS scores along bovine autosomes in Muturu cattle. Red dash line indicate threshold of iHS = 4.

Amongst the top 10 candidate selection signature regions, beside several known bovine genes, we also found many others yet to be annotated (Table 3.1). The strongest selection signal in Muturu was observed on BTA 23 (-10log (*iHS* P value) = 9.68) (Figure 3.2). Among the functionally annotated genes, we identified the bovine major histocompatibility complex (MHC) class I and II genes (*JSP.1*, *BOLA*, *BOLA-DRB3* and *BLA-DQB*), the olfactory receptor genes family (*OR5V1* and *OR12D3*), tripartite motif-containing (TRIM) genes (TRIM10, 15, 26 and 40) and the *NOTCH4*, *PBX2*, *RNF5*, *PPT2*, *PRRT1* and *AGER* genes.

The MHC region is known to contain several genes that have important roles in both innate and adaptive immune responses, the MHC and the TRIM genes particularly have roles in adaptive and innate immune responses respectively (Amills *et al.,* 1998, Si *et al.*, 2006, Ylinen *et al.*, 2006). The bovine MHC, also known as *BoLA* (bovine leucocyte antigen) complex, is of particular interest to animal breeders and veterinary geneticists because of its links to genetic resistance and susceptibility to several diseases. Other top selection signatures regions based on *iHS* test were detected on chromosomes BTA 12, 17, 26 and 13 (Table 3.1).

Table 3.1 | Top 10 candidate selection signature regions in Muturu

Chr	Start (bp)	End (bp)	-log10 (iHS)	Gene overlapping candidate selected region
23	26200001	26400000	9.69	*ENSBTAG00000026163*
12	73950001	74300000	8.17	*ENSBTAG00000047360, ENSBTAG00000047764*
23	26650001	27050000	7.03	*ENSBTAG00000046920, PBX2, AGER, RNF5, EGFL8, PRRT1, ENSBTAG00000046116, ENSBTAG00000023563, ENSBTAG00000048304, NOTCH4, AGPAT1, PPT2, ENSBTAG00000023541*
23	25700001	26200000	6.02	*BLA-DQB, ENSBTAG00000033979*
17	35800001	35950000	5.97	*ENSBTAG00000045738, ENSBTAG00000045938*
26	11450001	11600000	5.95	*KIF20B*
13	5300001	5450000	5.89	*SRY*
23	28400001	28650000	5.66	*JSP.1, BOLA, ENSBTAG00000037421, ENSBTAG00000044550, TRIM10, TRIM40, TRIM15, TRIM26*
23	29150001	29400000	5.50	*OR12D3, ENSBTAG00000039901, OR5V1, ENSBTAG00000013654, ENSBTAG00000031843, ENSBTAG00000039534, ENSBTAG00000046777, ENSBTAG00000047086, ENSBTAG00000046023, ENSBTAG00000038608, ENSBTAG00000037628, ENSBTAG00000031850, ENSBTAG00000000228*
14	71350001	71500000	5.49	*C8orf37*

3.3.3 Candidate signatures of selection revealed by *Rsb*

The EHH-based *Rsb* approached was further used to unravel the genomic signatures of selection in Muturu by comparing its genome to each of three other *B. taurus* breeds, namely N'Dama, Holstein and Jersey. The distribution of *Rsb* values along the bovine autosomes for the three comparisons are represented in the Manhattan plots shown in Figures 3.3 – 3.5. A total of 145 selection signature regions were identified, out of which 81 selection regions overlapped with a total of 206 candidate genes while the remaining 64 regions are gene desert regions (Supplementary Table S3.2).

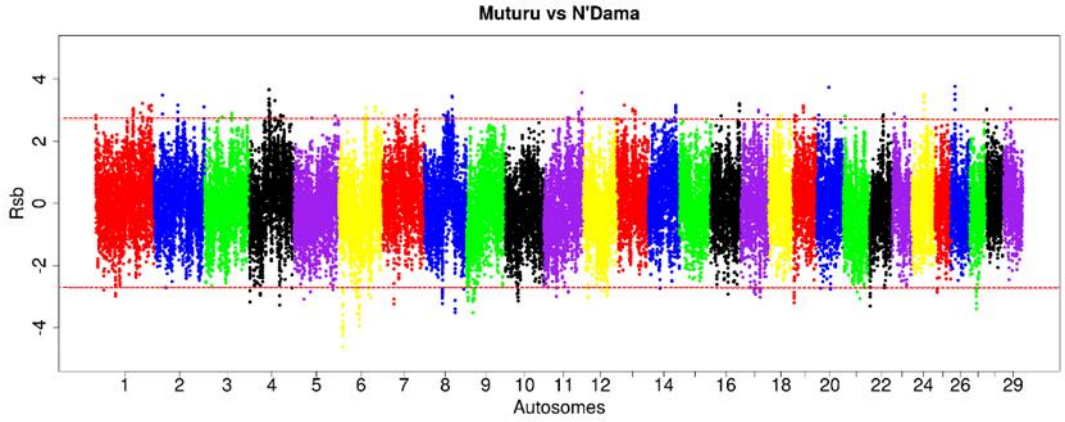


Figure 3.3 | Manhattan plot of the genome-wide distribution of Rsb values following the comparison of Muturu with N'Dama. Red dash lines indicate thresholds of selection windows in the top 0.5% *Rsb* scores.

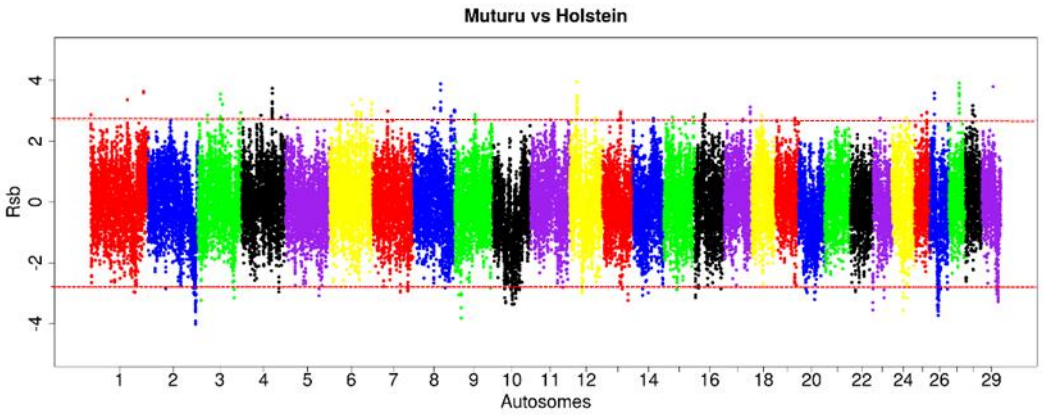


Figure 3.4 | Manhattan plot of the genome-wide distribution of Rsb values following the comparison of Muturu with Holstein. Red dash line indicate threshold of selection windows in the top 0.5% *Rsb* scores.

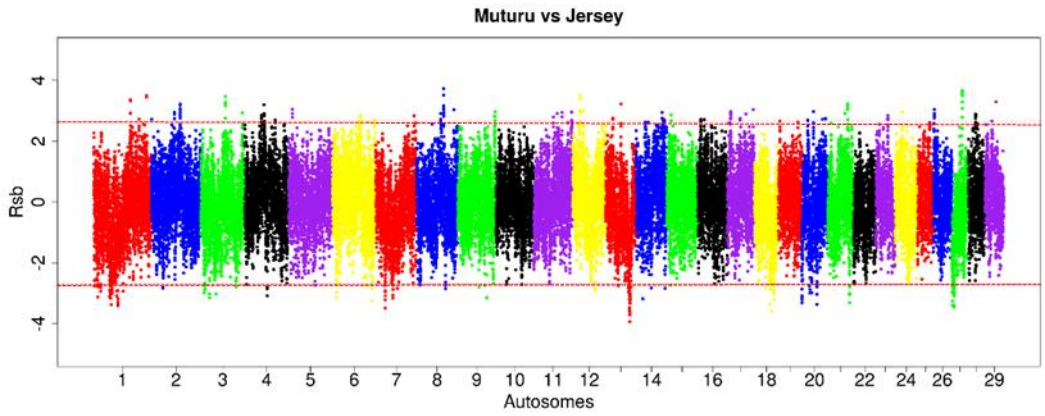


Figure 3.5 | Manhattan plot of the distribution of Rsb values following the comparison of Muturu with Jersey. Red dash lines indicate threshold of selection windows in the top 0.5% *Rsb* scores.

Ninety three, 73 and 66 genes were identified following the comparison of Muturu with N'Dama, Holstein and Jersey respectively (Figure 3.6). Overlaps among the three sets of candidate genes reveal two genes (*TFEC* and *OSBP2*) commonly detected in the three comparison tests. Twenty-one others were detected in two different comparison tests (Table 3.2). Among these, eight genes (*FAM124A*, *PIGK*, *ST6GALNAC5*, *WDR49*, *SERPINI2*, *CD38* and *ENSBTAG00000047597*) were found following comparison with the two European taurine breeds only, suggesting them to be West African taurine specific selected genes. These 23 candidate genes were found in ten candidate selected regions on chromosomes 1, 3, 4, 6, 8, 12, 13 and 17.

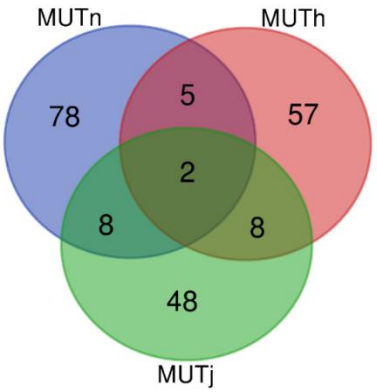


Figure 3.6 | Venn diagram of candidate genes identified in Muturu based on comparison with three *B. taurus* breeds. MUTn – Muturu versus N'Dama, MUTh – Muturu versus Holstein, MUTj – Muturu versus Jersey.

Table 3.2 | List of overlapping candidate genes detected by at least two different Rsb tests following the comparison of Muturu with each of three other taurine breeds (N'Dama, Holstein and Jersey).

Chr	Start (bp)	End (bp)	Gene	Reference populations
1	351708	362907	*RCAN1*	N'Dama, Holstein
1	100676575	100786489	*WDR49*	Holstein, Jersey
1	100791772	100827942	*SERPINI2*	Holstein, Jersey
3	67687802	67824632	*PIGK*	Holstein, Jersey
3	67863093	68073051	*ST6GALNAC5*	Holstein, Jersey
4	52589574	52812807	*TFEC*	N'Dama, Holstein, Jersey
4	52989410	52989499	*ENSBTAG00000045017*	N'Dama, Jersey
6	115802869	115858942	*CD38*	Holstein, Jersey
6	115940219	115943190	*FGFBP1*	N'Dama, Holstein
6	115981183	116105180	*PROM1*	N'Dama, Holstein
8	74834628	74913332	*PPP2R2A*	N'Dama, Jersey
8	74924184	74947549	*BNIP3L*	N'Dama, Jersey
8	75089346	75168436	*DPYSL2*	N'Dama, Jersey
8	75108590	75108708	*ENSBTAG00000029080*	N'Dama, Jersey
12	20704476	20778873	*FAM124A*	Holstein, Jersey
13	42398735	42399402	*NXT1*	N'Dama, Jersey
13	42407301	42413062	*GZF1*	N'Dama, Jersey
13	42416038	42453614	*NAPB*	N'Dama, Jersey
17	51217469	51219592	*ENSBTAG00000047597*	Holstein, Jersey
17	51260767	51262423	*ENSBTAG00000033558*	Holstein, Jersey
17	71750227	71755793	*DUSP18*	N'Dama, Holstein
17	71755943	71760583	*C17H5orf52*	N'Dama, Holstein
17	71771703	71894872	*OSBP2*	N'Dama, Holstein, Jersey

The strongest signal of positive selection in Muturu based on Rsb analysis was observed on BTA 12 (550 Mb region position 20550001 – 21100000, Rsb = 3.95) following comparison with the European taurine breeds. The candidate genes overlapping this region are FAM124A, SERPINE3, WDFY2 and INTS6 (Table 3.3). FAM124A was detected following comparison of Muturu with Holstein and Jersey, while the remaining three genes (SERPINE3, WDFY2 and INTS6) were detected following the comparison with Holstein only. In addition, other identified strong selection signatures in the top 15 *Rsb* scores are presented in Table 3.3. Amongst these, candidate selection regions identified on BTA 2 (*OLA1*), BTA 11 (*GFI1B, SPACA9, GTF3C5, CEL, RALGDS, AK8* and *TSC1*), BTA 24 (*ANKRD29, NPC1, RMC1* and *LAMA3*) and BTA 26 (*KIF20B*) are

the result of *Rsb* analysis between Muturu and N'Dama. Therefore, these genes may provide insights into the genetic differences between the shorthorn and longhorn taurine of West Africa.

Table 3.3 | List of the 15 top selection signatures in Muturu based on Rsb analysis

Chr	Start (bp)	End (bp)	*Rsb*	Genes nearest to the candidate selected region
12	20550001	21100000	3.95	*SERPINE3, FAM124A, WDFY2, INTS6*
27	28450001	29000000	3.90	*RNF122, DUSP26*
26	11450001	11800000	3.77	*KIF20B*
4	52650001	53100000	3.66	*TFEC*
11	102900001	103150000	3.56	*GFI1B, SPACA9, GTF3C5, CEL, RALGDS, AK8, TSC1*
24	33250001	33500000	3.52	*ANKRD29, NPC1, RMC1, LAMA3*
2	22650001	22800000	3.49	*OLA1*
3	67600001	68000000	3.47	*AK5, PIGK, ST6GALNAC5*
8	74850001	75300000	3.46	*PNMA2, BNIP3L, ADRA1A, PPP2R2A, DPYSL2*
6	86650001	86900000	3.37	*UGT2A1*
1	100750001	100900000	3.36	*WDR49, SERPINI2, ZBBX*
1	101000001	101200000	3.33	*ZBBX*
4	68650001	68750000	3.31	*JAZF1*
6	115750001	116050000	3.27	*FGFBP1, CD38, PROM1*
21	54650001	54900000	3.23	*FSCB*

3.3.4 Comparison between selection signatures in Muturu and selection signatures in other *B. taurus* breeds

To enable us to compare the selection signatures in Muturu with the ones in the other *B. taurus* breeds (N'Dama, Holstein and Jersey), we additionally performed selection scans of the genomes of each of the three *B. taurus* populations using both *iHS* and *Rsb* tests as described above. For the latter, each of the three *B. taurus* breeds was also compared to the other three breeds. The distributions of genome-wide iHS and Rsb scores in the three B. taurus breeds are presented in Figures 3.7 – 3.12.

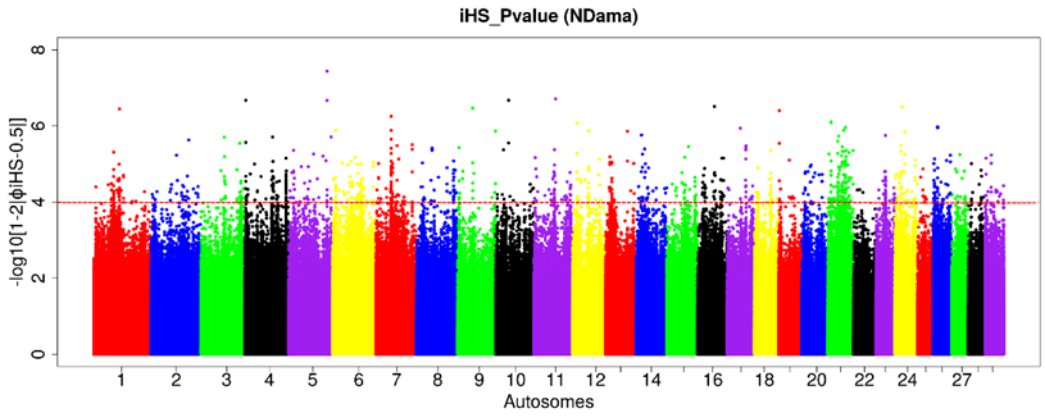


Figure 3.7 | Manhattan plot of the distribution of iHS scores along bovine autosomes in N'Dama cattle. Red dash line indicates threshold of iHS = 4

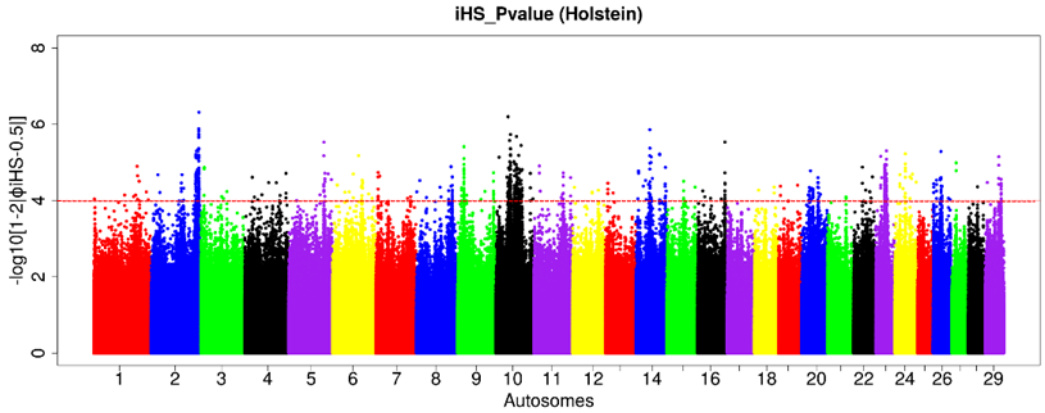


Figure 3.8 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Holstein cattle. Red dash line indicates threshold of iHS = 4

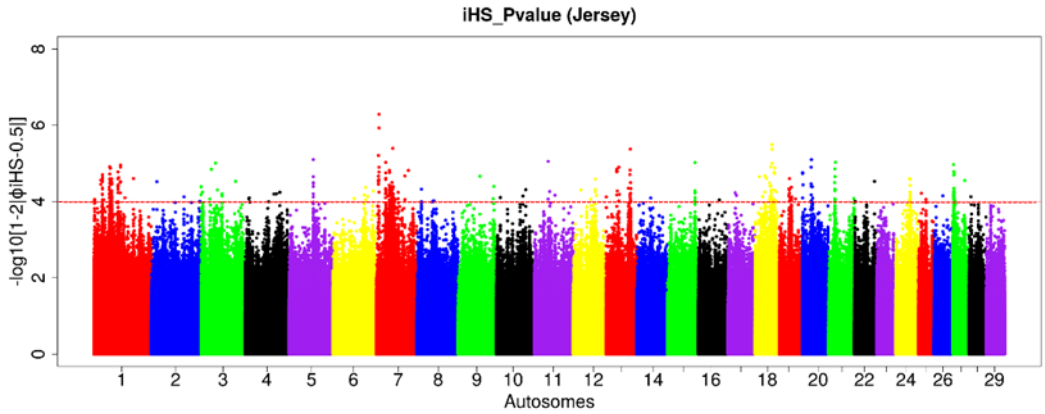


Figure 3.9 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Jersey cattle. Red dash line indicates threshold of iHS = 4

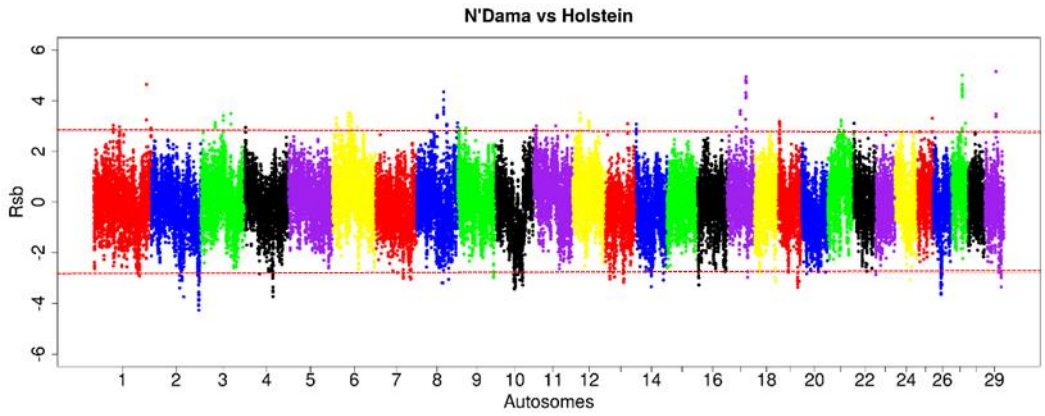


Figure 3.10 | Manhattan plot of the distribution of Rsb values following the comparison of N'Dama with Holstein. Red dash lines indicate threshold of selection windows in the top 0.5% Rsb scores.

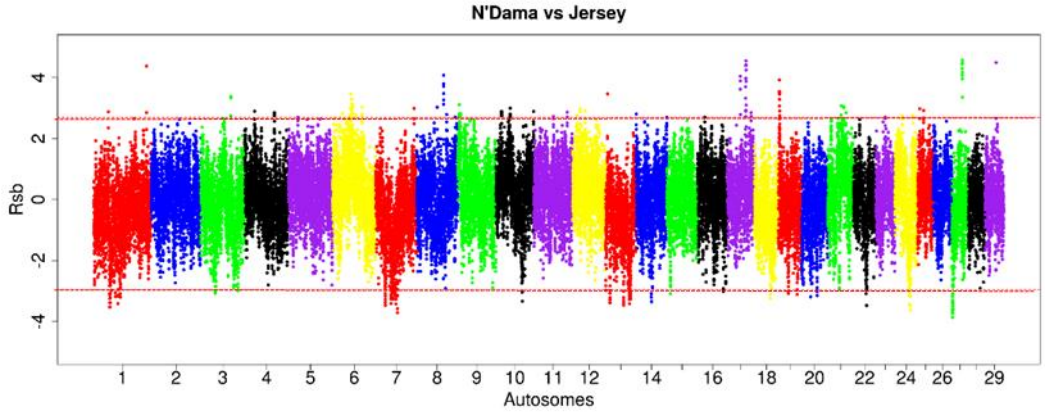


Figure 3.11 | Manhattan plot of the distribution of Rsb values following the comparison of N'Dama with Jersey. Red dash lines indicate threshold of selection windows in the top 0.5% Rsb scores.

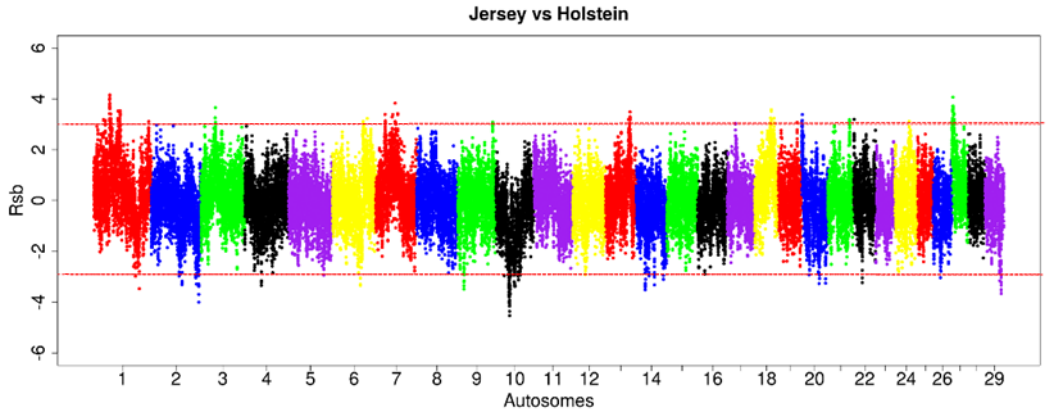


Figure 3.12 | Manhattan plot of the distribution of Rsb values following the comparison of Jersey with Holstein. Red dash lines indicate threshold of selection windows in the top 0.5% Rsb scores.

The numbers of candidate genes identified in N'Dama, Holstein and Jersey based on *iHS* analysis are 809, 557 and 346, respectively (Supplementary Tables S3.3 – S3.5). The total numbers of detected genes based on *Rsb* analyses of each of N'Dama, Holstein and Jersey with three other taurine breeds including Muturu are 123, 337 and 298, respectively (Supplementary Table S3.6 – S3.8). Following the two selection scan tests a total of 902, 761 and 563 distinct candidate genes, excluding duplicates, were identified in N'Dama, Holstein and Jersey respectively, the total was 607 in Muturu (Figure 3.13). The numbers of detected genes associated with fixed selective sweeps (*Rsb* tests) are higher in the European breeds than in Muturu (n = 206) and N'Dama (n = 123). Likewise, higher numbers of overlapped candidate genes between the two selection scan tests are observed in the European taurine than in the West African taurine breeds (Figure 3.13).

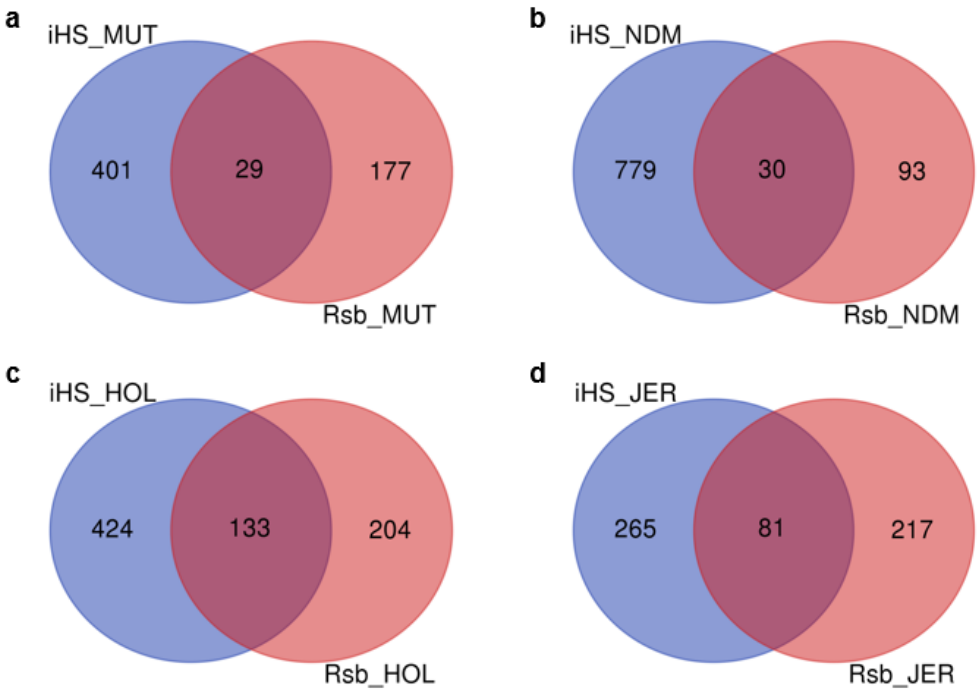


Figure 3.13 | Venn diagram showing the numbers of total and overlapping candidate genes detected in four *B. taurus* based on *iHS* and *Rsb* tests

A Venn diagram showing the numbers of unique and shared candidate genes in each of the four taurine breeds is shown in Figure 3.14. Furthermore, comparison of the total candidate genes sets identified in each of the taurine breeds reveals a moderate overlap between pairs of taurine breeds with no candidate gene common to the four gene sets (Figure 3.14). The numbers of candidate selected genes in Muturu which are also identified in the three other taurine breeds are 44, 29 and 19 in N'Dama, Holstein and Jersey, respectively. The full lists of these genes are presented in Supplementary Tables S3.9 – S3.11, respectively. Among the 44 genes shared with N'Dama is *BLA-DQB,* a member of the MHC class II genes. Additionally, seven candidate genes belonging to members of the bovine major histocompatibility complex (MHC) class II genes including *BOLA-DQA2, LOC100848815, BOLA-DQB, BOLA-DQA5, BOLA-DRB3, ENSBTAG00000003352, ENSBTAG00000000432* were commonly detected in Muturu and N'Dama and also in Holstein. These MHC genes are located on BTA 23 and are among the candidate genes in the top 10 signature of selection regions in Muturu identified by *iHS* test (Table 3.2).

Based on *Rsb* tests, *RNF122, SERPINE3, INTS6, FAM124A, DUSP26, PIGK* and *ST6GALNAC5* genes were commonly detected in Muturu and N'Dama and were also located in the top selection signature regions for Muturu (Table 3.4). One gene, *PPP2R2B*, is shared between Muturu, N'Dama and Jersey (Figure 3.12). We hypothesize that the genes shared with N'Dama could be related to tropical adaptation while those shared with the European taurine may related to production and/or morphological traits. On the other hand, the total of 507 candidate genes detected only in the Muturu breed could be related to the unique phenotypes of Muturu.

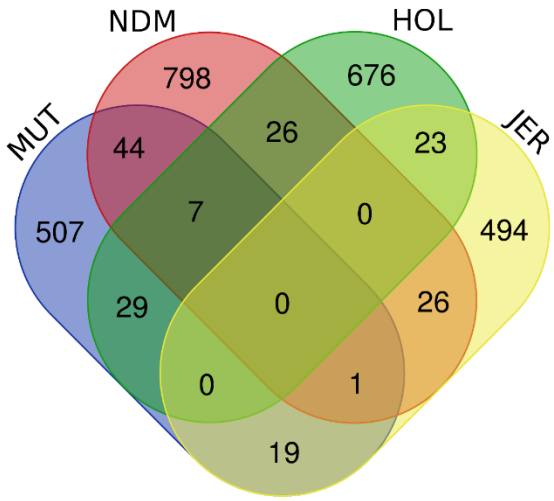


Figure 3.14 | Venn diagram showing unique and shared candidate genes identified in Muturu (MUT), N'Dama (NDM), Holstein (HOL) and Jersey (JER) breeds based on genome-wide iHS and Rsb selection scan tests across breeds.

3.3.5 Functional annotation of candidate genes

To map the genes under selection in Muturu to known biological processes and pathways, we performed gene enrichment (over-representation) analyses using the web-based DAVID database v. 6.8 (Huang *et al.*, 2009, Huang *et al.*, 2009). Four distinct functional annotation analyses were performed: Muturu candidate selected genes shared only with N'Dama (n = 44), Muturu candidate selected genes shared only with N'Dama and Holstein (n = 7), Muturu candidate selected genes shared only with Holstein (n = 29), Muturu candidate selected genes shared only with Jersey (n = 19), Muturu specific candidate selected genes (n = 507) (Figure 3.14).

The annotation of the 44 genes shared with N'Dama indicates the importance of these genes in three KEGG pathways, namely phagosome (*BLA-DQB, ATP6V0D2, COLEC11*), metabolic pathways (*ATP6VOD2, ST6GALNAC5, ADI1, ALLC* and *PIGK*) and olfactory transduction (*LOC787659, LOC507560* and *OR6C76*). The seven MHC class II genes common to Muturu and N'Dama, which are also shared with Holstein,

are involved in immune response (including the phagosome pathway). The twenty-nine Muturu candidate genes shared only with Holstein are involved in the innate immune response, GO:0045087 (*FBXO9, TRIM10*, and *TRIM15*), neuroactive ligand-receptor interaction (*GHR, GHRHR* and *ADCYAP1R1*) and olfactory transduction pathways, while the 19 genes shared only with Jersey are involved in metabolic and arachidonic acid metabolism (*ALOX12, ALOX15*, *ALOX12E and UGT2A1*) pathways.

Functional annotation of the 507 unique candidate genes identified in Muturu revealed roles in several biological processes, including the biological processes already mentioned above such as the olfactory transduction pathways (n = 31), phagosome (n = 5), innate immune response (n = 3), metabolic (n = 24), neuroactive ligand-receptor interaction (n = 5) and arachidonic acid metabolism (n = 3) pathways. We therefore repeated the functional annotation analysis using the entire 607 gene sets found under selection in Muturu. The complete list of functional annotation terms including GO term and KEGG pathways based on the 607 candidate genes is provided in Supplementary Table S3.12. Some randomly selected annotation terms and associated genes, which we believe are relevant to important Muturu phenotypes include immunity/disease resistance (e.g. immune response, innate immune response, Natural killer cell mediated cytotoxicity and Phagosome), heat tolerance (e.g. protein folding), tick resistance (e.g. inflammatory response), feeding behavior (e.g. Glucose homeostasis) and morphological traits (e.g. skeletal muscle fiber development) of Muturu, are presented in Table 3.4.

Table 3.4 | Candidate genes in selected pathways of relevance to important Muturu phenotypes.

S/No	Annotation terms	Gene count	Genes list	phenotypes
1	GO:0007608: Sensory perception of smell	25	*see Supplementary Table S3.12*	4.40E-10
2	BTA04740: Olfactory transduction	44	*see Supplementary Table S3.12*	6.80E-03
3	GO:0006955: Immune response	10	*FAS, JASP.1, BOLA, BLA-DQB, LOC100848815, OTUD7A, BOLA-DQA2, BOLA-DQA5, BOLA-DQB, BOLA-DRB3*	1.10E-01
4	BTA04145: Phagosome	13	*see Supplementary Table S3.12*	1.80E-03
5	BTA04916: Melanogenesis	8	*GNAI1, KIT, WNT1, WNT10A, WNT6, CTNNB1, FZD6, PLCB1*	2.10E-02
6	BTA04390: Hippo signaling pathway	12	*TEAD3, WNT1, WNT10A, WNT6, BMP5, CTNNB1, CRB2, FZD6, PPP2CB, PPP2R2A, PPP2R2B, PPP2R2C*	3.30E-03
7	BTA04310: Wnt signalling pathway	9	*SENP2, WNT1, WNT10A, WNT6, CSNK1A1, CTNNB1, FZD6, PLCB1, RAC1*	3.80E-02
8	BTA04919: Thyroid hormone signalling pathway	7	*ATP1A1, CTNNB1, NOTCH4, NCOA2, PLCB1, RCAN1, STAT1*	9.60E-02
9	BTA05143: African trypanosomiasis	3	*FAS, IDO2, PLCB1*	2.80E-01
10	GO_0006457: Protein folding	5	*DNLZ, FKBP11, FKBP5, PFDN2, PFDN4*	
11	GO_0042593: Glucose homeostasis	3	*ADGRF5, DBH, WFS1*	6.60E-01
12	GO_0045087: Innate immune response	6	*FBXO9, IRF5, TRIM10, TRIM14, TRIM15, TRIM62*	7.10E-01
13	BTA04080: Neuroactive ligand-receptor interaction	8	*ADCYAP1R1, ADRA1A, CHRM3, GABRA5, GABBR2, GRM7, GHR, GHRHR*	7.10E-01
14	BTA04650: Natural killer cell mediated cytotoxicity	3	*FAS, RAC1, VAV2*	8.40E-01
15	BTA01100: Metabolic pathways	33	*See Supplementary Table S3.12*	7.30E-01
16	GO:0006954: Inflammatory response	3	*FAS, KIT, AGER*	9.8E-1
17	GO:0048741: Skeletal muscle fibre development	3	*ACTA1, ACTA2, FAM65B*	9.5E-2

3.3.6 Muturu specific missense SNPs at candidate selected regions

We identified and selected for further analyses the non-synonymous (missense) SNPs within genes in candidate selected regions in Muturu (*iHS* and *Rsb* tests). Missense SNPs are particularly of interest because they lead to amino acid changes in proteins, and as such, polymorphisms may be associated with phenotypes, e.g. disease resistance/susceptibility (Stefl *et al.*, 2013). From 242,910 SNPs identified within the candidate selected regions in Muturu, 917 are missense SNPs (Table 3.5). These variants were detected in 273 genes, 152 of them containing between 2 to 10 missense SNPs, and for seven genes (*RBBP8, TEX15, LAMA3, TRIM40, LOC100848815, NID1* and *OR12D3*), numbers of varied missense SNPs between 11 in *OR12D3* and 32 in *RBBP8* (Supplementary Table S3.13). The *RBBP8* gene in particular contains up 28 novel missense variants, the observed allele frequencies are less than 50% (AF > 0.5) in the Muturu population (n = 10).

One hundred and sixty-three missense variants were found to be fixed (allele frequency = 1) in Muturu (n =10). These variants were therefore compared with other taurine breeds. Amongst these, 18 found in genes including *TFEC, DDN, PROM1, TBC1D2, CCDC187, ALOX15B, BOLA-DQB, RIPOR2* and three uncharacterized genes, are either missing or at frequencies less than 50% in the three other taurine breeds (N'Dama, Holstein and Jersey). 23 missense variants located in genes including *PIGK, TRAF3IP1, GHRHR, TNS2, SERPINE3, BOLA-DQA2*, and *NID1* are at frequencies between 50% and 100% in N'Dama but are either missing or at frequencies less than 50% in the European taurine breed, while Five variants found within two genes (ENSBTAG00000047597 and *BOLA-DQA5*) are at frequencies above 50% in the European taurine but are either at frequencies less 25% or missing in N'Dama. These

variants are list in Table 3.6. Lastly, the remaining 117 missense variants fixed in the Muturu population are also either fixed (100%) or present at frequencies more than 50% in the three other taurine breeds.

Table 3.5 | distribution of SNPs within the identified region under selection in Muturu

Variant type	Count	Percent
Splice acceptor variant	6	0.00%
Splice donor variant	5	0.00%
Stop gained	15	0.01%
Stop lost	1	0.00%
Start lost	2	0.00%
Missense variant	917	0.38%
Splice region variant	144	0.06%
Synonymous variant	1,164	0.48%
Stop retained variant	1	0.00%
Coding sequence variant	6	0.00%
Mature miRNA variant	3	0.00%
5 prime UTR variant	84	0.03%
3 prime UTR variant	358	0.15%
Non-coding transcript exon variant	93	0.04%
Intron variant	48,444	19.94%
Upstream gene variant	9,240	3.80%
Downstream gene variant	7,658	3.15%
Intergenic variant	174,769	71.95%
Total	**242910**	100.00%

Table 3.6 | List of selected genes containing Muturu unique and shared missense variants

Breed/population	Gene	Chr	Start	End	Allele	MUT	NDM	HOL	JER
Muturu and ET	*ENSBTAG00000047597*	17	51217737	51217739	A	1.00	NA	0.45	0.60
	BOLA-DQA5	23	25405216	25405218	C	1.00	0.2	0.625	0.4
	BOLA-DQA5	23	25405259	25405261	A	1.00	0.25	0.66667	0.4
	BOLA-DQA5	23	25405216	25405218	C	1.00	0.2	0.625	0.4
	uncharacterized	23	26354255	26354257	C	1.00	NA	0.85714	0.75
Muturu with N'Dama	*PIGK*	3	67740491	67740493	A	1.00	0.90	0.25	0.35
	TRAF3IP1	3	118280037	118280039	C	1.00	0.85	0.10	0.25
	TRAF3IP1	3	118296231	118296233	T	1.00	0.83	0.25	0.40
	TRAF3IP1	3	118279756	118279758	A	1.00	0.80	0.10	0.20
	TRAF3IP1	3	118279780	118279782	A	1.00	0.80	0.10	0.20
	GHRHR	4	65801935	65801937	C	1.00	1.00	NA	0.05
	TNS2	5	27072835	27072837	G	1.00	1.00	NA	0.10
	TNS2	5	27075001	27075003	C	1.00	1.00	NA	0.10
	TNS2	5	27074317	27074319	C	1.00	1.00	NA	0.05
	uncharacterized	12	77988600	77988600	G	1.00	1.00	NA	NA
	uncharacterized	12	77988609	77988609	A	1.00	1.00	NA	NA
	uncharacterized	12	77988681	77988681	C	1.00	1.00	NA	NA
	uncharacterized	12	77988731	77988733	G	1.00	1.00	0.14	NA
	uncharacterized	12	77988259	77988261	G	1.00	1.00	0.14	NA
	SERPINE3	12	20858123	20858125	C	1.00	0.95	0.25	0.30
	uncharacterized	13	423080	423082	A	1.00	0.70	0.10	NA
	uncharacterized	13	424740	424742	C	1.00	0.65	0.10	NA
	uncharacterized	13	423618	423620	G	1.00	0.5	0.1	NA
	OLFACTORY REC	15	81220648	81220650	G	1.00	0.8	0.45	0.1

	OLFACTORY REC	15	81220574	81220576	T	1.00	0.7	0.45	0.15
	OLFACTORY REC	15	81220232	81220234	G	1.00	0.5	0.38889	0.05556
	OLFACTORY REC	15	81220249	81220251	G	1.00	0.5	0.125	NA
	BOLA-DQA2	23	25351733	25351735	C	1.00	1.00	NA	0.33
	NID1	28	8798180	8798180	C	1.00	1.00	NA	NA
	NID1	28	8798200	8798200	G	1.00	1.00	NA	NA
	NID1	28	8798242	8798242	G	1.00	1.00	NA	NA
Muturu-specific	*TFEC*	4	52746522	52746524	A	1.00	0.20	0.35	0.40
	DDN	5	30986418	30986420	T	1.00	0.35	NA	0.45
	uncharacterized	5	82655125	82655127	G	1.00	0.25	0.10	0.35
	uncharacterized	5	82656132	82656134	T	1.00	0.25	0.10	0.35
	uncharacterized	5	82656195	82656197	G	1.00	0.15	0.10	0.35
	uncharacterized	5	82656274	82656276	A	1.00	0.15	0.10	0.35
	PROM1	6	116025806	116025808	A	1.00	0.25	0.40	0.05
	TBC1D2	8	63797826	63797828	A	1.00	0.10	0.30	NA
	CCDC187	11	103811792	103811794	T	1.00	0.45	0.20	0.17
	uncharacterized	12	74838297	74838297	A	1.00	NA	0.33	NA
	uncharacterized	12	74838323	74838323	C	1.00	NA	0.33	NA
	uncharacterized	12	74838341	74838341	T	1.00	NA	0.33	NA
	uncharacterized	17	35913849	35913849	T	1.00	NA	NA	NA
	ALOX15B	19	28314450	28314452	T	1.00	0.25	0.15	0.05
	BOLA-DQB	23	25375390	25375392	T	1.00	0.33	NA	0.20
	BOLA-DQB	23	25388099	25388099	G	1.00	0.43	NA	NA
	RIPOR2	23	32742809	32742811	A	1.00	0.40	0.25	0.20
	RIPOR2	23	32753726	32753728	C	1.00	0.05	0.25	0.15

3.4 Discussion

Genome-wide selection signatures were assessed in the endangered Muturu cattle in order to provide insights into and possibly unravel the genetic architecture and control of some of its important phenotypes. We have used two complementary EHH-based selection scan tests, within population iHS and between population Rsb. Both tests have substantial statistical power to detect loci under natural selection even when sample sizes are small (Pickrell *et al.,* 2009). In particular, the iHS test has the power to detect partial selective sweeps within a population, while Rsb can detect selected alleles that have risen to near-fixation or to fixation in a specific population but remain polymorphic in the other population (Biswas and Akey 2006, Voight *et al.,* 2006). To enable us to characterize the genomic signatures of Muturu, we have also carried out comparison with regions identified in three other taurine breeds following the same procedures.

3.4.1 Selection signatures shared with the European taurine breeds

Although, Muturu cattle and the two European breeds are both shorthorn taurine, they are believed to have diverged thousands of year ago following their domestication from the wild auroch at the Near East domestication center and distinct geographic dispersion (Loftus *et al.,* 1999). The breeds have evolved differently, following different demographic histories and human and environmental selection pressures. In particular, the European breeds have undergone strong human-mediated selection pressures particular for production traits such as dairy production. On the other hand, Muturu, a local breed from Africa have been subjected to natural selection pressures as a result of the tropical environments, in particular environmental selection in tsetse infected areas.

Nevertheless, some candidate selected genes were identified both in Muturu and each of the European breeds but there was no overlap identified between Muturu and both European breeds (Figure 3.12).

The majority of the 29 genes shared between Muturu and Holstein (Supplementary Table S3.10) belong to the olfactory receptor genes family, other genes include *BOLA, TRIM10, TRIM15, MAP3K5, GCM1, MAP7, CHAF1B, GHRHR* and *GHR*. The BOLA and TRIMs genes are known to have related function in adaptive and innate immune response respectively. Sequence variation at the Growth Hormone Releasing Hormone Receptor (*GHRHR*) and the Growth Hormone Receptor (*GHR*) have been reported to affect stature in humans (Goddard *et al.,* 1995, Ko *et al.,* 2009, Inoue *et al.,* 2011, Dauber *et al.,* 2014). Polymorphism within GHRHR gene have also been associated with growth traits in livestock in general, such as cattle and goat (Curi *et al.,* 2005, An *et al.,* 2011, Liu *et al.,* 2011). To the best of our knowledge, the *GHR* gene has been reported so far to be associated with milk production but not bovine stature in cattle (Blott *et al.,* 2003, Viitala *et al.,* 2006, Bouwman *et al.,* 2018). However, variants in this genes have been demonstrated to affect body size in other mammalian species, such as in the case of the Laron syndrome in humans (Cui *et al.,* 2015), in the miniature pig (Tian *et al.,* 2014) and in dog breeds (Rimbault *et al.,* 2013, Plassais *et al.,* 2017). Muturu cattle are particularly known for their dwarf stature which may explain the selection signals observed in our study at the *GHR* and *GHRHR* genes. In contrast, the selection signals at the same genes in Holstein might be as a consequence of selection for milk trait rather than stature. This may be further investigated by looking at and comparing the haplotype structure in the two breeds. Interestingly, a fixed non-synonymous T > C mutation at rs109390134 (exon 6) of the *GHRHR* genes was

identified in Muturu cattle as well as in N'Dama, with the latter breed also of smaller size compared to the Holstein, with the allele being absent in Holstein.

3.4.2 Selection signatures shared with the African taurine N'Dama

Muturu and N'Dama are two cattle breeds indigenous to West Africa. They are known for their resistance to disease (trypanosomiasis) and they are adapted to the humid and sub-humid tropical environments. Their trypanotolerance phenotype has been linked to their innate capacity to control anaemia and level of parasitemia (Murray and Morrison 1978), processes involving their innate and adaptive immune responses and hematopoietic control**.** We expected that the shared genomic signatures of these two African breeds would point towards their common distinct phenotypes.

Among the candidate genes under selection in the two west African taurine breeds are members of the MHC class II genes, although some of these genes are also under selection in the Holstein breed. In cattle and other ruminant species, the MHC region is highly polymorphic and is a major source of variation in immune responses. For instance, polymorphisms in *BOLA-DRB3* has been linked to resistance to bovine leukaemia virus (BLV) infection (Xu *et al.,* 1993, Amills *et al.,* 1998). Some other previous studies have linked it with tolerance/resistance to several cattle diseases such as mastitis, tick-borne infections (Sharif *et al.,* 1998, Martinez *et al.,* 2006, Kulberg *et al.,* 2007, Rupp *et al.,* 2007, Duangjinda *et al.,* 2013). Evolutionary factors such as positive selection, genetic drift and population bottlenecks may have also played a role in the extensive genetic diversity at the MHC loci (Mikko and Anderson 1995, Amills *et al.,* 1998, Lewin *et al.,* 1999). We also identified a fixed non-synonymous mutation

T > C mutation at rs208515389 (exon 3) of the *BOLA-DQA2* gene in both Muturu and N'Dama, this allele was absent in the Holstein cattle and at a low frequency (AF = 0.3) in Jersey. We may therefore hypothesize that a trait like the trypanotolerant trait of Muturu and N'Dama may be associated with this variation within the MHC gene region. Further genetic variation studies of the MHC gene region between trypanotolerant and susceptible cattle populations may provide further insight into the contribution of this region to trypanotolerance in the West African *B. taurus* population. A detailed investigation of the genetic architecture of trypanotolerance in West African Taurine population is presented in Chapter 4 of this thesis.

Considering the environmental challenges which may have shaped the genome of West African taurine, a gene of interest identified in a candidate selected region in both N'Dama and Muturu is *INTS6.* This gene is upregulated in peripheral blood leukocytes of cattle and buffalo exposed to heat stress (Kolli 2012) and following exposure to UV irradiation, with a switch from the expression of long mRNA isoforms to short isoforms and preferential use of short alternative last exon (ALE) transcript isoforms after exposure to UV irradiation (Williamson *et al.,* 2017). The signature of selection signal in the genomic region including *INTS6* may therefore be a consequence of enviromental climatic pressures such as heat (high temperature) and UV radiation in these two taurine cattle living in tropical areas.

The *DUSP26* gene was also identified in a selection region in both West African taurine breeds; the gene has been identified among highly expressed genes in a resistant group of goats exposed to gastrointestinal nematode (GIN) infections (Bhuiyan *et al.,* 2017).

3.4.3 Selection signatures unique to Muturu cattle.

A total of 507 genes representing approximately 83% of the identified candidate genes in Muturu are not shared with the other three *B. taurus* breeds studied. A similar pattern of high numbers of unique candidate genes is observed in the other breeds (Figure 3.12). The high proportions of unique candidate genes may be indicative of varying selection pressures encountered by the different breeds and/or different demographic histories. Among the Muturu unique candidate genes, *TFEC* and *OSBPS* genes were the only two genes identified in the *Rsb* comparisons of Muturu with each of the other breeds (Figure 3.4).

Of particular interest is the transcription factor EC (TFEC); this gene is also found within a candidate selected region detected by the *iHS* test and it contains a missense SNP (G>A, gCt/gTt) that is fixed in Muturu (Table 3.6). *TFEC* is a member of the MIT-TFE gene family of transcription factors, which also includes *MITF, TFE3* and *TFEB* (Hemesath *et al.,* 1994). TFEC is mainly expressed in cells of myeloid origin, MITF is predominantly expressed in melanocytes, osteoclasts, mast cells, macrophages, NK cell, B cells and heart, while *TFE3* and *TFEB* are thought to be more-ubiquitously expressed (Rehli *et al.,* 1999, Martina *et al.,* 2014). Homo-dimerization and hetero-dimerization within members of the MITF/TFE family are critical for binding to DNA and transcriptional activation of the target genes (Martina *et al.,* 2014). The microphthalmia-associated transcription factor (*MITF*) in particular has been shown to be involved in the development of melanocytes and melanoma and has been shown to regulate transcription of a broad range of genes, ranging from genes important for pigment production to genes involved in cell cycle regulation, migration and survival.

When *TFEC* acts like an isoform of *MITF* it is involved in retinal pigment epithelium (RPE) differentiation (Hemesath *et al.,* 1994, Levy *et al.,* 2006, Bharti *et al.,* 2012, Raviv *et al.,* 2014).

Another unique Muturu candidate gene is the *KIT* gene (*iHS* = 4.5), which also interact with the *MITF* gene and plays important roles in melanogenesis (Besmer *et al.,* 1993). While the *TFEC* gene in particular has not been related to coat colour in cattle, it may have a role in agouti coat color in mutant mice (Steingrímsson *et al.,* 2002). However, selection signatures spanning the regions of *KIT* and *MITF* genes have been associated with coat color variation in several animals species including pigs (Rubin *et al.,* 2012), sheep (Fariello *et al.,* 2014), goats (Benjelloun *et al.,* 2015, Wang *et al.,* 2016, Guo *et al.,* 2018), horses (Hauswirth *et al.,* 2012, Hauswirth *et al.,* 2013) and cattle populations (Fontanesi *et al.,* 2010, Hayes *et al.,* 2010, Whitacre 2014, Kim *et al.,* 2017). Nevertheless, positive selection signature in the regions overlapping *TFEC* and *KIT* genes in Muturu may also have influence on the different coat color patterns observed in the Nigerian Muturu (Adebambo 2001).

Interestingly, three other unique Muturu candidate genes, namely *FAS, IDO2* and *PLCB1* are involved in the African trypanosomiasis pathway as revealed by DAVID (Supplementary Table S3.13). These genes may have important roles in the tolerance of Muturu to trypanosome infection, they represent interesting candidates genes for further investigation. Other identified unique Muturu candidate genes include *ZRANB3* and *MAP3K5,* related to feed efficiency/residual feed intake in cattle and pigs (Bovine HapMap Consortium, 2009, Do *et al.,* 2014, Taye *et al.,* 2017), *ABCG2* and *LAP3,* related to milk production (Blott *et al.,* 2003, Cohen-Zinder *et al.,* 2005, Ron *et al.,*

2006, Flori *et al.,* 2009), *RFX2, SRY, LAP3* and *GPX5* genes, related to reproduction and fertility (Horvath *et al.,* 2004, Szreder and Zwierzchowski 2004, Chabory *et al.,* 2010, Mishra *et al.,* 2013, Braud *et al.,* 2017), *ACTA1* and *CTNNB1*, related to heat stress (Kecskés *et al.*, 2015, Guo *et al.,* 2016), and the *PRKAG3* gene, related to meat quality traits in domestic cattle and pigs (Reardon *et al.,* 2010, Ryan *et al.,* 2012, Zhang *et al.,* 2015, Bongiorni *et al.,* 2016). The latter gene has also recently been reported as a candidate domestication gene due to its extensive miRNA binding site polymorphisms between *B. taurus* and *B. primigenius* (Braud *et al.,* 2017).

4.3.4 Selection signatures as a result of population divergence between Muturu and N'Dama

Seventy eight genes out of the total 93 candidate genes were identified following the comparison between Muturu and N'Dama and not in the comparison of Muturu with the European taurine breeds (Figure 3.6). These genes may provide a glimpse into the breed divergence between West African shorthorn and longhorn taurine population. Among them are *GFI1B*, *NPC1* and *OLA1* genes.

GFI1B is essential for neutrophil differentiation and is required for the development of both erythroid and megakaryocytic lineages. Congenital mutations of the gene in humans and mouse have been related to abnormal platelet function and a decrease in erythropoiesis of embryonic stem cells, respectively (Karsunky *et al.,* 2002, Saleque *et al.,* 2002, Anguita *et al.,* 2016). *NPC1* has been reported to play an important role in subcellular lipid transport, homeostasis, platelet function and formation. Louwette and collaborators reported that an *NPC1* defect resulted in abnormal platelet formation and function, which implies that the *NPC1* gene may play a role in platelet function and

formation (Louwette *et al.,* 2012). Other studies have reported the roles of the *NPC1* gene in the control of appetite in mice (Xie *et al.,* 1999), obesity in humans (Meyre *et al.,* 2009, Mejía-Benítez *et al.,* 2013) and possible regulation of growth and body development in cattle (Dang *et al.,* 2014). *OLA1* (Obg-like ATPase 1) has important roles in the regulation of cellular stress responses such as oxidative stress (Zhang *et al.,* 2009) and improved thermal resistance in mammalian cells (Mao *et al.,* 2013). The latter role is mediated by the gene involvement in the stabilization and subsequent upregulation of the heat shock *HSP70* gene (Mao *et al.,* 2013).

The above mentioned important functions of *GFI1B, NPC1* and *OLA1* genes could be indicative of their roles in critical adaptation traits of Muturu to the tropics. The influence of *GFI1B* and NPC1 genes on erythropoiesis and feed intake, respectively, could be related to the ability of Muturu to tolorate trypanosome infection while the role of *OLA1* gene on thermotolerance in Muturu is also likely. However, both Muturu and N'Dama are known for their tolerance to trypanosomiasis and they both have intrinsic thermotolerance, thus, it is likely that they share common genetic control of these local adaptation traits. Nevertheless, it is envisaged that individual breed factors could also contribute to their adaptative phenotypes. Therefore, these genes may represent examples of genes contributing to common phenotypes but under positive selection in one breed and not the other.

CHAPTER 4:

Genome-wide sequences analysis of West African taurine reveal their unique trypanotolerant adaptations

4.1 Introduction

Indigenous West African *Bos taurus* cattle (WAT) comprise the longhorn taurine (N'Dama, Kuri) and shorthorn population, e.g. Lagune, Baoule, Borgou, Somba and Muturu (Rege *et al.,* 1994). They likely migrated from the centre(s) of cattle domestication in the Near East and first appeared on the African continent approximately 7,000 years ago (Lott and Hart 1979, Loftus *et al.,* 1994). It is unknown if the shorthorn and longhorn taurine have two distinct origins or if one of them originated from the other. They are uniquely adapted to sub-humid and humid West African environments which are often infested with tsetse fly *Glossina sp*, the intermediate host of the *Trypanosoma sp.* parasites and the causative agent of African trypanosomosis. The main *Trypanosoma sp*. known to infect livestock species include *T. congolense, T. vivax* and *T. brucei brucei* (Murray *et al.,* 1982). In cattle, trypanosomiasis, also called nagana, is a devastating disease characterized by severe anaemia, weight loss, infertility and abortion. It may be lethal during the acute phase especially if proper veterinary interventions are not provided (Nantulya 1990). The disease causes severe economic losses to the livestock sector on the African continent (Kristjanson *et al.,* 1999, Naessens 2006).

Indigenous West African *Bos taurus* are well known for their natural tolerance to trypanosomiasis infection in contrast to non-African *B. taurus* and *B. indicus* that are highly susceptible to the disease. Their trypanotolerance have been linked to their innate ability to control parasitemia and anaemia, to mount an adequate immune response and to maintain body weight. Compared to other cattle breeds, they are able to remain productive even after trypanosome infection (Murray and Morrison 1978, Murray *et al.,* 1982). The origin of the trypanotolerance characteristics of the WAT remains

speculative. It may have followed the arrival and subsequent selection of trypanosusceptible cattle in tsetse infected areas and/or introgression from trypanotolerant African auroch animals.

Amongst the two recognized West African longhorn, namely the N'Dama and the Kuri, only the former is reported as trypanotolerant. The N'Dama inhabits the forest-savannah region from Guinea, Senegal, Sierra Leone, southern Mali and the northwest of Côte d'Ivoire, spreading in pockets over the whole Sudano-Guinean savannah belt, as far as Nigeria and Central-West African countries**.** West African Shorthorn cattle are divided into two groups, according to their geographic origins and phenotypes, the Savannah Shorthorn (Savannah West African Shorthorn) and the Dwarf Shorthorn (Dwarf West African Shorthorn). The latter includes the trypanotolerant Muturu cattle, a generic Hausa word for humpless used for all indigenous taurine shorthorns in English-speaking countries (Nigeria, Ghana and Liberia) (Epstein 1971). The genetic basis of trypanotolerance in the West African shorthorn *B. taurus* have been poorly studied, the exception being the Baoule cattle in Burkina Faso (Aboagye *et al.,* 1994, Smetko *et al.,* 2015), compared to the longhorn N'Dama. No studies have attempted so far to compare the genetic control of trypanotolerance between longhorn and shorthorn West African taurine.

In N'Dama cattle, Hanotte *et al.* (2003) mapped QTL associated with trypanotolerance on 18 different cattle autosomes in an F2 crosses of trypanotolerant N'Dama and trypanosusceptible Boran cattle. They showed independent genetic control for parasitemia, PCV and body weight, with most QTLs with minor effect; the major QTLs were located on chromosomes 2, 4, 7, 16 and 27 (Hanotte *et al.,* 2003). Subsequently,

expression analyses in blood cells reported pathways and genes differentially regulated in trypanotolerant N'Dama and trypanosusceptible Kenyan Boran (O'Gorman *et al.,* 2009, Noyes *et al.,* 2011). Genes such as *TICAM1, ARHGAP15, SLC40A1, GFM1* and *INHBA* have been proposed as candidate trypanotolerant genes (Dayo *et al.,* 2009, O'Gorman *et al.,* 2009, Noyes *et al.,* 2011). More recently, full genome sequence analysis reported several candidate genome regions under positive selection in N'Dama cattle including genes with functions related to immunity, anaemia and feeding behaviours that may be linked to the trypanotolerance phenotypes (Kim *et al.,* 2017, Taye *et al.,* 2017). Whether the genome signatures and innate mechanisms controlling trypanotolerance in N'Dama are unique or shared with other trypanotolerance cattle breeds remain unknown.

In this chapter, we report for the first time using whole genome re-sequencing data signatures of positive selection within and between the two groups of trypanotolerant West African taurine populations, the shorthorn Muturu from Nigeria and the longhorn N'Dama from Guinea. Analyses were also performed between these trypanotolerant breeds and two groups of trypanosusceptible cattle, African *B. indicus* and European *B. taurus*. The study involves the use of multiple selection scan tests including integrated haplotype score (*iHS*) (Voight *et al.,* 2006), *Rsb* (Tang *et al.,* 2007), pooled heterozygosity (*Hp*) (Rubin *et al.,* 2010), cross-population extended haplotype homozygosity (XP-EHH) (Sabeti *et al.,* 2007), cross-population composite likelihood ratio (XP-CLR) (Chen *et al.,* 2010) and population differentiation (F_{ST}) (Weir and Cockerham 1984). We aim to identify candidate trypanotolerant genes and pathways across West African taurine population, their haplotypes and polymorphisms compared to trypanosusceptible breeds.

4.2 Materials and Methods

4.2.1 Cattle populations and genome-wide SNP data

Seventy-six cattle samples representing 8 different cattle populations were included in this study. They were categorized into three groups, namely (i) West African taurine (WAT), including Muturu (MUT, $n = 10$ sampled in Nigeria), and N'Dama (NDA, $n = 10$, sampled in the Fouta Djallon of Guinea), (ii) African zebu (AFZ), including Baggara (BGR, $n = 9$, Sudan), Gash (GSH, $n = 8$, Sudan), Kenyan Boran (BOR, $n = 10$, Kenya) and Ogaden (OGD, $n = 9$, Ethiopia), and (iii) European taurine (ET) including Holstein (HOL, $n = 10$) and Jersey (JER, $n = 10$). See Materials and Methods section in Chapter 2 for further details.

A total of approximately 35 million bi-allelic autosomal SNPs were used for analyses corresponding to the SNPs called on the 76 samples following joint genotyping and quality controls step (see Chapter 2, Materials and Methods section).

4.2.2 Selective sweep analyses

We used six selection scan approaches to investigate the genomic signatures in Muturu and N'Dama, two within-population tests (*Hp* and *iHS*) and three between-population comparison tests (*Rsb*, XPEHH, XPCLR and F_{ST}).

4.2.2.1 Pooled heterozygosity (*Hp*) approach

The *Hp* analysis involves counting of reads with the most and least abundantly observed alleles at every SNP position, within a window size and a sliding step size (Rubin *et al.*, 2010). The autosomal SNPs in each of Muturu and N'Dama were analysed separately within a window size of 500 kb and a sliding step of 100 kb. The distribution of *Hp*

values were then normalized by transforming *Hp* to Z-scores (*ZHp*) using (*ZHp* = (*Hp* - *μHp*)/*σHp*) (Rubin *et al.,* 2010, Rubin *et al.,* 2012, Axelsson *et al.,* 2013). Windows with less than 20 SNPs were excluded from the analysis, the ZHp values for all windows were ranked and the 0.5% ZHp extreme negative values were considered.

4.2.2.2 Integrated haplotype score (*iHS*) tests

See chapter 3.

4.2.2.3 *Rsb* tests

See chapter 3.

4.2.2.4 Cross-population extended haplotype homozygosity (XPEHH)

The XPEHH statistic (Sabeti *et al.*, 2007) is based on similar assumptions as *Rsb* and it involves contrasting the *iES* statistic between populations. For XPEHH analyses, haplotype phasing was inferred simultaneously on all detected bi-allelic SNPs in the eight cattle populations using BEAGLE (Browning and Browning 2007). Then, the phased genotypes were used for pairwise comparisons between four population sets: Muturu (MUT), N'Dama (NDM), European taurine (ET) and African zebu (AFZ). ET included Holstein and Jersey cattle samples, while AFZ comprised of Boran, Baggara, Gash and Ogaden zebu samples. Thus, the following four pairwise comparisons tests were performed: (i) MUT *versus* AFZ, (ii) MUT *versus* ET, (iii) NDM *versus* AFZ and, (iv) NDM *versus* ET.

XPEHH statistics were calculated for all SNPs in both directions using the default parameters of the Selscan software (Szpiech and Hernandez 2014). As in the *Rsb* statistic, extreme positive XPEHH values suggest positive selection in the target population (in our case Muturu and N'Dama), whereas negative values indicate selection in the reference population (Pickrell *et al.,* 2009). The raw positive XPEHH values were considered and summarized by dividing the genome into consecutive, non-overlapping 50 kb windows. The SNP with the maximum XPEHH value in each window was considered for the summary statistics for that window. To consider variation in SNP density in windows, we computed the empirical *P*-values for XPEHH summary statistics (Pickrell *et al.,* 2009, Granka *et al.,* 2012). Accordingly, empirical *P*-values for each of the summary statistics were computed by clustering the windows into ten bins of equal number of SNPs. Within each bin and for each window *i*, the fraction of windows with a value of the statistic greater than that in *i* is defined as the empirical *P*-value following previously reported methods (Pickrell *et al.,* 2009, Granka *et al.,* 2012).

4.2.2.5 Cross-population composite likelihood ratio (XPCLR)

The XP-CLR is a likelihood method that is based on multi-locus allele frequency differentiation between two populations. It increases resolution by enhancing the signals and localization of the position of a selected allele (Chen *et al.,* 2010). The XPCLR 1.0 software (https://reich.hms.harvard.edu/software) (last accessed January 2018) was used to calculate XPCLR scores within a non-overlapping window of 50 kb and a down-weighted correlation level from which SNPs contribute to the XPCLR score set to 0.95 (Chen *et al.,* 2010). A maximum of 600 SNPs was used within each

window in each of the comparison tests. Four pairwise comparison tests were performed as in XPEHH (see 4.2.2.4).

4.2.2.6 Population differentiation (F_{ST})

VCFtools *version 0.1.14* (Danecek *et al.,* 2011) was used to estimate sequence differentiation, F_{ST}, between the combined N'Dama and Muturu population and all the other cattle populations combined. F_{ST} values were estimated on non-overlapping 50 kb windows across the autosomes and the estimated weighted F_{ST} values were then *Z*-transformed. The weighted F_{ST} value is derived from the ratio of the separate per-locus sums of numerator and denominator values and is based on the the Weir and Cockerham 1984 estimator (Weir and Cockerham 1984).

4.2.3 Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway annotation and protein-protein interaction analysis

The 0.5% extreme window values were considered as candidate selection signature signals in all tests. Bedtools *version 2.25.0* (Quinlan and Hall 2010) was used to merge the overlapping selected windows. Cattle genes which overlapped genomic windows passing the significant selective sweep threshold were retrieved from *Ensembl Genes 92 database* using *Ensembl BioMart* online tool (http://www.ensembl.org/biomart) (Smedley *et al.,* 2009). The candidate genes were then processed in web-based Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (http://david.abcc.ncifcrf.gov/) (Huang *et al.,* 2009, Huang *et al.,* 2009) for functional annotation and identification of over-represented genes involved in biological process (GO) and KEGG pathways. The web-based PANTHER Classification System

(Dorrington *et al.,* 2013) was also used to map the candidate genes to known biological processes or pathways.

4.2.4 Haplotype diversity of candidate region

We investigated haplotype diversity within trypanotolerant candidate gene regions for evidence of distinct haplotypes that may be unique to the WAT. We used hapFLK software (Fariello *et al.,* 2013) and the scripts provided on the hapFLK webpage (https://forge-dga.jouy.inra.fr/projects/hapflk) for the estimation of haplotypes frequencies and visualization of haplotype clusters in the regions of selected candidate genes (Fariello *et al.,* 2013).

4.2.5 Fine mapping of candidate mutation within candidate selection regions

Variation within selected candidate genes was further investigated to identify unique variant within the WAT population which may have been selected for in relation to the trypanotolerant phenotypes of the breeds. To achieve this, we obtained the genotype files for BTA 5 and BTA 23 of over 100 worldwide cattle breeds, comprising more than 2,600 worldwide *Bos taurus* and *Bos indicus* cattle samples included in the 1000 Bull genomes project (Daetwyler *et al.,* 2013) Run 6. From these samples, we obtained four additional WAT cattle samples, Lagune (n = 1), Somba (n = 1) and N'Dama (n = 2), leaving us with a total of 24 trypanotolerant cattle samples for further analyses.

VCFtools *v. 1.14* (Danecek *et al.,* 2011) and the Integrated Selection of Allele Favored by Evolution (iSAFE) software (Akbari *et al.,* 2018) were used for the estimation of per-SNP F_{ST} values and SAFE scores at the candidate selected genes. In addition, iSAFE scores for polymorphisms within an expanded area of up to 0.5 Mb around a selective

sweep region were estimated. For all three estimates, the 24 WAT samples were grouped as a single population and compared to the over 2,600 non-WAT taurine worldwide cattle samples.

4.3 Results

4.3.1 Genome selection signatures in trypanotolerant Muturu and N'Dama

To investigate the candidate signatures for positive selection in our two trypanotolerant West African *B. taurus* cattle breeds (Muturu and N'Dama), we combined the results of our five selection scan methods (*iHS*, *Hp*, *Rsb*, XPEHH and XPCLR). When a test involved comparison of one population or group with several others, the significant regions were added to give the total number of significant regions for that particular population or group. For instance, the significant genome regions obtained after comparison of Muturu with the African zebu were merged with the significant regions obtained following the use of the same selection test for the comparison with the European taurines.

Following these approaches, the number of candidate selected regions in Muturu ranges from 47 (*ZHp*) to 432 (XPCLR), while it ranges from 45 (*ZHp*) to 507 (*iHS*) in N'Dama (Table 4.1). A total of 1,283 (Muturu) and 1500 (N'Dama) genes are annotated in the candidate regions (Figures 4.1a, b). The lists of these regions and the associated candidate genes for each test are provided as Supplementary Tables S4.1 – S4.6, with the exception of the results for *iHS* and *Rsb* for which they are provided in Chapter 3 (Supplemenatry Tables S3.1 – S 3.4). The genomic distributions (Manhattan plots) of the selected regions based on the different methods are shown in Figures 4.2 – 4.7.

Table 4.1 | Number of candidate selected regions and genes under positive selection in Muturu and N'Dama.

Selection scan tests	Number of windows analysed	Selection window criteria	Muturu			N'Dama		
			CR	TGR	GDR	CR	OCG	GDR
iHS	50,000	$P < 0.0001$	266	430	97	507	809	185
ZHp	25,000	lowest 0.5%	47	286	6	45	366	5
Rsb	50,000	top 0.5%	145	206	64	114	123	50
XPEHH	50,000	top 0.5%	324	346	129	465	359	212
XPCLR	50,000	top 0.5%	432	353	193	400	273	204

Note: CR – number of candidate regions**, TGR –** total number of genes within regions**, GDR –** Gene desert regions, **OCG** – Overlapping Candidate Genes

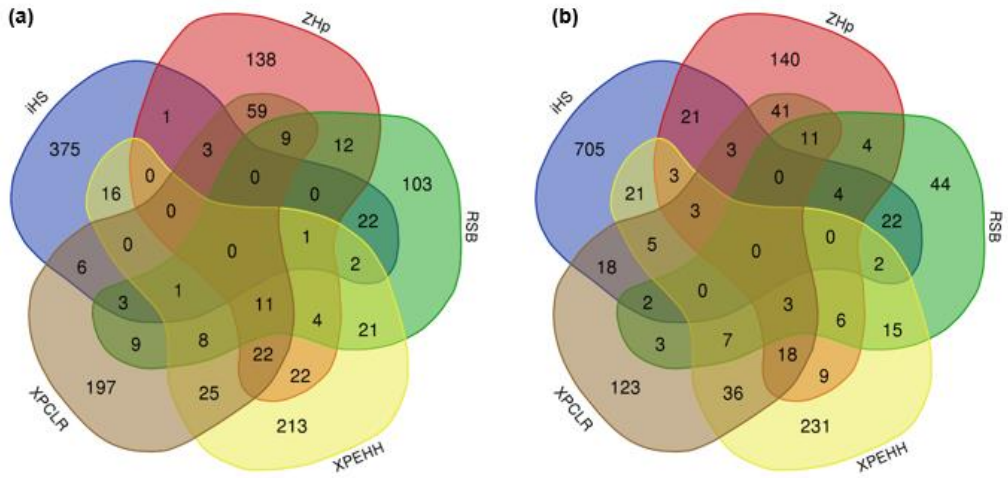


Figure 4.1 | Venn diagrams of the numbers of overlapped candidate genes detected by various selection scan tests in (a) Muturu, and (b) N'Dama.

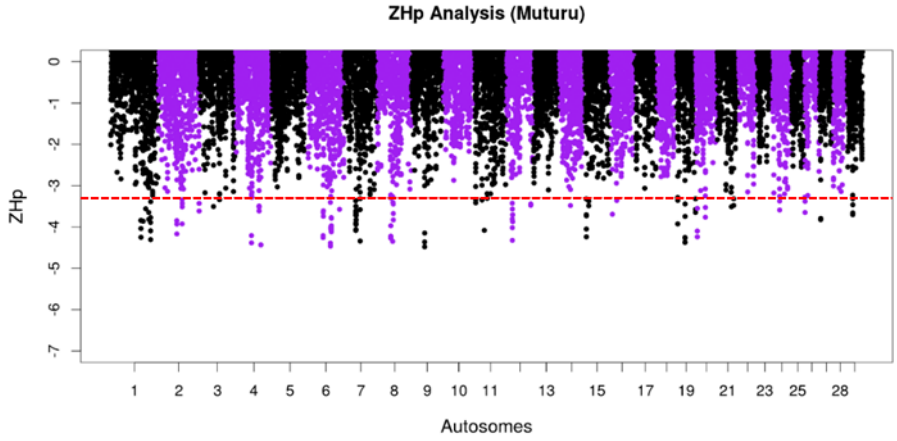


Figure 4.2 | Manhattan plots of the distribution of ZHp test scores along *B. taurus* autosomes in Muturu.

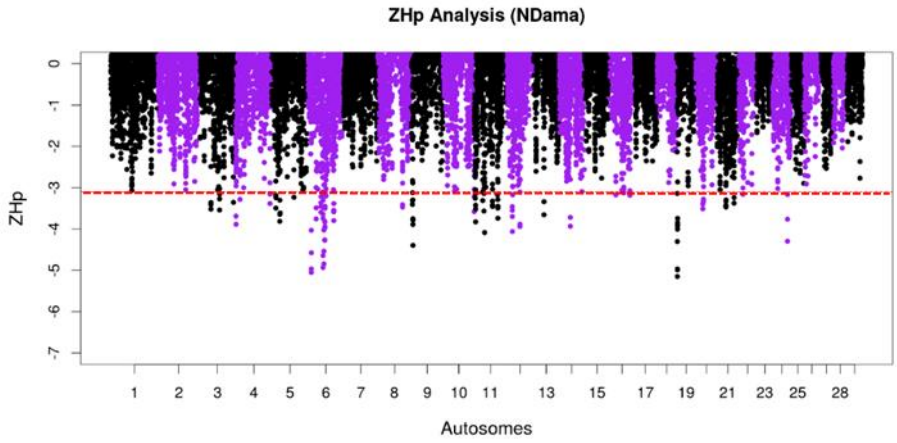


Figure 4.3 | Manhattan plots of the distribution of *ZHp* test scores along *B. taurus* autosomes in N'Dama.

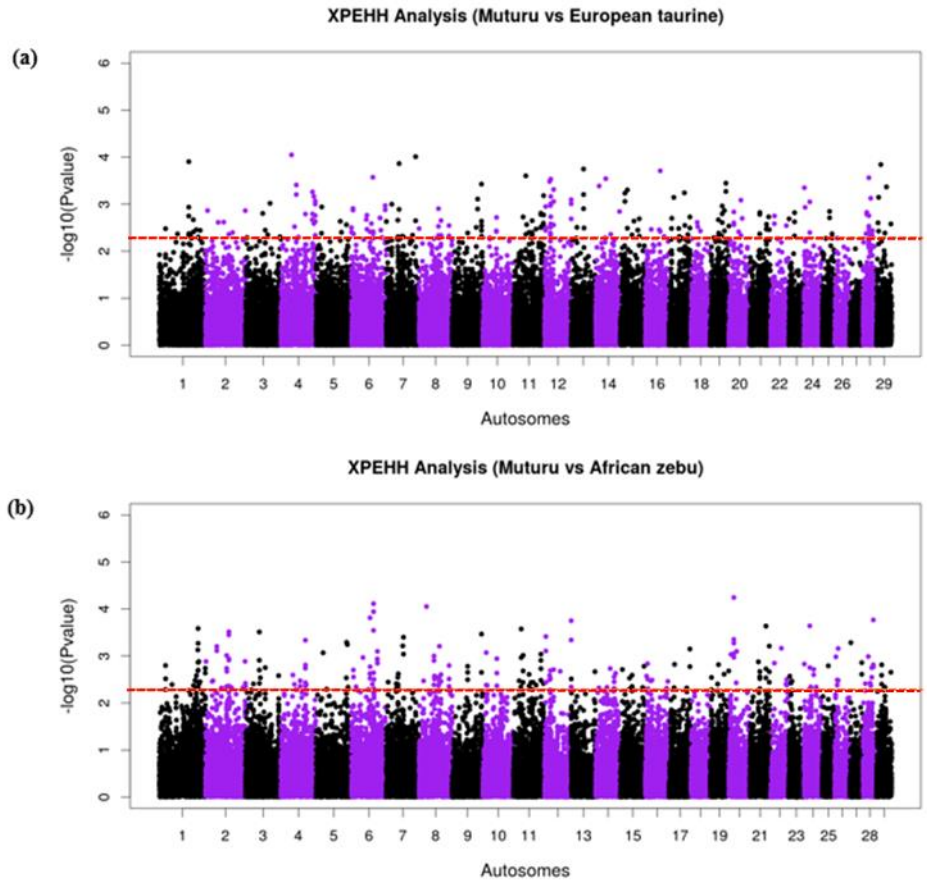


Figure 4.4 | Manhattan plots of the distribution of XPEHH scores along *B. taurus* autosomes following comparison of Muturu with (a) European taurine and (b) African zebu.

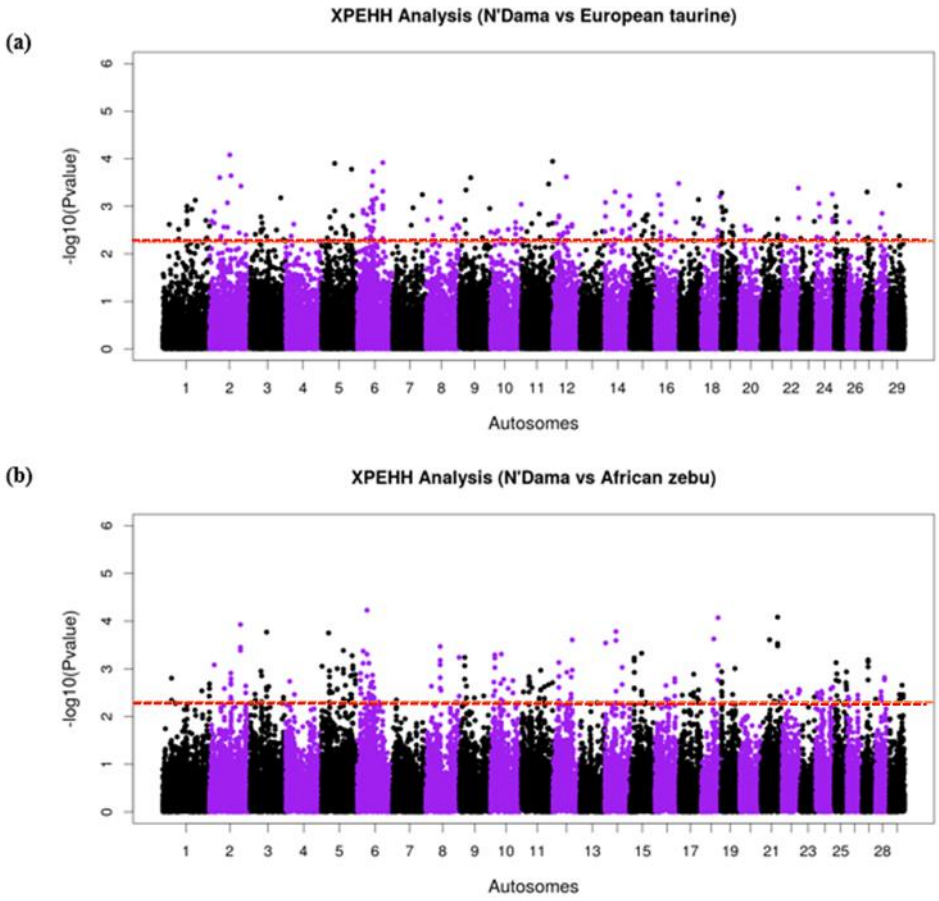


Figure 4.5 | Manhattan plots of the distribution of XPEHH scores along *B. taurus* autosomes following comparison of N'Dama with (a) European taurine and (b) African zebu.

Coincidently, 257 candidate genes are present in regions detected by more than one screening method in each West African taurine population. This number is equivalent to 20% and 17% of the total number of candidate genes in Muturu and N'Dama respectively. The remaining 80% (*n* = 1,026) and 83% (*n* = 1,243) genes are detected by only one test in Muturu and N'Dama, respectively. No genes were detected by all five selection tests in both cattle populations (Figure 4.1a, b). However, 13 candidate genes (*SERPINI2, WDR49, SLC9A9, ZBBX, STAT1, GLS, PIGK, TFEC, RORB, INTS6, FAM124A, CHD3* and *ENSBTAG00000046699*) in Muturu, and six candidate genes (*YTHDC1, RPS6KA5, COMMD1, INTS6, USH2A* and *C16H1orf21*) in N'Dama were detected by four different tests (Supplementary Tables S4.7 and S4.8). In Muturu, the

genes are located on chromosomes 1, 2, 3, 4, 8, 12 and 19. In N'Dama, they are located on chromosomes 6, 10, 11, 12 and 16. The only gene identified both in Muturu and N'Dama is located on BTA 12 (integrator complex subunit 6 protein *INTS6*) by four different tests. This gene has already discussed in Chapter 3. Several candidate selected regions, without annotated genes, i.e gene desert regions (GDR), were also detected (Table 4.1, Supplementary Tables S4.9 – S4.14); they were not investigated further within the scope of this study.

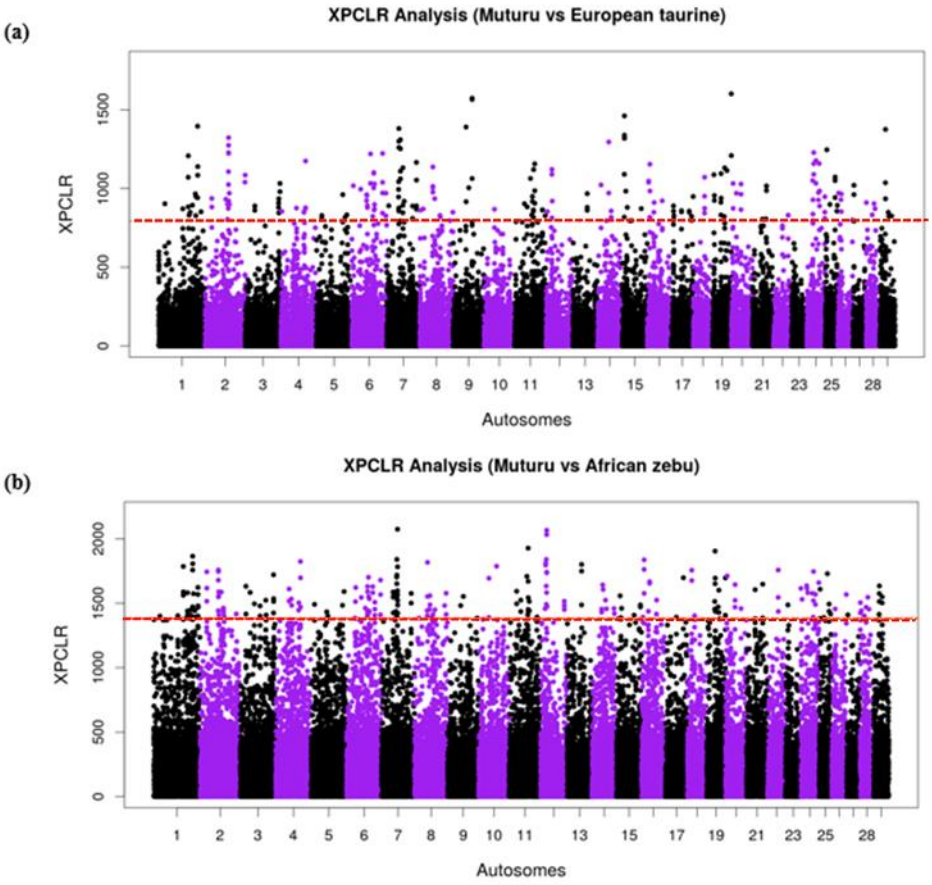


Figure 4.6 | Manhattan plots of the distribution of XPCLR scores along B. taurus autosomes following comparison of Muturu with (a) European taurine and (b) African zebu.

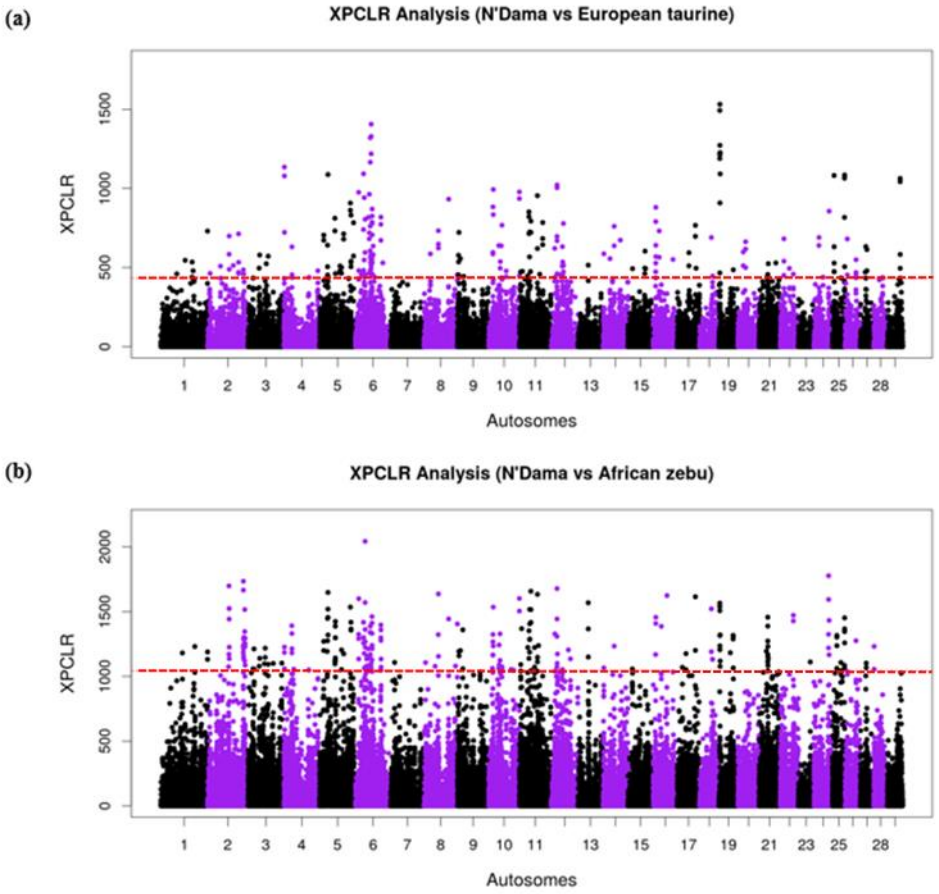


Figure 4.7 | Manhattan plots of the distribution of XPCLR scores along *B. taurus* autosomes following comparison of N'Dama with (a) European taurine and (b) African zebu.

4.3.2 Functional annotation of candidate genes in Muturu and N'Dama

To gain insight into the functions of the candidate genes, we performed functional annotation and enrichment analyses using DAVID v6.8 (Huang *et al.*, 2009, Huang *et al.*, 2009), and then compared the genes involved in biological processes (GO term) and pathways (KEGG). The numbers of mapped gene IDs in DAVID were 973 for Muturu, and 1120 for N'Dama. In Muturu, four annotation clusters were the most significantly enriched ones with enrichment scores > 1.3. They include the annotation terms such as skin barrier (GO:0061436), arachidonic acid metabolic process (GO:0019369), sensory perception of smell (GO:0007186), insulin secretion (KEGG PATHWAY:bta04911) and immune response (GO:0006955). In N'Dama, two annotation clusters were

significantly enriched (enrichment scores > 1.3) with the annotation terms retinol metabolism (bta00830), flavonoid biosynthesis process (GO:0009813) and sensory perception of smell (GO: 0007608). Immune response is also identified in N'Dama but in an annotation cluster with an enrichment score of 1.06. The complete lists of all biological processes and pathways are provided in supplementary tables S4.15 and S4.16 for Muturu and N'Dama, respectively.

4.3.3 Overlaps between candidate signatures of selection in Muturu and N'Dama and cattle transcriptome analysis (Noyes *et al.* 2011)

Noyes *et al.* (Noyes *et al.*, 2011) reported differentially expressed genes and their pathways following *T. congolese* infection in N'Dama and Kenyan Boran cattle in three tissues (liver, spleen and lymph node). We made a comparison between the genes identified in our study within candidate positively selected regions in N'Dama and Muturu with the genes involved in pathways responding to infection in (Noyes *et al.*, 2011) (Figure 4.8).

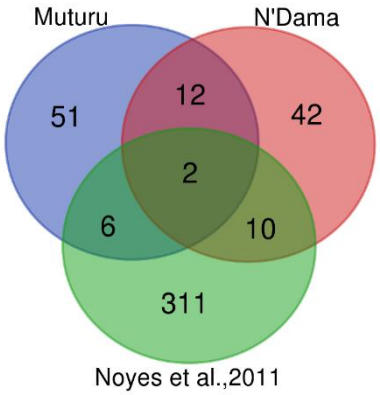


Figure 4.8 | Venn diagram of unique and shared candidate genes in Noyes et al. (2011) pathways involved in response to trypanosome infection.

Table 4.2 indicates the genes identified in our study involved in the (Noyes *et al.*, 2011) reported pathways. Among the identified pathways in Noyes *et al.* (2011) are five immune-related strong responders in the liver, namely natural killer (NK) cell-mediated cytotoxicity, T-cell receptor signalling pathway, B-cell receptor signalling pathway, Fc epsilon RI signalling pathway, and leukocyte transendothelial migration. Other pathways include cytokine-cytokine receptor interaction, MAPK signalling pathway, and neurodegenerative disease, etc (Table 4.2).

Table 4.2 | Overlaps between genes identified under signatures of selection in Muturu and N'Dama and major KEGG pathways of relevance to responses to trypanosomes infection following Noyes *et al*. (2011).

S/No	Biological processes and pathways	Muturu	N'Dama
1	Bta0230: Purine metabolism	*GUCY2D, PFAS, NT5E, PDE5A,* ***AK5, ALLC****, GUCY1A2, AK8, PDE9A, ADCY8*	*ADK,* ***AK5****, PRUNE, ADPRM, ENTPD8, POLE4, GDA, POLD3, NPR1, AK2, PDE3A, NUDT5, ATIC, PDE6A, AK6,* ***ALLC****, ENTPD2*
2	Bta04640: Hematopoietic cell lineage	*IL4R, CD4, KIT, CSF3, CD34*	*IL2RA, ITGA2, ITGA1*
3	Bta04650: Cell adhesion molecules (CAMs)	*CD4,* ***BLA-DQB, BOLA-DQA2, LOC100848815****, JSP.1, CD34, F11R,* ***BOLA-DQB, BOLA-DRB3, BOLA-DQA5****, SDC3, BOLA*	*CLDN23,* ***BLA-DQB, BOLA-DQA2****, ITGB7,* ***LOC100848815****, ITGB8, OCLN,* ***BOLA-DQB, BOLA-DRB3, BOLA-DQA5****, SDC1, NCAM1*
4	Bta04650: Natural killer cell mediated cytotoxicity	*VAV1, PIK3R5, RAC1, FAS, PRKCA,* ***PTPN6****, VAV2*	*HRAS,* ***PTPN6****, RAF1, PPP3CA, CARD11*
5	Bta04670: Leukocyte transendothelial migration	*PIK3R5, VAV1, RAC1, PRKCA, CTNNB1, F11R, VAV2, GNAI1, CTNNA1, NCF4*	
6	Bta04662: B cell receptor signaling pathway	*INPP5D, PIK3R5, RAC1, VAV1, VAV2,* ***CARD11, PTPN6***	*HRAS, RAF1, PPP3CA,* ***CARD11, PTPN6***
7	Bta04660: T cell receptor signalling pathway	*CD4, PIK3R5, VAVI, VAV2,* ***CARD11, PTPN6***	*HRAS, RASGRP1, RAF1, IL2, PRKCO, PPP3CA, ZAP70,* ***CARD11, PTPN6***
8	Bta04664 Fc epsilon RI signaling pathway	*VAV1, PIK3R5, RAC1, VAV2, GAB2, INPP5D*	

9	Bta04010: MAPK signaling pathway	*TAB1, RAC1, RPS6KA2, STK3,* ***MAP2K5****, MRAS,* ***MAP3K13****, CACNG4, FGFR4, PDGFRA,* ***GNA12****, FAS, PRKCA, CACNG5, FGF2, MAP3K5, DUSP5, CACNG1, ARRB2, MAP3K3*	*DUSP6, RASGRP1, RPS6KA5, MAP3K12,* ***MAP2K5****,* ***MAP3K13,*** *HRAS, BDNF, RAF1, NR4A1, PPP3CA,* ***GNA12****, FGF9, PTPRR, TGFBR2, CACNA1E*
10	Bta04060 Cytokine-cytokine Receptor interaction	*IL4R, FAS, KIT, CSF2RB, TNFRSF11A, CX3CR1, CXCL16, GHR, PDGFRA,* ***ACVR1****, CSF3, IL20RB*	*CXCL11, IL18RAP, CXCL10, TNFRSF9, IL18R1, ACVR1B, AMHR2, IL2, TGFBR2,* ***ACVR1****, CCL20, CXCL9, IL2RA*
11	Bta04110: Cell cycle	*SMC3, CDC16, CDC6, MCM6, CDC23, ORC5, CDKN1C*	*ESPL1, SMAD2, MCM5*

Note – The genes in bold are the 14 commonly identified in Muturu and N'Dama, while the underlined genes are the additional six genes in Muturu which overlap with the genes reported by Noyes *et al*. (2011).

4.3.4 West African taurine shared genetic control of trypanotolerance

In anticipation that the two WAT populations may share similar genetic control of trypanotolerance, we expected to identify common sets of candidate genes in pathways relevant to the pathophysiology of trypanosomiasis disease according to Noyes *et al.* (2011). We identified sixteen common candidate genes in Muturu and N'Dama in the reported pathways as candidate genes linked to trypanotolerance in WAT cattle population. Additional six genes which were reported in Noyes *et al*. (2011 (n = 6, underlined in Table 4.2) were identified among the candidate genes under selection Muturu but not N'Dama. In total, we considered all 20 WAT candidate genes (Figure 4.8, Table 4.2, Table 4.3) as putatively liniked to trypanotolerance in cattle. These genes are located on chromosomes BTA 1, 2, 3, 5, 6, 7, 8, 10, 19, 23 and 25. Functional annotation and enrichment analyses based on Reactome pathways in PANTHER *ver 13.1* (Thomas *et al.,* 2003) confirm their relevance in response to trypanosome infection pathways (Supplementary Table S4.17). Among the 20 genes, six are located on BTA 23. They are all members of the major histocompatibility complex (MHC) class II with related functions in immune responses. These genes are present in a BTA 23 genome region detected by *iHS* (discussed in Chapter 3) supporting on-going selection pressures, in contrast to the other genes, detected mainly by the other tests, where haplotypes are probably fixed or near fixation in Muturu and N'Dama.

Table 4.3 | West African taurine candidate trypanotolerant genes

chr	Position start-end (bp)	Candidate gene	Description	Selection tests	
				Muturu	N'Dama
1	82301321-82426795	*MAP3K13*	mitogen-activated protein kinase kinase kinase 13	iHS	iHS,
2	39287889-39361498	*ACVR1*	activin A receptor type 1	iHS	XPEHH
2	61821905-61852264	*MCM6*	minichromosome maintenance complex component	ZHp	Noyes et al.,2011
2	123237086-123244753	*SDC3*	syndecan 3	XPCLR	Noyes et al.,2011
3	67354501-67614628	*AK5*	adenylate kinase 5	iHS, Rsb, XPCLR	ZHp
5	75641217-75659769	*NCF4*	neutrophil cytosolic factor 4	iHS	Noyes et al.,2011
5	103862679-103872171	*PTPN6*	protein tyrosine phosphatase, non-receptor type 6	XPCLR	XPEHH, XPCLR
6	71373513-71421283	*PDGFRA*	platelet derived growth factor receptor alpha	XPCLR	Noyes et al.,2011
7	18866561-18937070	*VAV1*	vav guanine nucleotide exchange factor 1	XPEHH	Noyes et al.,2011
8	112837179-112860706	*ALLC*	allantoicase	iHS, Rsb	Rsb, XPEHH, XPCLR
10	14354690-14626128	*MAP2K5*	mitogen-activated protein kinase kinase 5	XPEHH	XPEHH, XPCLR
19	41185975-41196948	*CDC6*	cell division cycle 6	XPCLR	Noyes et al.,2011

23	25855146-25863045	*BLA-DQB*	boLa class II histocompatibility antigen, DQB*0101 beta chain precursor	iHS	Rsb
23	25351283-25356959	*BOLA-DQA2*	*Bos taurus* major histocompatibility complex, class II, DQ alpha 2	iHS	iHS
23	25404020-25411643	*BOLA-DQA5*	MHC class II member	iHS	iHS
23	25375270-25388620	*BOLA-DQB*	boLa class II histocompatibility antigen, DQB*0101 beta chain precursor	iHS	iHS
23	25458594-25476944	*BOLA-DRB3*	MHC class II member	iHS	iHS
23	25426330-25430097	*LOC100848815*	MHC class II member	iHS	iHS
25	40960556-41068746	*CARD11*	caspase recruitment domain family member 11	XPCLR	XPEHH, XPCLR
25	41099259-41171209	*GNA12*	G protein subunit alpha 12	XPCLR	XPEHH, XPCLR

We additionally performed a population differentiation (F_{ST}) test between WAT (combined Muturu and N'Dama sequences data) *versus* the other cattle populations (African zebu and European taurine). The genome-wide estimated weighted F_{ST} values ranged from 0 – 0.6 and following Z-transformation, ZF_{ST} values varied from 0 – 7.1. A Manhattan plot of the distribution of weighted F_{ST} values across the 29 BTA is presented in Figure 4.9. Two hundred fifty-one windows representing the top 0.5% windows with the highest weighted Fs_T values were considered here as common WAT candidate selected regions. One hundred forty-five include a total of 229 genes (Supplementary Table S4.18) including three of our 20 candidate commonly selected genes, *viz*. *MCM6* on BTA 2 (ZF_{ST} = 3.68), *PTPN6* on BTA 5 (ZF_{ST} = 3.49) and *GNA12* on BTA 25 (ZF_{ST} = 3.72).

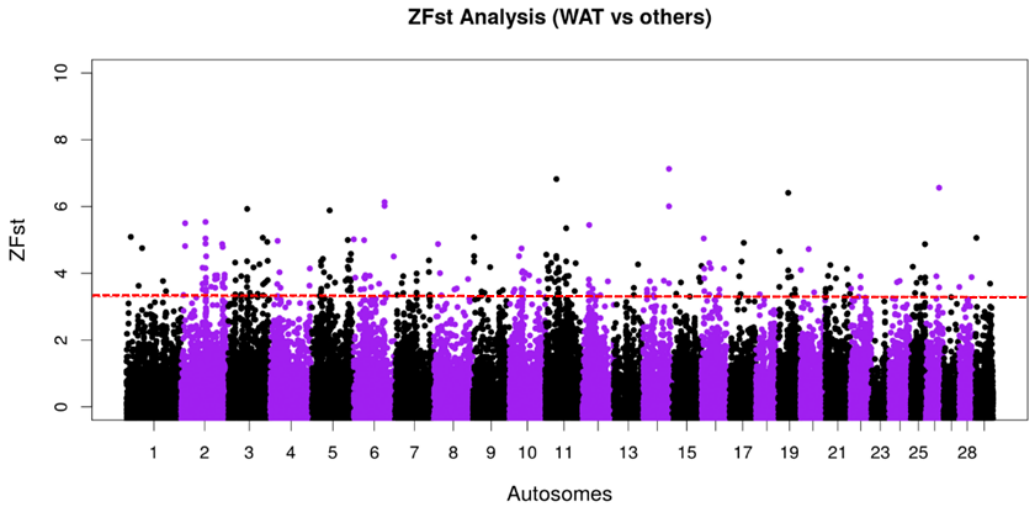


Figure 4.9 | Manhattan plots of the distribution of weighted ZF_{ST} values along B. taurus autosomes following comparison of West African taurine with other cattle populations (European taurine and African zebu sequences data combined).

Regulatory relationships among the 20 WAT candidate trypanotolerant genes were investigated using protein-protein interaction network analysis with STRING database *ver.* 10.5 (Szklarczyk *et al.*, 2016). At medium confidence level (0.400), the protein – protein interaction networks include 14 of the 20 genes, notably those involving members of the MHC class II genes, *PTPN6, VAV1, PDGFRA, MAP2K5*, *GNA12* and *NCF4* (Figure 4.10a). At the highest confidence level (0.900), the interactions between *VAV1*, *PTPN6* gene and the MHC class II genes remain (Figure 4.10b). Network analysis was repeated for Muturu and N'Dama for the 20 common genes plus each of their unique trypanotolerant related candidate genes listed in Table 4.3. All these networks of protein-protein interactions (Figures 4.10 – 4.12) reveal *PTPN6* at the centre of most protein interactions, particularly acting as an intermediary between the MHC class II genes and the other genes, with the green arrowheads indicating that *PTPN6* is activated by members of the MHC class II genes (*BOLA-DQB, BOLA-DRB3* and *BLA-DQB*).

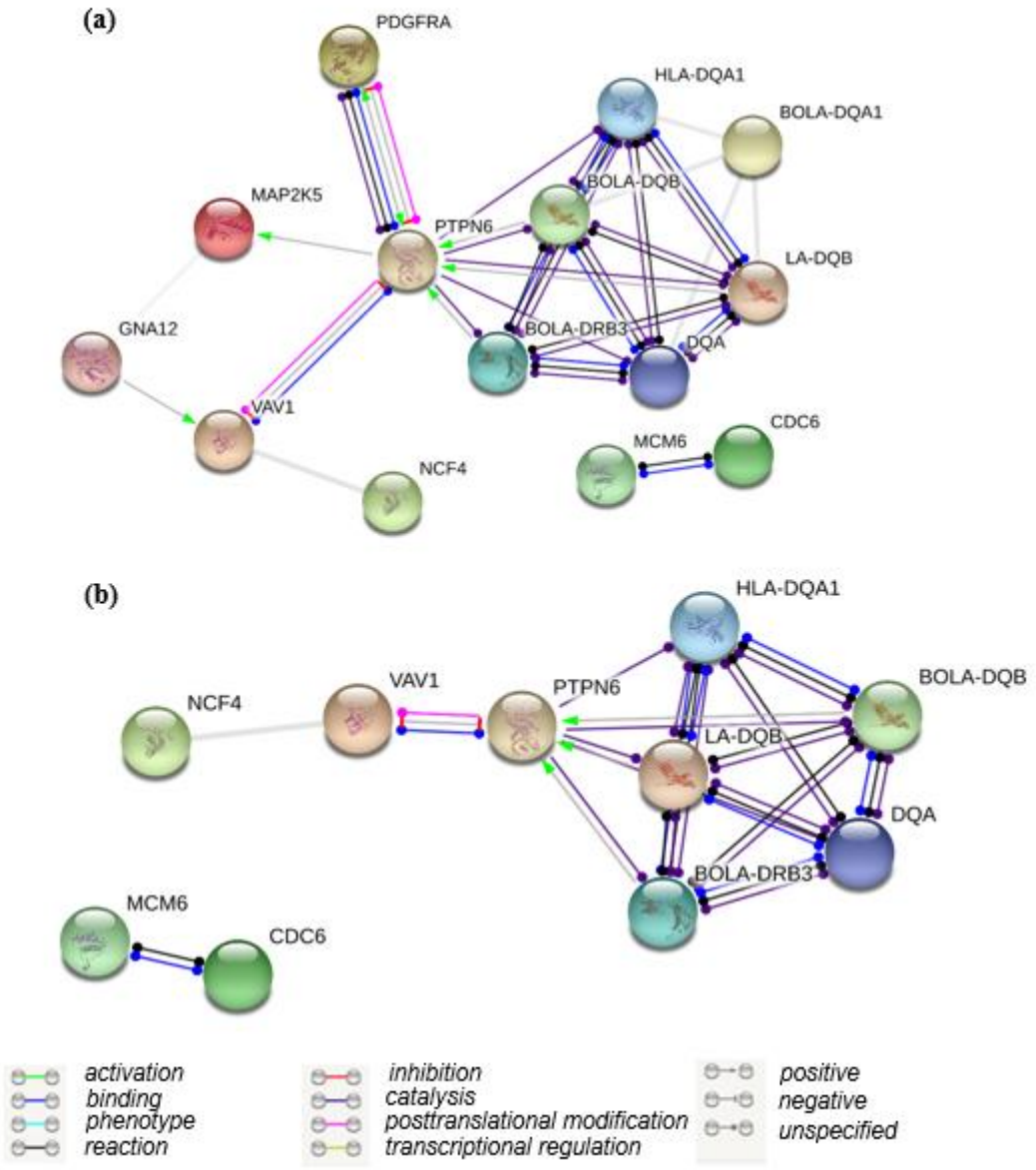


Figure 4.10 | STRING protein network view of strong regulatory relationships between PTPN6 gene and the MHC class II genes. Coloured nodes are proteins while coloured lines (edges) between the proteins indicate the type of molecular action between proteins, each of them can be positive ( ) negative ( ) or unspecified which mean no evidence for one way or the other. (a) Protein-protein interaction at medium confidence level (0.400). (b). Protein-protein interaction at the highest confidence level (0.900).

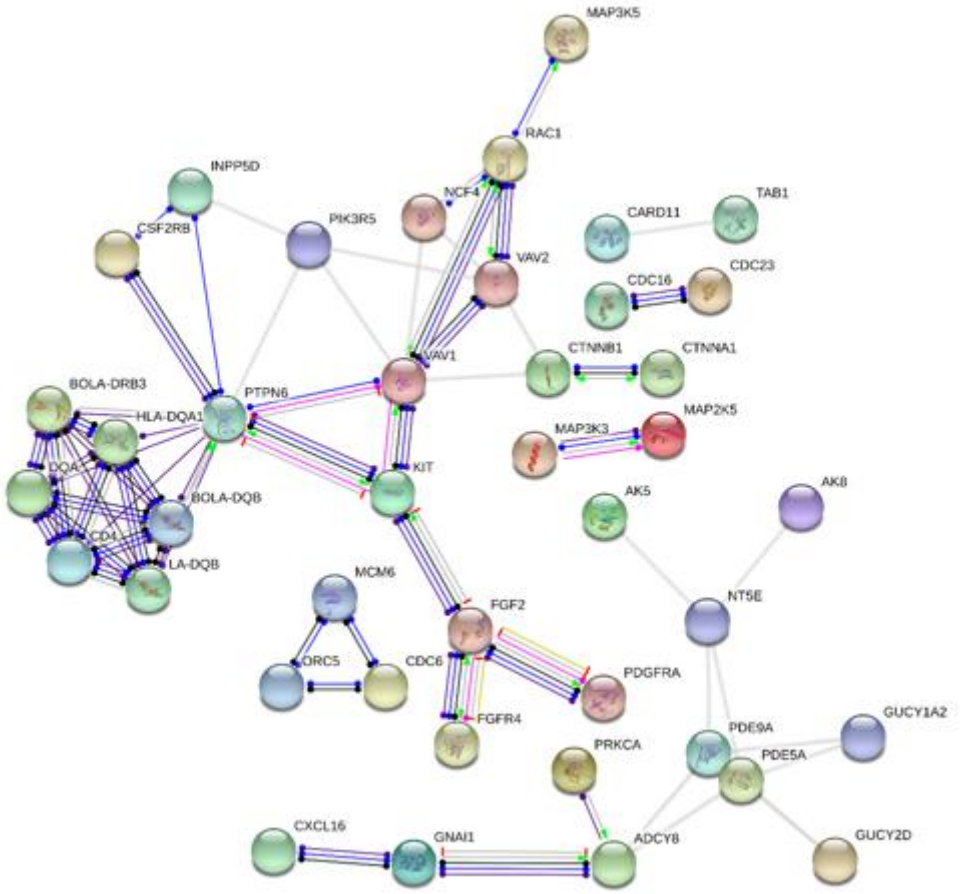


Figure 4.11 | STRING protein network view of strong regulatory relationships between PTPN6 gene and the MHC class II genes and other trypanotolerance related genes unique to Muturu. Coloured nodes are proteins while coloured lines (edges) between the proteins indicate the type of molecular action between proteins. Protein-protein interaction at the highest confidence level is shown (0.900). (Refer to Figure 4.10 for explanation).

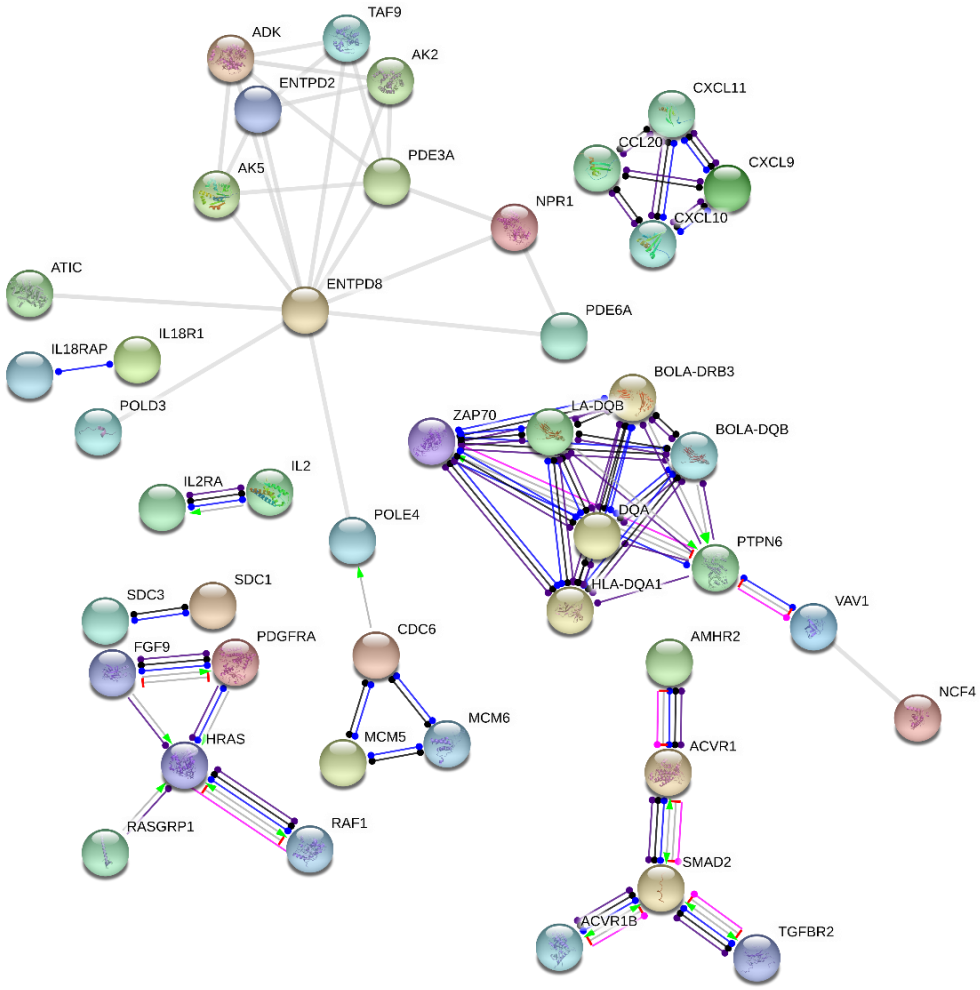


Figure 4.12 | STRING protein network view of strong regulatory relationships between PTPN6 gene and the MHC class II genes and other trypanotolerance related genes unique to N'Dama. Coloured nodes are proteins while coloured lines (edges) between the proteins indicate the type of molecular action between proteins. Protein-protein interaction at the highest confidence level is shown (0.900). (Refer to Figure 4.10 for explanation).

4.3.5 Haplotype differentiation at candidate signature of selection regions

The selective sweep in the *PTPN6* gene region is further supported by the sharing of a distinct haplotype in the two trypanotolerant WAT populations compared to the groups of trypanosusceptible cattle populations. The frequency of the haplotype appeared fixed in N'Dama and near complete fixation in Muturu, whereas the other cattle populations segregate for a number of different haplotypes in the region (Figure 4.13).

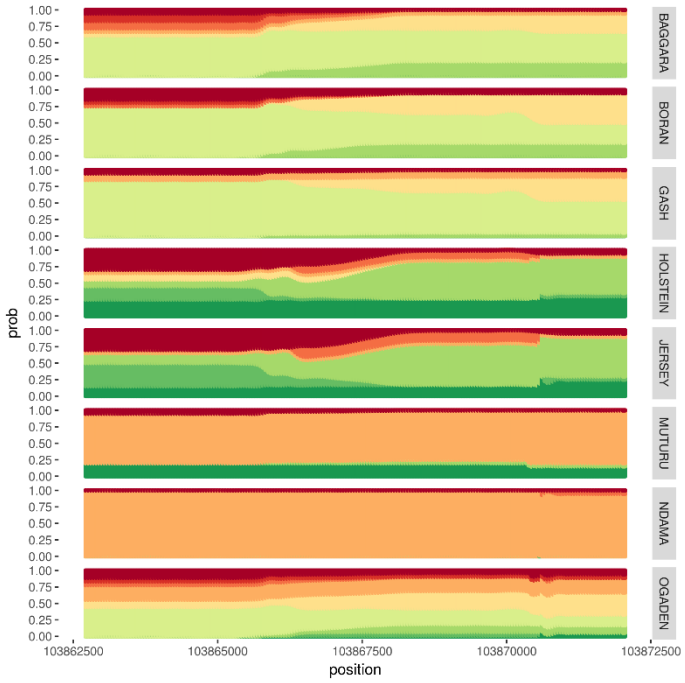


Figure 4.13 | Haplotype diversity of trypanotolerant and trypanosusceptible cattle populations at the candidate trypanotolerant *PTPN6* gene region (BTA 5).

In contrast, we did not identify any distinct haplotype in the MHC class II gene regions on BTA 23 among the different cattle populations (Supplementary Figure S4.1). We also investigated the haplotype diversity in the genome regions of previously reported "strong" candidate trypanotolerant genes. Among such previously reported trypanotolerant genes are *TICAM1, ARHGAP15, INHBA, GFM1, SLC40A1* and *HCRTR1* (Dayo *et al.*, 2009, O'Gorman *et al.*, 2009, Noyes *et al.*, 2011, Kim *et al.*, 2017). In our study *ARHGAP15, INHBA, SLC40A1* and *HCRTR1* genes were among

the detected candidate genes in N'Dama only, the *GFM1* gene was identified in Muturu only, while the *TICAM1* gene was not detected in either of the two WAT cattle populations. Surprisingly, we did not identify any distinct WAT population – specific haplotypes in the regions of any of these genes in either of the populations (Supplementary Figures S4.2 – S4.7).

More specifically, the regions of *HCRTR1* and *SLC40A1* genes contain haplotypes that are present in all eight cattle populations although the allele within the regions are fixed or nearly fixed in the *B. taurus* samples (Supplementary Figures S4.2 and S4.3). In *TICAM1* and *INHBA* gene regions, the African and European *B. taurus* populations share several haplotypes different from the *B. indicus* populations. The haplotype within *TICAM1* gene is fixed in Muturu and Jersey populations and nearly fixed in Holstein and N'Dama (Supplementary Figure S4.4), while in the *INHBA* gene, the allele is fixed in Muturu population only, nearly fixed in N'Dama and Holstein but a very low frequency in Jersey cattle (Supplementary Figure S4.5). Lastly, all eight cattle populations segregate for a number of different haplotypes in the regions of *ARHGAP15* and *GFM1* genes (Supplementary Figures S4.6 and S4.7). Our results of the analyses of these genes, particularly for the *TICAM1* and *INHBA* gene regions, are in agreement with the findings of Álvarez *et al*. (2016) who reported a lack of divergence among different cattle populations including West African taurine samples in these two trypanotolerant candidate gene regions (Álvarez *et al.,* 2016).

4.3.6 Fine-mapping of the variants at the *PTPN6* and *BoLA* genes

Variation within the *PTPN6* and the bovine *BoLA* genes was further investigated to identify unique variants within the WAT population which may have been selected for

in relation to the trypanotolerant phenotypes of the breeds. By comparing the 24 WAT samples grouped as a single population to 2,685 non-WAT taurine worldwide cattle samples, we calculated the per-SNP F_{ST} values and SAFE scores within 59,345 bp of the genome region which overlaps with *PTPN6* and other nearby candidate genes on BTA 5 (Chr5:103847447-103906801), and within 78,814 bp genome region which overlaps with the candidate *BoLA* genes on BTA 23 (Chr23:2535128-25430097). Also, the iSAFE scores of the polymorphisms within an expanded area of up to 0.5 Mb around each of the two selective sweep regions were estimated.

From a total of 1540 sequence variants analysed within the candidate gene regions on BTA 5, 19 variants were identified to have extreme population differentiation values (F_{ST} > 0.69) and were also among the most probable favoured variants based on SAFE > 0.2 and iSAFE > 0.1 scores (Figure 4.14). We considered these variants as the most probable favoured variants within the selective sweep region on BTA 5 (Table 4). Amongst these, two variants rs208442885 (Chr5:103870368, G > A) and rs209198174 (Chr5:103861777, G > A) are found in an intron and the downstream *PTPN6* gene region respectively. Variant rs208442885 is homozygous for the allele A in 18 WAT samples N'Dama (n = 11), Muturu (n = 6) and Lagune (n = 1), while the remaining six WAT samples are heterozygous at the position. The 2,685 cattle samples in 1000 Bull genome project are either homozygous reference or heterozygous at the same position. Specifically, this allele is fixed in N'Dama and nearly fixed in Muturu (Figure 4.13).

Other identified probable favoured variants are found within candidate genes such as *PHB2*, *ATN1*, *ENO2* and in gene desert areas. The *ATN1* gene contains six variants which include three intronic variants (rs136283001, rs447957799 and rs208250324),

two 3_prime_UTR variants (rs453707775, rs1115350837) and one synonymous variant (rs208956193). All six variants show high population differentiation ($F_{ST} > 0.95$) (Table 4.4).

Additionally, 32 variants were revealed based only on iSAFE estimates as probable favoured variants within an expanded region of 0.45 Mb around the candidate genes on BTA 5. These variants overlap with seven additional genes (*LRRC23*, *SPSB2*, *USP5*, *GNB3*, *P3H3*, *GPR162* and *PTMS*) not included among the trypanotolerant candidate genes of Muturu and N'Dama. However, three of the genes (*USP5*, *GNB3* and *P3H3*) were among the identified candidate genes in Muturu.

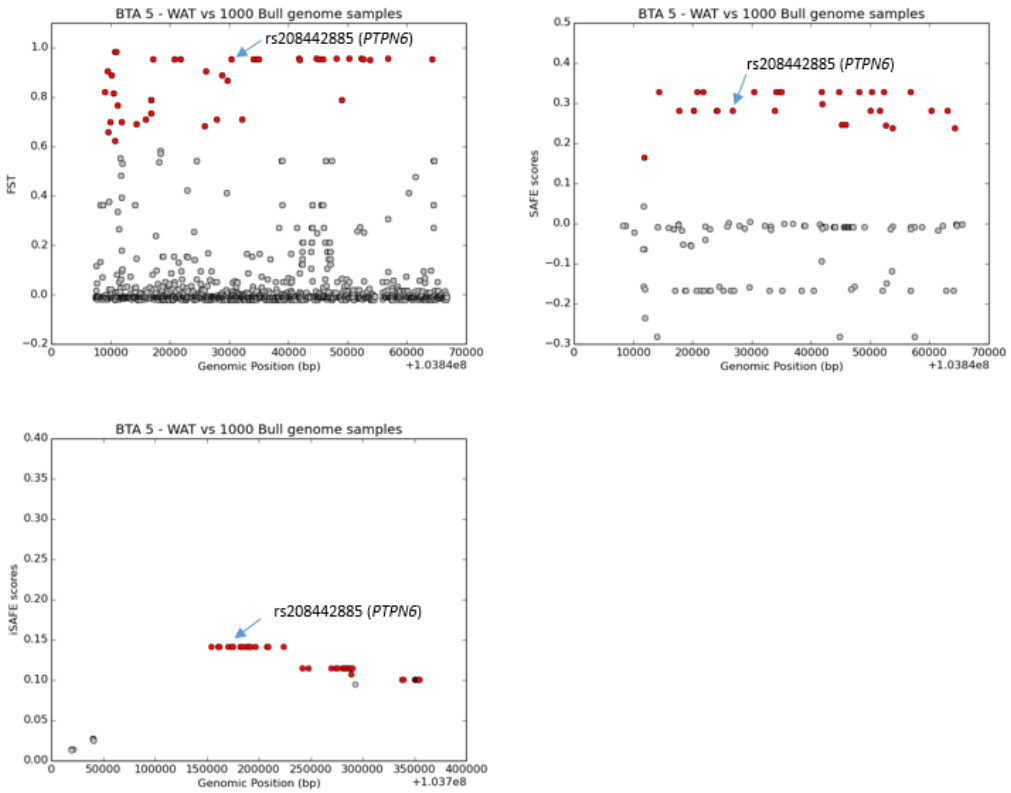


Figure 4.14 | FST and iSAFE plots showing excess population differentiation and identification of most probable favoured mutations within and around the selective sweep region on BTA 5.

Analysis of 3,597 variants within the 78.8 kb candidate gene region on BTA 23 (23:25351283-25430097) did not reveal any favoured variants or any particular variant with extreme population differentiation (F_{ST}) value. The only variant (BTA23:25359156) within this region with high *FST* value of 0.6 is located within an unannotated gene region. Within the *BOLA-DQA5* gene, we identified one variant (Chr 23:25405801) with F_{ST} value of up to 0.54 while two variants (Chr23: 25351607 and 23: 25352230) were identified within BOLA-DQA2 gene, with F_{ST} values of 0.36 and 0.38 respectively. However, within an expanded region of 0.5 Mb around the *BoLA* candidate genes (BTA23:25351283-25863045) containing 28,416 variants, we identified 55 variants with F_{ST} values of between 0.6 and 0.8 and another 431 variants ranked as most probable favoured mutations based on the iSAFE estimates. No overlap was identified between these two sets of variants, but the 55 high F_{ST} variants overlap with six genes including *BLA-DQB, BTNL1, BTNL2, BoLA DR_ALPHA* and two unannotated genes (ENSBTAG00000013919 and ENSBTAG00000038397), while the 431 favoured variants overlap with three genes (*BLA-DQB, BTNL2* and ENSBTAG00000038397).

Table 4.4 | List of putative favoured mutation within selective sweep region on BTA5

Gene	Candidate SNP Position	Candidate SNP ID	SNP function	REF	ALT	AF* (WAT) (*n* = 24)	AF*(Others) (*n* = 2,685)
PHB2	5:103854362	rs137443731	Intron variant	C	A	0.875	0.139
-	5:103860799	rs448269482	intergenic variant	G	A	0.875	0.017
PTPN6	5:103861777	rs209198174	downstream gene variant	G	A	0.875	0.017
	5:103870368	rs208442885	intron variant	G	A	0.875	0.017
-	5:103874070	rs472982890	intergenic variant	T	C	0.875	0.017
-	5:103874781	rs209573740	intergenic variant	C	T	0.875	0.017
-	5:103874993	rs137224767	intergenic variant	C	T	0.875	0.017
ATN1	5:103881805	0	3 prime UTR variant	A	T	0.875	0.016
	5:103881844	rs453707775	3 prime UTR variant	C	A	0.833	0.016
	5:103884656	rs136283001	intron variant	T	C	0.875	0.016
	5:103885190	rs447957799	intron variant	C	T	0.854	0.016
	5:103885874	rs208956193	synonymous variant	C	G	0.854	0.016
	5:103888100	rs208250324	intron variant	C	T	0.875	0.016
-	5:103890270	rs210971688	intergenic variant	G	T	0.875	0.016
-	5:103892234	rs443198317	intergenic variant	T	A	0.875	0.016
-	5:103892606	rs208527643	intergenic variant	C	T	0.854	0.016
-	5:103893743	rs467865242	intergenic variant	G	C	0.875	0.016
-	5:103896818	rs436396325	intergenic variant	T	C	0.875	0.016
ENO2	5:103904222	rs440651670	intron variant	A	C	0.833	0.015

*AF: Allele frequency of the alternate allele.

4.3.7 Overlaps of candidate selected genes with other signature of selection studies in N'Dama cattle

We investigated overlaps of our candidate genes under selection in Muturu and N'Dama with previously reported candidate selected genes in N'Dama signature of selection analyses. More specifically, we compared our candidate genes in Muturu and N'Dama with candidate genes in N'Dama cattle following high-density SNP data analysis (Xu *et al.,* 2014), and two more recent studies in which full genome sequences were used (Kim *et al.,* 2017, Taye *et al.*, 2017) (Figure 4.15).

Of the total 2,783 candidate genes identified in Muturu and N'Dama, 299 genes overlap with genes reported at least one of the three previous studies. These include 39 genes found both in Muturu and N'Dama, 222 genes found in N'Dama and 38 genes found in Muturu (Figure 4.15). These numbers represent 17.4% (N'Dama) and 6% (Muturu) of the candidate genes in our study.

Among our list of 20 trypanotolerant related candidate genes (Table 4.3), five genes (*PTPN6, GNA12, MAP2K5, CARD11* and *ALLC*) were also reported in at least two of these previous studies (Xu *et al.,* 2014, Taye *et al.*, 2017), in particular, *PTPN6* for which we have identified candidate favoured variants (Table 4.4)

Thirty-eight candidate genes detected in Muturu overlap with these previous studies in N'Dama, but were not detected in our N'Dama analysis. Twenty-one of these genes were detected by Xu and collaborators using HD SNP data (Xu *et al.,* 2014)

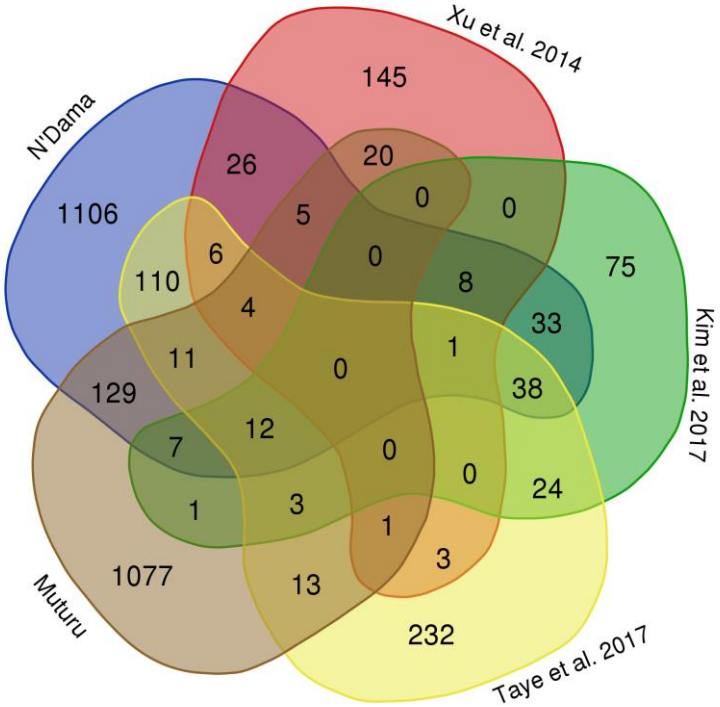


Figure 4.15 | Venn diagram plot of candidate genes in Muturu and N'Dama in comparison with the previous studies in N'Dama reported by Xu *et al*. (2014), Kim *et al*. (2017) and Taye *et al*. (2017).

4.4 Discussion

We have reported here the combined genome-wide signatures of selection scan analyses of two trypanotolerant West African taurine breeds, the shorthorn Muturu and longhorn N'Dama cattle, with the aim to identify common genetic control for their trypanotolerance traits. We then compare our findings with previous genome analysis and transcriptome studies in N'Dama cattle (Noyes *et al.,* 2011, Xu *et al.,* 2014, Kim *et al.,* 2017, Taye *et al.,* 2017).

Our primary objective was to identify a set of common candidate genes that control the innate mechanisms of trypanotolerance in West African *B. taurus* populations. Using six different selection scan tests, we obtained comprehensive lists of candidate genes under selection in trypanotolerant Muturu and N'Dama by performing within population selection scans (*iHS* and *ZHp* tests) and through comparison of their genomes with two different groups of trypanosusceptible cattle breeds (African zebu and European taurine) (Rsb, XPEHH, XPCLR and F_{ST} tests). We hypothesized that strong environmental selection pressures such as trypanosomiasis challenges would have leaded to fixation or near fixation of favourable haplotypes in WAT taurine as well as differentiation between WAT haplotypes and cattle breeds from other geographic areas (European taurine and African zebu). We then look for supportive evidence by crosschecking our findings with previous transcriptome (Noyes *et al.,* 2011) and genome scan analysis study (Xu *et al.,* 2014, Kim *et al.,* 2017, Taye *et al.,* 2017). Finally, we look for the most favoured variant (Akbari *et al.,* 2018) at the selected candidate genes.

Following these approaches, we have identified the Protein Tyrosine Phosphatase, Non-Receptor Type 6 protein encoded by the *PTPN6* gene as putatively the strongest candidate trypanotolerance gene in our study. The *PTPN6* gene show evidence of selection in both WAT cattle populations and it has also been detected under candidate selective sweeps in several other studies (Xu *et al.,* 2014, Kim *et al.,* 2017, Taye *et al.,* 2017). Also, we provide evidence of regulatory interactions at the highest confidence level between *PTPN6* and members of the bovine MHC class II genes by way of a protein-protein network (Figures 4.10 – 4.12). In addition, the role of the members of the MHC class II genes in immune response and their interactions as revealed here with the *PTPN6* gene are strongly suggestive of their contribution to the innate mechanism underlying the resistance to trypanosomiasis in WAT. *PTPN6* gene could play a central role among the different trypanotolerant candidate genes, acting as an intermediary between the MHC genes and other genes involved in response to trypanosomiasis infections (Figures 4.10 – 4.12).

The *PTPN6* gene belongs to the protein tyrosine phosphatase family and it is expressed primarily in the hematopoietic cells where it functions as an important regulator of multiple signalling pathways (Plutzky *et al.,* 1992, Beghini and Lazzaroni 2013). Functional annotation of this gene reveals its role in multiple important pathways including pathways that are relevant to the pathophysiology of trypanosomiasis disease. The pathways, including T cell and B cell signalling, natural killer cell mediated cytotoxicity, MAPK signalling and cytokine-cytokine receptor interaction (Table 4.2) have all been reported as strong responders to trypanosome infection following an expression experiment in infected cattle breeds (Noyes *et al.*, 2011).

Several other published reports have showed the relevance of these pathways in host susceptibility to trypanosomiasis (Magez *et al.,* 2008, Shigihara *et al.,* 2008, Wang *et al.,* 2008, Shrestha 2014, Stahl *et al.,* 2014, Wu *et al.,* 2017). For instance, it has been demonstrated that T cells have critical roles in the pathogenesis of trypanosome infections in humans and mice (Tarleton 1990, Brener and Gazzinelli 1997, Kumar and Tarleton 1998, Araújo *et al.,* 2005) as well as in cattle (Sileghem and Flynn 1994, Liu *et al.,* 2015). According to these studies, a significant response of either gamma delta or CD8+ T cells to trypanosome antigens was observed in trypanotolerant *B. taurus* whereas the response was negligible in the trypanosusceptible *B. indicus*. Also, experimental mice with CD8+ T cell function were observed to produce antibodies against *T. cruzi* and survived substantially longer than mice lacking CD8+ T cell function (Tarleton 1990)

Furthermore, one of the reported main host defence mechanisms against trypanosome infections is the presence of mature B-cells, which help to periodically reduce the level of parasitemia thereby contributing to the host survival during chronic infection (Magez *et al.,* 2008, Radwanska *et al.,* 2008, Amezcua Vesely *et al.,* 2012, Bermejo *et al.,* 2013). According to other studies in mice, the initial host defence mechanism is the response of T-cell-independent immunoglobulin M (IgM) antibody (Reinitz and Mansfield 1990), the host response following the generation of B cells is at a much later stage but can be strongly protective. Because the survival of B cells is limited by the parasite's ever changing variable surface glycoprotein (VSG), the susceptibility or resistance of a host to the infection is thus linked to its ability to continually generate new mature B-cells for constant interaction with the parasite VSG. The latter response was also reported as having been the case in Cape Buffalo *Syncerus caffer*, a reservoir

host with natural resistance to many parasitic diseases including trypanosomiasis (Guirnalda *et al.*, 2007).

The identification of a WAT population distinct haplotype within the *PTPN6* gene is further supportive evidence in favour of the pivotal role of *PTPN6* gene in trypanotolerance (Figure 4.13). Furthermore, we identified most probable favoured mutations (based on SAFE and iSAFE scores) with extreme population differentiation (F_{ST} = 0.95) within the *PTPN6* gene region when WAT was compared to the large group of non WAT cattle samples (over 100 worldwide different cattle breeds consisting of more than 2600 samples included in the 1000 Bull Genomes project run 6) (Table 4.4). Empirically, SAFE behaves like a standard normal random variable under neutrality and it can be used to test departure from neutrality. However, its real power appears during positive selection, when SAFE-scores change in a dramatic, but predictable manner. In addition, the iSAFE method exploits coalescent-based signals in the 'shoulders' of the selective sweep to rank all variants within 5 Mb around a region under selection. It is expected, therefore, that the mutation with the highest SAFE-score is the strongest candidate for the favoured mutation. This approach has been used to confirm causal variants within the region including the lactase gene (*LCT*) gene in humans (Akbari *et al.*, 2018).

However, we did not find any WAT specific haplotypes in the regions of several of the previously reported trypanotolerance related candidate genes (Supplementary Figures S4.2 – S4.7). It is possible that these genes may still contribute to trypanotolerance through the control of their level of expression rather than through selection for a specific haplotype.

Collectively, the results of this study provide compelling evidence that the *PTPN6* is indeed under selection in two WAT cattle populations and that the gene most likely through its interactions with other genes, influences a cascade of biological responses necessary to initiate and achieve long-term protection against trypanosomiasis infection in cattle.

CHAPTER 5:

Genomic signatures of positive selection in indigenous African zebu population

5.1 Introduction

Cattle are one of the most economically important livestock species. Since domestication, they have widely been kept worldwide as a means of livelihoods where they serve as a major source of draught power, hide, skin, milk and animal proteins as well as fulfilling socio-cultural roles, especially in Africa and Asia (Harris *et al.,* 1966, Barrett 1992, Sansoucy 1995, Thornton 2010). As a model organism, they have contributed to our understanding of mammalian physiology, including endocrine function, reproduction and growth (Gibbs *et al.,* 2002).

Worldwide cattle population have been classified into approximately 1,000 recognized breeds based on their origin, unique characteristics and adaptation to different environments (Mason 1969). At least 150 of these breeds, comprising of well over 192 million cattle, are indigenous to Africa (Rege 1999). Adapted to various environmental conditions, local African cattle have evolved their own physical and physiological adaptive traits, supported by a high genetic diversity following a rich history of introductions and movements within the African continent (Rege 1999, Hanotte *et al.,* 2000). Despite being poor producers compared to European breeds raised intensively in temperate environments, African cattle are well adapted and will thrive in their local environments, where exotic breeds will at best perform poorly or may not even survive. Thus, they represent a unique genetic resource that needs to be conserved, genetically characterized and utilized for the improvement of livestock productivity in the region (Loftus *et al.,* 1999, Troy *et al.,* 2001, Hanotte *et al.,* 2002).

The introduction of cattle into Africa occurred in different phases with the humpless longhorn taurine considered to be the first cattle brought, in approximately 8000 BC,

followed by the humpless shorthorn taurine approximately 2500 years later and the humped zebu introduced after another 1500 years (Payne and Hodges 1997). The rinderpest epidemic, which killed over 90% of all African cattle (mainly taurine breeds) in the 19th century, led to the current presence of predominantly humped zebu cattle types in West, Central and East Africa. Approximately, 60% of African zebu are found in East Africa (Rowe and Hødnebø 1994, Rege 1999, Spinage 2003, Felius *et al.,* 2014)

The majority of African zebu are kept for beef production due to their larger body sizes compared to taurine cattle. Some breeds, such as Kenana and Butana, are also kept for milk production. They show superior milk production traits compared to other African breeds and they have been referred to as African dairy zebu breeds (Musa *et al.*, 2005). Other indigenous African zebu breeds, like the white Fulani and Ogaden, are known to have both good dairy and beef characteristics (Pullan and Grindle 1980, Collier and Gunning 1999, Rege and Tawah 1999). However, in comparison with exotic cattle breeds, their productivity remains generally poor (Mwai *et al.,* 2015).

The challenge of hot and dry tropical environments is one of natural selection pressures faced by indigenous African cattle. Compared to the taurine breeds, zebu cattle show superior adaptation to high temperature and dry environmental conditions (Carvalho *et al.*, 1995, Hammond *et al.,* 1996, Gaughan *et al.,* 1999). Such ability of the zebu animal has been attributed to various physiological processes, including better thermoregulation, both in term of heat production and heat loss, and a low metabolic rate (Hansen 2004). For instance, zebu cattle have the ability to dissipate heat more efficiently than *B. taurus,* through their sweats glands, skin appendages and thoracic hump. Their sweat glands are larger, numerous, and superficially placed, producing

more sweat. The skin appendages and hump help to increase the skin surface area, thereby reducing the amount of heat produced by the body and increasing heat dissipation (Malayer *et al.,* 1990, Hansen 2004, Hernández-Cerón *et al.,* 2004, Lamy *et al.,* 2012).

Another advantage of thermal stress tolerance in zebu is due to their characteristic lighter colour, sleeker and shiny coat cover, opposed to the darker, denser and woolly coat of taurine cattle, especially in the European taurine (Hansen 2004, Lamy *et al.,* 2012). Of relevance here, is the recent understanding of the genetic control of the slick phenotype in the tropically adapted taurine Senepol cattle in which mutations within prolactin (*PRL*) and prolactin receptor (*PRLR*) genes have been associated with the slick coat phenotype in heat tolerant animals (Littlejohn *et al.,* 2014, Porto-Neto *et al.,* 2018).

It is believed that the genetic difference in heat tolerance between zebu and taurine also extends to the cellular level. For instance, there is evidence of variation of the effects of heat stress on lymphocytes from taurine (Angus) and zebu breeds (Brahman and Senepol) (Kamwanja *et al.,* 1994, Paula-Lopes *et al.,* 2003). In addition, the effect of high temperature on cellular functions such as feed intake, growth rate and milk yield as well as reproductive function appears more deleterious in taurine than zebu (Malayer *et al.,* 1990, Hernández-Cerón *et al.,* 2004).

Another major selection pressure on African cattle is tick infestation, in which the zebu are also believed to be more resistant than their taurine counterparts, especially the European taurine breeds (Frisch and O'Neill 1998, Mwangi *et al.,* 1998, Silva *et al.,*

2007, Piper *et al.,* 2009). Tick resistance has also been reported in other African breeds including the taurine N'Dama and the Sanga Ankole (Mattioli *et al.,* 2000). Tick resistance in indigenous cattle has been attributed to a long association between cattle and the ectoparasite, resulting a state of endemic stability. Generally, there is variation in the level of tick resistance both between and within cattle breeds and with regards to the species of tick (De Castro 1991, Mattioli *et al.,* 2000). The basis of the difference in level of tick resistance has been attributed to the innate immunity responses, a consequence therefore of genetic differences between cattle breeds. Also, the level of resistance is considered to be predominantly an acquired trait that becomes apparent only following a period of initial susceptibility to primary infestation (Seifert 1971, Piper *et al.,* 2009).

Tick-borne infections (TBIs) such as theileriosis, anaplasmosis, babesiosis, cowdriosis and dermatophilosis account for major production losses in the livestock industry worldwide. The magnitude of losses due to these diseases varies with the genotype of cattle, the species of infesting tick and the level of infestation (Mattioli *et al.,* 2000). *Bos indicus* breeds have been reported to show a higher level of resistance to some of the aforementioned TBIs than *Bos taurus* (Bock *et al.,* 1997, Jonsson *et al.,* 2008). For instance, in Africa, the Kenana zebu breed from Sudan has been reported to show some level of resistance to tropical theileriosis transmitted by the tick *Hyalomma anatolicum anatolicum* (Bakheit and Latif 2002).

In contrast to tick infestation, most cattle breeds including African zebu breeds are usually susceptible to trypanosomosis (Murray *et al.,* 1982), with a few exceptions such as the Nuba Mountain zebu cattle, Orma Boran and the Mursi, in which some level of

trypanotolerance have been reported most probably as a result of evolutionary adaptation to heavy tsetse fly challenge (Rege and Tawah 1999, Mwai *et al.,* 2015, Terefe *et al.,* 2015).

The unravelling of the genetic mechanisms underlying the aforementioned phenotypes in the African zebu cattle is no doubt an essential step towards the genetic characterization of the indigenous African cattle. In particular, the identification of the genes controlling the different production and adaptation traits in zebu may be useful in the improvement of livestock productivity in sub-Saharan Africa.

In this chapter, we aim to identify candidate genome regions and associated candidate genes that show strong evidence of positive selection in relation to the phenotypes of African zebu cattle. To achieve this, multiple signature of positive selection approaches including integrated haplotype homozygosity (iHS), pooled heterozygosity (ZHp), XPEHH, XPCLR and F_{ST} were applied in genome-wide selection scans of whole genome sequences from eight indigenous African zebu cattle breeds (Kenana, Butana, Aryashai, Gash, Baggara, African Fulani, Kenyan Boran and Ogaden). More specifically, the objectives of this chapter are threefold:

1. To pinpoint the loci under positive selection in each of the eight African zebu breeds, two within-populations tests, *iHS* and *ZHp*, were used to scan the genome of the individual cattle breeds. The candidate genes from all analyses were compared and analysed in relation to the different phenotypes of the African zebu such as growth and conformation traits, production traits, thermotolerance and tick resistance traits.

2. Tropical adaptation traits of the African zebu cattle were further investigated by comparing the genomes of combined African zebu breeds with non–African taurine breeds (Holstein, Jersey and Hanwoo) using XPCLR, XPEHH and F_{ST} tests

3. Lastly, the genetic basis of milk production characteristics of the African zebu dairy breeds (Kenana and Butana) was investigated by assessing the differences in haplotypes (using XPEHH test) and allele frequencies (using XPCLR) between each of the zebu dairy breeds with a combined group of African beef zebu including Baggara, Gash, Kenyan Boran and Ogaden.

Concurrently, selection signature scan analysis was also performed in the Gir breed, a dairy zebu breed from the Asian continent, in order identify genomic signatures shared with their African counterparts, particularly as related to morphology and milk production traits.

It is expected that the outcomes of this study will help to improve our understanding of the genetic basis of local adaptation, productivity (milk, meat), growth and conformation traits in indigenous African zebu cattle living in the dryland arid and semi-arid regions of sub-Saharan Africa. The results of this study are also expected to help the design of strategies aiming to conserve and use the diversity of these indigenous African zebu cattle breeds.

5.2 Materials and Methods

5.2.1 Study population and genome-wide SNP data

All cattle samples included in the present chapter are the ones described in chapter 2. In particular, the eight African zebu breeds include Aryshai (ARY, $n = 10$), Butana (BTN, $n = 8$), Kenana (KNN, $n = 10$), Baggara (BGR, $n = 9$), Gash (GSH, $n = 8$), Fulani (FLN, $n = 10$), Ogaden (OGD, $n = 9$) and Kenyan Boran (BOR, $n = 10$). Others breeds are one Asian zebu, Gir cattle (GIR, $n = 10$), two European taurine breeds, Holstein (HOL, $n = 10$) and Jersey (JER, $n = 10$), and one Asian taurine, Hanwoo (HWN, $n = 10$). The sequences and genotypes of all samples were processed as described in chapter 2 (See materials and methods section). Genome-wide autosomal bi-allelic SNPs were selected for further analyses after the quality control steps.

5.2.2 Selection sweep analyses

Several selection scan approaches were used to search the genomes of African zebu for signatures of positive selection. First, we used within population approaches *iHS* and *ZHp* to detected genomic signatures in each of the eight African zebu breeds and in the Gir breed. Next, XPEHH, XPCLR and F_{ST} tests were used to compare all African zebu cattle breeds (combined as a group) with a group of non-African breeds comprising Holstein, Jersey and Hanwoo breeds. Lastly, XPEHH and XPCLR approaches were applied, involving the comparison of each of the dairy zebu breeds (Kenana, Butana and Gir) with a group of non-dairy zebu breeds (comprising Baggara, Gash, Kenyan Boran and Ogaden breeds).

These methods were applied following the approaches described in chapters 3 and 4. However, haplotypes used for both *iHS* and XPEHH analyses were phased

simultaneously for all sample genotypes using Beagle software (Browning and Browning 2007). For all approaches, the test scores were summarized using overlapping window size of 100 kb with a 50 kb step size.

To consider selection signature, windows in the extreme 0.5% test values were considered for all tests, with the exception of *iHS*, where we considered a threshold of –10logiHS > 5 equivalent to $P < 0.00001$. Bedtools *v2.25.0* (Quinlan and Hall 2010) was used to merge the overlapping selected regions in all the tests.

5.2.3 Functional annotation of candidate genes

Genes which overlap the candidate selected regions were identified based on Ensembl Genes 92 database using *Ensembl BioMart* online tool (http://www.ensembl.org/biomart) (Smedley *et al.*, 2009) last accessed on August 2018). Candidate gene sets were processed using DAVID Bioinformatics online resources (Huang et al. 2009) for the identification of gene ontology (GO) biological processes. Further characterization of candidate selected regions was performed by finding overlaps with a list of priori genes within the cattle (UMD 3.1) Quantitative Trait Loci (QTL) (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index) database for selected phenotypes such as tick resistance and heat tolerance traits

5.3 Results

5.3.1 Genomic selection signatures of indigenous African zebu cattle

To identify the selective sweep regions in the samples of each of the African zebu breeds, we performed genome-wide scan of the 29 bovine autosomes using both *iHS* and *ZHp* tests. Similar tests were also performed in Gir zebu cattle to enable comparison

of results. The Manhattan plots of the standardized *iHS* P-values along the 29 bovine autosomes in each of the zebu breeds are presented in Supplementary Figures S5.1 – S5.9 while the of *ZHp* scores are presented in Supplementary Figures S5.10 – S5.18 for Kenyan Boran, Gash, Butana, Ogaden, Fulani, Baggara, Kenana, Aryashai and Gir breeds, respectively. The count of candidate selected regions and the associated candidate genes are presented in Table 5.1. For *iHS*, the numbers of candidate genes vary from 104 in Kenyan Boran to 251 in Gir, while for *ZHp*, they vary from 38 in Baggara to 313 in Gir. The full list of the detected candidate genes is provided in Supplementary Tables S5.1 – S5.18. In addition, several gene desert regions (GDR) were also detected based on the two selection scan tests in all the breeds (Table 5.1,) these regions are included in the Supplementary Tables S5.1 – S5.18, but were not investigated in the present study.

Based on the two selection tests, the total number across of the two gene sets in each breed reveal a total of 211, 223, 233, 290, 296, 300, 313, 342 and 517 candidate genes in Fulani, Baggara, Kenyan Boran Butana, Gash, Ogaden, Aryashai, Kenana and Gir, respectively. Also, the numbers of overlapping of candidate genes between the two selection scan tests vary from eight in Gash to 56 in Kenana (Figure 5.1)

Table 5.1 | Numbers of detected selection signature regions and associated genes in different zebu breeds

Breeds	*iHS*				*ZHp*		
	NCR	**NCG**	**GDR**		**NCR**	**NCG**	**GDR**
Baggara	116	204	37		29	38	7
Fulani	110	175	38		40	75	9
Butana	101	213	24		37	94	10
Aryashai	133	226	43		48	109	9
Gash	92	173	31		51	131	18
Ogaden	102	192	27		48	144	12
Kenyan Boran	69	104	18		64	160	12
Kenana	121	172	39		66	226	12
Gir	118	251	29		122	313	20

NCR – Numbers of candidate regions, **NCG** – Numbers of candidate genes and **GDR** – Gene desert region

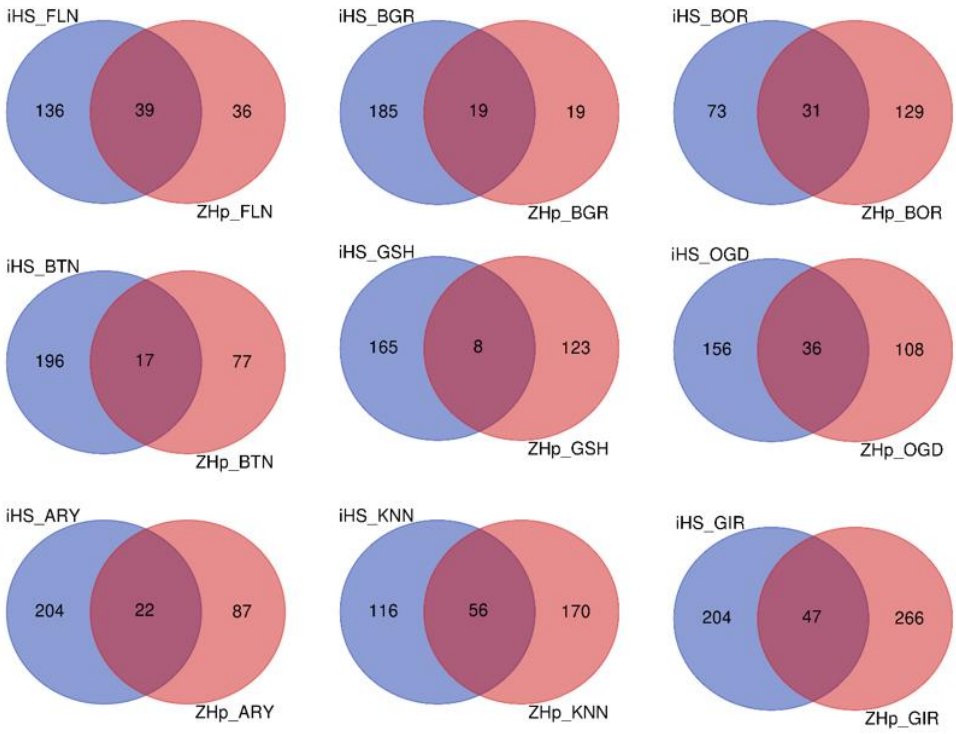


Figure 5.1 | Venn diagram showing overlap candidate genes identified in eight African zebu breeds and the Asian zebu Gir breed based on iHS and ZHp selection scan tests. FLN – Fulani, BGR – Baggara, BOR – Kenyan Boran, BTN – Butana, GSH – Gash, OGD – Ogaden, ARY – Aryashai, KNN – Kenana, GIR – Gir.

In total, 1,497 candidate genes were identified in the eight African zebu breeds, out of which 391 genes were detected in at least two breeds. Among the common candidate genes, 6, 4, 12, 26, 33 and 74 genes were shared by 8, 7, 6, 5, 4 and 3 breeds, respectively, while 235 were common between breed pairs. In addition, the number of candidate genes detected in only one breed vary from 90 in Fulani to 163 in Aryashai (Supplementary Table S5.19). A further comparison of all the detected candidate gene set in the eight African zebu breeds with the Gir breed revealed an overlap of 191 genes between African and Asian zebu, the list of these gene is provided in Supplementary Table S5.20).

To get insight on the genes under selection across the African zebu breeds, we considered genes shared across the populations. A total of 389 genes were shared by at least two African zebu breeds. We divided these genes in to two groups: group A (those shared by at least 5 breeds) and group B (those shared by between 2 to 4 breeds), these genes were also compared with the 191 genes shared with the Gir cattle as shown in Figure 5.2. The Group A genes consisting of 48 most shared genes across the African zebu populations, are presented in Table 5.2. Twenty of these genes were also shared with the Gir (Figure 5.2). Group B consists of 342 gene, of which 96 are shared with the Gir cattle (Figure 5.2).

Among the genes within candidate selected regions identified in the greatest number of African zebu populations are two olfactory receptor genes (*LOC100300085* and *LOC785149*) and the protein tyrosine phosphate, receptor type G (*PTPRG*), overlapping with selective sweeps regions on chromosome 7 (chr7: 44.35-44.5 Mb) and chromosome 22 (chr22: 39.85-40.05 Mb0, respectively. They are present in all eight

African zebu breeds. *PTPRG*, which may have function related to body weight and feed efficiency (Reyer *et al.,* 2015), is also shared with the Gir cattle. In addition, overlapping selective sweeps on chromosomes 5 (chr5: 48.05-48.4 Mb), 10 (chr10: 59.1-59.4 Mb), and 20 (chr20: 48.9-49.0 Mb) include genes *HMGA2*, *GLDN* and *CDH10*, respectively. These genes are detected in seven of the eight Africa zebu breeds, the exceptions being Ogaden (*HMGA2)*, Boran (*GLDN*) and Kenana (*CDH10*), respectively (Table 5.2).

HMGA2 and *GLDN* genes have related functions in growth traits, while *CHD10* is involved in the immune response (Fitzsimons *et al.,* 2014, Shin *et al.,* 2015, Mao *et al.,* 2016). Among the remaining highly shared candidate selected genes in the African zebu population, are eight genes: *KIF1C, INCA1, CAMTA2, SPAG7, ENO3, PFN1, RNF167* and *GP1BA* which overlap with a 150 kb selective sweep on chromosome 19 (chr19:27.0 – 27.15 Mb). These genes were detected by the *ZHp* test and were identified in five zebu breeds (Aryashai, Gash, Kenana, Kenyan Boran and Ogaden). Two previously reported genes, including *MITF* involved in coat colour and *PRKCZ* involved in feed intake, are also among the frequently identified genes across the African zebu. The coat colour gene is identified in six out the eight African zebu, the exception being the Baggara and Kenyan Boran, and also in the Gir cattle, while the *PRKCZ* gene is found in five of the six African zebu breeds from Sudan, the exception being the Fulani breed (Table 5.2).

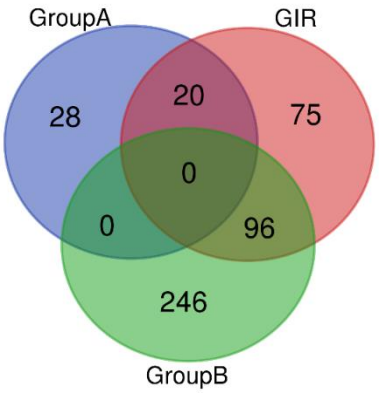


Figure 5.2 | Venn diagram showing overlap of highly shared (2 to 8) selected genes in African zebu breeds with selected gene in Asian zebu, based on iHS and ZHp analysis. GroupA consist of genes shared by at least 5 breeds while GroupB are those shared by between 2 to 4 zebu breeds.

Table 5.2 | Candidate genes under positive selection in eight African zebu and one Asian zebu breeds, identified in at least five breeds based on iHS and ZHp tests.

Chr	Position (Mb) Start	End	Breeds ARY	BGR	BTN	FLN	GSH	KNN	BOR	OGD	GIR	Method	Candidate gene
5	48.050	48.400	Y	Y	Y	Y	Y	Y	Y	N	Y	ZHp, iHS	*HMGA2*
5	47.950	48.400	Y	Y	Y	Y	Y	Y	N	N	N	ZHp, iHS	*ENSBTAG00000045454*
5	28.400	28.550	Y	N	Y	Y	N	Y	N	Y	N	ZHp, iHS	*SLC4A8, ENSBTAG00000045382*
5	58.350	58.450	N	N	Y	Y	Y	Y	N	Y	N	ZHp, iHS	*ENSBTAG00000042326*
7	44.350	44.500	Y	Y	Y	Y	Y	Y	Y	Y	N	ZHp, iHS	*LOC785149, LOC100300085*
7	44.150	44.300	Y	Y	N	Y	Y	Y	Y	Y	N	iHS	*LOC614592*
7	52.700	52.900	N	Y	Y	N	Y	Y	Y	Y	Y	ZHp, iHS	*NRG2, PSD2*
7	62.850	63.150	N	Y	Y	Y	N	Y	Y	Y	Y	ZHp, iHS	*CSNK1A1, bta-mir-145, bta-mir-143*
7	44.100	44.300	Y	Y	N	Y	Y	Y	N	Y	Y	HIS	*ENSBTAG00000046953*
7	7.700	7.850	Y	N	Y	Y	Y	Y	N	Y	N	iHS	*AP1M1, ENSBTAG00000042477*
7	6.650	6.800	Y	Y	Y	N	N	Y	Y	Y	N	iHS	*ENSBTAG00000047589*
7	56.050	56.400	Y	N	Y	N	Y	Y	N	Y	N	ZHp, iHS	*ARHGAP26, ENSBTAG00000045081*
7	53.850	54.050	Y	N	N	N	Y	N	Y	Y	Y	ZHp	*PCDHB1, ENSBTAG00000045858*
10	58.750	59.050	Y	N	N	Y	Y	Y	Y	Y	Y	ZHp, iHS	*SCG3, LYSMD2, TMOD2*
10	59.100	59.400	Y	Y	Y	Y	Y	Y	N	Y	N	ZHp, iHS	*GLDN, CYP19A1*
11	13.150	13.300	Y	N	N	Y	N	Y	Y	Y	N	ZHp	*ZNF638*
13	5.400	5.600	Y	Y	Y	Y	Y	Y	Y	Y	Y	ZHp, iHS	*SRY*
16	51.750	51.900	Y	Y	Y	N	Y	Y	N	N	N	ZHp, iHS	*SKI, FAAP20, PRKCZ*
17	35.850	36.000	Y	Y	Y	Y	Y	Y	Y	Y	Y	ZHp, iHS	*ENSBTAG00000045738, ENSBTAG00000045938*
19	27.000	27.150	Y	N	N	N	Y	Y	Y	Y	N	ZHp	*KIF1C, INCA1, CAMTA2, SPAG7, ENO3, PFN1, RNF167, GP1BA*
20	48.900	49.000	Y	Y	Y	Y	Y	N	Y	Y	N	ZHp	*CDH10*
21	58.950	59.100	Y	Y	N	Y	Y	Y	N	N	Y	HIS	*PRIMA1*
22	39.850	40.050	Y	Y	Y	Y	Y	Y	Y	Y	Y	ZHp, iHS	*PTPRG*
22	31.750	32.050	Y	N	Y	Y	Y	Y	N	Y	Y	ZHp, iHS	*MITF*
22	10.750	10.850	N	Y	Y	Y	N	Y	N	Y	Y	ZHp, iHS	*CDLK3*
27	5.950	6.100	Y	Y	Y	N	Y	Y	N	Y	N	iHS	*ENSBTAG00000038504*

5.3.2 Candidate genes related to heat tolerance

To gain insights into the candidate genes underlying tropical adaptation trait such as heat tolerance, we compared our list of 1,497 candidate genes with a priori lists of 85 genes that are located within cattle QTL regions for heat tolerance (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index). We found 5 genes that overlapped with cattle QTL for heat tolerance out of the total 1,497 genes. These genes are *DCLK3* identified in five breeds, *GOLGA4* and *TRANK1* identified in four breeds, while the remaining two are *ARL6IP1* and *ITGA9* genes, they were identified only in Gash and Baggara, respectively. The former three genes have previously been reported in a selective sweep analysis involving Brahman cattle (Ramey *et al.,* 2013). In addition, the *GOLGA4* and *ARL6IP1* genes have been associated with thermotolerance traits and cow conception rate in Holstein cattle (Dikmen *et al.,* 2015, Ortega *et al.,* 2016).

To investigate further the genomic signatures of the African zebu that may be associated with tropical adaptation in general, we performed additional genome-wide selection scan comparisons between African zebu population and a group of non-African cattle breeds including Holstein, Jersey and Hanwoo using the XPCLR, XPEHH and F_{ST} comparison tests. The genome-wide plots of each the three selection test scores are presented in Figures 5.3 - 5.5. We identified an additional 27 genes that may be involved in the control of heat tolerance, based on their overlap with cattle QTL database for heat tolerance/or and previously reported thermotolerance genes in cattle and other species (Kolli 2012, Hansen 2013, Howard *et al.*, 2014, Jin *et al.,* 2017, Kim *et al.,* 2017, Zeng *et al.,* 2018). The list of these genes is presented in Table 5.3

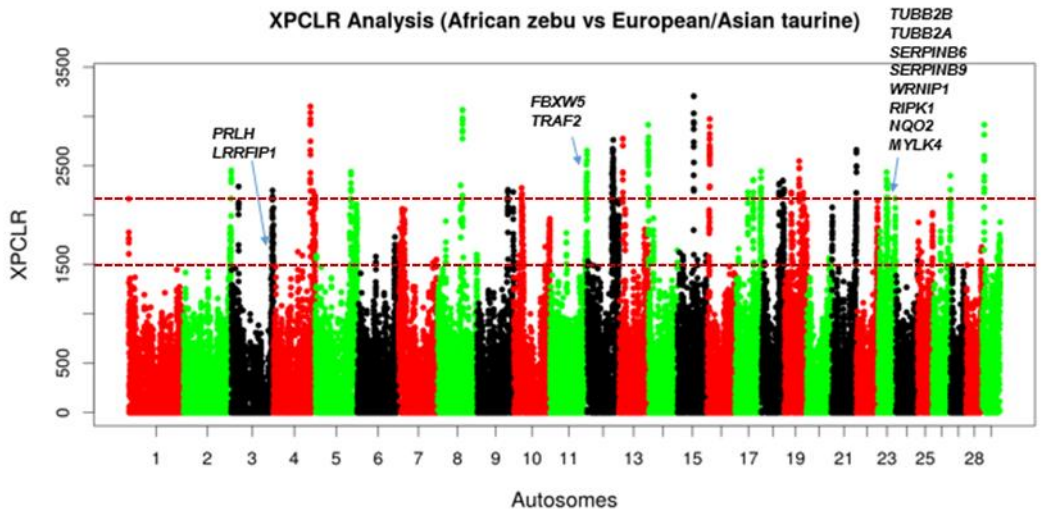


Figure 5.3 | Manhattan plot of the distribution of XPCLR scores following the comparison of African zebu population with a group of combined European and Asian *Bos taurus* breeds. Blue arrows indicate heat tolerance regions.

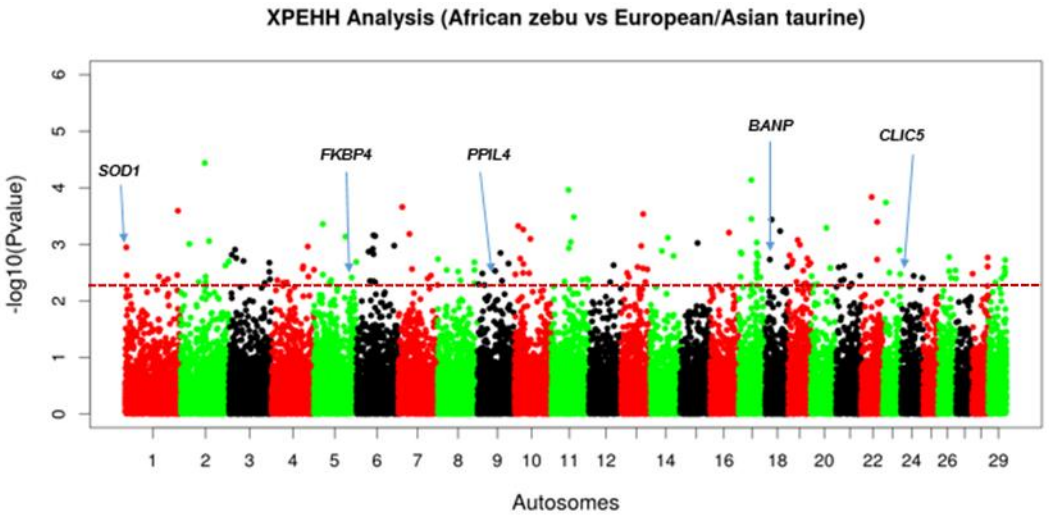


Figure 5.4 | Manhattan plot of the distribution of XPEHH scores following the comparison of African zebu population with a group of combined European and Asian *Bos taurus* breeds. Blue arrows indicate heat tolerance regions.

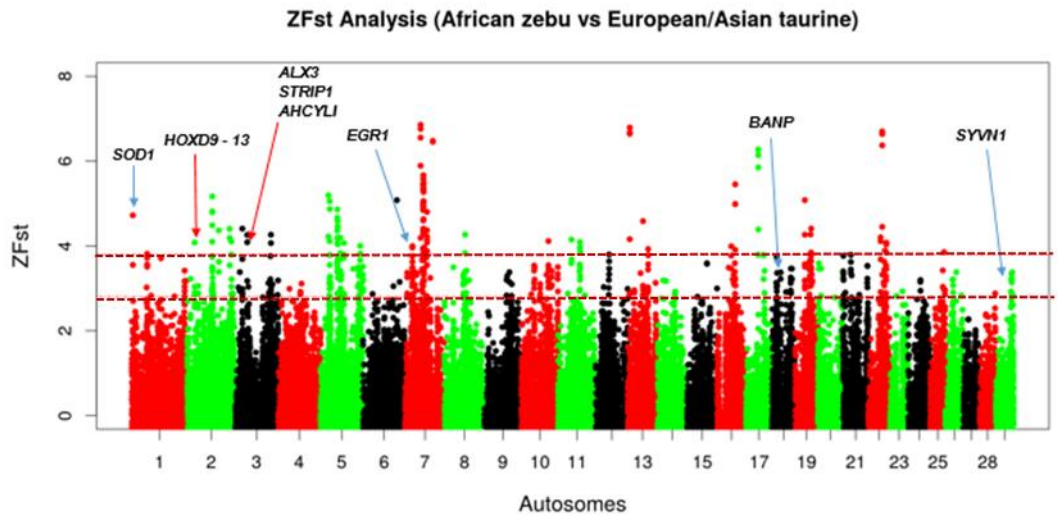


Figure 5.5 | Manhattan plot of the distribution of XPEHH scores following the comparison of African zebu population with a group of combined European and Asian *Bos taurus* breeds. Blue arrows indicates heat tolerance regions and genes, while red arrows indicate some of the identified tick resistance regions/genes.

Interestingly, among the identified genes are two well-known genes, *SOD1* and *PRLH* genes, in which haplotype variation have been associated with thermotolerance following their identification in selection signature regions in cattle (Hansen 2013, Kim *et al.,* 2017, Zeng *et al.,* 2018). Of particular here is the *PRLH* gene which may have regulatory relationships between other well-known thermotolerance related genes (*PRL* and *PRLR*), associated with cattle having the slick coat phenotype (Littlejohn *et al.,* 2014, Porto-Neto *et al.,* 2018). In addition, 13 genes among the overlapped genes have been previously reported as up-regulated genes in indigenous zebu cattle and buffalo exposed to heat stress (Kolli 2012). In another study, the *ENO3* gene has been identified among down-regulated genes in cattle during heat stress, likely due to its role in the down-regulation of carbohydrate breakdown for energy conservation during heat stress (Kolli *et al.,* 2014).

Table 5.3 | List of genes in African zebu breeds overlapping with heat tolerance QTL in cattle (UMD3.1) and previous reports.

Chr	Start (Mb)	End (Mb)	Gene	Reference(s)	Breed/Group	Method(s)
1	3.114	3.123	*SOD1*	Hansen 2013	African zebu	FST, XPEHH
3	20.771	20.774	*HIST2H2BE*	Kolli *et al*., 2012	KNN	ZHp
3	117.646	117.647	*PRLH*	Kim *et al*., 2017	African zebu	XPCLR
3	117.708	117.870	*LRRFIP1*	Kolli *et al*., 2012	African zebu	XPCLR
5	107.446	107.454	*FKBP4*	Kolli *et al*., 2012	African zebu	XPEHH
7	51.439	51.443	*EGR1*	Kolli *et al*., 2012	African zebu	FST
9	87.888	87.930	*PPIL4*	Kolli *et al*., 2012	African zebu	XPEHH
10	76.249	76.281	*SGPP1*	Kolli *et al*., 2012	African zebu	FST
11	106.287	106.291	*FBXW5*	Jin *et al*., 2017	African zebu	XPCLR
11	106.297	106.311	*TRAF2*	Jin et al., 2017	African zebu	XPCLR
13	74.068	74.088	*TOMM34*	Kolli *et al*., 2012	African zebu	XPCLR
14	83.552	83.639	*TAF2*	Kolli *et al*., 2012	ARY	ZHp
18	13.425	13.493	*BANP*	Kolli *et al*., 2012	African zebu	FST, XPEHH
19	27.073	27.079	*ENO3*	Kolli *et al*., 2014	ARY, BOR, GSH, KNN, OGD	ZHp
20	40.350	40.373	*TARS*	Kolli *et al*., 2012	ARY, BTN	ZHp
22	10.652	10.723	*TRANK1*	Cattle QTL	BGR, BTN, KNN, OGD	ZHp
22	10.775	10.790	*DCLK3*	Cattle QTL	BTN, FLN, KNN, OGD	ZHp
22	10.814	10.924	*GOLGA4*	Cattle QTL	BGR, BTN, FLN, KNN, OGD	ZHp
22	10.946	11.307	*ITGA9*	Cattle QTL	BGR	ZHp
23	19.167	19.333	*CLIC5*	Jin *et al*., 2017	African zebu	XPEHH
23	50.355	50.358	*TUBB2B*	Cattle QTL /Howard *et al*., 2014	African zebu	XPCLR
23	50.390	50.394	*TUBB2A*	Cattle QTL /Howard *et al*., 2014	African zebu	XPCLR
23	50.421	50.447	*RIPK1*	Cattle QTL /Howard *et al.*, 2014	African zebu	XPCLR
23	50.487	50.498	*NQO2*	Cattle QTL	African zebu	XPCLR
23	50.603	50.618	*SERPINB6*	Cattle QTL	African zebu	XPCLR

23	50.637	50.646	*SERPINB9*	Cattle QTL /Howard *et al*., 2014	African zebu	XPCLR
23	50.674	50.679	*SERPINB1*	Cattle QTL	African zebu	XPCLR
23	50.704	50.714	*WRNIP1*	Cattle QTL /Howard *et al*., 2014	African zebu	XPCLR
23	50.735	50.801	*MYLK4*	Cattle QTL	African zebu	XPCLR
25	16.544	16.550	*ARL6IP1*	Cattle QTL	GSH	ZHp
25	16.557	16.626	*SMG1*	Kolli *et al*., 2012	GSH	ZHp
29	43.997	44.003	*SYVN1*	Kolli *et al*., 2012	African zebu	FST

5.3.3 Candidate genes related to tick resistance

Tick resistance is another major adaptation phenotype of the African zebu breeds for which they are known to exhibit superior resistance compared to the *Bos taurus* breeds (Frisch and O'Neill 1998, Mwangi *et al.,* 1998, Silva *et al.,* 2007, Piper *et al.,* 2009). Similar to the investigation of heat tolerance traits, we compared entire gene sets detected in the African zebu population based on *iHS*, *ZHp*, XPCLR, *FST* and XPEHH with a priori lists of 361 genes that are located within regions of cattle QTL for tick resistance (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index). A total of 54 genes were found to overlap the QTL regions. Of these genes, 28 are detected by *ZHp* and *iHS* test, 10 genes are detected by XPCLR, one gene detected by XPEHH, seven genes detected by FST, and the remaining eight genes are detected by both *FST* and *ZHp* (Table 5.4).

Regions on chromosome 2 contain the highest number of the identified candidate tick resistance genes (n = 14). A majority of these genes are members of the homeobox D genes, including *HOXD1, 3, 4, 9 – 13.* These genes (*HOXD 9 – 13*) are detected following F_{ST} comparison between African zebu with European taurine and the within population *ZHp* tests in Kenana (*HOXD1, 3, 4, 9* – 13) and Ogaden (*HOXD9 – 13*). Other genes on chromosome 2 are *LNPX, ECE1, UBR4, ALDH4A1* and *TASIR2.* The latter three genes were detected based on the XPCLR selection scan test. The chromosomal distribution of the remaining candidate tick resistance genes are on chromosomes 3 (n = 7), 4 (n = 6), 11(n = 7), 13 (n = 5) and 20 (n = 8) (Table 5.4).

Interestingly, excluding genes detected by the breed group comparison tests, we observe that the genes found on are mostly associated with genomic signatures in one

zebu breed with an overlap between a maximum of two breeds. For instance, tick resistance genes on chromosome 2 are found in Kenana and Ogaden, chromosome 3; Aryashai and Ogaden, chromosome 10 in Gash, chromosome 11 in Aryashai, chromosome 13 in Butana, chromosome 20 in Fulani and Gash. Surprisingly, none of the identified tick resistance genes are found in Baggara or Kenyan Boran. Nevertheless, the role of these genes in the innate control of tick infestation requires further investigation.

Table 5.4 | List of detected genes in African zebu breeds that overlapped tick resistance QTL in cattle (UMD3.1).

Chr	Start	End	Gene	Breed/Group	Method	Chr	Start	End	Gene	Breed/Group	Method
2	20.76	20.761	*HOXD1*	KNN	ZHp	6	4.491	4.743	*PRDM5*	FLN, KNN	ZHp
2	20.776	20.781	*HOXD3*	KNN	ZHp	7	13.249	13.506	*CACNA1A*	ARY, BGR	ZHp
2	20.795	20.797	*HOXD4*	KNN	ZHp	10	46.007	46.131	*DAPK2*	GSH	ZHp
2	20.798	20.798	*bta-mir-10b*	KNN	ZHp	10	77.201	77.274	*SPTB*	GSH	FST, ZHp
2	20.824	20.826	*HOXD9*	KNN, OGD	FST, ZHp	11	20.691	20.759	*ATL2*	ARY	ZHp
2	20.829	20.832	*HOXD10*	KNN, OGD	FST, ZHp	11	20.93	20.967	*HNRNPLL*	ARY	ZHp
2	20.839	20.842	*HOXD11*	KNN, OGD	FST, ZHP	11	21.04	21.095	*GALM*	ARY	ZHp
2	20.848	20.849	*HOXD12*	KNN, OGD	FST, ZHp	11	21.105	21.111	*SRSF7*	ARY	ZHp
2	20.854	20.856	*HOXD13*	KNN, OGD	FST, ZHp	11	21.148	21.152	*GEMIN6*	ARY	ZHp
2	20.942	21.011	*LNPK*	KNN, OGD	FST, ZHp	11	101.593	101.645	*PRRC2B*	African zebu	XPEHH
2	131.935	132.06	*ECE1*	FLN	ZHp	11	104.408	104.417	*CACFD1*	African zebu	XPCLR
2	134.039	134.176	*UBR4*	African zebu	XPCLR	13	73.189	73.196	*TOX2*	OGD	ZHp
2	134.346	134.38	*ALDH4A1*	African zebu	XPCLR	13	73.67	73.75	*PKIG*	BTN	ZHp
2	134.391	134.405	*TAS1R2*	African zebu	XPCLR	13	73.75	73.774	*ADA*	BTN	ZHp
3	33.034	33.035	*KCNA10*	ARY, OGD	ZHp	13	73.833	73.847	*WISP2*	BTN	ZHp
3	33.056	33.066	*CYM*	ARY, OGD	ZHp	13	80.016	80.114	*NFATC2*	African zebu	XPCLR
3	33.081	33.086	*PROK1*	ARY, OGD	ZHp	16	40.119	40.472	*DNM3*	African zebu	FST
3	33.135	33.139	*LAMTOR5*	ARY, OGD	ZHp	19	16.353	17.562	*ASIC2*	BTN	ZHp
3	33.469	33.478	*ALX3*	African zebu	FST	20	71.395	71.42	*SLC12A7*	FLN, GSH	ZHp
3	33.483	33.504	*STRIP1*	African zebu	FST	20	71.427	71.446	*NKD2*	FLN, GSH	ZHp
3	33.514	33.556	*AHCYL1*	African zebu	FST	20	71.497	71.511	*TRIP13*	FLN, GSH	ZHp
4	66.697	66.756	*PLEKHA8*	African zebu	FST	20	71.511	71.53	*BRD9*	FLN, GSH	ZHp
4	113.562	113.563	*LRRC61*	African zebu	XPCLR	20	71.565	71.571	*TPPP*	GSH	ZHp
4	113.564	113.567	*RARRES2*	African zebu	XPCLR	20	71.602	71.645	*CEP72*	GSH	ZHp
4	113.602	113.604	*REPIN1*	African zebu	XPCLR	20	71.724	71.758	*SLC9A3*	GSH	FST, ZHp
4	113.675	113.698	*GIMAP8*	African zebu	XPCLR	20	71.766	71.79	*EXOC3*	African zebu	FST
4	113.721	113.73	*GIMAP7*	African zebu	XPCLR	22	45.095	45.556	*ERC2*	BTN	FST, ZHp

5.3.4 Candidate genes related to milk production traits

Among the indigenous African zebu breeds, the Kenana and Butana breeds have been recognized for their superior milk production capability when compared to other African zebu breeds (Cunningham, 1983). Their level of milk production was also reported to be comparable to those of the Red Sindhi from India (DAGRIS, 2017). Accordingly, some of the candidate selected region and their associated genes in these breeds could be associated with their milk production traits. To identify dairy related genes in the African zebu, particularly in Kenana and Butana as well as in the Gir breed, we compared the genomes of Kenana, Butana and Gir cattle separately with a group of four beef African cattle (African zebu non-dairy or AFZND). Several differentiated genome regions using XPCLR and XPEHH analysis were identified (Figures 5.6 – 5.7).

To then identify milk production traits related genes among the candidate genes, we considered the genes identified in all African zebu population and Gir based on *ZHp* and *iHS* analysis, as well as the genes detected following the comparison of the dairy zebu breeds with the AFZND (XPCLR and XPEHH selection scan tests). First, the entire gene sets were compared to a list of previously reported 44 genes associated with milk production and quality in cattle (Ogorevc *et al.,* 2009). Interestingly, two genes identified in Butana are the only genes found to overlap between the two gene sets. These are the milk fat globule-EGF factor 8 protein (*MFGE8*) and *CD14* genes, which overlap with candidate selected regions on chromosome 21 (XPCLR = 54.36) and chromosome 7 (XPCLR = 66.57), respectively. The former gene has role in milk production while the function of the later has been related to the lactating cow responses to mastitis (Ogorevc *et al.,* 2009).

To further our search for selected genes related to milk production traits, we made an additional comparison of our list of candidate genes with a previously reported gene set comprising of approximately 800 genes involved in pathways that are related to milk production in Polish Holstein dairy cattle (Suchocki *et al.*, 2016). These genes were classified into three categories including those related to milk yield (MY), protein Yield (PY) and fat yield (FY). We found an overlap of 78 genes between the two gene sets, which fell into two different categories: genes detected only in the dairy zebu breeds and not shared with others (n = 31), (list provided in Table 5.5) and genes also detected as well in non-dairy African zebu breeds (n = 47, Supplementary Table S5.21)

Of the 31 candidate milk genes identified in the dairy zebu breeds only, 7 are also present in candidate selected regions in the Gir, while the remaining 24 are shared equally between Butana (n =12) and Kenana (n = 12). Regarding the gene functions, four genes (*NUDT16, PAN2*, *ENO2* and *ZCCHC7*) have been linked to fat content, five (*RYR2, LEF1, TJP3, CACNA1D* and *CLDN4*) to milk volume and 19 genes in protein content (*MCM5, MAD2L1, CTSB, BUB1B, LAMP1, TFDP1, GRK1, CDC16, SMAD4, ACAP1, CDKN2A, RNF41, CDC25B, ATP6V0A1, SMURF2, NAGLU, CCR5, YWHAG* and *AP2A2*). The remaining three genes (*SRC, RHOA* and *PARD6G*) have been linked to both milk and protein yields; they were identified in the Gir breed only (Table 5.5).

Table 5.5 | Candidate genes related to milk production traits identified in African dairy zebu breeds (Butana and Kenana) and Asian dairy zebu (Gir)

Chr	Start (Mb)	End (Mb)	Gene	Breed	Method (s)	Function
1	140.045	140.046	*NUDT16*	BTN	ZHp, XPEHH	FY
5	57.339	57.353	*PAN2*	BTN	ZHp, XPCLR, XPEHH	FY
5	73.991	74.009	*MCM5*	BTN	ZHp	PY
5	103.900	103.907	*ENO2*	KNN	XPEHH	FY
5	57.428	57.454	*RNF41*	GIR	XPCLR	PY
6	6.013	6.020	*MAD2L1*	BTN	ZHp, XPCLR, XPEHH	PY
6	18.335	18.451	*LEF1*	KNN	XPCLR, XPEHH	MY
7	21.448	21.479	*TJP3*	BTN	XPEHH	MY
7	53.446	53.449	*CD14*	BTN	XPCLR	mastitis
8	7.415	7.423	*CTSB*	KNN	ZHp, XPCLR, XPEHH	PY
8	22.147	22.152	*CDKN2A*	BTN	XPEHH	PY
8	61.612	61.875	*ZCCHC7*	GIR	ZHp, XPCLR, XPEHH	FY
10	35.903	35.959	*BUB1B*	KNN	ZHp, XPCLR	PY
12	90.610	90.622	*LAMP1*	KNN	ZHp, XPCLR, XPEHH	PY
12	90.738	90.755	*TFDP1*	KNN	ZHp, XPCLR, XPEHH	PY
12	90.769	90.783	*GRK1*	KNN	ZHp, XPCLR, XPEHH	PY
12	91.039	91.066	*CDC16*	KNN	ZHp, XPCLR, XPEHH	PY
13	51.854	51.864	*CDC25B*	GIR	XPEHH	PY
13	66.986	67.005	*SRC*	GIR	XPEHH	MY, PY
19	27.673	27.685	*ACAP1*	BTN	ZHp	PY
19	43.201	43.252	*ATP6V0A1*	BTN	XPCLR	PY
19	49.384	49.429	*SMURF2*	BTN	XPCLR, XPEHH	PY
19	43.259	43.266	*NAGLU*	BTN	XPCLR	PY
21	20.890	20.905	*MFGE8*	BTN	XPCLR	FY
22	51.278	51.323	*RHOA*	GIR	XPEHH	MY, PY
22	53.599	53.602	*CCR5*	KNN	XPEHH	PY
22	47.721	48.067	*CACNA1D*	BTN	XPCLR, XPEHH	MY
24	50.994	51.046	*SMAD4*	BTN	ZHp, XPCLR	PY

24	0.583	0.660	*PARD6G*	GIR	ZHp, XPCLR	MY, PY
25	34.884	34.907	*YWHAG*	KNN	XPEHH	PY
25	33.988	33.990	*CLDN4*	KNN	XPCLR, XPEHH	MY
28	9.966	10.567	*RYR2*	KNN	ZHp	MY
29	50.407	50.445	*AP2A2*	GIR	XPEHH	PY

Note – MY – milk yield, PY – Protein yield and FY – fat yield.

Among the genes in the second category (n = 47) are 11 genes shared between at least one African non-dairy zebu and either Butana or Kenana or both. *PSD2*, *AP1M1*, *PRKCZ* and *GNS* are shared by at least 4 zebu breeds including Butana and Kenana and are all involved in milk protein content. *PSD2* is also shared with the Gir breed. In addition, *CDK2* and *RAB5B* are shared between Butana and Gash while *CDC6* and *ARAP2* genes are shared between Butana and Aryashai, the latter four genes are all involved in milk protein content. Overlap of milk related genes in the other African non-dairy zebu with Kenana include one milk yield gene, *CTNNA1*, identified in Aryashai, Fulani, Ogaden and Gir. *ENO3* was identified in four African non-dairy zebu breeds as well as in Kenana, and it has been linked to fat yield. This particular gene was also identified as a possible heat tolerance gene, suggesting a gene with pleiotropic effects; this requires further investigation. Lastly, the *MAGI1* gene is involved in milk yield and it was identified in Kenana and Aryashai (Supplementary table S5.22) (Suchocki *et al.*, 2016).

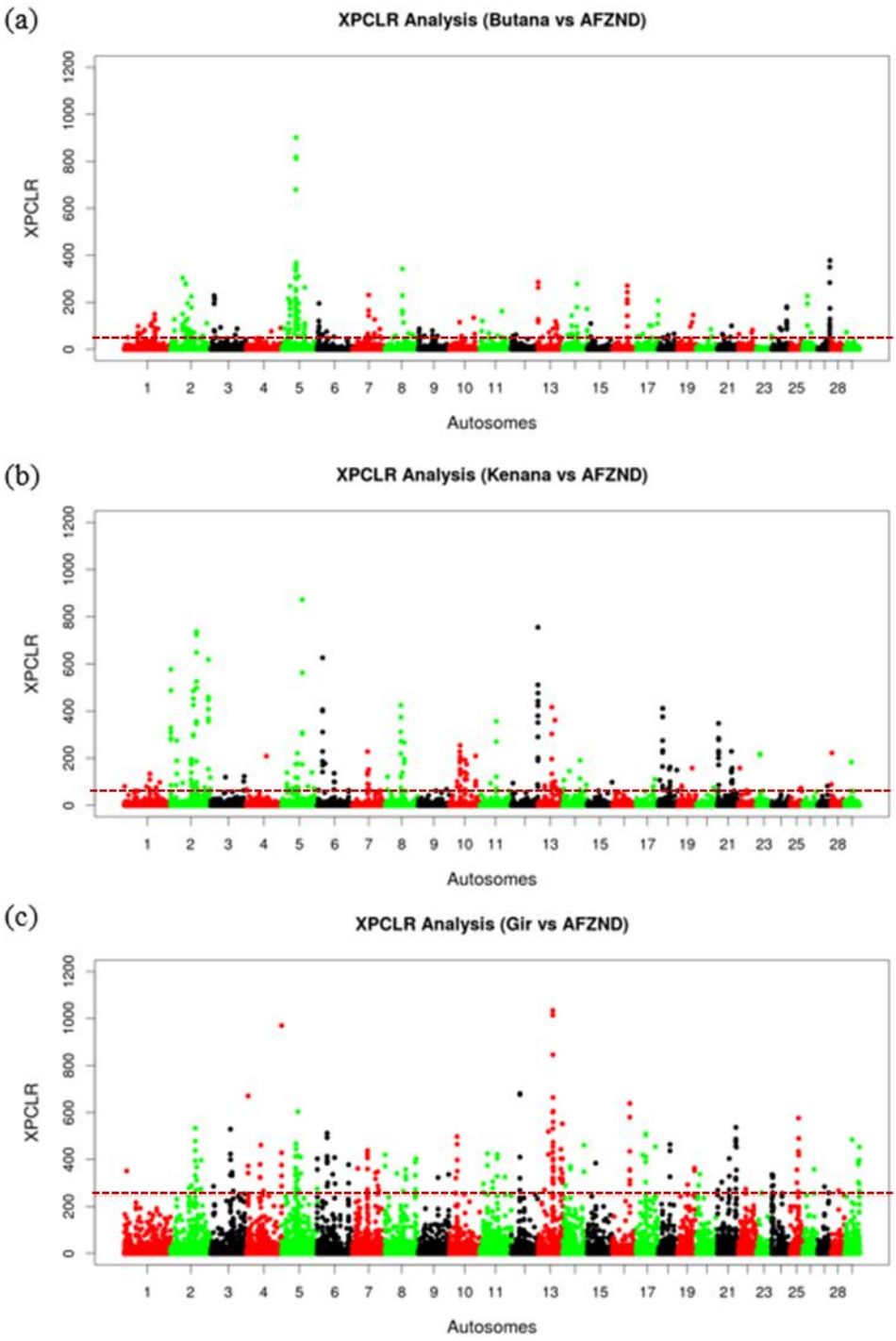


Figure 5.6 | Manhattan plot of the distribution of XPCLR scores along 29 bovine autosomes following the comparison of each of the zebu dairy breeds; (a) Butana, (b) Kenana and (c) Gir, with AFZND. The red dash line indicates the threshold of top 0.5% test scores considered here as evidence of selection.

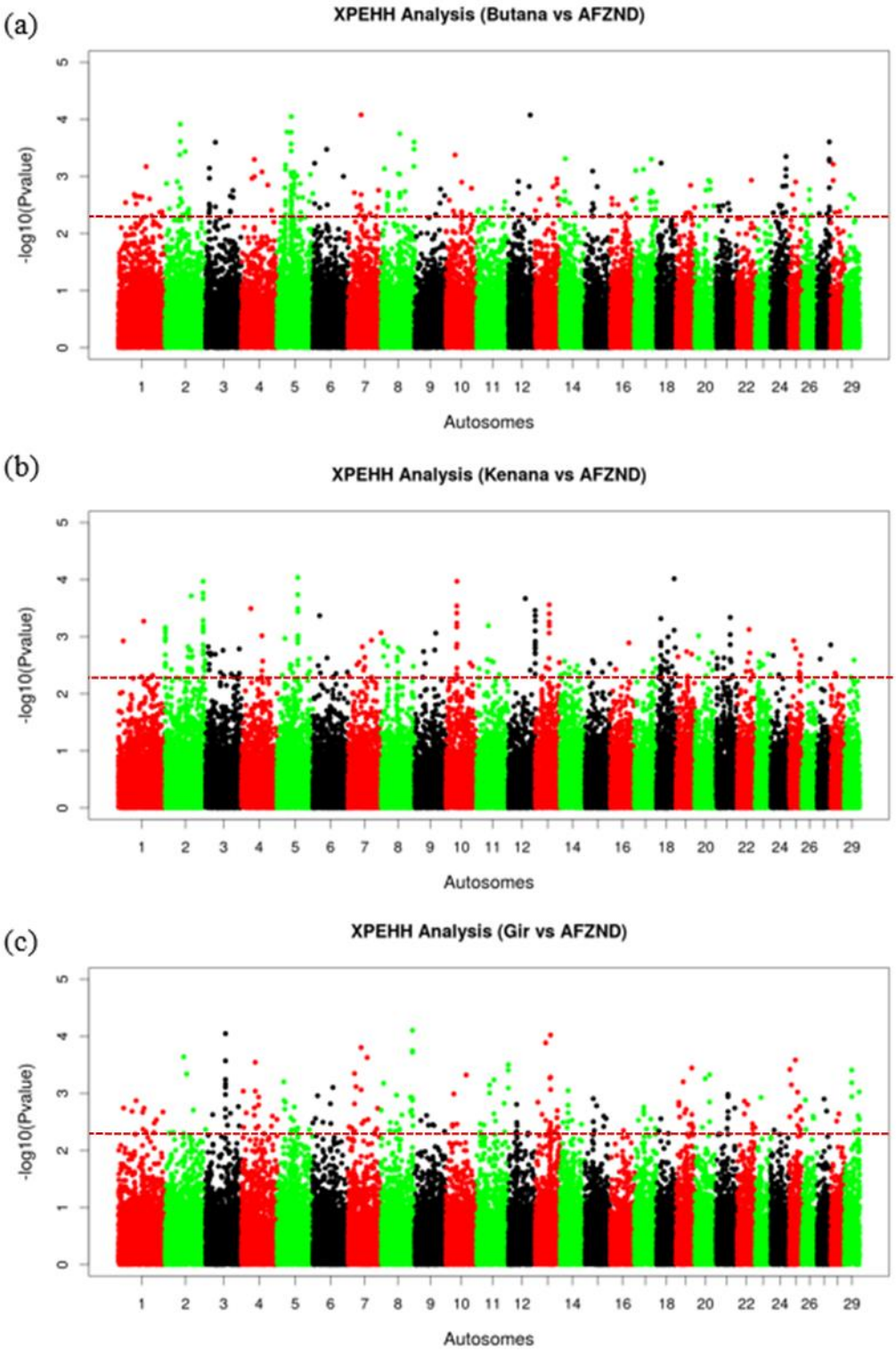


Figure 5.7 | Manhattan plot of the distribution of XPEHH scores along 29 bovine autosomes following the comparison of each of the zebu dairy breeds; (a) Butana, (b) Kenana and (c) Gir, with AFZND. The red dash line indicates the threshold of top 0.5% test scores considered here as evidence of selection.

5.3.5 Functional annotation of candidate genes

To explore further the functions of the candidate genes, we performed functional annotation of the entire gene sets detected by the three population differentiation tests (XPCLR, XPEHH and F_{ST}) using DAVID online bioinformatics tools. We particularly focused on Gene Ontology (GO) term/biological processes that are relevant to tropical adaptation. Enrichment analyses of the different gene sets revealed 540 GO-term biological processes, 25 of which were grouped into annotation clusters based on similar biological meaning and sharing of similar gene members (Supplementary Table S5.22) (Huang et al. 2009).

Of particular interest among the annotation clusters are the first and second clusters, with enrichment scores 2.7 and 2.6, respectively, which are above the recommended threshold of 1.3 (Huang et al. 2009). The first cluster consists of several biological processes related to immune responses. Some of these biological process are humoral immune response (GO:0006959), innate immune response (GO:0045087), adaptive immune response (GO:0002250), B cell proliferation (GO:0042100), T cell activation involved in immune response (GO:0002286), natural killer cell activation (GO:0002323) and cytokine mediated signaling pathway (GO:0019221), etc. The numbers of genes involved in each term ranged from 12 in GO:0002323 to 32 in GO: 0045087. A subset of 32 genes involved in innate immune response are present in several of these biological processes. They are presented in Table 5.6. Specifically, most of the genes are members of the interferon omega-1 genes and tripartite motif containing genes family (Table 5.6).

The second cluster consists of three biological processes: arachidonic acid secretion (GO:0050482), phospholipid metabolic process (GO:0006644) and lipid catabolic process (GO:0016042). The genes involved in this cluster are also presented in Table 5.6. They belong mostly to the phospholipase A2 group. Also included in Table 5.6 are several other randomly selected GO terms that we believed are relevant to the tropical adaptation of the African zebu. For instance, those relevant to immune response in general (e.g. immune response and defense response to bacterium), heat tolerance (e.g. insulin receptor signaling pathway, protein folding, response to hypoxia, response to protein stability, lipid homeostasis, lipid metabolic process), tick resistance (e.g. inflammatory response). Others also considered include feeding intake (e.g. response to glucose, glucose homeostasis, response to starvation and response to food availability), growth (e.g. positive regulation of multicellular organism growth) and reproduction/fertility (e.g. spermatogenesis, sperm capitation, mammary gland development and response to estrogen). The genes related to each of these terms are included in Table 5.6.

Table 5.6 | Some gene ontology term biological processes and associated genes related to the adaptive phenotypes of African zebu population

Phenotype	GO term	Genes	Count	P-value	Fold enrichment
Immunity	GO:0045087-innate immune response	*BPIFB1, FBXO9, FES, FADD, FCER1G, IFNAC, RELB, POLR3H, TNFAIP8L2, TRAF3, YES1, EIF4EBP3, IFNAH, LOC100337470, LOC100298530, LOC100298573, LOC618859, LOC618943, LOC618985, LOC781948, LOC104968438, IFNA16, IFN_TAU, KRT16, MIF, SLPI, TOLLIP, TMEM173, TRIM10, TRIM15, TRIM31, TRIM8*	32	0.00474	1.0
	GO:0006955 -immune response	*JSP.1, BOLA, BLA-DQB, LOC100848815, IGLL1, IL2RA, IL9, BoLA, BOLA-DQA2, BOLA-DQA5, BOLA-DQB, BOLA-DRA, BOLA-DRB3, VPREB1, VPREB2, LOC520402, PRG3, PRG3, SECTM1, LOC781977, SECTM1A, LOC100300442, SERPINB9, SUSD2, TMIGD2, TNFRSF10D*	26	0.05000	1.4
	GO:0042742-defense response to bacterium	*BCL3, FCER1G, APOA2, NOS2*	4	0.97000	0.6
	GO:0050911-detection of chemical stimulus involved in sensory perception of smell	*LOC787041, LOC100295806, LOC786202, LOC100847239, LOC100847281, LOC512296, LOC516132, LOC788246, LOC512948, LOC789504, OR7A17, LOC100847301, LOC533983, LOC786149, LOC787665, LOC618007*	16	0.00001	3.8
Heat tolerance	GO:0032496-response to lipopolysaccharide	*ADAM17, HNRNPA0, IRAK3, SLPI, TBXA2R, TNFRSF10D*	6	0.83646	0.9

	GO:0071222-cellular response to lipopolysaccharide	*AXL, CXCL16, NOS2, RARA*	4	0.94051	0.7
	GO:0008286-insulin receptor signaling pathway	*AKT1, BAIAP2L2, PTPN1*	3	0.84007	0.9
	GO:0006629-lipid metabolic process	*ATP5B, CNEP1R1, LCN12, PRLH*	4	0.84250	0.9
	GO:0016042-lipid catabolic process	*APOC2, PLA2G2D1, PLA2G2D4, IAH1, LIPE, PNPLA2, PLA2G2E, PLA2G2F, PLA2G3, PLA2G5, PLA2G2A, PLA2G2A, LOC615045*	13	0.01290	1.8
	GO:0001666-response to hypoxia	*ADAM17, CBFA2T3, AGER, EGR1, LOXL2, NOS2, P2RX3, RYR1*	8	0.49050	1.2
	GO:0006457-protein folding	*FKBP10, FKBP15, GRPEL2, ACTB, HSPA9, PPIL2, PPIL4, PFDN1, PFDN2, QSOX2, TBCD*	11	0.67443	1.3
	GO:0000209-protein polyubiquitination	*ARIH2, BTRC, MARCH8, PPIL2, RNF14, SOD1, UBE2C, UBE2L3, UBE3C*	9	0.33053	1.4
	GO:0031647-regulation of protein stability	*APOA2, DACT1, KAT2A, MAPK1*	4	0.79101	1.0
	GO:0050821-protein stabilization	*FBXW7, FLOT2, LAMP1, STK4*	4	0.99312	1.0
	GO:0055088-lipid homeostasis	*ABCB4, ACAD9, PNPLA2, PNPLA3*	4	0.44400	1.0
	GO:0009408-response to heat	*LRP11, ASIC3, CKM, P2RX3, SOD1*	5	0.20447	2.1
	GO:0045766-positive regulation of angiogenesis	*GATA6, ALOX12, ERAP1, GDF2, MYDGF, NOS3, PRKCA, RAMP2, RRAS, TMIGD2*	10	0.28270	1.4
Tick resistance	GO:0006954-inflammatory response	*AKT1, AXL, RELB, AGER, GGT5, HNRNPA0, IL17B, IL2RA, KRT16, MIF, NOS2, RARRES2, TBXA2R, TOLLIP, TGFB1, TNFRSF10D*	16	0.80810	0.9
Feeding habit	GO:0009749-response to glucose	*AP0A2, EGR1, PRLH*	3	0.65394	1.3

	GO:0042593-glucose homeostasis	*AKT1, BAD, DBH, SIRT6*	4	0.97921	
	GO:0009267-cellular response to starvation	*RALB, RRAGA, GAS6*	3	0.85052	0.9
	GO:0042755-eating behaviour	*ADORA2A, PYY, PRLH*	3	0.26666	1.0
	GO:0032094-response to food	*AKT1, GAST, GH1*	3	0.36246	2.4
Growth	GO:0040018-positive regulation of multicellular organism growth	*BBS2, GH1, HMGA2, PEX5, STAT5B*	5	0.26011	1.9
	GO:0071363-cellular response to growth factor stimulus	*CIB1, GAS6, AIPK1, STAT5B*	4	0.27965	2.2
Reproduction	GO:0007283-spermatogenesis	*AXL, GLI1, MOV10L1, ACE, AURKC, B4GALNT1, CDYL, CCDC155, GTSF1, GGT1, HMGA2, JAG2, KRT9, MEIOC, ODF3, PIWIL2, PAIP2, RARA, SOD1, TMEM203, TDRD9*	21	0.12128	1.3
	GO:0048240-sperm capacitation	*BSP1, BSP3, BSP5*	3	0.42372	2.1
	GO:0007286-spermatid development	*ADAD1, CIB1, KLHL10, OSBPS, TSSK1B, TSSK2, ZMYND15*	7	0.43638	1.3
	GO:0030879-mammary gland development	*FASN, HOXD9, NOTCH4*	3	0.26666	3.0
	GO:0043627-response to oestrogen	*GATA6, WNT7A, KRT19, SLC34A2*	4	0.27964	2.2

5.4 Discussion

5.4.1 Common genomic signatures within African zebu population

The majority of selected signatures and associated genes identified in this study based on *iHS* and *ZHp* analysis are unique to each breed. However, a number of selection regions are shared among different breeds. Indeed, we identified regions shared by all eight African breeds, some of which are also shared with the Gir zebu cattle from Asia. One of the most frequent selection signatures regions identified in all eight African zebu breeds is a 150 kb region on chromosome 7 (chr7:44.350-44.500) containing two olfactory receptor genes (*LOC785149 and LOC100300085*). The olfactory receptor (OR) proteins are encoded by the largest known multigene family in the mammalian genome (Niimura and Nei 2003).

They are expressed mainly in sensory neurons of the olfactory epithelium in nasal cavities. Although, the mechanism is still unknown, each olfactory neuron is believed to express only one OR gene (Chess *et al.,* 1994, Serizawa *et al.,* 2000, Kratz *et al.,* 2002). The role of these group of genes in mammals is in olfaction (the sense of smell) which is an adaptive trait critical for the survival of animals and relevant to the ability to avoid danger, search for food, identify mates and offspring and male fertility (Niimura and Nei 2003, Spehr *et al.,* 2003, Lee *et al.,* 2013). Genome regions overlapping OR genes are also commonly detected among loci under positive selection in African *Bos taurus* breeds (chapter 3). It is likely that these genes are under selection in indigenous African cattle breeds due largely to their production environments and husbandry systems, in which they are exposed to challenges such as searching for their edible food and water.

We also identified *PTPRG*, *HMGA2*, *CYPI9A1* and *GLDN* among the most frequent candidate genes under selection in the African zebu population. *PTPRG* is identified in all eight African zebu, *HMGA2* is identified in all breeds excluding Ogaden, while the *CYPI9A1* and *GLDN* genes are identified in all breeds, excluding Kenyan Boran. With the exception of *PTPRG*, these genes have been reported to have functions related to growth traits in cattle or in the case of *PTPRG*, in humans and other domestic species (Mao *et al.*, 2016, Bouwman *et al.*, 2018). *PTPRG* has been associated also with fatty acid composition in beef cattle (Zhu *et al.*, 2017).

Specifically, the *PTPRG* gene has been associated with body weight and bone mineral density in mouse (Keightley *et al.*, 1996, Klein *et al.*, 1998) and feed conversion efficiency and body weight in broiler chicken (Reyer *et al.*, 2015). Polymorphisms within the *HMGA2* gene have been associated with heights in humans (Allen *et al.*, 2010, N'Diaye *et al.*, 2011), body size in dogs (Boyko *et al.*, 2010), body composition in pigs (Kim *et al.*, 2004) as well as body size variation, conformation and locomotive traits in horses (Makvandi-Nejad *et al.*, 2012, Frischknecht *et al.*, 2015, Sevane *et al.*, 2017). Recent studies have also shown that the *HMGA2* gene could be associated with stature in cattle due to its role in the control of the expression of the cattle stature–associated *PLAG1* gene (Klemke *et al.*, 2014), and also following the identification of a candidate causal missense variant in a genome-wide association study for cattle stature involving a large number of worldwide cattle samples (Bouwman *et al.*, 2018).

The conformation traits of most African zebu cattle include the presence of a thoracic hump, long legs, narrow body, large body sizes with the average weight of an adult bull varying between 200 kg to more than 400 kg (Cunningham and Syrstad 1987). The

presence of selection signatures in the regions of *PTPRG* and *HMGA2* genes in the majority of African zebu breeds is suggestive of the roles of these genes in the conformational traits of the African zebu breeds which are mostly kept for beef production. These genes are also in candidate region under selection in the Gir zebu cattle from Asia, which has a large body size with an average adult body weight of 313 kg (Gaur *et al.,* 2003). Interestinly, the absence of selection signature overlapping the region of *HMGA2* gene in the Ogaden breed is in agreement with the classification of the breed as a Small East African zebu breed (Rege and Tawah 1999).

Other genes identified in a majority of African zebu cattle include genes with related functions in coat colour (e.g. *MITF*, *n* = 6, (Seitz *et al.*, 1999, Hayes *et al.*, 2010, Kim *et al.,* 2017), residual feed intake (e.g. *PRKCZ,* (Fitzsimons *et al.*, 2014), and immune response genes (e.g. *CDH10,* (Shin *et al.,* 2015). The *PRKCZ* is a particularly interesting gene which plays a role in insulin signaling pathway and has been implicated in glucose uptake (Farese *et al.,* 2005). Periods of dietary restriction in cattle and other species have been reported to coincide with increased insulin sensitivity and glucose uptake (Cartee *et al.,* 1994, Röpke *et al.,* 1994, Sternbauer and Luthman 2002, Barretero-Hernandez *et al.,* 2010). During this period, *PRKCZ* functions in the signaling pathway by regulating the translocation of the glucose transporter *GLUT4* to the cell surface for the uptake of glucose (Keogh *et al.,* 2015). These genes have observed to be differentially expressed in fasted rats but the expression not observed after re-feeding (Charron and Kahn 1990, Kahn 1994).

The immunity-related role of the *CDH10* gene may be in the pathophysiology of tuberculosis, as reported in birds following infection with *Mycobacterium avium*

subspecies *paratuberculosis* (Shin *et al.,* 2015). This gene was identified in all African zebu except Kenana. However, it is noteworthy to mention here another gene, *MARCO,* with related functions in immune responses against bacterial infection, in particular the causative agent of tuberculosis. The MARCO gene was detected in a selective sweep region on chromosome 2 in four Sudanese zebu breeds including Aryashai, Butana, Gash and Kenana. It is likely that this gene is associated with tuberculosis infection in cattle. *MARCO* is a member of the class A scavenger receptor family which have important role in inducing phagocytosis by binding microbial ligands and activating immune responses.

Specifically, *MARCO* play a role in the phagocytosis of bacteria including *Mycobacterium species* and *Salmonella species* as demonstrated in previous studies involving microbial infections in mice and zebra fish (Dorrington *et al.,* 2013, Benard *et al.,* 2014). Polymorphisms in the gene have also been reported to be associated with susceptibility to pulmonary tuberculosis (TB) in humans caused by *Mycobacterium tuberculosis* (Ma *et al.,* 2011, Bowdish *et al.,* 2013, Thuong *et al.,* 2016). Although, *M. tuberculosis* is the causative agent of TB in humans, the incidence of this anthropogenic disease is frequently identified in cattle, and particularly in Africa where the prevalence of the disease is also high in humans. It is believed that the transmission between man and cattle is as a result of the close association of cattle farmers to their animals (Ocepek *et al.,* 2005, Thoen *et al.,* 2008). Sudan happens to be one of the regions in Africa where very high prevalence of TB in cattle herd has been reported (Sulieman and Hamid 2002, Ibrahim *et al.,* 2016). The role of this gene in TB susceptibility in humans and its detection in this study in a candidate region under positive selection in African cattle support cattle as asymptomatic carriers of the humans *Mycobacterium tuberculosis*.

Complementing this, *MARCO* is also identified as a candidate gene in the Gir cattle from India, likely coinciding with the high prevalence of tuberculosis in the region (Konch *et al.*, 2017).

5.4.2 Genomic signatures related to tropical adaptation

Among the environmental characteristics of the pastoral sub-Saharan Africa dryland are high temperature and tick infestation, which negatively affect livestock productivity. High temperature causes heat stress in animals and ticks are vectors of tick-borne diseases. Both decrease animal performance and live weight gain and subsequently lead to loss of revenue to livestock keepers (Jonsson 2006). Indigenous zebu cattle are known to have varying levels of innate tick resistance and superior thermotolerance in the dryland compared to their taurine counterparts (Hansen 2004). The level of tick resistance, however, varies among the different breeds and the physical characteristics of the animals could affect the level of heat stress response. As discussed in details below, out results for breed-specific candidate genes for these two major adaptations are in agreement with these breeds' phenotypes. Both heat tolerance and tick resistance traits could be described as complex traits under the control of several genes, each contributing small effects on the phenotype.

5.4.3 Genetic control of heat tolerance involve complex mechanism

Following our selection scan approaches, we have identified heat tolerance genes based on the overlap with cattle QTL for heat tolerance and with genes previously reported in cattle and other species (Hansen 2004, Kolli 2012, Kolli *et al.*, 2014, Dikmen *et al.*, 2015, Kim *et al.*, 2017).

The genetic control of heat stress response is a complex trait. Accordingly, it will most likely consist of a cascade of biological processes and it will involve a large number of gene networks. The majority of the detected heat tolerance genes in the African zebu population are detected by the comparison tests, XPCLR, XPEHH and F_{ST}. The very few which were detected by within population *ZHp* test are also shared by a majority of the African breeds, indicative a common set of mechanism of heat tolerance in the African zebu population. This is an important finding of our study.

Exposure to heat stress can induce the transcription of a set of stress proteins which enhance cell survival under such conditions (Pickart 1999). Heat stress may lead to the damage of cellular proteins, which in turn triggers the synthesis of heat shock proteins (HSPs) at sub-heat shock temperatures (Goff and Goldberg 1985, Ananthan *et al.*, 1986). Consequently, a key cellular heat stress response is to reduce intracellular levels of denatured proteins, which can be highly toxic or lead to the aggregation of other protein through the exposure of hydrophobic amino acid side chains (Pickart 1999). Generally, the levels of intracellular damaged proteins can be reduced by degradation and/or refolding of the stress-damaged proteins (Parsell and Lindquist 1993). A group of well-known genes with these related protein functions are the HSPs and their co-factors (DnaJ, GroEL and Trigger Factor) (Parsell and Lindquist 1993, Sherman and Goldberg 1996). The induction of HSPs is extremely rapid and very strong during heat stress emergency. It correlates with tolerance to extreme heat in a wide variety of cells and organisms (Parsell and Lindquist 1993).

Based on different population differentiation selection tests, we identified seven members of the heat shock proteins and their co-factors of relevance to heat tolerance

in African zebu. These genes are *DNAJC7*, *HSPB9*, and *HSPB7* (XPCLR test), *HSF5* (XPEHH test), *DNAJ18, HSPA9* and *DNAJC11* (F_{ST} test).

Also, in the GO term "response to heat" (GO:0009408), we identified the protein superoxide dismutase 1 encoded by *SOD1*, a previously reported gene associated with heat tolerance in cattle (Hansen 2013, Kim *et al.,* 2017, Zeng *et al.,* 2018). CuZn-SOD (*SOD1*) expression is localized in the cytoplasm and partially in the intermembrane space of mitochondria, where its functions include catalyzation of the detoxification of denatured protein. Decrease in *SOD1* expression has been reported to increase vulnerability to heat stress (Ishii *et al.,* 2005, Fujii and Imai 2014).

Furthermore, heat stress has been implicated in insulin homeostasis. Being a potent antilipolytic signal, it is frequently elevated in heat stressed animals (Baumgard and Rhoads Jr 2013). A study involving a pig model reported increased insulin secretion during heat stress but at the same time observed decreases in pancreatic insulin content and insulin receptor gene expression at the adipose tissue level. This observation suggests that the increased circulating insulin might be triggered by a secondary signal rather than direct pancreatic hyperthermia (Victoria Sanz Fernandez *et al.,* 2015). However, increases in prolactin and liposaccharide are presumably associated with an increase in circulating insulin in heat stressed animals. During heat stress, prolactin increases β cell proliferation and glucose-stimulated insulin secretion, this has been observed both *in vivo* and *in vitro* (BRELJE and SORENSON 1991, Hughes and Huang 2011). In addition, the increase in liposaccharides observed during heat stress has been linked to increase intestinal permeability to luminal content due to redistribution of blood flow from the viscera to the periphery in an attempt to maximize radiant heat loss

(Lambert *et al.,* 2002). The decrease in intestinal perfusion also leads to mucosal hypoxia which compromises the intestinal barrier function (Hall *et al*. 1999).

The specific role of prolactin in heat stress adaptation is not fully understood but its involvement in water homeostasis and the sweating response has been reported (Alamer 2011). The stress response includes an increase in prolactin expression, as has been reported in a variety of species including humans (Iguchi *et al.,* 2012), pigs (Sanz Fernandez *et al.,* 2012), camels (Kataria and Kataria 2010), sheep (Kataria and Kataria 2011) and cattle (Yayou *et al.,* 2010). In cattle, genetic variants in the prolactin (*PRL*) and prolactin releasing receptor (*PRLR*) genes have been associated with the slick skin phenotype in cattle (Littlejohn *et al.,* 2014, Porto-Neto *et al.,* 2018). Interestingly, within some of the identified biological processes relevant to the aforementioned cellular responses to heat stress, we also identified the prolactin releasing hormone encoded by the *PRLH* gene. With an important function involved in stimulating the release of *PRL* and the regulation of *PRL* expression through binding to the *PRLR*, the *PRLH* gene no doubt is another promising heat tolerance candidate gene within African zebu population (Kim *et al.,* 2017).

The results of this chapter indicate that genetic variants at the *PRLH* and the previously mentioned eight members of the heat shock proteins and their co-factors may represent the major genes contributing to the adaptation of African zebu to the dryland environment. These results pave way for further investigation leading to the identification of the favourable allele - haplotypes within these genes in African zebu,

5.4.4 Evidence of variation in different host responses to tick infestation

Several genomic regions associated with tick burden in cattle have been reported (Martinez *et al.,* 2006, Machado *et al.,* 2010, Porto Neto *et al.,* 2011, Makina *et al.,* 2015). In the study of Machado *et al* (2010), six genomic regions (QTLs) associated with bovine tick resistance were reported following a whole genome scan with microsatellite markers. They also classified the identified QTLs based on dry and rainy seasons as they believed different gene sets could be involved in resistance mechanism. Hence, they reported dry season specific QTL on chromosomes 2 and 10 and rainy season specific QTL on chromosomes 5, 11 and 27. While on chromosomes 23, a QTL was reported to be involved in tick resistance during both seasons (Machado *et al.,* 2010). In this study, following the comparison of the genes detected in the eight African zebu breeds (*iHS* and *ZHp* tests), we identified chromosomal regions with several genes overlapping the candidate tick resistance QTL on chromosomes 2, 3, 4, 10, 11, 13 and 20. The highest number of our overlapping genes was identified on chromosome 2, where a dry season tick resistance specific QTL is found Machado *et al.* (2010) in agreement with the largely dry across the year environment of the African zebu studied here.

However, contrary to the large number of heat tolerance genes within the zebu population, there is minimum overlap of candidate tick resistance genes across African zebu breeds. This is in agreement with previous studies reporting variation in host responses to tick infestation and tick species within and between different cattle breeds (Sutherst *et al.,* 1983, Lemos *et al.,* 1985, De Castro 1991, Norval *et al.,* 1997, Mattioli *et al.,* 2000).

The inflammatory response is believed to be a key defence mechanism exhibited by tick resistant host. Exposure of the infesting tick to the inflammatory response in the host causes significantly lower expression of genes encoding salivary proteins that suppress host immunity, inflammation and coagulation, thereby making the resistant host less attractive to tick larvae (Franzin *et al.,* 2017). Interestingly, functional annotation of the genes detected in the African zebu population revealed 16 genes involved in inflammatory response (GO). One of the genes, *RARRES2* overlap the tick resistance QTL on chromosome 4 (Machado *et al.,* 2010). Retinoic acid receptor responder 2 (*RARRES2*), encoding the novel adipokine chemerin, has been linked to adipose tissue inflammation through activation of the chemokine-like receptor 1 (Tönjes *et al.,* 2014, Helfer and Wu 2018), although, the genetic mechanism underlining chemerin release from adipose tissue are yet unknown. Aside inflammatory response, the role of immune response genes in resistance to tick burden is likely equally important. We identified several immunity-related genes on chromosome 23 consisting of several bovine MHC class II genes including the well-known *BOLA-DRB3* that has been associated with tick resistance in cattle (Martinez *et al.*, 2006). Both the *RARRES2* and *BOLA-DRB3* genes were detected based on XPCLR test, indicative of the possibility of population-wide selection of these genes. We hypothesize that genetic variations within these two genes may contribute significantly to host resistance to tick infestation in African zebu breeds in addition to the breed specific genetic mechanisms.

5.4.5 Towards the unravelling of the genetic basis of milk production trait in African dairy zebu

Milk is an essential part of the diet consumed by humans worldwide. With the increasing human population, the demand for milk continues to be on the rise and it

may become a food security issue if the demand is not met. As the milk consumed by man comes mainly from cows, scientific research aimed at improving the productivity of cows in order to meet the demand for consumption has been ongoing for decades. In the developed nations, successful breeding has led to the development of exotic cattle breeds like the Holstein Friesian, Jersey, among many others which are specialized for milk production. Hence the demand for milk consumption is not exclusively a food security issue in developed nations.

In sub-Saharan Africa, the demand for milk is still a major food security issue because of low local production of milk, which is a direct consequence of the low level of productivity of the indigenous breeds. Generally, the milk production capacity of the indigenous African cattle is very low compared to the specialized dairy breeds in the developed nations. Past efforts to increase milk production in Africa has involved importing exotic dairy breeds into the African continent. A major setback of this endeavour is the inability of the exotic breeds to thrive in the tropics due various environmental challenges including diseases and heat stress.

Looking forwards, an interesting option is to genetically improve the productivity of indigenous breeds, which are already adapted to their local environments. Among the indigenous breeds, Kenana and Butana zebu breeds have been recognized as promising dairy breeds because of their superior milk production capability when compared to other breeds from the region (Musa *et al.,* 2005). The documented levels of their milk production averages 538.26 and 598.73 kg of milk per lactation for Kenana and Butana breeds respectively, under local management (Musa *et al.,* 2006). However, more than three times the local production capacity have been achieved for animals kept within

improved production facilities e.g, 1400 – 2100 kg for Kenana, and up 1662 kg for Butana (Musa *et al.,* 2005, Yousif and El-Moula 2006). These two African dairy breeds are believed to have milk production capacity comparable to the Asian dairy zebu like Gir, Sahiwal and Red Sindhi (Rege *et al.*, 2006)

Until this study, they was no genomic information available in relation to the dairy characteristics of these two dairy zebu breeds, from Africa. Previous reports have largely focused on phenotypic characterization; while a few genetic reports were available they were based on less informative genetic data (Ahmed *et al.,* 2017, Bahbahani *et al.,* 2018). Also, while Kim et al. (2017)'s pioneering study on African cattle genome did include the Kenana, this study does not identify any region specifically related to milk production. To the best of our knowledge, the work reported here is the first whole genome sequence analysis on the identification of genome-wide selection signatures and associated candidate gene of the two breeds with a focus on their dairy characteristics.

An essential components of milk are lipids, which are present in the form of milk fat globules (MFG) and are produced in the milk secreting cells, alveoli, in the internal walls of mammary gland cavities. The polar lipids fraction of the milk fat confers the structural constituents of the fat globule and ensure the stability of the milk emulsion system. However, there is variation in milk yield and composition of milk in different species as well as within species based on a number of factors including season of calving, stage of lactation, lactation order, milking interval, age of animal, diseases and abnormal conditions, feed and nutritional levels.

Considering the dry and hot production environment of the African dairy zebu, their milk production level and milk composition are expected to be affected by environmental factors such as heat, feed and nutritional level and diseases stresses. Heat stress has been reported to affect milk yield as well as dry matter intake in dairy cattle. The effect on milk yield is attributed to endocrine imbalance such as altering the levels of prolactin, thyroid hormones, glucocorticoid, growth hormone, estrogen, progesterone and oxytocin (Pragna *et al.,* 2017). Various studies have also reported declines in total protein content and total fat content as a result of heat stress (Dunn *et al.,* 2014, Yano *et al.*, 2014, Zhu *et al.,* 2017).

In this study, we identified several genes affecting milk production in Butana, Kenana and Gir. Among the most promising previously identified milk production genes, we found *MFGE8* and *CD14* genes in the Butana breeds, but not in Kenana and Gir. While the *CD14* antigen could have functions related to immune response to mastitis, the specific role of *MFGE8* in milk production in not understood but the gene expression is influenced by prolactin (Franchi *et al.,* 2010). The majority of the remaining identified genes have functions related to protein yield, with few genes having roles in fat yield. As both selection for milk traits and heat tolerance affects lipid metabolism, it is possible that a higher selection pressure due to heat stress may have restricted selection with regards to milk fat content. Also, our results suggest that the protein composition of the milk of African zebu could be affected by heat stress and lack of water, with the lack of overlap of candidate milk production genes among the three zebu breeds a result of breed differences in milk composition. This however needs to be further investigated.

CHAPTER 6:

Conclusions and future directions

6.1 Conclusions

The diversity which exists among various species is notably observed as variation in their phenotypes such as physical appearance (morphology), productivity traits, environmental adaptations including heat tolerance and disease resistance. The ultimate goal of any evolutionary scientist is to understand the underlying genetic variation contributing to phenotypic diversity in different species. Since Darwin's seminal work on "The Variation of Plants and Animals under Domestication" (Darwin 1868), livestock breeds diversity have increasingly become models illustrating and supporting the mechanisms of evolution and selection. This is now achieved by accurate identification of loci under selection in the genome and the associated causal mutation(s) through signature of selection studies.

African cattle continue to play a major role in the socioeconomic development and nutritional security in human population in sub-Saharan Africa. The availability of whole genome sequences and the development of new statistical selection scan methodologies for the mining of the genome have offered the opportunity to understand at genome-wide level, the adaptation and productivity traits of all domesticates including the African cattle. Until now, African cattle have been poorly studied. This work aimed to fill this gap by providing deeper genome analyses across two West African taurine and eight African zebu cattle.

The main motivations of the present study were the unravelling of the genetic architecture including genome signatures, associated candidate genes and genetic variants, related to the phenotypic variation observed within indigenous African cattle and in comparison to other non-African cattle breeds. In particular, we focused on

disease resistance trait in trypanotolerant West African taurine breeds, milk production traits in African dairy zebu breeds and tick resistance and heat tolerance traits in African zebu breeds. We adopted a combination of different selection scan approaches including within-population and between population methods in order to identify comprehensive lists of candidate genes associated with signatures of positive selection in the different cattle genomes studied. The between population approach involved the comparison of combined cattle breeds sharing similar phenotype with a group of other cattle breeds having an opposing phenotype in order to identify candidate genes related to the specific phenotype under consideration. We performed functional annotation of the detected candidate genes and specifically strived to elucidate candidate genes putatively linked to the phenotypes studied.

By analysing a very large number of breeds and comparing them with reference breeds, we have been able to unravel candidate genome regions linked to key environmental adaptation and productivity traits. In particular, all results showed that while we do find some breed/population specific adaptation, it seems that there is at least some commonality across breeds for major adaptation traits to environmental challenge. For example, in the investigation of trypanotolerance in West African taurine (WAT), *PTPN6* gene was revealed in both WAT studied breeds as one of the commonly selected genes involved in relevant pathways responding to trypanosome infection in cattle. Also, a protein-protein interaction network indicated its pivotal role through its interactions with the bovine MHC class II genes and other genes, possibly to initiate a cascade of biological processes necessary to confer protective immunity for the survival of infected cattle.

The genetic control of thermotolerance is another adaptive trait, which is shared across many African zebu breeds. Based on this study, several candidate genes related to this trait were identified in the eight African zebu breeds studied including some previously reported, e.g. *HSPA9*, *SOD1*, *PRLH*, as well as a new one, e.g. *MYLK4*.

6.2 Future Directions

While we were able to identify a lot of candidate regions and associated candidate genes, the precise genetic control underlying this adaptation is still unknown. In future, further studies such as functional genomic studies and genome-wide associations especially involving a large cattle sample cohort are required to confirm part of the results reported here. Importantly, precise phenotype measurements will be required to support the roles of candidate causative mutation(s) in our quest to unravel the genetic architecture of environmental adaptation and productivity traits of indigenous African livestock.

Nevertheless, we believe that this study represents a milestone contributing to the understanding of the genetic control of adaptation and productivity traits in African indigenous cattle, with most population now endangered following introduction of exotic breeds and crossbreeding. African cattle diversity is fading away. There is an urgent need to document it and to conserve it through utilisation for the benefit of the farmers. Identifying genomic regions related to these important traits and related causative marker will lead to the possibility of doing marker assisted selection or introgression within breed and in crossbreeding programmes even in the absence of identification of the causative mutation. The findings here thus set the pace for further engagement in defining the biological functions and phenotypes influenced by the

reported genes and it will pave the way for translational livestock productivity projects that will affect the livelihood of livestock keepers in sub-Saharan Africa and contribute to global food security.

References

Aboagye, G., C. Tawah and J. Rege (1994). "Shorthorn cattle of West and Central Africa III. Physical, adaptive and special genetic characteristics." World Anim. Rev **78**(1): 22-32.

Achilli, A., S. Bonfiglio, A. Olivieri, A. Malusa, M. Pala, et al. (2009). "The multifaceted origin of taurine cattle reflected by the mitochondrial genome." PLoS One **4**(6): e5753.

Adebambo, O. A. (2001). "The Muturu: A rare sacred breed of cattle in Nigeria." Animal Genetic Resources/Resources Génétiques Animales/Recursos Genéticos Animales **31**: 27-36.

Ahmed, A. S., S. Rahmatalla, R. Bortfeldt, D. Arends, M. Reissmann, et al. (2017). "Milk protein polymorphisms and casein haplotypes in Butana cattle." Journal of Applied Genetics **58**(2): 261-271.

Ai, H., L. Huang and J. Ren (2013). "Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers." PLoS One **8**(2): e56001.

Akbari, A., J. J. Vitti, A. Iranmehr, M. Bakhtiari, P. C. Sabeti, et al. (2018). "Identifying the favored mutation in a positive selective sweep." Nature Methods.

Alamer, M. (2011). "The role of prolactin in thermoregulation and water balance during heat stress in domestic ruminants." Asian Journal of Animal and Veterinary Advances **6**(12): 1153-1169.

Alexander, D. H., J. Novembre and K. Lange (2009). "Fast model-based estimation of ancestry in unrelated individuals." Genome Research **19**(9): 1655-1664.

Allen, H. L., K. Estrada, G. Lettre, S. I. Berndt, M. N. Weedon, et al. (2010). "Hundreds of variants clustered in genomic loci and biological pathways affect human height." Nature **467**(7317): 832.

Almal, S. H. and H. Padh (2012). "Implications of gene copy-number variation in health and diseases." Journal of Human Genetics **57**.

Alsiddig, M., S. Babiker, M. Galal and A. Mohammed (2010). "Phenotypic characterization of Sudan Zebu cattle (Baggara type)." Research Journal of Animal and Veterinary Sciences **5**: 10-17.

Álvarez, I., L. Pérez-Pardal, A. Traoré, I. Fernández and F. Goyache (2016). "Lack of haplotype structuring for two candidate genes for trypanotolerance in cattle." Journal of Animal Breeding and Genetics **133**(2): 105-114.

Amaral, A. J., L. Ferretti, H.-J. Megens, R. P. Crooijmans, H. Nie, et al. (2011). "Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA." PLoS One **6**(4): e14782.

Amezcua Vesely, M. C., D. A. Bermejo, C. L. Montes, E. V. Acosta-Rodríguez and A. Gruppi (2012). "B-cell response during protozoan parasite infections." Journal of Parasitology Research **2012**.

Amills, M., V. Ramiya, J. Nonmine and H. Lewin (1998). "The major histocompatibility complex." Rev. sci. tech. Off. int. Epiz **17**(1): 108-120.

An, X., L. Wang, J. Hou, G. Li, Y. Song, et al. (2011). "Novel polymorphisms of goat growth hormone and growth hormone receptor genes and their effects on growth traits." Molecular Biology Reports **38**(6): 4037-4043.

Ananthan, J., A. L. Goldberg and R. Voellmy (1986). "Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes." Science **232**(4749): 522-524.

Andersson, L. (2001). "Genetic dissection of phenotypic diversity in farm animals." Nature Reviews Genetics **2**(2): 130-138.

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

Anguita, E., R. Gupta, V. Olariu, P. J. Valk, C. Peterson, et al. (2016). "A somatic mutation of GFI1B identified in leukemia alters cell fate via a SPI1 (PU. 1) centered genetic regulatory network." Developmental Biology **411**(2): 277-286.

Araújo, A. F., B. C. de Alencar, J. R. C. Vasconcelos, M. I. Hiyane, C. R. Marinho, et al. (2005). "CD8+-T-cell-dependent control of Trypanosoma cruzi infection in a highly susceptible mouse strain after immunization with recombinant proteins based on amastigote surface protein 2." Infection and immunity **73**(9): 6017-6025.

Axelsson, E., A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster, et al. (2013). "The genomic signature of dog domestication reveals adaptation to a starch-rich diet." Nature **495**(7441): 360-364.

Bahbahani, H., B. Salim, F. Almathen, F. Al Enezi, J. M. Mwacharo, et al. (2018). "Signatures of positive selection in African Butana and Kenana dairy zebu cattle." PLoS One **13**(1): e0190446.

Bahbahani, H., A. Tijjani, C. Mukasa, D. Wragg, F. Almathen, et al. (2017). "Signatures of Selection for Environmental Adaptation and Zebu× Taurine Hybrid Fitness in East African Shorthorn Zebu." Frontiers in Genetics **8**: 68.

Bakheit, M. and A. Latif (2002). "The innate resistance of Kenana cattle to tropical theileriosis (Theileria annulata infection) in the Sudan." Annals of the New York Academy of Sciences **969**(1): 159-163.

Barbato, M., P. Orozco-terWengel, M. Tapio and M. W. Bruford (2015). "SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data." Frontiers in Genetics **6**: 109.

Barendse, W., B. E. Harrison, R. J. Bunch, M. B. Thomas and L. B. Turner (2009). "Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits." BMC Genomics **10**(1): 178.

Barretero-Hernandez, R., M. Galyean and J. Vizcarra (2010). "The effect of feed restriction on plasma ghrelin, growth hormone, insulin, and glucose tolerance in pigs." The Professional Animal Scientist **26**(1): 26-34.

Barrett, J. C. (1992). The economic role of cattle in communal farming systems in Zimbabwe, Overseas Development Institute London.

Bashir, H. and I. El Zubeir (2013). "Milk Production and Reproduction Performance of Baggara Cattle Raised Under Extensive and Semi-Extensive Systems in South Kordofan State, Sudan." Journal of Animal Production Advances **3**(5): 192-202.

Baumgard, L. H. and R. P. Rhoads Jr (2013). "Effects of heat stress on postabsorptive metabolism and energetics." Annual Review of Animal Biosciences **1**(1): 311-337.

Beaumont, M. A. and D. J. Balding (2004). "Identifying adaptive genetic divergence among populations from genome scans." Molecular Ecology **13**(4): 969-980.

Beaumont, M. A. and R. A. Nichols (1996). "Evaluating loci for use in the genetic analysis of population structure." Proceedings of the Royal Society of London. Series B: Biological Sciences **263**(1377): 1619-1626.

Beghini, A. and F. Lazzaroni (2013). "PTPN6 (protein tyrosine phosphatase, non-receptor type 6)."

Benard, E. L., S. J. Roobol, H. P. Spaink and A. H. Meijer (2014). "Phagocytosis of mycobacteria by zebrafish macrophages is dependent on the scavenger receptor Marco, a key control factor of pro-inflammatory signalling." Developmental & Comparative Immunology **47**(2): 223-233.

Benjelloun, B., F. J. Alberto, I. Streeter, F. Boyer, E. Coissac, et al. (2015). "Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data." Frontiers in Genetics **6**: 107.

Berglund, J., E. M. Nevalainen, A.-M. Molin, M. Perloski, C. Andre, et al. (2012). "Novel origins of copy number variation in the dog genome." Genome Biology **13**.

Bermejo, D. A., S. W. Jackson, M. Gorosito-Serran, E. V. Acosta-Rodriguez, M. C. Amezcua-Vesely, et al. (2013). "Trypanosoma cruzi trans-sialidase initiates a program independent of the transcription factors RORγt and Ahr that leads to IL-17 production by activated B cells." Nature Immunology **14**(5): 514.

Besmer, P., K. Manova, R. Duttlinger, E. J. Huang, A. Packer, et al. (1993). "The kit-ligand (steel factor) and its receptor c-kit/W: pleiotropic roles in gametogenesis and melanogenesis." Development **119**(Supplement): 125-137.

Bharti, K., M. Gasper, J. Ou, M. Brucato, K. Clore-Gronenborn, et al. (2012). "A regulatory loop involving PAX6, MITF, and WNT signaling controls retinal pigment epithelium development." PLoS Genetics **8**(7): e1002757.

Bhuiyan, A. A., J. Li, Z. Wu, P. Ni, A. A. Adetula, et al. (2017). "Exploring the Genetic Resistance to Gastrointestinal Nematodes Infection in Goat Using RNA-Sequencing." International Journal of Molecular Sciences **18**(4): 751.

Bickhart, D. M., Y. Hou, S. G. Schroeder, C. Alkan, M. F. Cardone, et al. (2012). "Copy number variation of individual cattle genomes using next-generation sequencing." Genome Res **22**.

Bijma, P. and J. Woolliams (2002). "Design of sustainable breeding programs in developed countries."

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

Bittante, G. (2011). "Italian animal genetic resources in the Domestic Animal Diversity Information System of FAO." Italian Journal of Animal Science **10**(2): e29.

Blench, R. M. (1998). "Le West African Shorthorn au Nigeria." Des taurins et des hommes: Cameroun, Nigeria.: 249-292.

Blott, S., J.-J. Kim, S. Moisio, A. Schmidt-Küntzel, A. Cornet, et al. (2003). "Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition." Genetics **163**(1): 253-266.

Bock, R., A. De Vos, T. Kingston and D. McLellan (1997). "Effect of breed of cattle on innate resistance to infection with Babesia bovis, B bigemina and Anaplasma marginale." Australian Veterinary Journal **75**(5): 337-340.

Boettcher, P., A. Caroli, A. Stella, S. Chessa, E. Budelli, et al. (2004). "Effects of casein haplotypes on milk production traits in Italian Holstein and Brown Swiss cattle." Journal of Dairy Science **87**(12): 4311-4317.

Bongiorni, S., C. Gruber, S. Bueno, G. Chillemi, F. Ferre, et al. (2016). "Transcriptomic investigation of meat tenderness in two Italian cattle breeds." Animal Genetics **47**(3): 273-287.

Bosse, M., H.-J. Megens, O. Madsen, Y. Paudel, L. A. Frantz, et al. (2012). "Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape." PLoS Genetics **8**(11): e1003100.

Bourn, D., W. Wint, R. Blench and E. Woolley (1994). "Nigerian livestock resources survey." World Animal Review **78**(1): 49-58.

Bouwman, A. C., H. D. Daetwyler, A. J. Chamberlain, C. H. Ponce, M. Sargolzaei, et al. (2018). "Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals." Nature Genetics **50**(3): 362-367.

Bowdish, D. M., K. Sakamoto, N. A. Lack, P. C. Hill, G. Sirugo, et al. (2013). "Genetic variants of MARCO are associated with susceptibility to pulmonary tuberculosis in a Gambian population." BMC Medical Genetics **14**(1): 47.

Boyko, A. R., P. Quignon, L. Li, J. J. Schoenebeck, J. D. Degenhardt, et al. (2010). "A simple genetic architecture underlies morphological variation in dogs." PLoS Biology **8**(8): e1000451.

Bradley, D., D. MacHugh, R. Loftus, R. Sow, C. Hoste, et al. (1994). "Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations." Animal Genetics **25**(S2): 7-12.

Bradley, D. G., D. E. MacHugh, P. Cunningham and R. T. Loftus (1996). "Mitochondrial diversity and the origins of African and European cattle." Proceedings of the National Academy of Sciences **93**(10): 5131-5135.

Bradley, D. G. and D. A. Magee (2006). "Genetics and the origins of domestic cattle." Documenting domestication: new genetic and archaeological paradigms: 317-328.

Braud, M., D. A. Magee, S. D. Park, T. S. Sonstegard, S. M. Waters, et al. (2017). "Genome-wide microRNA binding site variation between extinct wild aurochs and modern cattle identifies candidate microRNA-regulated domestication genes." Frontiers in Genetics **8**: 3.

BRELJE, T. C. and R. L. SORENSON (1991). "Role of prolactin versus growth hormone on islet BCell proliferation in vitro: implications for pregnancy." Endocrinology **128**(1): 45-57.

Brener, Z. and R. T. Gazzinelli (1997). "Immnunological Control of Trypanosoma cruzi infection and pathogenesis of Chagas' disease." International archives of allergy and Immunology **114**(2): 103-110.

Browning, S. R. and B. L. Browning (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.

Buchmann, R. and S. Hazelhurst (2014). "Genesis manual." University of the Witwatersrand, Johannesburg.(available at: http://www. bioinf. wits. ac. za/software/genesis/Genesis. pdf.

Cañón, J., P. Alexandrino, I. Bessa, C. Carleos, Y. Carretero, et al. (2001). "Genetic diversity measures of local European beef cattle breeds for conservation purposes." Genetics Selection Evolution **33**(3): 311.

Cartee, G., E. Kietzke and C. Briggs-Tung (1994). "Adaptation of muscle glucose transport with caloric restriction in adult, middle-aged, and old rats." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **266**(5): R1443-R1447.

Carvalho, F., M. Lammoglia, M. Simoes and R. Randel (1995). "Breed affects thermoregulation and epithelial morphology in imported and native cattle subjected to heat stress." Journal of Animal Science-Menasha then Albany then Champaign Illinois-**73**: 3570-3573.

Carvalho, F., M. Lammoglia, M. Simoes and R. Randel (1995). "Breed affects thermoregulation and epithelial morphology in imported and native cattle subjected to heat stress." Journal of Animal Science **73**(12): 3570-3573.

Casas, E., S. White, D. Riley, T. Smith, R. Brenneman, et al. (2005). "Assessment of single nucleotide polymorphisms in genes residing on chromosomes 14 and 29 for association with carcass composition traits in *Bos indicus* cattle." Journal of Animal Science **83**(1): 13-19.

Chabory, E., C. Damon, A. Lenoir, J. Henry-Berger, P. Vernet, et al. (2010). "Mammalian glutathione peroxidases control acquisition and maintenance of spermatozoa integrity 1." Journal of Animal Science **88**(4): 1321-1331.

Charron, M. J. and B. B. Kahn (1990). "Divergent molecular mechanisms for insulin-resistant glucose transport in muscle and adipose cells in vivo." Journal of Biological Chemistry **265**(14): 7994-8000.

Chen, C., R. Qiao, R. Wei, Y. Guo, H. Ai, et al. (2012). "A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits." BMC Genomics **13**(1): 733.

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

Chess, A., I. Simon, H. Cedar and R. Axel (1994). "Allelic inactivation regulates olfactory receptor gene expression." Cell **78**(5): 823-834.

Choi, J.-W., B.-H. Choi, S.-H. Lee, S.-S. Lee, H.-C. Kim, et al. (2015). "Whole-genome resequencing analysis of Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection." Molecules and Cells **38**(5): 466.

Cicconardi, F., G. Chillemi, A. Tramontano, C. Marchitelli, A. Valentini, et al. (2013). "Massive screening of copy number population-scale variation in *Bos taurus* genome." BMC Genomics **14**.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, 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.

Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loor, A. Everts-van der Wind, et al. (2005). "Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle." Genome Research **15**(7): 936-944.

Collier, P. and J. W. Gunning (1999). "Explaining African economic performance." Journal of Economic Literature **37**(1): 64-111.

Collins, F. S., M. Morgan and A. Patrinos (2003). "The Human Genome Project: lessons from large-scale biology." Science **300**(5617): 286-290.

Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, et al. (2010). "Origins and functional impact of copy number variation in the human genome." Nature **464**.

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

Corbin, L. J., A. Liu, S. Bishop and J. Woolliams (2012). "Estimation of historical effective population size using linkage disequilibria with marker data." Journal of Animal Breeding and Genetics **129**(4): 257-270.

Cosart, T., A. Beja-Pereira, S. Chen, S. B. Ng, J. Shendure, et al. (2011). "Exome-wide DNA capture and next generation sequencing in domestic and wild species." BMC Genomics **12**(1): 347.

Courtin, F., V. Jamonneau, G. Duvallet, A. Garcia, B. Coulibaly, et al. (2008). "Sleeping sickness in West Africa (1906–2006): changes in spatial repartition and lessons from the past." Tropical Medicine & International Health **13**(3): 334-344.

Cui, D., F. Li, Q. Li, J. Li, Y. Zhao, et al. (2015). "Generation of a miniature pig disease model for human Laron syndrome." Scientific Reports **5**: 15603.

Cunningham, E. and O. Syrstad (1987). "Types and breeds of tropical and temperate cattle." Crossbreeding *Bos indicus* and *Bos taurus* for milk production in the Tropics. FAO Paper **68**.

Curi, R. A., H. De Oliveira, A. C. Silveira and C. Lopes (2005). "Association between IGF-I, IGF-IR and GHRH gene polymorphisms and growth and carcass traits in beef cattle." Livestock Production Science **94**(3): 159-167.

da Silva, J. M., P. F. Giachetto, L. O. da Silva, L. C. Cintra, S. R. Paiva, et al. (2016). "Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits." BMC Genomics **17**(1): 454.

Daetwyler, H., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, et al. (2013). "The 1000 bull genomes project." Nature Genetics.

Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen, et al. (2014). "Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle." Nature Genetics **46**(8): 858-865.

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

Dang, Y., M. Li, M. Yang, X. Cao, X. Lan, et al. (2014). "Identification of bovine NPC1 gene cSNPs and their effects on body size traits of Qinchuan cattle." Gene **540**(2): 153-160.

Das, A., F. Panitz, V. R. Gregersen, C. Bendixen and L.-E. Holm (2015). "Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes." BMC Genomics **16**(1): 1043.

Dauber, A., R. G. Rosenfeld and J. N. Hirschhorn (2014). "Genetic evaluation of short stature." The Journal of Clinical Endocrinology & Metabolism **99**(9): 3080-3092.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, et al. (2011). "Genome-wide genetic marker discovery and genotyping using next-generation sequencing." Nature Reviews Genetics **12**(7): 499-510.

Dayo, G. K., S. Thevenon, D. Berthier, K. MOAZAMI-GOUDARZI, C. Denis, et al. (2009). "Detection of selection signatures within candidate regions underlying trypanotolerance in outbred cattle populations." Molecular Ecology **18**(8): 1801-1813.

De Castro, J. (1991). "Resistance to ixodid ticks in cattle with an assessment of its role in tick control in Africa." Breeding for disease resistance in farm animals. CABI, Wallingford, UK: 244-262.

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

Delaneau, O., Zagury, J. F., & Marchini, J. (2012). Improved whole-chromosome phasing for disease and population genetic studies. Nature Methods, 10(1), 5.

Denbow, J. R. and E. N. Wilmsen (1986). "Advent and course of pastoralism in the Kalahari." Science **234**(4783): 1509-1515.

Dikmen, S., X. z. Wang, M. S. Ortega, J. Cole, D. Null, et al. (2015). "Single nucleotide polymorphisms associated with thermoregulation in lactating dairy cows exposed to heat stress." Journal of Animal Breeding and Genetics **132**(6): 409-419.

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

Dorrington, M. G., A. M. Roche, S. E. Chauvin, Z. Tu, K. L. Mossman, et al. (2013). "MARCO is required for TLR2-and Nod2-mediated responses to Streptococcus pneumoniae and clearance of pneumococcal colonization in the murine nasopharynx." The Journal of Immunology **190**(1): 250-258.

Duangjinda, M., Y. Jindatajak, W. Tipvong, J. Sriwarothai, V. Pattarajinda, et al. (2013). "Association of BoLA-DRB3 alleles with tick-borne disease tolerance in dairy cattle in a tropical environment." Veterinary Parasitology **196**(3): 314-320.

Dunn, R. J., N. E. Mead, K. M. Willett and D. E. Parker (2014). "Analysis of heat stress in UK dairy cattle and impact on milk yields." Environmental Research Letters **9**(6): 064006.

Eck, S. H., A. Benet-Pagès, K. Flisikowski, T. Meitinger, R. Fries, et al. (2009). "Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery." Genome Biology **10**(8): R82.

Edea, Z., H. Dadi, S.-W. Kim, T. Dessie, T. Lee, et al. (2013). "Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers." Frontiers in Genetics **4**: 35.

Ellegren, H. and B. C. Sheldon (2008). "Genetic basis of fitness differences in natural populations." Nature **452**(7184): 169-175.

Epstein, H. (1971). The origin of the domestic animals of Africa, Africana Publishing Corporation.

Epstein, H. (1984). The Origin ofthe Domestic Animals of Africa, Vol. 1, Africana Publishing Corporation, New York. Epstein H., Mason IL.

Erhardt, G. and C. Weimann (2007). "Use of molecular markers for evaluation of genetic diversity and in animal production." Archivos Latinoamericanos de Produccion Animal **15**(S1): 63-66.

Excoffier, L., T. Hofer and M. Foll (2009). "Detecting loci under selection in a hierarchically structured population." Heredity **103**(4): 285-298.

Fadista, J., M. Nygaard, L.-E. Holm, B. Thomsen and C. Bendixen (2008). "A snapshot of CNVs in the Pig genome." PLoS One **3**.

FAO (2013). "In Vivo Conservation of Animal Genetic Resources."

Farese, R., M. Sajan and M. Standaert (2005). Atypical protein kinase C in insulin action and insulin resistance, Portland Press Limited.

Fariello, M.-I., B. Servin, G. Tosser-Klopp, R. Rupp, C. Moreno, et al. (2014). "Selection signatures in worldwide sheep populations." PLoS One **9**(8): e103813.

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

Felius, M. (1995). Cattle breeds: An encyclopedia, C Misset bv.

Felius, M., M.-L. Beerling, D. S. Buchanan, B. Theunissen, P. A. Koolmees, et al. (2014). "On the history of cattle genetic resources." Diversity **6**(4): 705-750.

Fitzsimons, C., D. Kenny, S. Waters, B. Earley and M. McGee (2014). "Effects of phenotypic residual feed intake on response to a glucose tolerance test and gene expression in the insulin signaling pathway in longissimus dorsi in beef cattle." Journal of Animal Science **92**(10): 4616-4631.

Flori, L., S. Fritz, F. Jaffrézic, M. Boussaha, I. Gut, et al. (2009). "The genome response to artificial selection: a case study in dairy cattle." PLoS One **4**(8): e6595.

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

Foll, M. and O. Gaggiotti (2008). "A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective." Genetics **180**(2): 977-993.

Fontanesi, L., F. Beretti, P. L. Martelli, M. Colombo, S. Dall'olio, et al. (2011). "A first comparative map of copy number variations in the sheep genome." Genomics **97**.

Fontanesi, L., P. L. Martelli, F. Beretti, V. Riggio, S. Dall'olio, et al. (2010). "An initial comparative map of copy number variations in the goat (Capra hircus) genome." BMC Genomics **11**.

Fontanesi, L., M. Tazzoli, V. Russo and J. Beever (2010). "Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus." Animal Genetics **41**(3): 295-303.

Franchi, A., S. Bocca, S. Anderson, R. Riggs and S. Oehninger (2010). "Expression of milk fat globule EGF-factor 8 (MFG-E8) mRNA and protein in the human endometrium and its regulation by prolactin." Molecular Human Reproduction **17**(6): 360-371.

Franzin, A. M., S. R. Maruyama, G. R. Garcia, R. P. Oliveira, J. M. C. Ribeiro, et al. (2017). "Immune and biochemical responses in skin differ between bovine hosts genetically susceptible and resistant to the cattle tick Rhipicephalus microplus." Parasites & Vectors **10**(1): 51.

Freeman, A., D. Bradley, S. Nagda, J. Gibson and O. Hanotte (2006). "Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle." Animal Genetics **37**(1): 1-9.

Frisch, J. and C. O'Neill (1998). "Comparative evaluation of beef cattle breeds of African, European and Indian origins. 2. Resistance to cattle ticks and gastrointestinal nematodes." Animal Science **67**(1): 39-48.

Frischknecht, M., V. Jagannathan, P. Plattet, M. Neuditschko, H. Signer-Hasler, et al. (2015). "A non-synonymous HMGA2 variant decreases height in Shetland ponies and other small horses." PLoS One **10**(10): e0140749.

Fujii, J. and H. Imai (2014). "Redox reactions in mammalian spermatogenesis and the potential targets of reactive oxygen species under oxidative stress." Spermatogenesis **4**(2): e979108.

Garrigan, D. and P. W. Hedrick (2003). "Perspective: detecting adaptive molecular polymorphism: lessons from the MHC." Evolution **57**(8): 1707-1722.

Gaughan, J., T. Mader, S. Holt, M. Josey and K. Rowan (1999). "Heat tolerance of Boran and Tuli crossbred steers." Journal of Animal Science **77**(9): 2398-2405.

Gaur, G., S. Kaushik and R. Garg (2003). "The Gir cattle breed of India-characteristics and present status." Animal Genetic Resources/Resources Génétiques animales/Recursos Genéticos Animales **33**: 21-29.

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

Gautier, M. and M. Naves (2011). "Footprints of selection in the ancestral admixture of a New World Creole cattle breed." Molecular Ecology **20**(15): 3128-3143.

Gibbs, R., G. Weinstock, S. Kappes, L. Schook, L. Skow, et al. (2002). Bovine genomic sequencing initiative.

Goddard, A. D., R. Covello, S.-M. Luoh, T. Clackson, K. M. Attie, et al. (1995). "Mutations of the growth hormone receptor in children with idiopathic short stature." New England Journal of Medicine **333**(17): 1093-1098.

Goff, S. A. and A. L. Goldberg (1985). "Production of abnormal proteins in E. coli stimulates transcription of Ion and other heat shock genes." Cell **41**(2): 587-595.

Goodnight, K. and D. Queller (1999). "Computer software for performing likelihood tests of pedigree relationship using genetic markers." Molecular Ecology **8**(7): 1231-1234.

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

Granka, J. M., B. M. Henn, C. R. Gignoux, J. M. Kidd, C. D. Bustamante, et al. (2012). "Limited evidence for classic selective sweeps in African populations." Genetics **192**(3): 1049-1064.

Gray, I. C., D. A. Campbell and N. K. Spurr (2000). "Single nucleotide polymorphisms as tools in human genetics." Human Molecular Genetics **9**(16): 2403-2408.

Grigson, C. (1978). "The craniology and relationships of four species of Bos: 4. The Relationship between *Bos primigenius* Boj. and *B. taurus* L. and its implications for the Phylogeny of the Domestic Breeds." Journal of Archaeological Science **5**(2): 123-152.

Grigson, C. (1980). "The craniology and relationships of four species of Bos 5. *Bos iudicus* L." Journal of Archaeological Science **7**(1): 3-32.

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

Groeneveld, L., J. Lenstra, H. Eding, M. Toro, B. Scherf, et al. (2010). "Genetic diversity in farm animals–a review." Animal Genetics **41**(s1): 6-31.

Guirnalda, P., N. B. Murphy, D. Nolan and S. J. Black (2007). "Anti-Trypanosoma brucei activity in Cape buffalo serum during the cryptic phase of parasitemia is mediated by antibodies." International Journal for Parasitology **37**(12): 1391-1399.

Guo, J., H. Tao, P. Li, L. Li, T. Zhong, et al. (2018). "Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds." Scientific Reports **8**(1): 10405.

Guo, Q., D. Miller, H. An, H. Wang, J. Lopez, et al. (2016). "Controlled Heat Stress Promotes Myofibrillogenesis during Myogenesis." PLoS One **11**(11): e0166294.

Gwakisa, P., S. Kemp and A. Teale (1994). "Characterization of Zebu cattle breeds in Tanzania using random amplified polymorphic DNA markers." Animal Genetics **25**(2): 89-94.

Gwaza, D. and O. Momoh (2016). "Endangered indigenous cattle Breeds of Nigeria a case for their conservation and management." World Scientific News **30**: 68-88.

Hammond, A., T. Olson, C. Chase, E. Bowers, R. Randel, et al. (1996). "Heat tolerance in two tropically adapted *Bos taurus* breeds, Senepol and Romosinuano, compared with Brahman, Angus, and Hereford cattle in Florida." Journal of Animal Science **74**(2): 295-303.

Hanotte, O., D. G. Bradley, J. W. Ochieng, Y. Verjee, E. W. Hill, et al. (2002). "African pastoralism: genetic imprints of origins and migrations." Science **296**(5566): 336-339.

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

Hanotte, O., Y. Ronin, M. Agaba, P. Nilsson, A. Gelhaus, et al. (2003). "Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle." Proceedings of the National Academy of Sciences **100**(13): 7443-7448.

Hanotte, O., C. Tawah, D. Bradley, M. Okomo, Y. Verjee, et al. (2000). "Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds." Molecular Ecology **9**(4): 387-396.

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

Hansen, P. (2013). "Cellular and molecular basis of therapies to ameliorate effects of heat stress on embryonic development in cattle." Animal Reproduction. **10**(3): 322-333.

Harris, M., N. K. Bose, M. Klass, J. P. Mencher, K. Oberg, et al. (1966). "The Cultural Ecology of India's Sacred Cattle [and Comments and Replies]." Current Anthropology **7**(1): 51-66.

Hartl, D. L., A. G. Clark and A. G. Clark (1997). Principles of population genetics, Sinauer associates Sunderland.

Hauswirth, R., B. Haase, M. Blatter, S. A. Brooks, D. Burger, et al. (2012). "Mutations in MITF and PAX3 cause “splashed white” and other white spotting phenotypes in horses." PLoS Genetics **8**(4): e1002653.

Hauswirth, R., R. Jude, B. Haase, R. R. Bellone, S. Archer, et al. (2013). "Novel variants in the KIT and PAX 3 genes in horses with white-spotted coat colour phenotypes." Animal Genetics **44**(6): 763-765.

Hayes, B., A. Chamberlain, S. Maceachern, K. Savin, H. McPartlan, et al. (2009). "A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle." Animal Genetics **40**(2): 176-184.

Bovine HapMap Consortium, (2009). "Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds."

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman and M. E. Goddard (2010). "Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits." PLoS Genetics **6**(9): e1001139.

Hayes, B. J., P. M. Visscher, H. C. McPartlan and M. E. Goddard (2003). "Novel multilocus measure of linkage disequilibrium to estimate past effective population size." Genome Research **13**(4): 635-643.

Hedrick, P. W. (2001). "Conservation genetics: where are we now?" Trends in Ecology & Evolution **16**(11): 629-636.

Helfer, G. and Q.-F. Wu (2018). "Chemerin: a multifaceted adipokine involved in metabolic disorders." Journal of Endocrinology: JOE-18-0174.

Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. Taylor, R. Ogden, et al. (2011). "Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges." Molecular Ecology Resources **11**(s1): 123-136.

Hemesath, T. J., E. Steingrímsson, G. McGill, M. J. Hansen, J. Vaught, et al. (1994). "microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family." Genes & Development **8**(22): 2770-2780.

Hernández-Cerón, J., C. Chase and P. Hansen (2004). "Differences in heat tolerance between preimplantation embryos from Brahman, Romosinuano, and Angus breeds." Journal of Dairy Science **87**(1): 53-58.

Horvath, G. C., W. S. Kistler and M. K. Kistler (2004). "RFX2 is a potential transcriptional regulatory factor for histone H1t and other genes expressed during the meiotic phase of spermatogenesis." Biology of Reproduction **71**(5): 1551-1559.

Hou, Y., G. E. Liu, D. M. Bickhart, L. K. Matukumalli, C. Li, et al. (2012). "Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle." Functional & Integrative Genomics **12**.

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

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

Huang, D. W., B. T. Sherman and R. A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nature Protocols **4**(1): 44-57.

Hughes, E. and C. Huang (2011). "Participation of Akt, menin, and p21 in pregnancy-induced β-cell proliferation." Endocrinology **152**(3): 847-855.

Ibeagha-Awemu, E. M., O. C. Jann, C. Weimann and G. Erhardt (2004). "Genetic diversity, introgression and relationships among West/Central African cattle breeds." Genetics Selection Evolution **36**(6): 673-690.

Ibeagha-Awemu, E. and G. Erhardt (2005). "Genetic structure and differentiation of 12 African *Bos indicus* and *Bos taurus* cattle breeds, inferred from protein and microsatellite polymorphisms." Journal of Animal Breeding and Genetics **122**(1): 12-20.

Ibrahim, S., U. Abubakar, S. Danbirni, A. Usman, F. Ballah, et al. (2016). "Molecular identification of Mycobacterium tuberculosis transmission between cattle and man: a case report." Journal of Microbiology & Experimentation **3**(Suppl 3): 00091.

Iguchi, M., A. E. Littmann, S.-H. Chang, L. A. Wester, J. S. Knipper, et al. (2012). "Heat stress and cardiovascular, hormonal, and heat shock proteins in humans." Journal of Athletic Training **47**(2): 184-190.

Inoue, H., N. Kangawa, A. Kinouchi, Y. Sakamoto, C. Kimura, et al. (2011). "Identification and functional analysis of novel human growth hormone–releasing hormone receptor (GHRHR) gene mutations in Japanese subjects with short stature." Clinical Endocrinology **74**(2): 223-233.

Ionita-Laza, I., A. J. Rogers, C. Lange, B. A. Raby and C. Lee (2009). "Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis." Genomics **93**.

Ishii, T., S. Matsuki, Y. Iuchi, F. Okada, S. Toyosaki, et al. (2005). "Accelerated impairment of spermatogenic cells in SOD1-knockout mice under heat stress." Free Radical Research **39**(7): 697-705.

Jabbar, M. A., B. Swallow, G. d'Iteren and A. Busari (1998). "Farmer preferences and market values of cattle breeds of West and Central Africa." Journal of Sustainable Agriculture **12**(2-3): 21-47.

Jiang, L., J. Jiang, J. Wang, X. Ding, J. Liu, et al. (2012). "Genome-wide identification of copy number variations in chinese Holstein." PLoS One **7**.

Jin, Y., T. Zhou, X. Geng, S. Liu, A. Chen, et al. (2017). "A genome-wide association study of heat stress-associated SNPs in catfish." Animal Genetics **48**(2): 233-236.

Jonsson, N. (2006). "The productivity effects of cattle tick (Boophilus microplus) infestation on cattle, with particular reference to *Bos indicus* cattle and their crosses." Veterinary Parasitology **137**(1-2): 1-10.

Jonsson, N. N., R. Bock and W. Jorgensen (2008). "Productivity and health effects of anaplasmosis and babesiosis on *Bos indicus* cattle and their crosses, and the effects of differing intensity of tick control in Australia." Veterinary Parasitology **155**(1-2): 1-9.

Kahn, B. B. (1994). "Dietary regulation of glucose transporter gene expression: tissue specific effects in adipose cells and muscle." The Journal of Nutrition **124**(suppl_8): 1289S-1295S.

Kamwanja, L., C. Chase, J. Gutierrez, V. Guerriero, T. Olson, et al. (1994). "Responses of bovine lymphocytes to heat shock as modified by breed and antioxidant status." Journal of Animal Science **72**(2): 438-444.

Kantanen, J., I. Olsaker, L.-E. Holm, S. Lien, J. Vilkki, et al. (2000). "Genetic diversity and population structure of 20 North European cattle breeds." Journal of Heredity **91**(6): 446-457.

Karsunky, H., I. Mende, T. Schmidt and T. Möröy (2002). "High levels of the onco-protein Gfi-1 accelerate T-cell proliferation and inhibit activation induced T-cell death in Jurkat T-cells." Oncogene **21**(10): 1571.

Kataria, N. and A. K. Kataria (2010). "Can prolactin be a measurable marker of stress in dromedaries." Slovenian Veterinary Research **47**(3): 133-138.

Kataria, N. and A. K. Kataria (2011). "Use of prolactin as an indicator of stress in Marwari sheep from arid tracts in India." Comparative Clinical Pathology **20**(4): 333-336.

Kecskés, M., G. Jacobs, S. Kerselaers, N. Syam, A. Menigoz, et al. (2015). "The Ca 2+-activated cation channel TRPM4 is a negative regulator of angiotensin II-induced cardiac hypertrophy." Basic Research in Cardiology **110**(4): 43.

Keightley, P. D., T. Hardge, L. May and G. Bulfield (1996). "A genetic map of quantitative trait loci for body weight in the mouse." Genetics **142**(1): 227-235.

Keogh, K., D. A. Kenny, A. K. Kelly and S. M. Waters (2015). "Insulin secretion and signaling in response to dietary restriction and subsequent re-alimentation in cattle." Physiological Genomics **47**(8): 344-354.

Kilian, B. and A. Graner (2012). "NGS technologies for analyzing germplasm diversity in genebanks." Briefings in Functional Genomics: elr046.

Kim, J., O. Hanotte, O. A. Mwai, T. Dessie, S. Bashir, et al. (2017). "The genome landscape of indigenous African cattle." Genome Biology **18**(1): 34.

Kim, K. S., H. Thomsen, J. Bastiaansen, N. T. Nguyen, J. Dekkers, et al. (2004). "Investigation of obesity candidate genes on porcine fat deposition quantitative trait loci regions." Obesity **12**(12): 1981-1994.

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

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

Klein, R. F., S. R. Mitchell, T. J. Phillips, J. K. Belknap and E. S. Orwoll (1998). "Quantitative trait loci affecting peak bone mineral density in mice." Journal of Bone and Mineral Research **13**(11): 1648-1656.

Klemke, M., M. H. Müller, W. Wosniok, D. N. Markowski, R. Nimzyk, et al. (2014). "Correlated expression of HMGA2 and PLAG1 in thyroid tumors, uterine leiomyomas and experimental models." PLoS One **9**(2): e88126.

Ko, J. M., J. Y. Park and H. W. Yoo (2009). "Common exon 3 polymorphism of the GH receptor (GHR) gene and effect of GH therapy on growth in Korean children with idiopathic short stature (ISS)." Clinical Endocrinology **70**(1): 82-87.

Kolli, V. (2012). Gene expression profiling in indigeneous cattle and buffalo peripheral blood leukocytes in response to heat stress, NDRI, Karnal.

Kolli, V., R. Upadhyay and D. Singh (2014). "Peripheral blood leukocytes transcriptomic signature highlights the altered metabolic pathways by heat stress in zebu cattle." Research in Veterinary Science **96**(1): 102-110.

Konch, P., B. Dutta, S. Goswami, A. Barua and G. Saikia (2017). "Prevalence of bovine tuberculosis in Assam, India." International Journal of Chemical Studies **5**(3): 143-146.

Kratz, E., J. C. Dugas and J. Ngai (2002). "Odorant receptor gene regulation: implications from genomic organization." Trends in Genetics **18**(1): 29-34.

Kreitman, M. (2000). "Methods to detect selection in populations with applications to the human." Annual Review of Genomics and Human Genetics **1**(1): 539-559.

Kristjanson, P. M., B. M. Swallow, G. Rowlands, R. Kruska and P. De Leeuw (1999). "Measuring the costs of African animal trypanosomosis, the potential benefits of control and returns to research." Agricultural Systems **59**(1): 79-98.

Kulberg, S., B. Heringstad, O. Guttersrud and I. Olsaker (2007). "Study on the association of BoLA-DRB3. 2 alleles with clinical mastitis in Norwegian Red cows." Journal of Animal Breeding and Genetics **124**(4): 201-207.

Kumar, S., M. Nei, J. Dudley and K. Tamura (2008). "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences." Briefings in Bioinformatics **9**(4): 299-306.

Kumar, S., G. Stecher and K. Tamura (2016). "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets." Molecular Biology and Evolution **33**(7): 1870-1874.

Kumar, S. and R. L. Tarleton (1998). "The relative contribution of antibody production and CD8+ T cell function to immune control of Trypanosoma cruzi." Parasite Immunology **20**(5): 207-216.

Lai, S.-J., Y.-P. Liu, Y.-X. Liu, X.-W. Li and Y.-G. Yao (2006). "Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation." Molecular Phylogenetics and Evolution **38**(1): 146-154.

Lambert, G., C. V. Gisolfi, D. J. Berg, P. L. Moseley, L. W. Oberley, et al. (2002). "Selected contribution: Hyperthermia-induced intestinal permeability and the role of oxidative and nitrosative stress." Journal of Applied Physiology **92**(4): 1750-1761.

Lamy, E., S. van Harten, E. Sales-Baptista, M. M. M. Guerra and A. M. de Almeida (2012). Factors influencing livestock productivity. Environmental Stress and Amelioration in Livestock Production, Springer**:** 19-51.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Larkin, D. M., H. D. Daetwyler, A. G. Hernandez, C. L. Wright, L. A. Hetrick, et al. (2012). "Whole-genome resequencing of two elite sires for the detection of haplotypes

under selection in dairy cattle." Proceedings of the National Academy of Sciences **109**(20): 7693-7698.

Latif, A., D. Punyua, S. Nokoe and P. Capstick (1991). "Tick infestations on Zebu cattle in western Kenya: individual host variation." Journal of Medical Entomology **28**(1): 114-121.

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

Lee, K., D. T. Nguyen, M. Choi, S.-Y. Cha, J.-H. Kim, et al. (2013). "Analysis of cattle olfactory subgenome: the first detail study on the characteristics of the complete olfactory receptor repertoire of a ruminant." BMC Genomics **14**(1): 596.

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

Lemos, A., R. Teodoro, G. Oliveira and F. Madalena (1985). "Comparative performance of six Holstein-Friesian× Guzera grades in Brazil 3. Burdens of Boophilus microplus under field conditions." Animal Science **41**(2): 187-191.

Leutenegger, A.-L., B. Prum, E. Génin, C. Verny, A. Lemainque, et al. (2003). "Estimation of the inbreeding coefficient through use of genomic data." The American Journal of Human Genetics **73**(3): 516-523.

Levy, C., M. Khaled and D. E. Fisher (2006). "MITF: master regulator of melanocyte development and melanoma oncogene." Trends in Molecular Medicine **12**(9): 406-414.

Lewin, H. A., G. C. Russell and E. J. Glass (1999). "Comparative organization and function of the major histocompatibility complex of domesticated cattle." Immunological Reviews **167**(1): 145-158.

Lewontin, R. and J. Krakauer (1973). "Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms." Genetics **74**(1): 175-195.

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

Liao, X., F. Peng, S. Forni, D. McLaren, G. Plastow, et al. (2013). "Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection 1." Genome **56**(10): 592-598.

Lin, B. Z., S. Sasazaki and H. Mannen (2010). "Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations analyzed by SNP markers." Animal Science Journal **81**(3): 281-289.

Littlejohn, M. D., K. M. Henty, K. Tiplady, T. Johnson, C. Harland, et al. (2014). "Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle." Nature Communications **5**: 5861.

Liu, G., D. Sun, H. Wu, M. Zhang, H. Huan, et al. (2015). "Distinct contributions of CD4+ and CD8+ T cells to pathogenesis of Trypanosoma brucei infection in the context of gamma interferon and interleukin-10." Infection and Immunity **83**(7): 2785-2795.

Liu, G., C. Van Tassell, T. Sonstegard, R. Li, L. Alexander, et al. (2008). Detection of germline and somatic copy number variations in cattle. Animal Genomics for Animal Health, Karger Publishers. **132:** 231-237.

Liu, G. E. and D. M. Bickhart (2012). "Copy number variation in the cattle genome." Functional & Integrative Genomics **12**.

Liu, J., L. Zhang, L. Xu, H. Ren, J. Lu, et al. (2013). "Analysis of copy number variations in the sheep genome using 50 K SNP BeadChip array." BMC Genomics **14**.

Liu, Y., X. Lan, Y. Qu, Z. Li, Z. Chen, et al. (2011). "Effects of genetic variability of the dairy goat growth hormone releasing hormone receptor (GHRHR) gene on growth traits." Molecular Biology Reports **38**(1): 539-544.

Liu, Y., X. Qin, X.-Z. H. Song, H. Jiang, Y. Shen, et al. (2009). "*Bos taurus* genome assembly." BMC Genomics **10**(1): 180.

Loftus, R., O. Ertugrul, A. Harba, M. El-Barody, D. MacHugh, et al. (1999). "A microsatellite survey of cattle from a centre of origin: the Near East." Molecular Ecology **8**(12): 2015-2022.

Loftus, R., D. Mac Hugh, L. Ngere, D. Balain, A. Badi, et al. (1994). "Mitochondrial genetic variation in European, African and Indian cattle populations." Animal Genetics **25**(4): 265-271.

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

Lott, D. F. and B. L. Hart (1979). "Applied ethology in a nomadic cattle culture." Applied Animal Ethology **5**(4): 309-319.

Louwette, S., L. Régal, C. Wittevrongel, C. Thys, G. Vandeweeghde, et al. (2012). "NPC1 defect results in abnormal platelet formation and function: studies in Niemann–Pick disease type C1 patients and zebrafish." Human Molecular Genetics **22**(1): 61-73.

Lynch, M. and K. Ritland (1999). "Estimation of pairwise relatedness with molecular markers." Genetics **152**(4): 1753-1766.

Ma, M.-J., H.-B. Wang, H. Li, J.-H. Yang, Y. Yan, et al. (2011). "Genetic variants in MARCO are associated with the susceptibility to pulmonary tuberculosis in Chinese Han population." PLoS One **6**(8): e24069.

Machado, M. A., A. L. S. Azevedo, R. L. Teodoro, M. A. Pires, M. G. C. Peixoto, et al. (2010). "Genome wide scan for quantitative trait loci affecting tick resistance in cattle (*Bos taurus* × *Bos indicus*)." BMC Genomics **11**(1): 280.

MacHugh, D. E. (1996). Molecular biogeography and genetic structure of domesticated cattle, Trinity College Dublin.

MacHugh, D. E., M. D. Shriver, R. T. Loftus, P. Cunningham and D. G. Bradley (1997). "Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*)." Genetics **146**(3): 1071.

Magez, S., A. Schwegmann, R. Atkinson, F. Claes, M. Drennan, et al. (2008). "The role of B-cells and IgM antibodies in parasitemia, anemia, and VSG switching in trypanosoma brucei–infected mice." PLoS Pathogens **4**(8): e1000122.

Makina, S. O., F. C. Muchadeyi, E. Marle-Köster, J. F. Taylor, M. L. Makgahlela, et al. (2015). "Genome-wide scan for selection signatures in six cattle breeds in South Africa." Genetics Selection Evolution **47**(1): 92.

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

Makvandi-Nejad, S., G. E. Hoffman, J. J. Allen, E. Chu, E. Gu, et al. (2012). "Four loci explain 83% of size variation in the horse." PLoS One **7**(7): e39929.

Malayer, J., P. Hansen, T. Gross and W. Thatcher (1990). "Regulation of heat shock-induced alterations in the release of prostaglandins by the uterine endometrium of cows." Theriogenology **34**(2): 219-230.

Mao, R., V. Rubio, H. Chen, L. Bai, O. Mansour, et al. (2013). "OLA1 protects cells in heat shock by stabilizing HSP70." Cell Death & Disease **4**(2): e491.

Mao, X., G. Sahana, D.-J. De Koning and B. Guldbrandtsen (2016). "Genome-wide association studies of growth traits in three dairy cattle breeds using whole-genome sequence data." Journal of Animal Science **94**(4): 1426-1437.

Martín-Burriel, I., E. García-Muro and P. Zaragoza (1999). "Genetic diversity analysis of six Spanish native cattle breeds using microsatellites." Animal Genetics **30**(3): 177-182.

Martina, J. A., H. I. Diab, H. Li and R. Puertollano (2014). "Novel roles for the MiTF/TFE family of transcription factors in organelle biogenesis, nutrient sensing, and energy homeostasis." Cellular and Molecular Life Sciences **71**(13): 2483-2497.

Martinez, M., M. Machado, C. Nascimento, M. Silva, R. Teodoro, et al. (2006). "Association of *BoLA-DRB3*. 2 alleles with tick (Boophilus microplus) resistance in cattle." Genetics & Molecular Research **5**(3): 513-524.

Mason, I. L. (1969). "A world dictionary of livestock breeds, types and varieties." Technical Communications. Commonwealth Bureau of Animal Breeding and Genetics(8).

Mattioli, R., M. Bah, J. Faye, S. Kora and M. Cassama (1993). "A comparison of field tick infestation on N'Dama, Zebu and N'Dama× Zebu crossbred cattle." Veterinary Parasitology **47**(1): 139-148.

Mattioli, R., M. Bah, S. Kora, M. Cassama and D. Clifford (1995). "Susceptibility to different tick genera in Gambian N'Dama and Gobra zebu cattle exposed to naturally occurring tick infestations." Tropical Animal Health and Production **27**(2): 95-105.

Mattioli, R. C., V. S. Pandey, M. Murray and J. L. Fitzpatrick (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.

Matukumalli, L., C. Lawley, M. Georges, W. Coppieters, K. Gietzen, et al. (2011). "Analyzing LD blocks and CNV segments in cattle: novel genomic features identified using the BovineHD BeadChip." San Diego, CA: Illumina Inc: 370-002.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, et al. (2009). "Development and characterization of a high density SNP genotyping assay for cattle." PLoS One **4**(4): e5350.

Maule, J. (1990). "The cattle of the tropics. Centre for Tropical Veterinary Medicine." University of Edinburgh, Great Britain **225**.

Mbole-Kariuki, M., T. Sonstegard, A. Orth, S. Thumbi, B. de C Bronsvoort, et al. (2014). "Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya." Heredity **113**(4): 297-305.

Mbole-Kariuki, M. N., T. Sonstegard, A. Orth, S. Thumbi, B. de C Bronsvoort, et al. (2014). "Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya." Heredity **113**(4): 297.

McKay, S. D., R. D. Schnabel, B. M. Murdoch, L. K. Matukumalli, J. Aerts, et al. (2008). "An assessment of population structure in eight breeds of cattle using a whole genome SNP panel." BMC Genetics **9**(1): 37.

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

McTavish, E. J., J. E. Decker, R. D. Schnabel, J. F. Taylor and D. M. Hillis (2013). "New World cattle show ancestry from multiple independent domestication events." Proceedings of the National Academy of Sciences **110**(15): E1398-E1406.

Meadow, R. H. (1984). "Notes on the faunal remains from Mehrgarh, with a focus on cattle (Bos)." South Asia archaeology. Cambridge University Press, Cambridge: 30-40.

Meghen, C., D. MacHugh and D. Bradley (1994). "Genetic characterization and West African cattle." World Animal Review **78**(1): 59-66.

Mei, C., H. Wang, Q. Liao, L. Wang, G. Cheng, et al. (2017). "Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing." Molecular Biology and Evolution: msx322.

Mejía-Benítez, A., M. Klünder-Klünder, L. Yengo, D. Meyre, C. Aradillas, et al. (2013). "Analysis of the contribution of FTO, NPC1, ENPP1, NEGR1, GNPDA2 and MC4R genes to obesity in Mexican children." BMC Medical Genetics **14**(1): 21.

Metta, M., S. Kanginakudru, N. Gudiseva and J. Nagaraju (2004). "Genetic characterization of the Indian cattle breeds, Ongole and Deoni (*Bos indicus*), using microsatellite markers–a preliminary study." BMC Genetics **5**(1): 16.

Metzker, M. L. (2010). "Sequencing technologies—the next generation." Nature Reviews Genetics **11**(1): 31-46.

Meyre, D., J. Delplanque, J.-C. Chèvre, C. Lecoeur, S. Lobbens, et al. (2009). "Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations." Nature Genetics **41**(2): 157.

Mikko, S. and L. Anderson (1995). "Extensive MHC class II DRB3 diversity in African and European cattle." Immunogenetics **42**(5): 408-403.

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

Mishra, C., T. K. Palai, L. N. Sarangi, B. R. Prusty and B. R. Maharana (2013). "Candidate gene markers for sperm quality and fertility in bulls." Veterinary World **6**(6): 905-910.

Moazami-Goudarzi, K., J. Furet, F. Grosclaude and D. Laloë (1997). "Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites." Animal Genetics **28**(5): 338-345.

Msalya, G., E.-S. Kim, E. L. Laisser, M. J. Kipanyula, E. D. Karimuribo, et al. (2017). "Determination of Genetic Structure and Signatures of Selection in Three Strains of Tanzania Shorthorn Zebu, Boran and Friesian Cattle by Genome-Wide SNP Analyses." PLoS One **12**(1): e0171088.

Mukesh, M., M. Sodhi, S. Bhatia and B. Mishra (2004). "Genetic diversity of Indian native cattle breeds as analysed with 20 microsatellite loci." Journal of Animal Breeding and Genetics **121**(6): 416-424.

Murray, M. (1988). "Trypanotolerance, its criteria and genetic and environmental influences." Livestock Production in Tsetse Affected Areas of Africa, Eds ILCA/ILRAD: 133-151.

Murray, M. (1988). "Trypanotolerance, its criteria and genetic and environmental influences." Livestock Production in Tsetse Affected Areas in Africa: 133-151.

Murray, M. and W. Morrison (1978). Parasitemia and host susceptibility to African trypanosomiasis. Pathogenicity of trypanosomes: proceedings of a workshop... IDRC, Ottawa, ON, CA.

Murray, M., W. Morrison and D. Whitelaw (1982). "Host susceptibility to African trypanosomiasis: trypanotolerance." Advances in Parasitology **21**: 1-68.

Murray, M., W. Morrison and D. Whitelaw (1982). Host susceptibility to African trypanosomiasis: trypanotolerance. Advances in Parasitology, Elsevier. **21:** 1-68.

Musa, L.-A., M.-K. Ahmed, K. J. Peters, B. Zumbach and K. E. Gubartalla (2005). "The reproductive and milk performance merit of Butana cattle in Sudan." Archives Animal Breeding **48**(5): 445-459.

Musa, L., K. Peters and M. Ahmed (2006). "On farm characterization of Butana and Kenana cattle breed production systems in Sudan." Livestock Research for Rural Development **18**(12): 2006.

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

Mwangi, E., P. Stevenson, J. Ndung'u, M. Stear, S. Reid, et al. (1998). "Studies on host resistance to tick infestations among trypanotolerant *Bos indicus* cattle breeds in east Africa." Annals of the New York Academy of Sciences **849**(1): 195-208.

N'Diaye, A., G. K. Chen, C. D. Palmer, B. Ge, B. Tayo, et al. (2011). "Identification, replication, and fine-mapping of Loci associated with adult height in individuals of african ancestry." PLoS Genetics **7**(10): e1002298.

Naessens, J. (2006). "Bovine trypanotolerance: a natural ability to prevent severe anaemia and haemophagocytic syndrome?" International Journal for Parasitology **36**(5): 521-528.

Nantulya, V. (1990). "Trypanosomiasis in domestic animals: the problems of diagnosis." Revue Scientifique et Technique **9**(2): 357-367.

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

Niimura, Y. and M. Nei (2003). "Evolution of olfactory receptor genes in the human genome." Proceedings of the National Academy of Sciences **100**(21): 12235-12240.

Nijman, I. J., D. C. Van Boxtel, L. M. Van Cann, Y. Marnoch, E. Cuppen, et al. (2008). "Phylogeny of Y chromosomes from bovine species." Cladistics **24**(5): 723-726.

Norval, R., R. Sutherst, O. Jorgensen and J. Kerr (1997). "The effects of the bont tick, Amblyomma hebraeum, on milk production of Sanga and Sanga x Brahman cattle." Medical and Veterinary Entomology **11**(2): 143-147.

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

Noyes, H., A. Brass, I. Obara, S. Anderson, A. L. Archibald, et al. (2011). "Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection." Proceedings of the National Academy of Sciences **108**(22): 9304-9309.

O'Gorman, G. M., S. D. Park, E. W. Hill, K. G. Meade, P. M. Coussens, et al. (2009). "Transcriptional profiling of cattle infected with Trypanosoma congolense highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility." BMC Genomics **10**(1): 207.

O'Reilly, P. F., E. Birney and D. J. Balding (2008). "Confounding between recombination and selection, and the Ped/Pop method for detecting selection." Genome Research.

Ocepek, M., M. Pate, M. Žolnir-Dovč and M. Poljak (2005). "Transmission of Mycobacterium tuberculosis from human to cattle." Journal of Clinical Microbiology **43**(7): 3555-3557.

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

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

Ortega, M. S., A. Denicol, J. Cole, D. Null and P. Hansen (2016). "Use of single nucleotide polymorphisms in candidate genes associated with daughter pregnancy rate for prediction of genetic merit for reproduction in Holstein cows." Animal Genetics **47**(3): 288-297.

Pareek, C. S., R. Smoczynski and A. Tretyn (2011). "Sequencing technologies and genome sequencing." Journal of Applied Genetics **52**(4): 413-435.

Parsell, D. and S. Lindquist (1993). "The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins." Annual Review of Genetics **27**(1): 437-496.

Paula-Lopes, F., C. Chase, Y. Al-Katanani, C. Krininger, R. Rivera, et al. (2003). "Genetic divergence in cellular resistance to heat shock in cattle: differences between breeds developed in temperate versus hot climates in responses of preimplantation embryos, reproductive tract tissues and lymphocytes to increased culture temperatures." Reproduction **125**(2): 285-294.

Pavlopoulos, G. A., A. Oulas, E. Iacucci, A. Sifrim, Y. Moreau, et al. (2013). "Unraveling genomic variation from next generation sequencing data." BioData mining **6**(1).

Payne, W. J. A. (1970). "Cattle production in the Tropics. Volume 1. General introduction and breeds and breeding." Cattle production in the Tropics. Volume 1. General introduction and breeds and breeding.

Payne, W. J. A. and J. Hodges (1997). Tropical cattle: origins, breeds and breeding policies, Blackwell Science.

Pickart, C. (1999). Ubiquitin and the stress response. Stress Proteins, Springer**:** 133-152.

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

Pickrell, J. K. and J. K. Pritchard (2012). "Inference of population splits and mixtures from genome-wide allele frequency data." PLoS Genetics **8**(11): e1002967.

Pinto, D., A. T. Pagnamenta, L. Klei, R. Anney, D. Merico, et al. (2010). "Functional impact of global rare copy number variation in autism spectrum disorders." Nature **466**.

Piper, E. K., N. N. Jonsson, C. Gondro, A. E. Lew-Tabor, P. Moolhuijzen, et al. (2009). "Immunological profiles of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, Rhipicephalus (Boophilus) microplus." Clinical and Vaccine Immunology **16**(7): 1074-1086.

Plassais, J., M. Rimbault, F. J. Williams, B. W. Davis, J. J. Schoenebeck, et al. (2017). "Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness." PLoS Genetics **13**(3): e1006661.

Plutzky, J., B. Neel, R. Rosenberg, R. Eddy, M. Byers, et al. (1992). "Chromosomal localization of an SH2-containing tyrosine phosphatase (*PTPN6*)." Genomics **13**(3): 869-872.

Porto-Neto, L. R., D. M. Bickhart, A. J. Landaeta-Hernandez, Y. T. Utsunomiya, M. Pagan, et al. (2018). "Convergent Evolution of Slick Coat in Cattle through Truncation Mutations in the Prolactin Receptor." Frontiers in Genetics **9**: 57.

Porto-Neto, L. R., T. S. Sonstegard, G. E. Liu, D. M. Bickhart, M. V. Da Silva, et al. (2013). "Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping." BMC Genomics **14**(1): 876.

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

Porto Neto, L., R. Bunch, B. Harrison and W. Barendse (2011). "DNA variation in the gene ELTD1 is associated with tick burden in cattle." Animal Genetics **42**(1): 50-55.

Pragna, P., P. Archana, J. Aleena, V. Sejian, G. Krishnan, et al. (2017). "Heat stress and dairy cow: impact on both milk yield and composition." International Journal of Dairy Science **12**(1): 1-11.

Prezeworski, M., G. Coop and J. D. Wall (2005). "The signature of positive selection on standing genetic variation." Evolution **59**(11): 2312-2323.

Pritchard, J. K., J. K. Pickrell and G. Coop (2010). "The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation." Current Biology **20**(4): R208-R215.

Pullan, N. and R. Grindle (1980). "Productivity of white Fulani cattle on the Jos plateau, Nigeria. IV. Economic factors." Tropical Animal Health and Production **12**(3): 161-170.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, 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.

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

Qanbari, S., E. Pimentel, J. Tetens, G. Thaller, P. Lichtner, et al. (2010). "The pattern of linkage disequilibrium in German Holstein cattle." Animal Genetics **41**(4): 346-356.

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

Qanbari, S., T. M. Strom, G. Haberer, S. Weigend, A. A. Gheyas, et al. (2012). "A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens." PLoS One **7**(11): e49525.

Quilez, J., A. D. Short, V. Martínez, L. J. Kennedy, W. Ollier, et al. (2011). "A selective sweep of> 8 Mb on chromosome 26 in the Boxer genome." BMC Genomics **12**(1): 339.

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

Radwanska, M., P. Guirnalda, C. De Trez, B. Ryffel, S. Black, et al. (2008). "Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses." PLoS Pathogens **4**(5): e1000078.

Rahman, I. (2007). "Sudanese cattle resources and their productivity: a review." Agricultural Reviews **28**(4): 305-308.

Ramayo-Caldas, Y., A. Castello, R. N. Pena, E. Alves, A. Mercade, et al. (2010). "Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip." BMC Genomics **11**.

Ramey, H. R., J. E. Decker, S. D. McKay, M. M. Rolf, R. D. Schnabel, et al. (2013). "Detection of selective sweeps in cattle using genome-wide SNP data." BMC Genomics **14**(1): 382.

Ramos, A. M., R. P. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald, et al. (2009). "Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology." PLoS One **4**(8): e6524.

Randhawa, I. A., M. S. Khatkar, P. C. Thomson and H. W. Raadsma (2016). "A meta-assembly of selection signatures in cattle." PLoS One **11**(4): e0153013.

Randhawa, I. A. S., M. S. Khatkar, P. C. Thomson and H. W. Raadsma (2014). "Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep." BMC Genetics **15**(1): 1.

Raviv, S., K. Bharti, S. Rencus-Lazar, Y. Cohen-Tayar, R. Schyr, et al. (2014). "PAX6 regulates melanogenesis in the retinal pigmented epithelium through feed-forward regulatory interactions with MITF." PLoS Genetics **10**(5): e1004360.

Reardon, W., A. Mullen, T. Sweeney and R. Hamill (2010). "Association of polymorphisms in candidate genes with colour, water-holding capacity, and composition traits in bovine M. longissimus and M. semimembranosus." Meat science **86**(2): 270-275.

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., G. Aboagye and C. Tawah (1994). "Shorthorn cattle of West and Central Africa. IV. Production characteristics." World Animal Rev.(FAO) **78**(1): 33-48.

Rege, J., W. Ayalew, E. Getahun, O. Hanotte and T. Dessie (2006). DAGRIS (Domestic Animal Genetic Resources Information System). International Livestock Research Institute, Addis Ababa, Ethiopia.

Rege, J. and C. Tawah (1999). "The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains." Animal Genetic Resources/Resources génétiques animales/Recursos Genéticos Animales **26**: 1-25.

Rehli, M., A. Lichanska, A. I. Cassady, M. C. Ostrowski and D. A. Hume (1999). "TFEC is a macrophage-restricted member of the microphthalmia-TFE subfamily of basic helix-loop-helix leucine zipper transcription factors." The Journal of Immunology **162**(3): 1559-1565.

Reinitz, D. M. and J. M. Mansfield (1990). "T-cell-independent and T-cell-dependent B-cell responses to exposed variant surface glycoprotein epitopes in trypanosome-infected mice." Infection and Immunity **58**(7): 2337-2342.

Reyer, H., R. Hawken, E. Murani, S. Ponsuksili and K. Wimmers (2015). "The genetics of feed conversion efficiency traits in a commercial broiler line." Scientific Reports **5**: 16387.

Rimbault, M., H. C. Beale, J. J. Schoenebeck, B. C. Hoopes, J. J. Allen, et al. (2013). "Derived variants at six genes explain nearly half of size reduction in dog breeds." Genome Research **23**(12): 1985-1995.

Rincon, G., K. Weber, A. Van Eenennaam, B. Golden and J. Medrano (2011). "Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys." Journal of Dairy Science **94**(12): 6116-6121.

Rischkowsky, B. and D. Pilling (2007). The state of the world's animal genetic resources for food and agriculture, Food & Agriculture Org.

Roberts, C. and A. Gray (1973). "Studies on trypanosome-resistant cattle. II. The effect of trypanosomiasis on N'Dama, Muturu and Zebu cattle." Tropical Animal Health and Production **5**(4): 220-233.

Ron, M., M. Cohen-Zinder, C. Peter, J. Weller and G. Erhardt (2006). "A polymorphism in ABCG2 in *Bos indicus* and *Bos taurus* cattle breeds." Journal of Dairy Science **89**(12): 4921-4923.

Röpke, R., D. Schams, F. Schwarz and M. Kirchgessner (1994). "Growth-related hormones in plasma of bulls, steers and heifers given food with two different energy levels." Animal Science **59**(3): 367-377.

Rowe, J. A. and K. Hødnebø (1994). "Rinderpest in the Sudan 1888-1890: The Mystery of the Missing Panzootic." Sudanic Africa **5**: 149-178.

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

Rubin, C.-J., M. C. Zody, J. Eriksson, J. R. Meadows, E. Sherwood, et al. (2010). "Whole-genome resequencing reveals loci under selection during chicken domestication." Nature **464**(7288): 587-591.

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

Ryan, M. T., R. M. Hamill, A. M. O'Halloran, G. C. Davey, J. McBryan, et al. (2012). "SNP variation in the promoter of the PRKAG3 gene and association with meat quality traits in pig." BMC Genetics **13**(1): 66.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, et al. (2002). "Detecting recent positive selection in the human genome from haplotype structure." Nature **419**(6909): 832-837.

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

Saleque, S., S. Cameron and S. H. Orkin (2002). "The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages." Genes & Development **16**(3): 301-306.

Sansoucy, R. (1995). "Livestock-a driving force for food security and sustainable development." World **3074**(5389): 1035.

Santana, M., P. Oliveira, V. Pedrosa, J. Eler, E. Groeneveld, et al. (2010). "Effect of inbreeding on growth and reproductive traits of Nellore cattle in Brazil." Livestock Science **131**(2): 212-217.

Sanz Fernandez, M. V., N. Upah, S. Pearce, A. Nayeri, E. Sucu, et al. (2012). "Prolactin's role during acute and chronic heat stress in growing pigs." Animal Industry Report **658**(1): 69.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**.

Seifert, G. (1971). "Variations between and within breeds of cattle in resistance to field infestations of the cattle tick (Boophilus microplus)." Australian Journal of Agricultural Research **22**(1): 159-168.

Seitz, J. J., S. M. Schmutz, T. D. Thue and F. C. Buchanan (1999). "A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle." Mammalian Genome **10**(7): 710-712.

Serizawa, S., T. Ishii, H. Nakatani, A. Tsuboi, F. Nagawa, et al. (2000). "Mutually exclusive expression of odorant receptor transgenes." Nature Neuroscience **3**(7): 687.

Sevane, N., S. Dunner, A. Boado and J. Cañon (2017). "Polymorphisms in ten candidate genes are associated with conformational and locomotive traits in Spanish Purebred horses." Journal of Applied Genetics **58**(3): 355-361.

Sharif, S., B. Mallard, B. Wilkie, J. Sargeant, H. Scott, et al. (1998). "Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle." Animal Genetics **29**(3): 185-193.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nature Biotechnology **26**(10): 1135-1145.

Sherman, M. and A. Goldberg (1996). Involvement of molecular chaperones in intracellular protein breakdown. Stress-inducible cellular responses, Springer**:** 57-78.

Sherry, S. T., M.-H. Ward, M. Kholodov, J. Baker, L. Phan, et al. (2001). "dbSNP: the NCBI database of genetic variation." Nucleic Acids Research **29**(1): 308-311.

Shigihara, T., M. Hashimoto, N. Shindo and T. Aoki (2008). "Transcriptome profile of Trypanosoma cruzi-infected cells: simultaneous up-and down-regulation of proliferation inhibitors and promoters." Parasitology Research **102**(4): 715-722.

Shin, M.-K., H.-T. Park, S. W. Shin, M. Jung, Y. B. Im, et al. (2015). "Whole-blood gene-expression profiles of cows infected with Mycobacterium avium subsp. paratuberculosis reveal changes in immune response and lipid metabolism." Journal of Microbiology & Biotechnology **25**: 255-267.

Shrestha, D. (2014). "Production and expression of inflammation and angiogenic parameters triggered by different genetic population of Trypanosoma cruzi."

Si, Z., N. Vandegraaff, C. O'hUigin, B. Song, W. Yuan, et al. (2006). "Evolution of a cytoplasmic tripartite motif (TRIM) protein in cows that restricts retroviral infection." Proceedings of the National Academy of Sciences **103**(19): 7454-7459.

Sileghem, M. and J. Flynn (1994). "Involvement of gamma-delta T cells in immunity to trypanosomiasis."

Silva, A. M. d., M. M. d. Alencar, L. C. d. A. Regitano, M. C. d. S. Oliveira and W. Barioni Júnior (2007). "Artificial infestation of Boophilus microplus in beef cattle heifers of four genetic groups." Genetics and Molecular Biology **30**(4): 1150-1155.

Smedley, D., S. Haider, B. Ballester, R. Holland, D. London, et al. (2009). "BioMart–biological queries made easy." BMC Genomics **10**(1): 22.

Smetko, A., A. Soudre, K. Silbermayr, S. Müller, G. Brem, et al. (2015). "Trypanosomosis: potential driver of selection in African cattle." Frontiers in Genetics **6**: 137.

Snelling, W. M., R. Chiu, J. E. Schein, M. Hobbs, C. A. Abbey, et al. (2007). "A physical map of the bovine genome." Genome Biology **8**(8): R165.

Spehr, M., G. Gisselmann, A. Poplawski, J. A. Riffell, C. H. Wetzel, et al. (2003). "Identification of a testicular odorant receptor mediating human sperm chemotaxis." Science **299**(5615): 2054-2058.

Spickett, A. (1994). "Genetic resistance of livestock to infectious diseases and parasites." Infectious diseases of livestock with special reference to Southern Africa **1**: 143-150.

Spinage, C. A. (2003). Cattle plague: a history, Springer Science & Business Media.

Stafuzza, N. B., A. Zerlotini, F. P. Lobo, M. E. B. Yamagishi, T. C. S. Chud, 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.

Stahl, P., V. Ruppert, R. T. Schwarz and T. Meyer (2014). "Trypanosoma cruzi evades the protective role of interferon-gamma-signaling in parasite-infected cells." PLoS One **9**(10): e110512.

Stefl, S., H. Nishi, M. Petukh, A. R. Panchenko and E. Alexov (2013). "Molecular mechanisms of disease-causing missense mutations." Journal of Molecular Biology **425**(21): 3919-3936.

Steingrímsson, E., L. Tessarollo, B. Pathak, L. Hou, H. Arnheiter, et al. (2002). "Mitf and Tfe3, two members of the Mitf-Tfe family of bHLH-Zip transcription factors, have important but functionally redundant roles in osteoclast development." Proceedings of the National Academy of Sciences **99**(7): 4477-4482.

Sternbauer, K. and J. Luthman (2002). "Insulin sensitivity of heifers on different diets." Acta Veterinaria Scandinavica **43**(2): 107.

Stock, F. and D. Gifford-Gonzalez (2013). "Genetics and African cattle domestication." African Archaeological Review **30**(1): 51-72.

Stothard, P., X. Liao, A. S. Arantes, M. De Pauw, C. Coros, et al. (2015). "A large and diverse collection of bovine genome sequences from the Canadian Cattle Genome Project." GigaScience **4**(1): 49.

Suchocki, T., K. Wojdak-Maksymiec and J. Szyda (2016). "Using gene networks to identify genes and pathways involved in milk production traits in Polish Holstein dairy cattle." Czech Journal of Animal Science **61**(11): 526-538.

Sudmant, P. H., J. O. Kitzman, F. Antonacci, C. Alkan, M. Malig, et al. (2010). "Diversity of human copy number variation and multicopy genes." Science **330**.

Suh, Y. and J. Vijg (2005). "SNP discovery in associating genetic variation with human disease phenotypes." Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **573**(1): 41-53.

Sulieman, M. and M. Hamid (2002). "Identification of acid fast bacteria from caseous lesions in cattle in Sudan." Zoonoses and Public Health **49**(9): 415-418.

Sunnucks, P. (2000). "Efficient genetic markers for population biology." Trends in Ecology & Evolution **15**(5): 199-203.

Sutherst, R., G. Maywald, J. Kerr and D. Stegeman (1983). "The effect of cattle tick (Boophilus microplus) on the growth of *Bos indicus* × *Bos taurus* steers." Australian Journal of Agricultural Research **34**(3): 317-327.

Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, et al. (2016). "The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible." Nucleic Acids Research: gkw937.

Szpiech, Z. A. and R. D. Hernandez (2014). "selscan: an efficient multithreaded program to perform EHH-based scans for positive selection." Molecular Biology and Evolution **31**(10): 2824-2827.

Szreder, T. and L. Zwierzchowski (2004). "Polymorphism within the bovine estrogen receptor-alpha gene 5'-region." Journal of Applied Genetics **45**(2): 225-236.

Tang, K., K. R. Thornton and M. Stoneking (2007). "A new approach for using genome scans to detect recent positive selection in the human genome." PLoS Biology **5**(7): e171.

Tarleton, R. (1990). "Depletion of CD8+ T cells increases susceptibility and reverses vaccine-induced immunity in mice infected with Trypanosoma cruzi." The Journal of Immunology **144**(2): 717-724.

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

Taye, M., W. Lee, S. Jeon, J. Yoon, T. Dessie, et al. (2017). "Exploring evidence of positive selection signatures in cattle breeds selected for different traits." Mammalian Genome **28**(11-12): 528-541.

Terefe, E., A. Haile, W. Mulatu, T. Dessie and O. Mwai (2015). "Phenotypic characteristics and trypanosome prevalence of Mursi cattle breed in the Bodi and Mursi districts of South Omo Zone, southwest Ethiopia." Tropical Animal Health and Production **47**(3): 485-493.

Thoen, C. O., J. H. Steele and M. J. Gilsdorf (2008). Mycobacterium bovis infection in animals and humans, John Wiley & Sons.

Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, et al. (2003). "PANTHER: a library of protein families and subfamilies indexed by function." Genome Research **13**(9): 2129-2141.

Thornton, P. K. (2010). "Livestock production: recent trends, future prospects." Philosophical Transactions of the Royal Society B: Biological Sciences **365**(1554): 2853-2867.

Thuong, N. T. T., T. T. B. Tram, T. D. Dinh, P. V. K. Thai, D. Heemskerk, et al. (2016). "MARCO variants are associated with phagocytosis, pulmonary tuberculosis susceptibility and Beijing lineage." Genes and Immunity **17**(7): 419.

Tian, Y., M. Yue, Y. Gu, W. Gu and Y. Wang (2014). "Single-nucleotide polymorphism analysis of GH, GHR, and IGF-1 genes in minipigs." Brazilian Journal of Medical and Biological Research **47**(9): 753-758.

Tönjes, A., M. Scholz, J. Breitfeld, C. Marzi, H. Grallert, et al. (2014). "Genome wide meta-analysis highlights the role of genetic variation in RARRES2 in the regulation of circulating serum chemerin." PLoS Genetics **10**(12): e1004854.

Troy, C. S., D. E. MacHugh, J. F. Bailey, D. A. Magee, R. T. Loftus, et al. (2001). "Genetic evidence for Near-Eastern origins of European cattle." Nature **410**(6832): 1088-1091.

Utsunomiya, Y. T., A. M. P. O'Brien, T. S. Sonstegard, C. P. Van Tassell, A. S. do Carmo, et al. (2013). "Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods." PLoS One **8**(5): e64280.

Veneroni, G., S. Meirelles, D. Grossi, G. Gasparin, A. Ibelli, et al. (2010). "Prospecting candidate SNPs for backfat in Canchim beef cattle." Genetics and Molecular Research: 1997-2003.

Victoria Sanz Fernandez, M., J. S. Johnson, M. Abuajamieh, S. K. Stoakes, J. T. Seibert, et al. (2015). "Effects of heat stress on carbohydrate and lipid metabolism in growing pigs." Physiological Reports **3**(2): e12315.

Viitala, S., J. Szyda, S. Blott, N. Schulman, M. Lidauer, et al. (2006). "The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle." Genetics.

Vitalis, R., K. Dawson and P. Boursot (2001). "Interpretation of variation across marker loci as evidence of selection." Genetics **158**(4): 1811-1823.

Vitalis, R., K. Dawson, P. Boursot and K. Belkhir (2003). "DetSel 1.0: a computer program to detect markers responding to selection." Journal of Heredity **94**(5): 429-431.

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

Volker, M., N. Backstrom, B. M. Skinner, E. J. Langley, S. K. Bunzey, et al. (2010). "Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution." Genome Research **20**.

Wambura, P., P. Gwakisa, R. Silayo and E. Rugaimukamu (1998). "Breed-associated resistance to tick infestation in *Bos indicus* and their crosses with *Bos taurus*." Veterinary Parasitology **77**(1): 63-70.

Wang, X., J. Liu, G. Zhou, J. Guo, H. Yan, et al. (2016). "Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits." Scientific Reports **6**: 38932.

Wang, X., S. Nahashon, T. K. Feaster, A. Bohannon-Stewart and N. Adefope (2010). "An initial map of chromosomal segmental copy number variations in the chicken." BMC Genomics **11**.

Wang, Y., J. Utzinger, J. Saric, J. V. Li, J. Burckhardt, et al. (2008). "Global metabolic responses of mice to Trypanosoma brucei brucei infection." Proceedings of the National Academy of Sciences **105**(16): 6127-6132.

Weikard, R., P. Widmann, J. Buitkamp, R. Emmerling and C. Kuehn (2012). "Revisiting the quantitative trait loci for milk production traits on BTA6." Animal Genetics **43**(3): 318-323.

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." Evolution **38**(6): 1358-1370.

Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, et al. (2008). "The complete genome of an individual by massively parallel DNA sequencing." Nature **452**(7189): 872-876.

Whitacre, L. (2014). Structural variation at the KIT locus is responsible for the piebald phenotype in Hereford and Simmental cattle, University of Missouri-Columbia.

Wiener, P. and S. Wilkinson (2011). "Deciphering the genetic basis of animal domestication." Proceedings of the Royal Society of London B: Biological Sciences: rspb20111376.

Williamson, L., M. Saponaro, S. Boeing, P. East, R. Mitter, et al. (2017). "UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene." Cell **168**(5): 843-855. e813.

Wright, S. (1943). "Isolation by distance." Genetics **28**(2): 114.

Wu, H., G. Liu and M. Shi (2017). "Interferon Gamma in African Trypanosome Infections: Friends or Foes?" Frontiers in Immunology **8**: 1105.

Xie, C., S. D. Turley, P. G. Pentchev and J. M. Dietschy (1999). "Cholesterol balance and metabolism in mice with loss of function of Niemann-Pick C protein." American Journal of Physiology-Endocrinology and Metabolism **276**(2): E336-E344.

Xu, A., M. Van Eijk, C. Park and H. A. Lewin (1993). "Polymorphism in BoLA-DRB3 exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus." The Journal of Immunology **151**(12): 6977-6985.

Xu, L., D. M. Bickhart, J. B. Cole, S. G. Schroeder, J. Song, et al. (2014). "Genomic signatures reveal new evidences for selection of important traits in domestic cattle." Molecular Biology and Evolution **32**(3): 711-725.

Xu, L., D. M. Bickhart, J. B. Cole, S. G. Schroeder, J. Song, 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, L., J. B. Cole, D. M. Bickhart, Y. Hou, J. Song, et al. (2014). "Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins." BMC Genomics **15**.

Yang, Z. (2007). "PAML 4: phylogenetic analysis by maximum likelihood." Molecular Biology and Evolution **24**(8): 1586-1591.

Yano, M., H. Shimadzu and T. Endo (2014). "Modelling temperature effects on milk production: a study on Holstein cows at a Japanese farm." SpringerPlus **3**(1): 129.

Yayou, K.-i., S. Ito, N. Yamamoto, S. Kitagawa and H. Okamura (2010). "Relationships of stress responses with plasma oxytocin and prolactin in heifer calves." Physiology & Behavior **99**(3): 362-369.

Ylinen, L. M., Z. Keckesova, B. L. Webb, R. J. Gifford, T. P. Smith, et al. (2006). "Isolation of an active Lv1 gene from cattle indicates that tripartite motif protein-mediated innate immunity to retroviral infection is widespread among mammals." Journal of Virology **80**(15): 7332-7338.

Yousif, I. and A. F. El-Moula (2006). "Characterisation of Kenana cattle breed and its production environment." Animal Genetic Resources/Resources Génétiques Animales/Recursos Genéticos Animales **38**: 47-56.

Zarrei, M., J. R. MacDonald, D. Merico and S. W. Scherer (2015). "A copy number variation map of the human genome." Nature Reviews Genetics **16**.

Zeng, L., N. Chen, Q. Ning, Y. Yao, H. Chen, et al. (2018). "PRLH and SOD 1 gene variations associated with heat tolerance in Chinese cattle." Animal Genetics.

Zhan, B., J. Fadista, B. Thomsen, J. Hedegaard, F. Panitz, et al. (2011). "Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping." BMC Genomics **12**.

Zhang, C., Z. Wang, H. Bruce, R. A. Kemp, P. Charagu, et al. (2015). "Genome-wide association studies (GWAS) identify a QTL close to PRKAG3 affecting meat pH and colour in crossbred commercial pigs." BMC Genetics **16**(1): 33.

Zhang, J., V. Rubio, M. W. Lieberman and Z.-Z. Shi (2009). "OLA1, an Obg-like ATPase, suppresses antioxidant response via nontranscriptional mechanisms." Proceedings of the National Academy of Sciences **106**(36): 15356-15361.

Zhang, L., S. Jia, M. Yang, Y. Xu, C. Li, et al. (2014). "Detection of copy number variations and their effects in Chinese bulls." BMC Genomics **15**.

Zhu, B., H. Niu, W. Zhang, Z. Wang, Y. Liang, et al. (2017). "Genome wide association study and genomic prediction for fatty acid composition in Chinese Simmental beef cattle using high density SNP array." BMC Genomics **18**(1): 464.

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

Supplementary Tables

Chapter 2

Supplementary Table S2.1 | Proportions of ancestry and inbreeding level for individual cattle samples **(file is provided as an electronic format)**

Chapter 3

Supplementary Table S3.1 | List of candidate selected regions and associated genes detected in Muturu based on iHS test **(file is provided as an electronic format)**

Supplementary Table S3.2 | List of candidate selected regions and associated genes detected in Muturu based on Rsb test **(file is provided as an electronic format)**

Supplementary Table S3.3 | List of candidate selected regions and associated genes detected in N'Dama cattle based on iHS test **(file is provided as an electronic format)**

Supplementary Table S3.4 | List of candidate selected regions and associated genes detected in Holstein cattle based on iHS test **(file is provided as an electronic format)**

Supplementary Table S3.5 | List of candidate selected regions and associated genes detected in Jersey cattle based on iHS test **(file is provided as an electronic format)**

Supplementary Table S3.6 | List of candidate selected regions and associated genes detected in N'Dama cattle based on Rsb test **(file is provided as an electronic format)**

Supplementary Table S3.7 | List of candidate selected regions and associated genes detected in Holstein cattle based on Rsb test **(file is provided as an electronic format)**

Supplementary Table S3.8 | List of candidate selected regions and associated genes detected in Jersey cattle based on Rsb test **(file is provided as an electronic format)**

Supplementary Table S3.9 | List of candidate genes shared between Muturu and N'Dama

Chr	Gene start (bp)	Gene end (bp)	Gene name	Chr	Gene start (bp)	Gene end (bp)	Gene name
1	82243015	82278080	*LIPH*	12	20856479	20887977	*BLA-DQB*
1	82176394	82210749	*SENP2*	12	74816045	74978179	*ATP6V0D2*
1	82301321	82426795	*INTS6*	12	20704476	20778873	ENSBTAG00000023309
1	82285913	82294596	*C5orf51*	14	78908052	78964199	ENSBTAG00000026030
1	82250680	82250794	*SERPINE3*	16	48081092	48137067	ENSBTAG00000026163
3	67863093	68073051	*ADI1*	16	48181539	48314477	*RF00001*
3	67687802	67824632	*OR6C76*	16	48024410	48077736	*DNAJC25*
5	58849295	58850254	*OTUD7A*	17	50734879	50736971	*PIGK*
5	58751279	58752217	*ST6GALNAC5*	18	45337460	45339631	ENSBTAG00000031111
5	58796713	58797757	*KCNAB2*	18	45406679	45406784	*OXCT1*
5	58781761	58782696	*NPHP4*	18	45438596	45440643	ENSBTAG00000033979
6	78991885	79313522	*GNG10*	20	32633116	32650605	*BOD1*
8	112948194	112954511	*MBIP*	20	32683996	32848723	*DCDC2C*
8	102784501	102790524	*ALLC*	20	5226483	5238709	ENSBTAG00000038521
8	112837179	112860706	*ADGRL3*	21	30667805	30753716	*CHD5*
8	112896493	112901718	*RPS7*	21	46916937	46935846	*C8H9orf84*
8	112869671	112892968	*COLEC11*	21	9617226	9635374	ENSBTAG00000045737
8	112911793	112920759	*RNASEH1*	23	25855146	25863045	*RF00026*
8	102759008	102776534	*RNF122*	23	26354054	26354780	*FAM124A*
8	112796601	112836172	*DUSP26*	23	25966734	25980368	ENSBTAG00000046439
8	102796661	102872318	*MAP3K13*	27	28450185	28459083	ENSBTAG00000046611
12	20888781	20973823	*TMEM41A*	27	28963365	28969954	ENSBTAG00000047571

Supplementary Table S3.10 | List of candidate genes shared between Muturu and Holstein

Chr	Gene start (bp)	Gene end (bp)	Gene name	Chr	Gene start (bp)	Gene end (bp)	Gene name
1	150360988	150386586	*ENSBTAG00000000228*	23	25030303	25051787	*ENSBTAG00000031843*
1	150439867	150440580	*GHR*	23	29223821	29224756	*RF00001*
4	65670543	65729276	*MAP7*	23	25080359	25080477	*TRIM26*
4	45594206	45682229	*BOLA*	23	28563086	28579480	*TRIM40*
4	65793129	65809031	*TRIM10*	23	28626339	28638869	*ENSBTAG00000037421*
9	75364392	75519232	*TRIM15*	23	28530978	28540122	*ENSBTAG00000038608*
9	75549283	75778418	*GCM1*	23	29314810	29315824	*ENSBTAG00000039534*
10	50282171	50283148	*MAP3K5*	23	29241755	29242690	*OR12D3*
20	31890736	32199996	*CHAF1B*	23	29149938	29150906	*ORC5*
23	29327118	29328065	*ENSBTAG00000013654*	23	28601333	28601408	*ENSBTAG00000044550*
23	28502524	28506312	*FBXO9*	23	29299942	29300874	*ENSBTAG00000046023*
23	28613833	28622421	*FOXB1*	23	29280761	29282962	*ENSBTAG00000046777*
23	28600596	28610740	*ADCYAP1R1*	23	29290572	29291495	*ENSBTAG00000047086*
23	25066915	25087445	*RF00003*	24	28822883	28823043	*GHRHR*
23	29206927	29207889	*CLDN14*				

Supplementary Table S3.11 | List of candidate genes shared between Muturu and Jersey

Chr	Gene start (bp)	Gene end (bp)	Gene name
6	86810496	86835733	*UGT2A1*
6	86676447	86713666	*ALOX15*
13	338154	339068	*ENSBTAG00000015047*
13	422889	424810	*MED11*
19	27331075	27339653	*PELP1*
19	27253471	27255303	*EPB41L3*
19	27281925	27303514	*BCL6B*
19	27447122	27451108	*C19H17orf49*
19	27438490	27441264	*RNASEK*
19	27436503	27438226	*ALOX12*
19	27421125	27434258	*ENSBTAG00000024891*
19	27441652	27441763	*ENSBTAG00000025720*
19	27441343	27441429	*bta-mir-497*
19	27361427	27370256	*bta-mir-195*
19	27249528	27253097	*ALOX12E*
19	27268438	27276816	*CXCL16*
19	27298367	27298439	*ARRB2*
19	27388083	27388880	*bta-mir-2338*
24	39312037	39404644	ENSBTAG00000047925

Supplementary Table S3.12 | list of functional annotation terms including GO term and KEGG pathways based on the 607 selected genes in Muturu **(file is provided as an electronic format)**

Supplementary Table S3.13 | Count of missense variants in genes overlapping selected regions in Muturu **(file is provided as an electronic format)**

Chapter 4

Supplementary Table S4.1 | List of candidate selected genome regions and related genes in Muturu based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S4.2 | List of candidate selected genome regions and related genes in N'Dama based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S4.3 | List of candidate selected genome regions and related genes in Muturu based on XPEHH test **(file is provided as an electronic format)**

Supplementary Table S4.4 | List of candidate selected genome regions and related genes in N'Dama based on XPEHH test **(file is provided as an electronic format)**

Supplementary Table S4.5 | List of candidate selected genome regions and related genes in Muturu based on XPCLR test **(file is provided as an electronic format)**

Supplementary Table S4.6 | List of candidate selected genome regions and related genes in N'Dama based on XPCLR test **(file is provided as an electronic format)**

Supplementary Table S4.7 | Comparison of candidate selected genes identified in Muturu following five different selection scan tests **(file is provided as an electronic format)**

Supplementary Table S4.8 | Comparison of candidate selected genes identified in N'Dama following five different selection scan tests **(file is provided as an electronic format)**

Supplementary Table S4.9 | Identified Gene desert regions in Muturu based on ZHp test

Chr	Start (bp)	End (bp)	ZHp scores
3	47200001	47700000	-3.50
4	83800001	84400000	-4.43
6	75300001	75900000	-4.04
6	76900001	78100000	-4.12
8	50200001	50700000	-4.35
16	8500001	9000000	-3.69

Supplementary Table S4.10 | Identified Gene desert regions in N'Dama based on ZHp test

Chr	Start (bp)	End (bp)	ZHp scores
6	52500001	53200000	-4.84
9	3100001	3700000	-3.41
9	4400001	5300000	-4.40
11	34000001	34500000	-4.09
11	76900001	77400000	-3.17

Supplementary Table S4.11 | Identified Gene desert regions in Muturu based on XPEHH test

Chr	Start (bp)	End (bp)	XPEHH	P-value	log10(P-value)	Chr	Start (bp)	End (bp)	XPEHH	P-value	log10(P-value)
1	21050001	21150000	3.64	0.0033	2.48	11	53750001	53800000	3.56	0.0040	2.40
1	21250001	21350000	3.36	0.0038	2.42	11	53950001	54100000	3.82	0.0042	2.38
1	101100001	101150000	3.55	0.0042	2.37	11	63100001	63150000	2.98	0.0051	2.30
1	101150001	101200000	3.25	0.0044	2.36	11	63700001	63750000	3.05	0.0040	2.40
1	106050001	106100000	3.00	0.0047	2.33	11	64600001	64650000	3.25	0.0045	2.35
1	120800001	120950000	4.03	0.0040	2.39	11	77700001	77750000	3.38	0.0034	2.47
1	122000001	122100000	3.78	0.0038	2.42	12	3150001	3200000	3.57	0.0036	2.44
1	132600001	132850000	4.30	0.0001	3.97	12	3450001	3500000	3.87	0.0011	2.95
1	133000001	133150000	3.85	0.0030	2.53	12	4050001	4300000	4.24	0.0003	3.50
2	1450001	1600000	3.93	0.0023	2.64	12	17100001	17250000	4.43	0.0003	3.48
2	6800001	6850000	2.90	0.0014	2.86	12	19800001	19850000	3.16	0.0031	2.51
2	22700001	22850000	3.66	0.0030	2.53	12	19850001	19900000	3.05	0.0041	2.39
2	25300001	25400000	3.30	0.0043	2.37	12	19900001	20050000	3.60	0.0011	2.97
2	92650001	92700000	1.92	0.0040	2.40	12	20800001	20850000	4.06	0.0010	3.01
2	135400001	135550000	4.19	0.0024	2.63	12	21000001	21050000	3.45	0.0010	3.01
3	5200001	5250000	3.46	0.0051	2.29	12	30750001	30800000	3.41	0.0028	2.56
3	47250001	47550000	4.22	0.0005	3.27	13	19750001	19800000	3.63	0.0031	2.51
3	76000001	76050000	3.53	0.0045	2.35	13	42300001	42350000	4.21	0.0002	3.75
4	36400001	36450000	4.24	0.0001	4.05	13	81400001	81550000	3.84	0.0034	2.47
4	41050001	41200000	3.82	0.0043	2.37	14	32450001	32500000	1.80	0.0052	2.29
4	46300001	46400000	3.76	0.0043	2.36	14	51550001	51600000	3.54	0.0044	2.36
4	46750001	46800000	2.70	0.0030	2.52	15	23700001	23750000	3.63	0.0032	2.49
4	52850001	52900000	3.58	0.0004	3.41	15	23750001	23800000	4.05	0.0005	3.30
4	54350001	54400000	2.56	0.0053	2.27	16	7650001	7750000	3.91	0.0029	2.53
4	83750001	83900000	4.13	0.0011	2.96	16	8550001	8650000	3.77	0.0040	2.40

4	84000001	84150000	4.15	0.0007	3.13
4	84150001	84250000	4.16	0.0028	2.55
4	108300001	108350000	4.22	0.0016	2.79
5	18250001	18300000	1.89	0.0042	2.38
5	19100001	19150000	2.33	0.0011	2.94
5	23850001	24000000	3.92	0.0016	2.79
6	400001	450000	3.28	0.0039	2.41
6	5200001	5250000	3.58	0.0013	2.87
6	15800001	15900000	3.62	0.0041	2.38
6	39400001	39550000	3.96	0.0020	2.71
6	50900001	50950000	3.71	0.0021	2.67
6	52850001	52900000	2.17	0.0017	2.76
6	53200001	53250000	3.13	0.0034	2.47
6	63550001	64000000	4.28	0.0002	3.60
6	73650001	73700000	3.21	0.0053	2.27
6	74550001	74600000	3.30	0.0038	2.42
6	74650001	74700000	3.26	0.0042	2.38
6	75550001	75850000	4.62	0.0001	4.10
6	98800001	98850000	3.92	0.0020	2.71
6	100700001	100750000	1.95	0.0036	2.45
7	40850001	40900000	3.22	0.0050	2.30
7	54800001	54850000	2.58	0.0048	2.32
7	56550001	56700000	3.78	0.0011	2.97
7	100600001	100650000	3.78	0.0001	4.01
7	100650001	100700000	3.71	0.0022	2.65
8	15350001	15500000	3.83	0.0038	2.42
8	66800001	66850000	3.52	0.0046	2.34
8	68550001	68600000	3.46	0.0052	2.28
16	51150001	51200000	3.06	0.0038	2.42
16	51200001	51250000	3.76	0.0002	3.71
16	77350001	77500000	3.77	0.0039	2.41
17	8300001	8350000	3.80	0.0033	2.48
17	9250001	9300000	3.74	0.0020	2.70
17	16150001	16200000	2.48	0.0007	3.14
17	32950001	33000000	2.68	0.0033	2.48
17	52450001	52500000	2.54	0.0006	3.24
18	21250001	21300000	3.19	0.0024	2.61
18	23100001	23150000	2.01	0.0030	2.52
18	32200001	32250000	1.95	0.0037	2.43
19	4200001	4350000	3.98	0.0046	2.34
19	12150001	12200000	3.22	0.0052	2.29
19	63350001	63500000	4.04	0.0012	2.90
20	26300001	26450000	3.92	0.0015	2.81
20	41150001	41200000	3.06	0.0008	3.09
20	45600001	45650000	2.11	0.0020	2.70
21	29100001	29250000	3.77	0.0023	2.65
21	54850001	55200000	4.27	0.0004	3.41
21	64350001	64400000	2.11	0.0019	2.73
22	36300001	36450000	3.93	0.0014	2.87
23	1350001	1400000	2.85	0.0021	2.69
23	20150001	20200000	3.70	0.0024	2.63
24	1250001	1400000	3.87	0.0004	3.35
26	8450001	8500000	3.21	0.0053	2.28
27	3550001	3700000	4.26	0.0007	3.16
28	1700001	1800000	3.94	0.0050	2.30
28	12850001	12900000	3.78	0.0015	2.82

8	102700001	102750000	3.41	0.0029	2.54	28	19900001	19950000	3.41	0.0029	2.53
9	87400001	87450000	3.30	0.0013	2.90	28	19950001	20000000	3.27	0.0018	2.76
9	87450001	87500000	3.51	0.0008	3.11	28	20000001	20050000	3.26	0.0003	3.56
9	100600001	100800000	4.06	0.0004	3.43	28	20250001	20300000	3.60	0.0034	2.47
9	100800001	100850000	3.92	0.0023	2.64	28	20550001	20600000	3.72	0.0039	2.40
9	100850001	100900000	3.66	0.0029	2.54	28	26150001	26200000	3.32	0.0037	2.43
9	101150001	101200000	3.22	0.0048	2.32	28	32450001	32500000	3.33	0.0037	2.44
10	46650001	46700000	2.85	0.0019	2.72	28	32500001	32550000	3.22	0.0049	2.31
11	33400001	33550000	3.92	0.0018	2.75	29	7250001	7300000	3.57	0.0037	2.43
11	42600001	42650000	4.18	0.0002	3.60	29	14650001	14700000	3.66	0.0001	3.84
11	42700001	42750000	3.00	0.0044	2.36	29	34150001	34200000	2.75	0.0004	3.37
11	45700001	45750000	2.04	0.0027	2.57						

Supplementary Table S4.12| Identified Gene desert regions in N'Dama based on XPEHH test

Chr	Start (bp)	End (bp)	XPEHH	P-value	log10(P-value)	Chr	Start (bp)	End (bp)	XPEHH	P-value	log10(P-value)
1	22050001	22100000	1.70	0.0024	2.62	11	65650001	65700000	3.16	0.0027	2.57
1	29450001	29500000	3.32	0.0045	2.34	11	78400001	78450000	3.22	0.0051	2.29
1	29500001	29550000	3.63	0.0016	2.80	11	91950001	92000000	3.55	0.0021	2.67
1	98500001	98550000	1.90	0.0012	2.93	12	19650001	19700000	3.20	0.0007	3.13
1	149350001	149400000	3.05	0.0046	2.33	12	19800001	19850000	2.80	0.0017	2.77
1	157750001	157800000	3.13	0.0026	2.59	12	19850001	19900000	2.60	0.0040	2.40
1	157850001	157900000	3.66	0.0026	2.59	12	19900001	19950000	2.83	0.0016	2.80
1	157950001	158000000	3.56	0.0021	2.68	12	19950001	20000000	3.30	0.0014	2.84
2	15950001	16000000	3.64	0.0008	3.08	12	44650001	44700000	3.33	0.0052	2.28
2	37700001	37750000	1.50	0.0048	2.32	12	44700001	44750000	3.79	0.0002	3.62
2	48550001	48600000	1.66	0.0027	2.56	12	46600001	46650000	3.54	0.0012	2.93
2	50250001	50300000	3.36	0.0026	2.59	12	49700001	49750000	3.08	0.0035	2.45
2	60650001	60700000	2.97	0.0028	2.56	12	54050001	54100000	3.06	0.0049	2.31
2	68450001	68500000	2.62	0.0001	4.08	12	61100001	61150000	3.42	0.0034	2.47
2	71900001	71950000	3.10	0.0039	2.41	12	61250001	61300000	3.64	0.0015	2.83
2	72450001	72500000	3.08	0.0044	2.35	12	63450001	63500000	1.67	0.0027	2.58
2	72600001	72650000	3.02	0.0002	3.64	12	63750001	63800000	3.15	0.0024	2.62
2	72650001	72700000	3.29	0.0016	2.79	12	64800001	64850000	4.00	0.0002	3.61
2	72850001	72900000	3.24	0.0046	2.34	12	64900001	64950000	3.58	0.0011	2.97
2	106250001	106300000	2.61	0.0004	3.42	14	32550001	32600000	2.10	0.0050	2.30
2	123000001	123050000	2.13	0.0040	2.40	14	37600001	37650000	4.04	0.0002	3.78
3	17950001	18000000	1.54	0.0039	2.41	14	37850001	37900000	3.35	0.0030	2.53
3	37000001	37050000	3.56	0.0011	2.95	14	57400001	57450000	3.39	0.0037	2.44
3	37050001	37100000	2.92	0.0035	2.46	14	60600001	60650000	1.96	0.0010	3.00
3	37100001	37150000	2.94	0.0031	2.50	14	74200001	74250000	1.50	0.0050	2.30

3	37150001	37200000	3.29	0.0017	2.78
3	39350001	39400000	3.38	0.0049	2.31
3	39400001	39450000	3.64	0.0014	2.85
3	55650001	55700000	2.88	0.0043	2.36
3	55750001	55800000	3.20	0.0025	2.61
3	56100001	56150000	3.10	0.0038	2.42
3	56200001	56250000	3.15	0.0028	2.55
3	63350001	63400000	3.40	0.0024	2.63
4	12850001	12900000	3.46	0.0018	2.74
4	28950001	29000000	2.12	0.0034	2.46
5	1700001	1750000	3.63	0.0009	3.05
5	23900001	23950000	3.96	0.0002	3.75
5	44450001	44500000	2.87	0.0013	2.90
5	52850001	52900000	3.77	0.0010	3.00
5	52900001	52950000	3.48	0.0026	2.58
5	53000001	53050000	3.47	0.0027	2.56
5	73750001	73800000	3.25	0.0043	2.37
5	73800001	73850000	3.79	0.0004	3.38
5	80900001	80950000	3.55	0.0022	2.65
5	80950001	81000000	3.08	0.0039	2.41
5	98550001	98600000	3.50	0.0014	2.85
6	10950001	11000000	3.65	0.0012	2.91
6	11000001	11050000	3.59	0.0017	2.76
6	11350001	11400000	3.42	0.0033	2.48
6	19650001	19700000	2.35	0.0014	2.86
6	19700001	19750000	3.65	0.0008	3.11
6	19750001	19800000	3.50	0.0004	3.37
6	32500001	32550000	3.90	0.0014	2.84
14	81650001	81700000	1.86	0.0015	2.83
15	14850001	14900000	3.76	0.0006	3.23
15	14900001	14950000	3.42	0.0007	3.17
15	14950001	15000000	3.37	0.0025	2.60
15	15300001	15350000	3.35	0.0012	2.92
15	15450001	15500000	3.17	0.0025	2.60
15	40000001	40050000	3.30	0.0035	2.45
15	76950001	77000000	1.50	0.0051	2.30
15	78900001	78950000	2.72	0.0027	2.56
16	11650001	11700000	2.19	0.0006	3.24
16	36850001	36900000	3.35	0.0043	2.37
16	49100001	49150000	3.33	0.0042	2.38
16	80000001	80050000	2.30	0.0003	3.48
17	39150001	39200000	3.41	0.0035	2.46
17	51550001	51600000	2.16	0.0038	2.42
17	61200001	61250000	3.27	0.0041	2.39
18	39850001	39900000	3.25	0.0044	2.36
18	42200001	42250000	2.82	0.0052	2.28
19	2650001	2700000	3.37	0.0056	2.25
19	2700001	2750000	3.81	0.0005	3.28
19	2750001	2800000	3.94	0.0010	2.98
19	2800001	2850000	3.44	0.0030	2.53
19	3500001	3550000	3.54	0.0032	2.50
19	3550001	3600000	3.56	0.0019	2.73
19	3600001	3650000	3.50	0.0037	2.43
19	3700001	3750000	3.07	0.0045	2.35
19	4200001	4250000	2.17	0.0036	2.44
19	6750001	6800000	1.88	0.0013	2.88

6	32550001	32600000	4.24	0.0001	4.23
6	32600001	32650000	2.98	0.0025	2.60
6	32650001	32700000	3.86	0.0005	3.30
6	39000001	39050000	3.40	0.0008	3.12
6	39300001	39350000	3.27	0.0040	2.40
6	43250001	43300000	3.29	0.0011	2.96
6	43350001	43400000	3.40	0.0031	2.50
6	47050001	47100000	2.68	0.0033	2.48
6	49250001	49300000	3.10	0.0033	2.48
6	49750001	49800000	3.63	0.0007	3.13
6	49800001	49850000	3.80	0.0004	3.43
6	50050001	50100000	3.32	0.0047	2.33
6	50100001	50150000	3.40	0.0036	2.44
6	51450001	51500000	3.30	0.0011	2.94
6	52800001	52850000	2.83	0.0051	2.30
6	52850001	52900000	2.72	0.0028	2.55
6	53050001	53100000	2.97	0.0029	2.54
6	53100001	53150000	3.60	0.0009	3.03
6	53150001	53200000	3.81	0.0002	3.73
6	53200001	53250000	2.85	0.0014	2.87
6	53250001	53300000	2.40	0.0008	3.12
6	55800001	55850000	3.37	0.0025	2.61
6	55950001	56000000	3.02	0.0021	2.69
6	56150001	56200000	3.37	0.0040	2.40
6	56300001	56350000	3.35	0.0042	2.38
6	64100001	64150000	3.02	0.0056	2.25
6	64150001	64200000	2.90	0.0007	3.17
6	72350001	72400000	3.22	0.0050	2.30
19	8350001	8400000	2.31	0.0019	2.73
19	30700001	30750000	2.18	0.0035	2.45
20	18050001	18100000	2.73	0.0026	2.58
20	19400001	19450000	1.48	0.0052	2.28
20	22700001	22750000	1.60	0.0032	2.49
21	9400001	9450000	3.02	0.0054	2.27
21	44550001	44600000	3.23	0.0047	2.33
21	56100001	56150000	3.82	0.0003	3.53
21	56200001	56250000	4.06	0.0001	4.08
21	56250001	56300000	3.30	0.0050	2.30
21	56300001	56350000	2.13	0.0041	2.38
21	57500001	57550000	1.53	0.0040	2.40
21	65500001	65550000	3.39	0.0038	2.42
21	65750001	65800000	2.81	0.0054	2.27
22	3250001	3300000	3.32	0.0046	2.33
22	56550001	56600000	2.32	0.0018	2.76
22	59350001	59400000	3.60	0.0029	2.54
23	1350001	1400000	1.50	0.0047	2.32
23	47550001	47600000	3.27	0.0039	2.41
24	5000001	5050000	3.13	0.0032	2.49
24	9750001	9800000	3.43	0.0040	2.39
24	10950001	11000000	2.38	0.0009	3.06
24	13350001	13400000	1.82	0.0017	2.78
24	19400001	19450000	2.85	0.0047	2.33
24	19600001	19650000	3.14	0.0030	2.53
24	19750001	19800000	3.08	0.0043	2.36
24	21800001	21850000	3.31	0.0034	2.47
24	55850001	55900000	3.71	0.0006	3.25

7	68900001	68950000	1.94	0.0011	2.97	24	56900001	56950000	1.78	0.0018	2.74
7	100600001	100650000	2.90	0.0006	3.25	24	57600001	57650000	3.38	0.0024	2.62
8	48550001	48600000	3.50	0.0015	2.81	25	5350001	5400000	2.88	0.0010	2.99
8	48600001	48650000	3.85	0.0007	3.18	25	6050001	6100000	3.84	0.0007	3.13
8	105650001	105700000	1.69	0.0025	2.60	25	6100001	6150000	3.88	0.0016	2.81
9	18050001	18100000	3.97	0.0009	3.06	25	8700001	8750000	3.29	0.0037	2.44
9	18200001	18250000	3.86	0.0006	3.24	25	14950001	15000000	3.58	0.0018	2.74
9	18250001	18300000	3.63	0.0017	2.78	25	15000001	15050000	3.30	0.0048	2.32
9	38400001	38450000	2.66	0.0003	3.60	25	40650001	40700000	3.41	0.0023	2.64
9	102600001	102650000	3.45	0.0011	2.95	25	40750001	40800000	3.30	0.0049	2.31
10	9550001	9600000	3.11	0.0037	2.44	26	3800001	3850000	3.38	0.0039	2.41
10	11250001	11300000	3.57	0.0019	2.72	26	36550001	36600000	3.33	0.0045	2.35
10	15250001	15300000	3.46	0.0029	2.54	26	36600001	36650000	3.40	0.0036	2.45
10	23650001	23700000	2.19	0.0034	2.47	26	36700001	36750000	3.07	0.0041	2.39
10	24450001	24500000	2.19	0.0033	2.49	27	7700001	7750000	2.54	0.0050	2.30
10	84700001	84750000	3.00	0.0022	2.66	27	16450001	16500000	2.56	0.0005	3.30
10	84750001	84800000	2.90	0.0039	2.41	27	19800001	19850000	3.48	0.0017	2.77
11	7050001	7100000	3.44	0.0031	2.51	27	19850001	19900000	3.75	0.0006	3.19
11	34950001	35000000	1.71	0.0023	2.63	27	19900001	19950000	3.78	0.0009	3.04
11	35900001	35950000	3.52	0.0025	2.61	27	21400001	21450000	2.53	0.0051	2.29
11	60300001	60350000	3.10	0.0014	2.84	27	45400001	45403592	2.10	0.0043	2.37
11	60600001	60650000	2.81	0.0049	2.31	28	21900001	21950000	1.87	0.0014	2.85
11	63700001	63750000	1.51	0.0045	2.35	29	33900001	33950000	3.07	0.0043	2.37
11	65550001	65600000	3.46	0.0028	2.55	29	34350001	34400000	3.52	0.0004	3.44
11	65600001	65650000	3.73	0.0011	2.97	29	34400001	34450000	3.45	0.0035	2.46

Supplementary Table S4.13 | Identified Gene desert regions in Muturu based on XPCLR test

Chr	Start (bp)	End (bp)	XPCLR	No of SNPs	Chr	Start (bp)	End (bp)	XPCLR	No of SNPs
1	21000238	21050237	1400.48	600	8	44252553	44302552	1553.38	600
1	103600238	103650237	830.53	600	8	50552553	50602552	935.39	600
1	114850238	114900237	1440.66	600	8	59002553	59052552	1390.39	600
1	114900238	114950237	894.01	600	8	68552553	68602552	829.19	600
1	114950238	115000237	1407.91	600	8	69402553	69452552	1450.35	600
1	115000238	115050237	1582.21	600	8	107902553	107952552	1578.45	600
1	124850238	124900237	1446.17	600	8	108002553	108052552	1414.28	600
1	124950238	125000237	965.75	600	9	44100647	44150646	1390.62	600
1	126750238	126800237	1405.67	600	9	44250631	44300630	1482.84	600
1	132500238	132550237	940.92	600	10	35050086	35100085	1389.97	600
1	132600238	132650237	1472.70	600	10	35100086	35150085	1694.81	600
1	132650238	132700237	1395.67	600	10	61350086	61400085	1788.17	600
1	132750238	132800237	1865.97	600	11	33350050	33400049	903.00	600
1	132800238	132850237	1758.96	600	11	39750105	39800104	1436.14	600
1	132950238	133000237	1084.27	600	11	39850105	39900104	1524.81	600
1	133050238	133100237	1807.71	600	11	42650050	42700049	882.87	600
1	135100238	135150237	1574.81	600	11	54000050	54050049	1063.91	600
1	139600238	139650237	1393.01	600	11	58350050	58400049	831.58	600
1	139950238	140000237	1442.37	600	11	63500105	63550104	1410.94	600
1	150550238	150600237	1401.10	600	11	63800050	63850049	864.25	600
2	22750028	22800027	936.74	600	11	63950050	64000049	816.08	600
2	22900028	22950027	1618.10	600	11	64600105	64700049	1928.59	600
2	62200028	62250027	1759.37	600	11	64950105	65000104	1452.30	600
2	64050028	64100027	1409.89	600	11	65000105	65050104	1543.91	600

2	70500028	70550027	931.57	600
2	76350028	76400027	901.00	600
2	76550028	76600027	804.52	600
2	78900028	78950027	1443.50	600
2	78950028	79000027	873.05	600
2	79000028	79050027	1462.83	600
2	79050028	79100027	1449.65	600
2	79550028	79600027	916.84	600
2	79600028	79650027	1459.89	600
2	79650028	79700027	1025.11	600
2	80050028	80100027	920.94	600
2	81950028	82000027	1400.29	600
2	82300028	82350027	969.91	600
2	82450028	82500027	899.02	600
3	33650050	33700049	1584.46	600
3	48150050	48200049	1496.23	600
3	87350050	87400049	1523.84	600
3	88450050	88500049	1409.20	600
3	114350443	114400442	934.73	600
4	3300334	3350333	854.27	600
4	45350050	45400049	1517.70	550
4	45500050	45550049	1611.40	550
4	52850334	52950049	1533.62	600
4	54300050	54350049	1484.33	550
4	78500334	78550333	848.38	600
4	78550334	78600333	852.29	600
4	83900334	83950333	874.67	600
4	83950334	84000333	877.41	600
12	4150334	4200333	853.29	600
12	17400227	17450226	1797.00	600
12	17500227	17550226	1810.67	600
12	19650334	19700333	1089.90	600
12	19800334	19850333	1121.28	600
12	20600227	20650226	1583.45	600
13	48800088	48850087	1487.56	600
13	48850088	48900087	861.37	600
13	48900088	48950087	1750.25	600
13	48950088	49000087	879.87	600
13	49000088	49050087	1802.09	600
14	30650430	30700429	795.11	600
14	46450430	46550429	1524.64	600
14	49650297	49700296	1457.45	600
15	7200550	7250549	1319.01	600
15	11850505	11900504	1558.36	600
15	11900550	11950549	983.36	600
15	25800550	25850549	794.74	600
15	31350550	31400549	795.76	600
16	8552189	8602188	1763.53	600
16	8602189	8652188	816.81	600
16	8652189	8702188	1838.09	600
16	8802189	8852188	1577.19	600
16	8852189	8902188	815.56	600
16	22302189	22352188	848.88	600
16	51352189	51402188	1524.96	600
17	9100328	9150327	825.35	600
17	37150493	37200492	1392.14	600

4	84100050	84150333	1824.93	600
4	84300050	84350049	1698.36	550
4	84350050	84400049	1470.79	550
5	11600045	11650044	1491.95	600
5	18550055	18600054	810.05	600
5	18650055	18700054	828.61	600
5	53300045	53350044	1434.07	600
5	90200055	90250054	960.82	600
5	97950055	98000054	816.53	600
6	5100466	5150465	1017.44	600
6	30200420	30250465	1513.28	600
6	31150420	31200419	1622.61	600
6	50250420	50300465	1545.30	600
6	50400466	50450465	960.84	600
6	53000466	53050465	891.28	600
6	53050466	53100465	878.72	600
6	63700466	63800465	1462.35	600
6	63850466	63900465	1032.60	600
6	71300420	71350419	1420.38	600
6	71550420	71600419	1591.13	600
6	74400466	74450465	1003.95	600
6	74500420	74550419	1702.35	600
6	74750466	74800465	1101.15	600
6	74800466	74850465	1096.55	600
6	75000420	75050419	1551.99	600
6	75600420	75650419	1429.00	600
6	75700420	75750465	1442.21	600
6	77100466	77150465	898.19	600
17	61150328	61200492	1698.93	600
17	69350328	69400327	822.39	600
18	38500649	38550648	1071.47	600
18	38600649	38650648	873.95	600
19	4400507	4450506	1087.40	600
19	5950507	6000506	924.26	600
19	6000507	6050506	955.35	600
19	39300507	39350506	821.96	600
19	63350474	63400473	1407.30	580
20	5800629	5850628	873.62	600
20	32350467	32400466	1470.48	600
20	52800467	52850466	1459.26	600
21	29150015	29200014	1605.86	600
21	54950015	55000014	1648.60	600
22	36251676	36301675	1546.08	600
22	36301676	36351675	1758.80	600
22	36601676	36651675	1436.87	600
24	51014	101013	1616.34	600
24	19751014	19801675	1636.91	600
24	39451676	39501675	927.07	600
24	39551014	39601675	1577.68	600
24	43051014	43101013	1521.10	600
24	49651014	49701013	1490.60	600
24	61301676	61351675	947.82	600
25	5450251	5500250	1245.96	600
25	16000251	16100031	1389.64	600
25	39850251	39900250	856.29	600
26	2750244	2800243	970.79	600

6	77150466	77250419	1432.87	600		26	11600428	11650427	1456.28	600
6	77250466	77300465	1029.02	600		26	11850244	11900243	853.31	600
6	77350420	77400419	1535.30	600		26	12000244	12050243	965.06	600
6	96550420	96600419	1508.55	600		27	3400580	3500465	1410.83	600
6	104700466	104750465	971.52	600		27	3550466	3600465	1020.15	600
7	40500729	40550728	857.40	600		28	800478	850477	1424.93	600
7	41000729	41100728	1555.91	600		28	850478	900477	1505.37	600
7	52150396	52200395	1840.73	600		28	14150478	14200477	1402.54	600
7	87550729	87600728	809.31	600		28	20250478	20300477	1470.06	600
7	87750729	87800728	888.10	600		28	20401017	20451016	832.90	600
7	100750729	100800728	1165.83	600		29	19200011	19250025	1575.84	600
7	100950729	101000728	889.12	600		29	19300011	19350010	1514.03	600
7	101000729	101050728	1053.34	600		29	19400011	19450025	1634.81	600
7	101100396	101200395	1576.83	600		29	19450026	19500025	934.55	600
8	43202553	43252552	1444.18	600		29	30850026	30900025	823.75	600
8	44102553	44152552	1013.38	600		29	40150026	40200025	825.21	600
8	44202553	44252552	977.00	600						

Supplementary Table S4.14 | Identified Gene desert regions in N'Dama based XPCLR test

Chr	Start (bp)	End (bp)	XPCLR	No of SNPs	Chr	Start (bp)	End (bp)	XPCLR	No of SNPs
1	105700238	105750237	439.08	600	9	5150647	5200646	440.57	600
1	105750238	105800237	534.62	600	9	9750631	9800630	1199.30	600
1	109000238	109050237	479.11	600	9	18000631	18050630	1059.12	600
1	114800238	114850237	1230.65	600	9	18050631	18100630	1359.76	600
1	157850238	157900237	1131.42	600	9	20800647	20850646	443.65	600
1	157950238	158000237	1188.64	600	10	15250086	15300125	1535.29	600
2	8200028	8250027	462.51	600	10	15300126	15400085	1234.87	600
2	43300028	43350027	508.59	600	10	35050086	35100085	1242.73	600
2	70800028	70850027	1074.05	600	10	35100086	35150085	1088.08	600
2	72800028	72850027	1442.76	600	10	38750126	38800125	495.55	600
2	72850028	72900027	698.85	600	10	39250126	39300125	537.07	600
2	126550028	126600027	1294.50	600	10	42300126	42350125	637.10	600
3	37100443	37150442	495.69	600	10	44000126	44050125	435.48	600
3	37250050	37300049	1078.51	600	10	59000126	59050125	477.96	600
3	39050443	39100442	578.68	600	10	59150126	59200125	441.11	600
3	39300050	39350049	1139.57	600	10	75650086	75700085	1053.74	600
3	61350443	61400442	523.49	600	11	4050105	4100104	1091.26	600
4	1350334	1400333	1078.91	600	11	7050105	7100104	1368.57	600
4	1450050	1500049	1058.87	590	11	31850050	31900049	851.44	600
4	1550334	1600333	722.03	600	11	31900050	31950104	1352.61	600
4	26300050	26350049	1391.47	590	11	32000050	32050104	1138.62	600
4	26650050	26700049	1181.40	590	11	33700050	33750049	567.15	600
4	26750050	26800049	1330.74	590	11	34000105	34050104	1186.41	600
4	26850334	26950333	1214.23	600	11	34150105	34200104	1284.24	600

4	27000334	27050333	454.34	600		11	34350105	34450049	1517.51	600
4	32950050	33000049	1055.55	590		11	34450105	34500104	1406.55	600
4	83850050	83900049	1049.97	590		11	34500105	34550104	1288.83	600
4	85000334	85050333	437.63	600		11	36950050	37000049	474.01	600
5	13700045	13750044	1199.65	600		11	50500105	50550104	1222.51	600
5	13800055	13850054	704.36	600		11	50750105	50800104	1071.92	600
5	52950045	53000044	1387.84	600		11	64150105	64200104	1238.03	600
5	53000045	53050044	1298.26	600		12	12100334	12150333	437.79	600
5	53050045	53100044	1142.17	600		12	19650334	19700333	1021.65	600
5	53100045	53150044	1119.72	600		12	19800334	19850333	1004.64	600
5	53150045	53200054	1424.35	600		12	19850334	19900333	460.36	600
5	53300045	53350044	1298.45	600		12	19950227	20000226	1100.79	600
5	74050045	74100044	1204.24	600		12	26450334	26500333	455.00	600
5	81000045	81100054	1174.61	600		12	37550334	37600333	508.55	600
6	10800420	10850419	1325.21	600		12	37800334	37900226	1052.60	600
6	10850420	10900465	1291.52	600		12	38850334	38900333	449.64	600
6	11050420	11100419	1134.70	600		12	38900334	38950333	534.95	600
6	11200420	11250419	1600.42	600		12	41150334	41200333	472.50	600
6	11250466	11300465	480.43	600		12	41300334	41350333	778.22	600
6	19800466	19850465	453.96	600		12	41550334	41600333	512.13	600
6	29600420	29650465	1177.59	600		12	57950227	58000226	1205.07	600
6	29650466	29700465	941.48	600		12	64850227	64900226	1135.71	600
6	29700466	29750465	438.62	600		13	33950088	34000087	514.81	600
6	30100420	30150419	1064.27	600		13	34000088	34050087	1568.99	600
6	32450466	32500465	594.79	600		15	15350550	15400549	492.43	600
6	32500466	32600419	2043.76	600		15	57600550	57650549	603.05	600
6	32600420	32650419	1570.72	600		16	8152189	8202188	1168.86	600
6	32650420	32700419	1362.68	600		16	8302189	8352188	472.66	600

6	32700420	32750419	1266.82	600		16	8602189	8652188	527.42	600
6	33250466	33300465	453.84	600		16	8652189	8702188	1409.22	600
6	34050420	34100419	1071.56	600		16	8702189	8752188	433.17	600
6	38150420	38200419	1133.90	600		16	8802189	8852188	1456.56	600
6	38200420	38250419	1081.22	600		16	8852189	8902188	640.29	600
6	43350420	43400465	1213.15	600		16	8902189	8952188	568.61	600
6	46650466	46700465	464.87	600		16	46652189	46702188	1624.75	600
6	49750466	49800465	605.58	600		17	17150493	17200492	1069.91	600
6	49800466	49850465	480.72	600		17	39200328	39250327	510.18	600
6	49900466	50000419	1140.86	600		17	39250328	39300327	593.34	600
6	50000420	50100465	1387.10	600		18	44250149	44300148	1131.35	600
6	50150420	50200419	1339.83	600		19	2650474	2750506	1565.97	600
6	50250420	50300465	1169.09	600		19	2750507	2800506	465.39	600
6	52350466	52400465	490.69	600		19	2800507	2850506	1493.16	600
6	52400466	52450465	455.63	600		19	2900474	2950473	1507.97	590
6	52550466	52600465	552.88	600		19	2950474	3000473	1206.16	590
6	52650466	52700465	582.43	600		19	3000474	3050506	1531.90	600
6	52850466	52950419	1066.09	600		19	3050507	3150506	1315.33	600
6	52950420	53000465	1218.91	600		19	3150507	3200506	908.24	600
6	53000466	53050465	1328.72	600		19	3550474	3600473	1077.51	590
6	53050466	53100465	1406.28	600		19	3750474	3850506	1320.45	600
6	53100466	53150465	511.48	600		20	18050629	18100628	511.22	600
6	55700420	55800419	1412.58	600		20	18100629	18150628	598.06	600
6	55850420	55900419	1461.90	600		20	25800629	25850628	616.62	600
6	55900420	55950465	1332.02	600		20	25900629	25950628	662.60	600
6	56000420	56050419	1239.82	600		21	29100015	29150014	1388.65	600
6	56050466	56100465	824.81	600		21	29150015	29200014	1457.61	600
6	56150466	56200465	778.64	600		21	29400015	29450014	1167.78	600

6	56300466	56350465	706.33	600		21	56350015	56400014	528.85	600
6	56350466	56450419	1214.48	600		22	44001676	44051675	1428.31	600
6	56650466	56700465	531.74	600		23	40050076	40100075	1111.29	580
6	56700466	56750465	458.88	600		24	16901676	16951675	639.06	600
6	63750466	63800465	441.23	600		24	17251676	17301675	689.55	600
6	86450420	86500419	1133.78	600		24	49651014	49701013	1594.34	600
6	86500420	86550419	1071.23	600		24	51201014	51251675	1166.89	600
7	13000396	13050395	1107.68	600		25	5450251	5500250	1081.74	600
8	21452553	21502552	586.26	600		25	5500251	5550250	630.94	600
8	37402553	37452552	1078.15	600		25	5550251	5600250	527.75	600
8	48152553	48202552	1322.82	600		25	5950251	6000250	473.35	600
8	48552553	48602552	1637.28	600		25	14800032	14850031	1303.11	600
8	48602553	48652552	623.05	600		25	14900032	14950031	1316.39	600
8	83402553	83452552	932.29	600		25	14950032	15000031	1237.02	600
8	106552553	106602552	1079.92	600		25	40650032	40750250	1085.04	600
8	112702553	112752552	453.25	600		26	36550244	36600243	434.88	600
9	3250647	3300646	453.54	600		26	36600428	36700243	1275.79	600
9	3350647	3400646	577.30	600		27	18500466	18550465	631.95	600
9	3700647	3750646	535.44	600		27	19850580	19900579	1102.41	580
9	3750647	3800646	446.01	600		27	19900580	19950579	1066.98	580
9	4750631	4800646	1192.87	600		27	22000466	22050465	613.05	600
9	5000631	5050630	1158.97	600		27	22050466	22100465	479.41	600

Supplementary Table S4.15 | List of functional annotation terms as revealed by DAVID and the related candidate genes identified in Muturu **(file is provided as an electronic format)**

Supplementary Table S4.16 | List of functional annotation terms as revealed by DAVID and the related candidate genes identified in N'Dama **(file is provided as an electronic format)**

Supplementary Table S4.17 | Functional annotation of candidate genes linked to trypanotolerance in WAT using PANTHER **(file is provided as an electronic format)**

Supplementary Table S4.18 | Candidate selected regions and related genes identified following population differentiation F_{ST} test between trypanotolerant WAT and trypanosusceptible other cattle breeds **(file is provided as an electronic format)**

Chapter 5

Supplementary Table S5.1 | Genes overlapping candidate selective sweep region in Kenyan Boran based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.2 | Genes overlapping candidate selective sweep region in Gash based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.3 | Genes overlapping candidate selective sweep region in Butana based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.4 | Genes overlapping candidate selective sweep region in Ogaden based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.5 | Genes overlapping candidate selective sweep region in Fulani based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.6 | Genes overlapping candidate selective sweep region in Baggara based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.7 | Genes overlapping candidate selective sweep region in Kenana based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.8 | Genes overlapping candidate selective sweep region in Aryashai based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.9 | Genes overlapping candidate selective sweep region in Gir based on iHS test **(file is provided as an electronic format)**

Supplementary Table S5.10 | Genes overlapping candidate selective sweep region in Kenyan Boran based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.11 | Genes overlapping candidate selective sweep region in Gash based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.12 | Genes overlapping candidate selective sweep region in Butana based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.13 | Genes overlapping candidate selective sweep region in Ogaden based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.14 | Genes overlapping candidate selective sweep region in Fulani based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.15 | Genes overlapping candidate selective sweep region in Baggara based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.16 | Genes overlapping candidate selective sweep region in Kenana based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.17 | Genes overlapping candidate selective sweep region in Aryashai based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.18 | Genes overlapping candidate selective sweep region in Gir based on ZHp test **(file is provided as an electronic format)**

Supplementary Table S5.19 | List of unique and shared candidate genes detected by ZHp and iHS in eight African zebu breeds **(file is provided as an electronic format)**

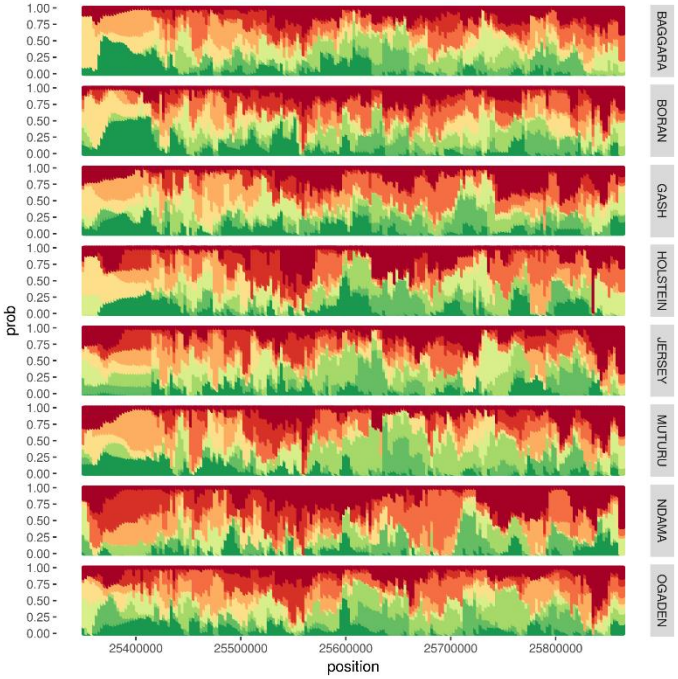
Supplementary Table S5.20 | List of selected genes shared by at least one African and Asian zebu breeds **(file is provided as an electronic format)**

Supplementary Table S5.21 | List of milk production related genes identified in African zebu breeds including the African zebu non-dairy breeds (AFZND) **(file is provided as an electronic format)**

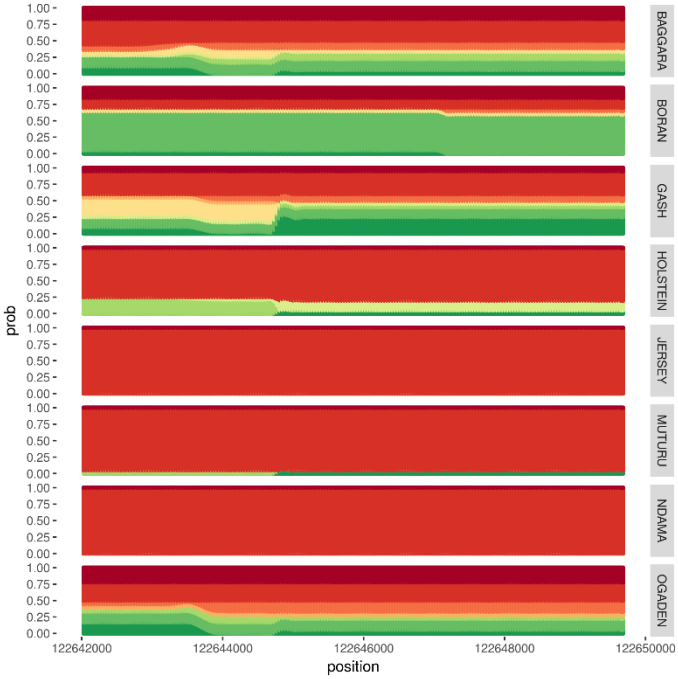
Supplementary Table S5.22 | List of functional annotation terms as revealed by DAVID and the related candidate genes following the comparison of African zebu to non-African cattle breeds **(file is provided as an electronic format)**

Supplementary Figures

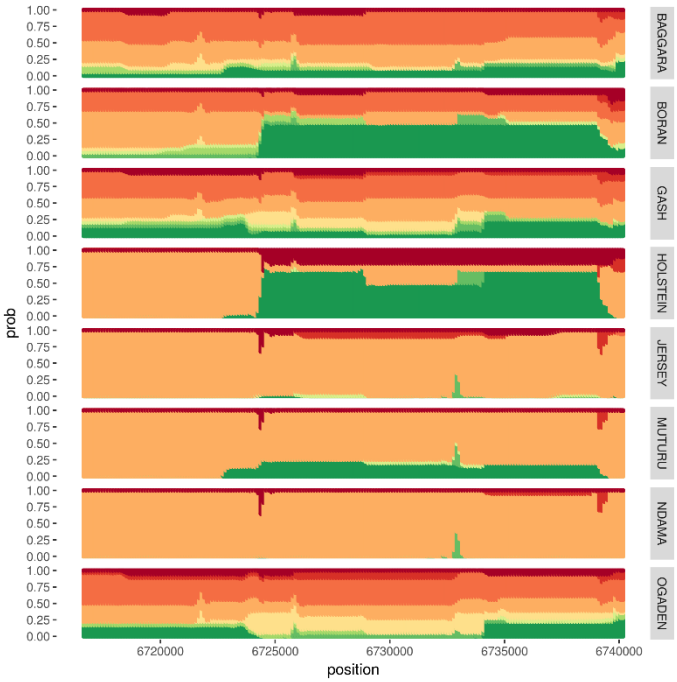
Chapter 4



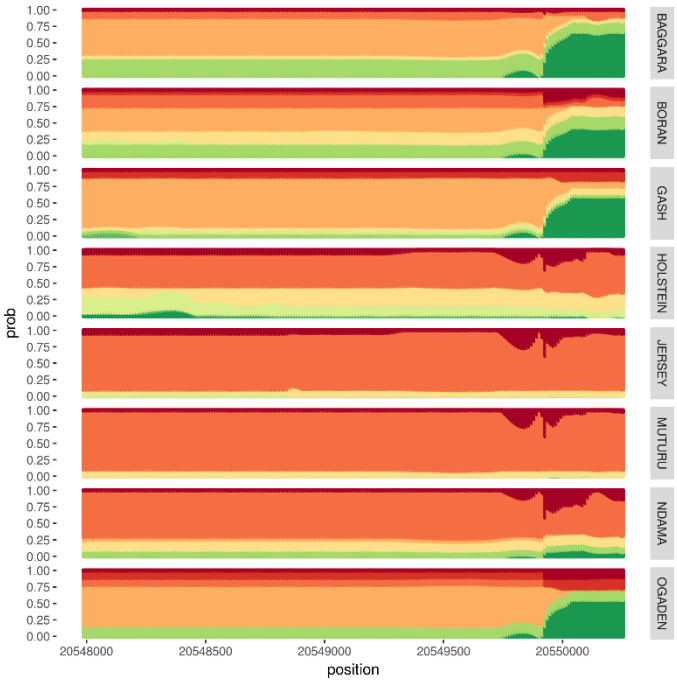
Supplementary Figure S4.1 | Haplotype diversity among trypanotolerant and trypanosusceptible cattle populations at the MHC class II gene region



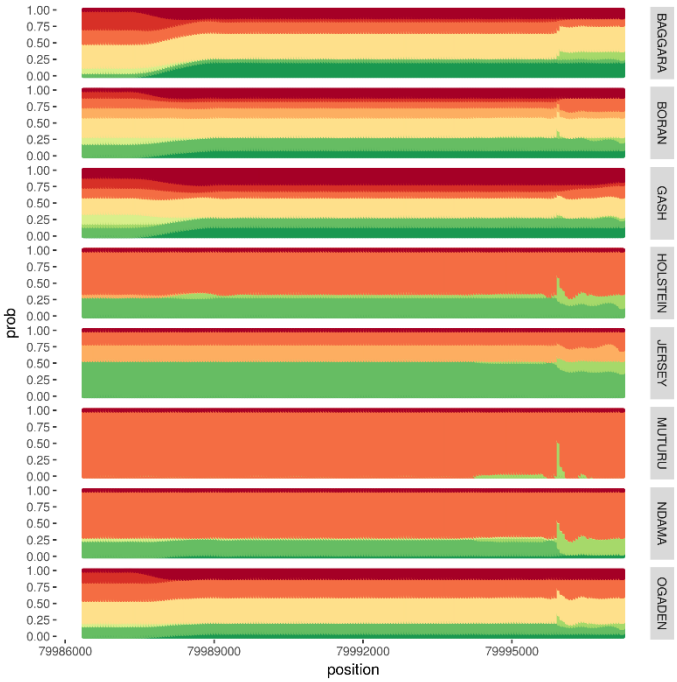
Supplementary Figure S4.2 | Haplotype diversity among trypanotolerant and trypanosusceptible cattle populations at the *HCRTR1* gene region



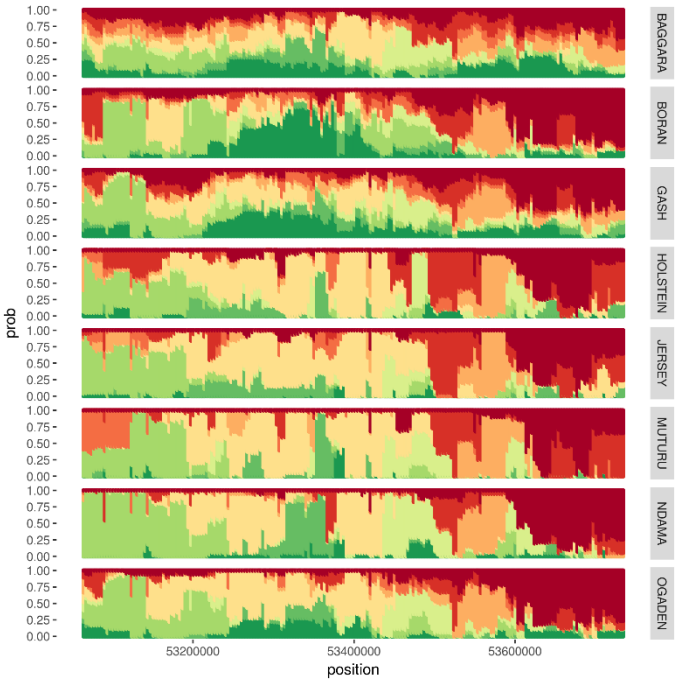
Supplementary Figure S4.3 | Haplotype diversity among trypanotolerant and trypanosusceptible cattle populations at the *SLC40A1* gene region



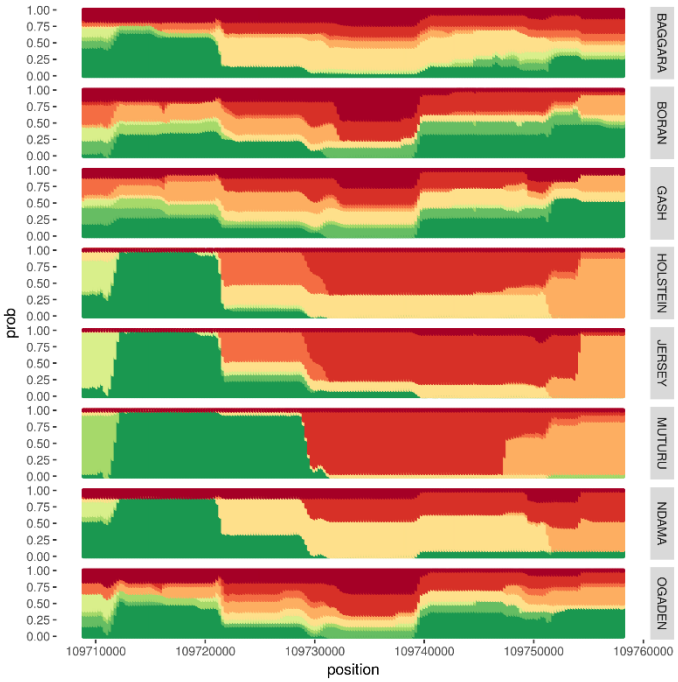
Supplementary Figure S4.4 | Haplotype diversity among trypanotolerant and trypanosusceptible cattle populations at the *TICAM1* gene region



Supplementary Figure S4.5 | Haplotype diversity among trypanotolerant and trypanosusceptible cattle populations at the *INHBA* gene region

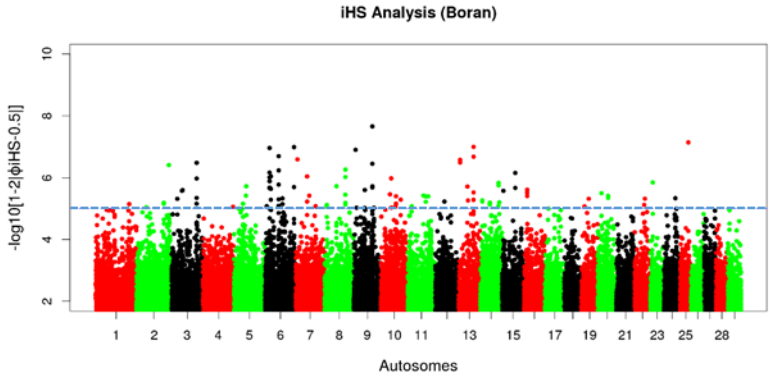


Supplementary Figure S4.6 | Haplotype diversity among trypanotolerant and trypanosusceptible cattle populations at the *ARHGAP15* gene region

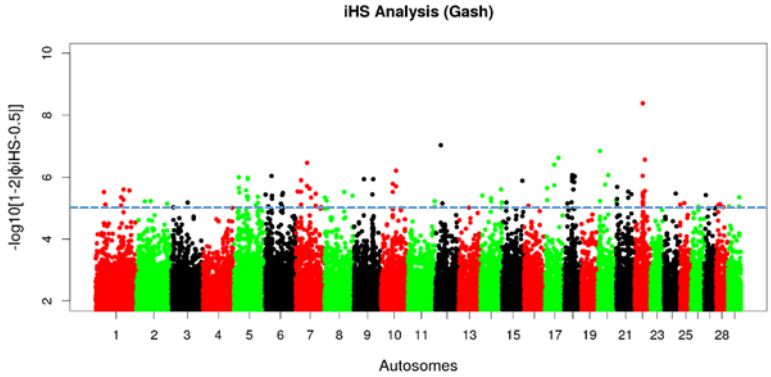


Supplementary Figure S4.7 | Haplotype diversity among trypanotolerant and trypanosusceptible cattle populations at the *GFM1* gene region

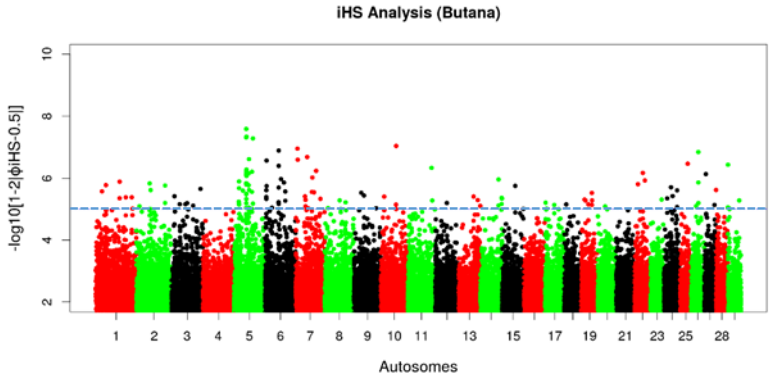
Chapter 5



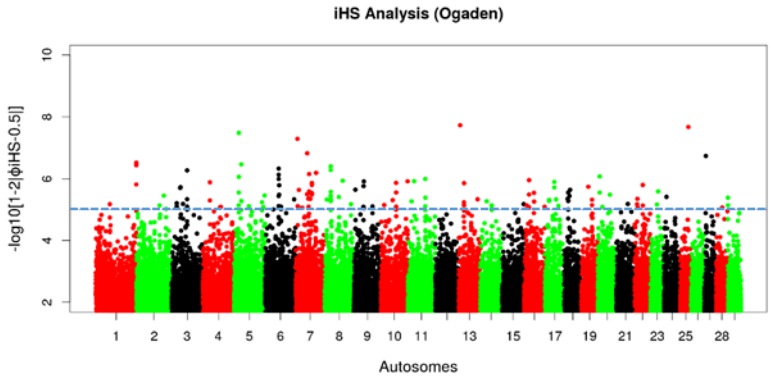
Supplementary Figure S5.1 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Kenyan Boran



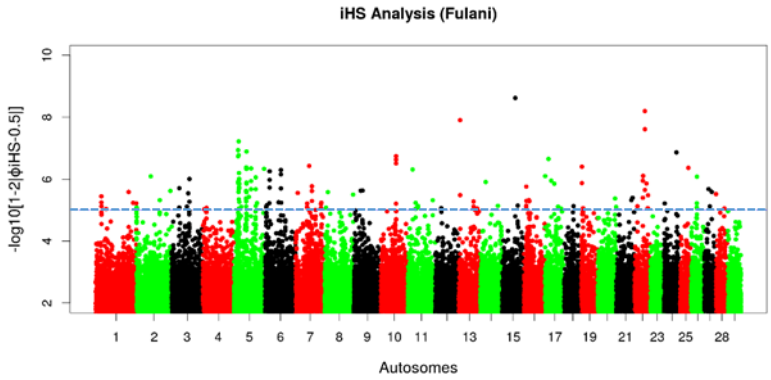
Supplementary Figure S5.2 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Gash



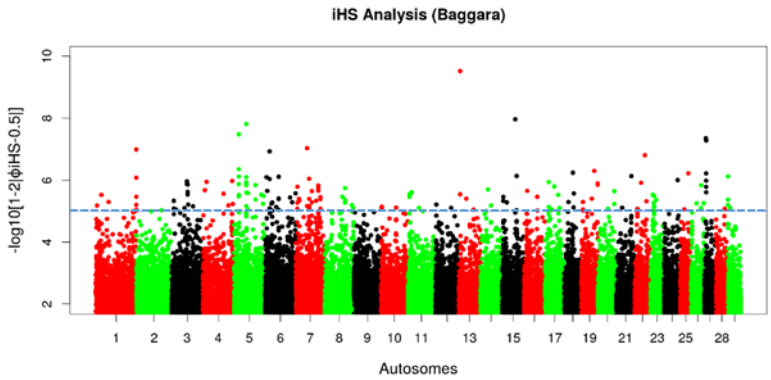
Supplementary Figure S5.3 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Butana



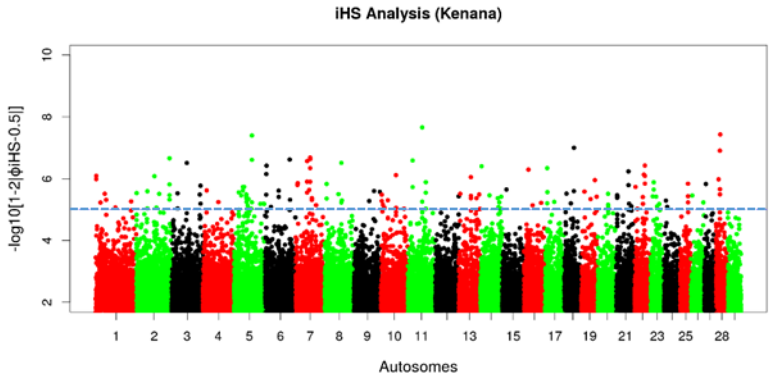
Supplementary Figure S5.4 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Ogaden



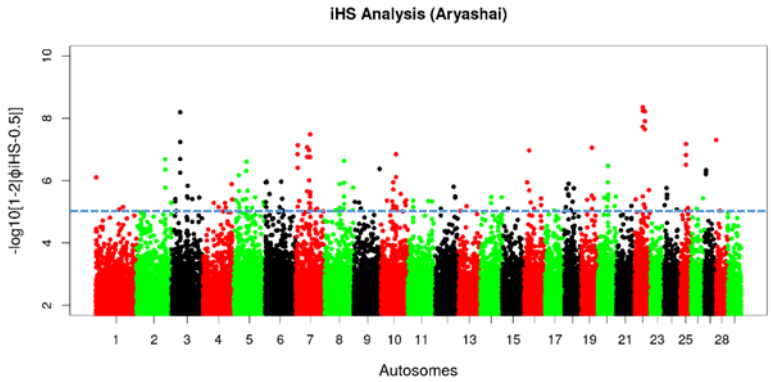
Supplementary Figure S5.5 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Fulani



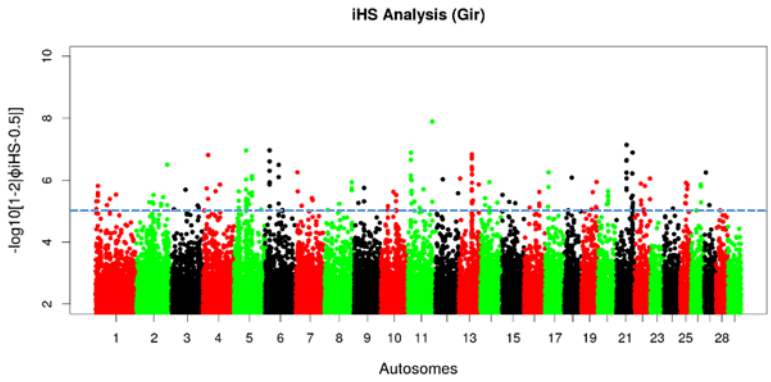
Supplementary Figure S5.6 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Baggara



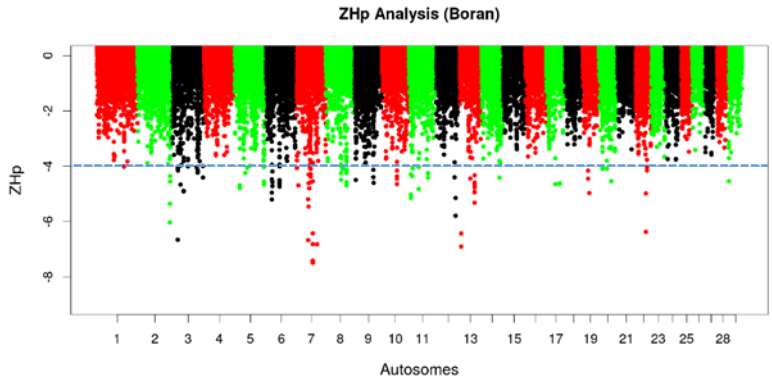
Supplementary Figure S5.7 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Kenana



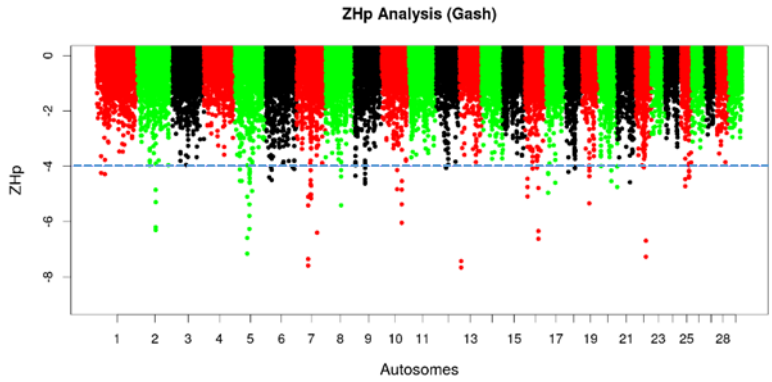
Supplementary Figure S5.8 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Aryashai



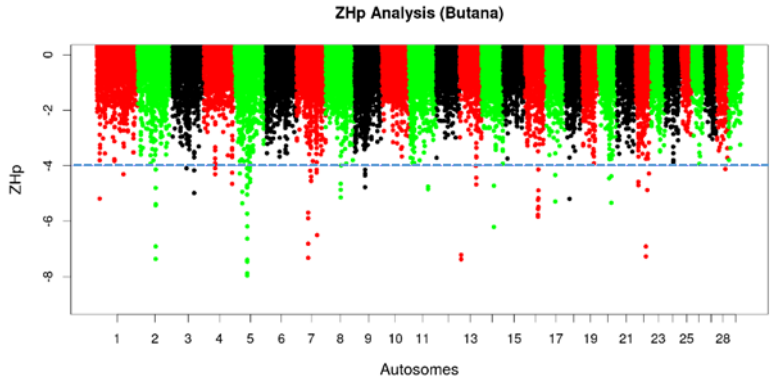
Supplementary Figure S5.8 | Manhattan plot of the distribution of iHS scores along bovine autosomes in Gir



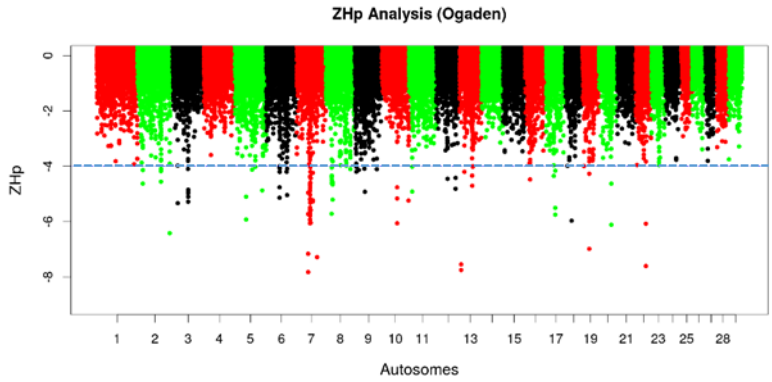
Supplementary Figure S5.10 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Kenyan Boran



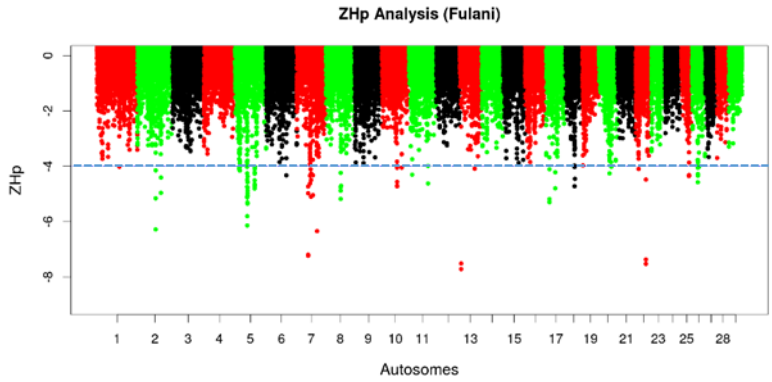
Supplementary Figure S5.11 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Gash



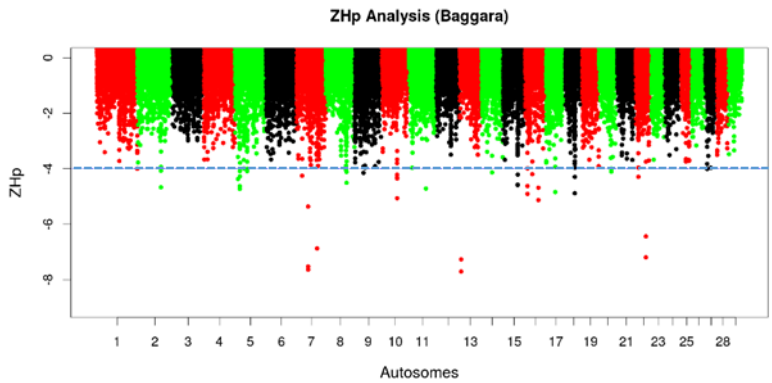
Supplementary Figure S5.12 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Butana



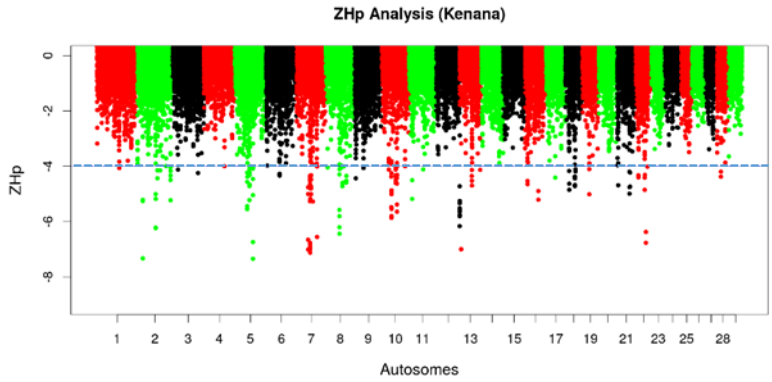
Supplementary Figure S5.13 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Ogaden



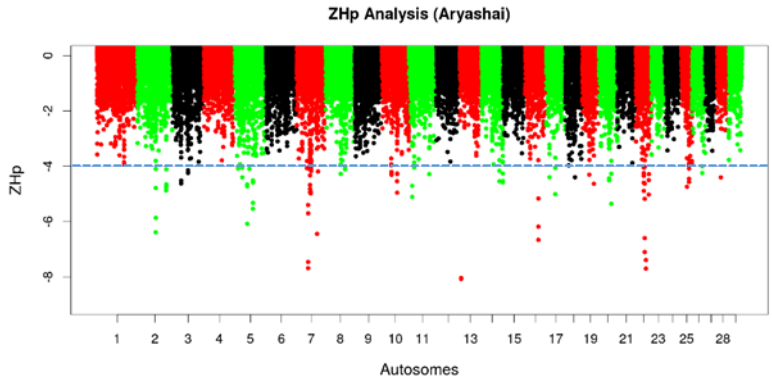
Supplementary Figure S5.14 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Fulani



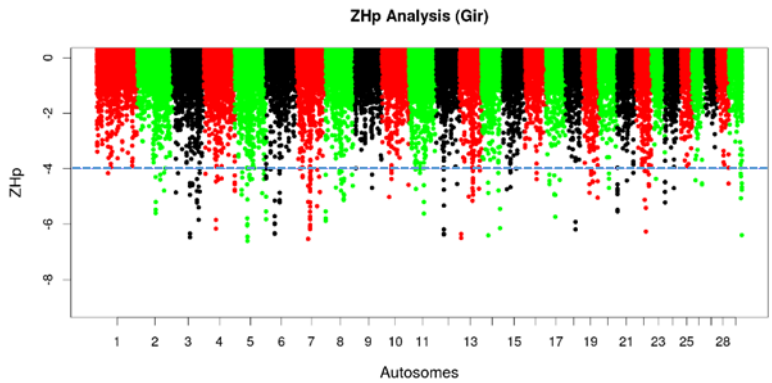
Supplementary Figure S5.15 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Baggara



Supplementary Figure S5.16 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Kenana



Supplementary Figure S5.17 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Aryashai



Supplementary Figure S5.18 | Manhattan plot of the distribution of *ZHp* scores along bovine autosomes in Gir