Whole genome analysis of African indigenous cattle breeds to assess genetic diversity, demographic history and selection signatures

By

Maano Bryton Malima

(20744049)

Supervisor: Prof Fourie Joubert

Co-Supervisor: Dr Avhashoni Agnes Zwane

Submitted in partial fulfilment of the requirements of the degree

Masters in Bioinformatics

Centre for Bioinformatics and Computational Biology

Department of Biochemistry, Microbiology, and Genetics

Faculty of Natural and Agricultural Sciences

University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

29 September 2022

Declaration

I, Maano Bryton Malima, hereby declare that the work on this dissertation is my original work, except where acknowledgments indicate otherwise. This work has not previously been submitted by me for a degree at this or any other University.

Signature: 

Date: 2022/09/27

Acknowledgements

This work would not have been possible without the help of the following people and organizations.

My supervisors, **Dr Avhashoni Agnes Zwane** and **Prof Fourie Joubert.** For giving me an opportunity to do what I love. I will forever be grateful for your support. I could not have done this without you and I learnt a lot!

The **Meat Industry Trust** – For funding me.

The **Agricultural Research Council** – for funding and supporting the project.

Dr Abdulfatai Tijjani and the **Agricultural Research Council-Animal Production (ARC-AP)** for sharing the sequence data for the project.

The students and personnel of **Animal Breeding and Genetics**, ARC-AP, for all your support and encouragement during this project.

The students and personnel of the **Centre for Bioinformatics and Computational Biology**, University of Pretoria, for the motivation and support.

My friends for the love and support during the project. Thank you for always believing in me and being by my side throughout the project.

My special thanks to my family, for always being a source of inspiration!

I THANK YOU!

Table of Contents

Declaration i

Acknowledgements ii

Abbreviations v

List of Figures vii

List of Tables xii

Abstract xiii

Chapter 1 1

1.1 General Introduction 1

1.2 Aim of the study 3

1.3 Specific aims: 3

Chapter 2 5

2. Literature review 5

2.1 Introduction 5

2.2 Importance of livestock farming in African communities 6

2.3 History of African indigenous cattle breeds 7

2.4 Current genomic research on African indigenous cattle breeds 17

2.5 Brief history of genomic markers and availability of genome-wide data 19

2.6 Benefits of applying bioinformatics in livestock genomics research 21

Chapter 3 22

3. Material and Methods 22

3.1 Introduction 22

3.2 Sampling and data collection 22

3.3 Bioinformatics tools and availability of high computing environment 24

3.4 Data quality control, and mapping 26

3.5 Variant calling 26

3.6 Population genomics analysis 27

3.7 Introgression, migration, admixture and gene flow analysis of African breeds using D-statistics and Treemix 29

Chapter 4 33

4. Results: Evolutionary changes and relationships of African cattle populations 33

4.1 Introduction 33

4.2 Summary statistics from mapping of the whole-genome data 33

4.3 Variants statistics 37

4.4 Heterozygosity estimates and nucleotide diversity 37

4.5 Genetic relationships and population structure 39

4.6 Introgression and gene flow analysis (ABBA-BABA Statistics) and Treemix 42
4.7 D-statistics follow up-using a more distant out-group 45
4.8 Demographic analysis 46
4.9 Identification of candidate genomic regions under selection 49
Chapter 5 61
5. General Discussion and Conclusions 61
5.1 Introduction 61
5.2 Sequencing, mapping and variants calling 61
5.3 Genomic diversity and population structure of African cattle breeds 62
5.4 Introgression signatures, admixture, and migrations events 64
5.5 Demographic changes 66
5.6 Selection signatures 68
5.8 Conclusions and future perspectives 70
5.7 Limitation of this study 70
References 72

Abbreviations

AFLPs……..……...…Amplified Fragment Length Polymorphism

AFR…….....Afrikaner

AnGR……..……Animal Genetic Resources for food and agriculture

ANGSD…….....Analysis of Next Generation Sequencing Data

ANK…….....Ankole

ARC……..…….....Agricultural Research Council

BON……...……..Bonsmara

bp…….……...……...Base pair

CHPC……..……Centre for High Performance Computing

Chr……..Chromosome

dadi…….Diffusion Approximations for Demographic Inference

dbSNP……..……Single Nucleotide Polymorphism Database

DNA ……...…….Deoxyribonucleic acid

D-stats…….D-statistics

ENA…….. ……..European Nucleotide Archives

FAO……..……Food and Agriculture Organization

GATK…….……Genome Analysis Toolkit

HOL……...…….Holstein

HWE……..……...…….……Hardy Weinberg Equilibrium

IG……Interglacial

iHS …….....integrated Haplotype scores

InDels…….Insertion–deletion mutations

KEN……...……..Kenana

kya……. ……...Thousand years ago

LGM….. ……..Last Glacial Maximum

MPT……...…….Mid-Pleistocene transition

mya..……..Million Years Ago

NCBI ……National Center for Biotechnology Information

ND…….....N'Dama

NGI……..…….……...Nguni

NGS …….…….….Next Generation Sequencing

PCA…….…..Principal Component Analysis

PCR……..…….……...Polymerase Chain Reaction

PPT ……Plio-Pleistocene transition

PSMC…….....The Pairwise Sequentially Markovian Coalescent

RAPDs…….…..Random Amplified Polymorphic DNA

RFLPs…….…..……...Restriction Fragment Length Polymorphisms

RNA ……...……..……Ribonucleic acid

SA……..…….……..South Africa

SFS…… …….....Site Frequency Spectrum

SNP…….....Single Nucleotide Polymorphism

UK……United Kingdom

US……..United State of America

WGS ……Whole Genome Sequencing

XP-EHH…….……...Cross-population Externded Haplotype Homozygosity

µ……...…...……Micro

List of Figures

Figure 2.1: The depiction of how genetic diversity, directly and indirectly, affect the populations ……..…….6

Figure 2.2: Inferred domestication areas and possible migrations that shaped the creation of African cattle breeds……...8

Figure 2.3: The picture of Kenana (zebu) cattle with its unique features. Visible unique features are grey colour, large ears, loose skin, and a hump above shoulders……..……11

Figure 2.4: Unique features of the Ankole cattle breed, visible unique features are; spotted colour, medium size, long and asymmetrical horn…….……....12

Figure 2.5: The above picture is showing unique features of the Afrikaner cattle breed, visible unique features are; light tan red colour, medium size, loose skin, with the picture showing the polled type …….……13

Figure 2.6: The above picture is showing unique features of the N'Dama cattle breed, visible unique features are; short legs, sandy color, small sized, thick and deep neck, horns are curved upwards……..…...14

Figure 2.7: The above picture is showing unique features of the Nguni cattle breed, visible unique features are; spotted colour, strong legs and hoofs, smaller in size, small sized horns and brush-like tail ……..……15

Figure 2.8: The above picture is showing unique features of the Bonsmara cattle breed, visible unique features are; red in colour, strong legs and a larger mediam size body showing blend of *Bos indicus* and *Bos taurus*, and de-horned as per breed standard……..……16

Figure 3.1: The map showing the distribution of the breeds used in this study according to their agro-ecological zones ……..……23

Figure 3.2: The analysis workflow from raw data processing, mapping, variant calling and population genomics analysis. Tools and software packages are highlighted in bold and analysis methods are shown is circles…….……..25

Figure 3.3: Showing the ABBA/BABA patterns and positions of individuals (genomes, P1, P2, P3, and O) on the tree topology. Note: The red arrow between P2 and P3 is showing an example of a possible gene flow between P2 and P3 with significant ABBA……..……29

Figure 4.1: Summary statistics of variants identified using GATK best practices, with the highest number of variants identified being SNPs, followed by InDels, multiallelic sites and multiallelic SNPs……37

Figure 4.2: Genome-wide heterozygosity estimates in all the populations. Kenana (KEN) cattle had the highest variation, followed by Ankole (ANK), Nguni (NGI), Afrikaner (AFR), Bonsmara (BON), N'Dama (ND) and Holstein (HOL). These estimates reflect the diversity within each population and they may be used to inform breeding and conservation programs.38

Figure 4.3 Estimations of nucleotide diversity between breeds. These estimates were produced by nucleotide diversity index (Pi) with a non-overlapping window of 100 kb using VCFtools and they are showing variation between the studied populations. Interestingly, Bonsmara (a composite) is showing to have a higher genetic diversity compared to N'Dama as observed in (Figure 4.2)……39

Figure 4.4: Pairwise genetic distances reconciled into a neighbour joining species tree from whole African cattle's genomes. SA breeds are denoted as NGI*, BON* and AFR* and they showed relationship with other African breeds Ankole (ANK*) and Kenana (KEN*), also Bonsmara (BON) showed relationship with Holstein (HOL*) and N'Dama (NDA*)……40

Figure 4.5: Comparative analysis of African indigenous cattle breeds based on the SNP data. SA breeds are denoted by the Red circle and their relationships were being compared to other African breeds (Purple; Kenana, Blue; N'Dma, Yellow; Holstein, Green; Ankole)……41

Figure 4.6: The STRUCTURE analysis of studied breeds; the lowest standard error was K=2 with CV error: 0.60915. Although the plot is not showing clearly the ancestral relationships between populations, it managed to capture some relationships, e.g. K=3 was informative in terms of the geographic location's relationships……..……42

Figure 4.7: Inference of population splits and migration events in different African populations. These are possible migration events that shaped the genomes of the studied populations. This plots presents novel relationships that has not been reported in previous studies e.g. NGI and KEN……..44

Figure 4.8: Population sizes inferred from SA breeds. The X-axis is the time scaled in (kya), and the Y-axis is representing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$……..…….47

Figure 4.9: Population sizes inferred from West (N'Dama), East (Kenana), and East (Ankole) African breeds. The X-axis is representing the time scaled in (kya), and the Y-axis is showing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$…….……47

Figure 4.10: Population sizes inferred from SA (Nguni), East (Ankole) African breeds. The X-axis is the time scaled in (kya), and the Y-axis is representing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$. This plot also show that the trajectory is not affected by the number of populations plotted……..……48

Figure 4.11: Population sizes inferred from, East (Kenana), and SA (Nguni) African breeds. The X-axis the time scaled in thousand years (kya), and the Y-axis is representing the effective population size. g is the generation time in a mutation rate of $1.1x10^{-8}$……...48

Figure 4.12: Population sizes inferred from East (Kenana), and SA (Afrikaner) African breeds. The X-axis is showing the time scaled in (kya), and the Y-axis is showing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$……...……49

Figure 4.13: The Manhattan plot for the iHS test based on the autosomes of the Nguni cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)…….……...50

Figure 4.14: The Manhattan plot for the iHS test based on the autosomes of the Afrikaner cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)…….50

Figure 4.15: The Manhattan plot for the iHS test based on the autosomes of the Kenana cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)……51

Figure 4.16: The Manhattan plot for the iHS test based on the autosomes of the N'Dama cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)……51

Figure 4.17: The distribution of standardized cross-population extended haplotype homozygosity (XP-EHH) scores of the Ankole and Afrikaner populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)……52

Figure 4.18: The distribution of standardized cross-population extended haplotype homozygosity (XP-EHH) scores of the Ankole and Nguni populations. The horizontal dotted line is separating significant candidate regions using threshold of (–log10 (p-value) > 4)……52

Figure 4.19: The distribution of standardized cross-population extended haplotype homozygosity (XP-EHH) scores of the Kenana and Afrikaner populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)……53

Figure 4.20: The distribution of standardized cross-population extended haplotype homozygosity (XP-EHH) scores of the Kenana and Nguni populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)……53

Figure 4.21: The distribution of standardized cross-population extended haplotype homozygosity (XP-EHH) scores of the N'Dama and Afrikaner populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)……54

Figure 4.22: The distribution of standardized cross-population extended haplotype homozygosity (XP-EHH) scores of the N'Dama and Nguni populations. The dotted horizontal lines is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4)……54

Figure 4.23: A representation of biological processes from the iHS significant genes. (Generated from Panther site)……...……..55

Figure 5.1: This figure is showing different domestication times for different species (including cattle) based on archaeological data…….……67

List of Tables

Table 2.1: The geographic information of breeds found in Africa including breeds that are part of this study. The information used to distinguish the breeds was generated from (https://www.thecattlesite.com/)……...10

Table 2.2**:** Examples of available commercial bovine SNP arrays and their sizes……20

Table 3.1: List of analysis tools and software packages used in the study……24-25

Table 4.1: The summary statistics of the mapped reads and identified SNPs and InDels...35-36

Table 4.2: Relationships of inter-regional gene flow/introgression events generated by the D-statistics/ABBA-BABA test……..…..43

Table 4.3: Relationships of inter-regional gene flow events generated by D-statistics using Buffalo as an out-group……....45-46

Table 4.4: Significant candidate regions selected by the two statistical test (iHS). Genes coloured in red are shared between populations……55-58

Table 4.5: List of significant genes from XP-EHH cross population analysis…….59-60

Abstract

The development and application of next-generation sequencing technologies have enabled the investigation of genomic data better and more efficiently. This progress has led to population genomic analysis of many African indigenous breeds such as Muturu, N'Dama, Sheko, Ankole, Afar, and Fogera. However, no study has used whole-genome sequence data to understand relationships between South African (SA) cattle breeds such as Nguni, Afrikaner, and Bonsmara, and African breeds such as Ankole, Kenana, and N'Dama, as a result, information such as genomic diversity, effective population changes, and adaptations remain unclear and this negatively impacts efforts of conservation and breed improvement. This study aimed to investigate the genomic relationships between three SA and three African cattle breeds, to assess genomic diversity, demographic history, introgression, and to identify regions of selection signatures. A total of 15 animals from SA cattle breeds Nguni (5), Afrikaner (5), and Bonsmara (5), were sequenced using Illumina Hiseq 2500 platform at 10X coverage. Data for Ankole (5), Kenana (5), and N'Dama (5) were obtained from the study of Kim et al. (2017). Variant calling was done using GATK and a total of 37,482,988 (SNPs) and 4,931,938 (InDels) were obtained across the breeds. Analysis of Next Generation Sequencing Data (ANGSD) was used to generate phylogeny, heterozygosity estimates, and introgression events using ABBA/BABA patterns. Principal component analysis, nucleotide diversity, ancestral admixtures, and Treemix were applied to unveil relationships and gene flow events. Then evidence of selection signatures was explored using two statistical methods iHS and XP-EHH. Kenana cattle exhibited higher levels of genetic diversity, followed by Ankole, Nguni, Afrikaner, Bonsmara, and N'Dama, and surprisingly Bonsmara, a SA composite, had higher genetic diversity than N'Dama. Relatedness, introgression and migration analysis supported findings of previous studies which indicated close relationships between SA indigenous cattle breeds and further unearth novel relationships between Nguni, Ankole, and Kenana cattle. This analysis also revealed the shared ancestry between Nguni and N'Dama, as well as their contribution to Ankole's genetic makeup, because of close relatedness between Bonsmara and Holstein, Water Buffalo was used to validate observed relationships. Moreover, we also observed Bonsmara ancestry and its relationship with taurine breeds. The demographic history analysis revealed how the effective population sizes of African breeds changed over different climatic epochs. Notably, we observed two contractions and two population expansions which are consistent with previous findings. The timing of the population sizes overlapped with the recorded ancient human activities such as migration and domestication. Selection signature

analysis identified 112 iHS and 120 XP-EHH candidate regions in the study populations. The annotation of candidate regions revealed potential genes associated with reproduction, growth, milk production, meat quality, diseases, and disease resistance. In particular genes such as ***CNTN6, KCNIP4, APP, MAP4K4, CDH13, PLCB4*** **and** ***AGO2*** showed strong positive selection. These findings provide important genomic information on genetic relationships between local and African indigenous cattle breeds, as well as the understanding of selection and adaptation events that will help in the improvement of these breeds.

Keywords: African cattle, whole-genome sequencing, genomics, genetic diversity, adaptation, demographic history, selection signatures.

Chapter 1

1.1 General Introduction

Since the human genome draft was published in April 2003, there has been a significant shift in genomic studies. Due to this success, understanding the fundamentals of animal health and production are now possible for most species (Garcia et al*.,* 2012). This progress has played an important role in advancing bovine and other livestock research. In 2004, the genome of domestic chicken (*Gallus gallus*) was sequenced with the size of ~1 billion base pairs, which is about 40% of the human genome. In this genome, 23,000 genes were identified (Pereira et al., 2013). In 2009, Ramos and colleagues used the Illumina's Genome Analyzer to study the genomes of four pig breeds (Wild boar, Duroc, Large White and Pietrain) from United States of America (USA), Netherlands, Denmark and Japan, and discovered about 272,000 single-nucleotide polymorphisms (SNPs). The *Bos taurus* assembly was completed in 2009 and was employed in the genome-wide association survey of over 37 000 SNPs from 19 cattle breeds (Tellam et al., 2009; Liu et al., 2009). Since then, many other genomes have been successfully assembled, e.g. turkey (*Meleagris gallopavo*), camel (*Camelus bactrianus ferus*), yak (*Bos grunniens*), sheep (*Ovis aries*), goat (*Caprahircus*), dog (*Canisfamiliaris*) and donkey (*Equus asinus asinus*) (Dalloul et al., 2010; Wang et al., 2012; Qiu et al., 2012; Jiang et al., 2014; Dong et al., 2012; Lindblad-Toh et al., 2005; Huang et al., 2015). These genomes are continually being modified to improve genomic assemblies, annotations and to overcome biases. They are also being applied in the development of efficient genomic tools which are used in the characterization of valuable traits (Andersson et al., 2015). This progress was made possible by the development of algorithms and software's for alignment and variant discovery, especially in humans. The remarkable turning point in cattle genomics came after the publishing of full genome assembly, which gave insights into the genetic composition of cattle genes and structural variations (Bovine HapMap Consortium, 2009). These findings were instrumental in unravelling the potential of cattle genomic research, adding information to mammalian evolution and biological processes (Baes et al., 2014; Bovine HapMap Consortium, 2009; Tellam et al., 2009; Kanzi et al., 2020). They also, enabled scientists to explore and discover the underlying mechanisms of species structures, biological functions, and the role of phenotypes in production, health, and diseases using new data analysis techniques (Garcia et al., 2012).

Advances in computer science is improving research in several ways, especially in the area of biology (Mariani et al., 2016). Genetics has shifted from the characterization of small numbers of important molecules (such as the highly expressed genes and proteins) to focusing on the whole ensemble of the molecules (metabolites, proteins, and clusters of genes) (Hocquette et al., 2007). Moreover, processes that used to take longer to run have become easier to apply in day-to-day operations, especially in bioinformatics and medical genomics (Baichoo et al., 2018). This success is caused by the application of high-throughput strategies in data acquisition and the use of bioinformatics to handle huge amounts of biological data (Hocquette et al., 2007). Since its inception, bioinformatics has improved biological research significantly by providing packages for data analysis and modeling. It has made it possible to manage and analyze structural and functional data (Chiusano et al., 2008; Bostan and Chiusano 2015; Esposito et al., 2016). Some of the tools have been applied in milestone projects such as the rice project in 2002, the chicken genome project in 2004 (the first farm animal to be sequenced) and the 1000 Bull Genome Project in 2012 (Hayes et al., 2012; Hayes et al., 2019). In African cattle studies, these tools were applied in understanding signatures of selection, adaptation, and determining genetic diversity from cohort data of African indigenous cattle populations (Taye et al., 2017; Kim et al., 2017; Cheruiyot et al., 2018; Kim et al., 2020).

The African farming systems are undergoing rapid loss of indigenous livestock breeds (FAO, 2007; Mwai et al., 2015; Kim et al., 2017). These breeds are utilized for food and other important products such as fuel, clothing, ceremonial objects, ploughing, and transportation, especially in rural communities (Dessie and Mwai, 2019). Recent studies reported that some unique features of African cattle breeds are under threat and important breeds such as Ankole are continuously declining and if nothing is done to address the rapid decline they might disappear altogether (Mwai et al., 2015; Dessie and Mwai, 2019). The sudden disappearance of some African indigenous breeds is mainly due to the introduction of exotic breeds into the African livestock farming systems, lack of proper farming strategies, as well as minimal or no studies to understand their genomic composition (Dessie and Mwai, 2019). The loss of indigenous livestock means a loss of unique genetic traits that may be needed in the near future (FAO, 2007). As the environmental conditions change (e.g. climate change), the genetic composition of African indigenous breeds will serve as an important resource for their survival (Mwai et al., 2015). Until recently, African indigenous cattle have not been studied intensively at the genomic level (Mwai et al., 2015). However, the situation is expected to change due to several studies that are being conducted. Most published work (Qwabe et al., 2013; Makina et

al., 2014; Zwane et al., 2016; Makina et al., 2016; Aliloo et al., 2018) was done using genotype data such as BovineSNP50 V2 BeadChip, GeneSeek Genomic Profiler (GGP-80K bead-chips), BovineHD chip with 777,000 SNPs and BovineLD v1.1 with 6,912 SNPs to characterize breeds, however it is only recently where whole-genome sequence is being used to further interrogate African genomic data (Taye et al., 2017; Kim et al. 2017; Zwane et al., 2019; Kim et al., 2020; Dutta et al., 2020). These studies focused on understanding population structure, adaptation, and identification of genomic regions harbouring economic important traits on African indigenous cattle breeds using short and long reads of whole-genome sequence data. Moreover, they have generated huge amounts of data that can be accessed from public databases such as European Nucleotide Archives (ENA) and National Centre for Biotechnology Information (NCBI). Chen et al. (2020) reported that huge amounts of data from next-generation sequencing has been produced and can be applied in genomic assessments such as genetic diversity and identification of regions of natural and artificial selection. However, many African cattle breeds have not been studied (Hanotte and Jianlin, 2005; Mwacharo et al., 2006; Mwai et al., 2015). This study used whole-genome sequence data from African cattle breeds to understand their relationships, demographic history, introgression, as well as selection signatures at whole-genome level. This will generate information that can be used to improve African cattle breeds, especially indigenous breeds.

1.2 Aim of the study

This project focused on the population genomics of African indigenous cattle breeds Afrikaner, Nguni, Bonsmara, Ankole, N'Dama, and Kenana to gather insights on their genomic diversity, demographic history, environmental adaptations, and identifying selection signatures. This information is critical in understanding their genetic variation, ancestral history, and relationships that exist between African indigenous cattle populations. It will also provide a framework for the design of breeding strategies to improve breeds performance while adapted to their local environments.

1.3 Specific aims:

1. To determine genomic relationships of African breeds using whole-genome sequence data.
2. To model gene flow and introgression events that have occurred between African breeds using D-stats and Treemix.

3. To determine changes in effective population size (demographic history) over time using the Pairwise Sequencially Markovian Coalescent (PSMC) method.
4. To gain insights into the patterns of signatures of selection using integrated Haplotype scores (iHS) and Cross-population Extended Haplotype Homozygosity (XP-EHH) statistical tests.

Chapter 2

2. Literature review

2.1 Introduction

The advent of whole-genome sequencing and genome-wide investigations has ushered in a new era of conservation and industrial applications (Yum et al., 2018; Kim et al., 2020), and has enabled both phenotypes and genomic information to be obtained through high-throughput genotyping and sequencing. This has created new opportunities for conducting efficient genomic studies (Habier et al., 2007; MacLeod et al., 2013). Furthermore, they enabled the analysis of evolutionary history of breeds to understand their ancestral lineages and also characterization of breeds for formulation of new breeding and management strategies (Makina et al., 2016). Genetic characterization also plays a role in the strategic prioritization of Animal Genetic Resources (AnGR) for food and agriculture (FAO, 2007). The number of SNPs generated by next-generation sequencing technologies has decreased the cost of sequencing and this offer scientists a unique opportunity to assess in detail how breeds are adapted to different environmental conditions (Cheruiyot et al., 2018). While many studies have attempted to investigate the relationships and adaptations of African cattle, analyses such as demographic history, relatedness, selection signatures, and admixture have not been fully conducted to understand relationships between SA and other African breeds.

African breeds carry unique genetic traits that should be studied and conserved for future breeding or improvement strategies. Genetic diversity is an essential element in farm animals necessary for genetic improvement for production in various environments. Moreover, it shapes and characterizes individuals, populations, species, and subspecies of life on earth (Notter, 1999; Laikre et al., 2010). The intra-population diversity reflects the number of various alleles present in the gene pool, which shows the presence of genotypes within populations (Laikre et al., 2010). Since the beginning of the 20th century, many studies have been conducted in evolutionary and conservation biology, and metrics such as genetic variance and heritability introduced important standards for understanding evolutionary processes (Fisher, 1930; Wright, 1931; Kim et al., 2021; Mauki et al., 2022). The current livestock diversity is the product of natural and artificial selection processes and understanding the existing diversity will help in breed's improvement efforts (Groeneveld et al., 2010).

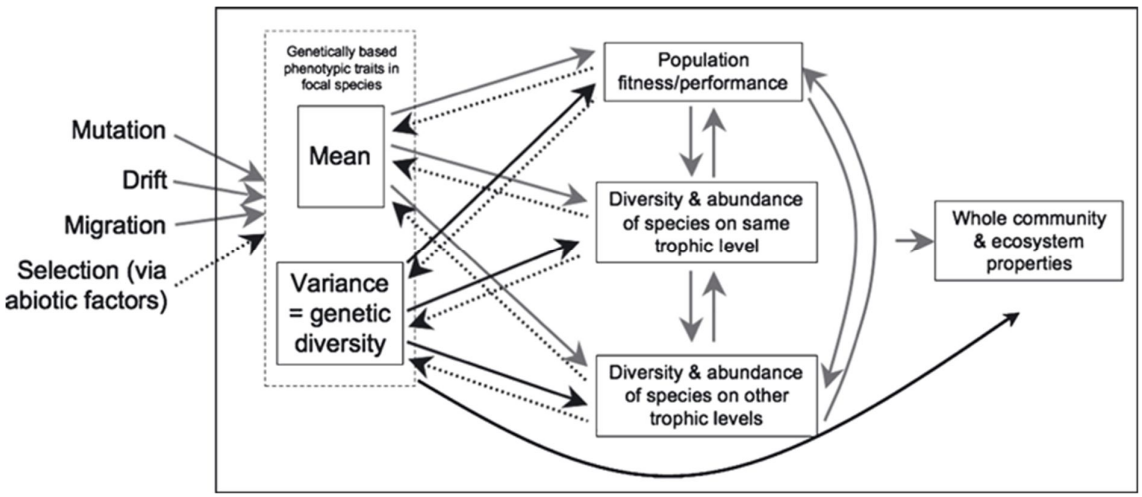

Figure 2.1: The depiction of how genetic diversity, directly and indirectly, affect the populations (Hughes et al., 2008).

The presence of genetic diversity (**Figure** 2.1) within and between livestock breeds is important for animal effectiveness in responding to climate changes (Makina et al., 2016). Furthermore, processes such as domestication, adaptation, and selective breeding have shaped the effective population size (demographic history) of livestock breeds, and have also left traceable genetic footprints, "selection signatures" in the genome of many biological species (Weldenegodguad, 2021). Therefore, genomic analysis of variants can unveil information about the structure of demographic histories, genomic diversity, and imprints left by selection pressures on any given group of populations (Weldenegodguad, 2021). Currently, there are bioinformatics packages or tools that are capable of reconstructing evolutionary events and detecting patterns left by past evolutionary events such as climate change and selection processes (Li and Durbin, 2011; Schiffels et al., 2020).

2.2 Importance of livestock farming in African communities

Livestock farming in Africa is characterized by domesticated animals such as chickens, goats, sheep, cattle, and pigs that are kept within communities for food, fibre, cultural and religious activities, and other economic reasons (Ntshepe, 2011; United Nations, 2012; Telugu et al., 2017). Livestock also plays a pivotal role in supporting economic affairs in developed and underdeveloped countries (Salmon et al., 2020). Its production accounts for one-fourth of the global land where one-third is occupied by the feed production. On average, it contributes about 40% of worldwide agricultural production (Stanford University, 2010). The global livestock sector is composed of separate entities that differ in terms of yield or outputs. These include

smallholder and pastoralists (600 million farmers) that keep animals primarily for food security and livelihoods maintenance, and to sustain essential commercial production processes (Thornton, 2010; Hoffmann, 2010). The African southernmost region is home to 90% of livestock farmers who are categorized as smallholders and form part of the non-formal sector. They account for 75% of farm animals where majority of these animals are native breeds (Nyamushamba et al., 2016). Livestock demand is increasing more rapidly than how cereal did during its dominating period (from its early stages of domestication 10 500 years ago where it was a preferred diet till recent times) (Madella et al., 2014). The most notable progress is the major increase in livestock flow and production, shown by the amount of trade between countries. The trade has grown from 4 % in 1980's to around 10 % recently (FAO, 2021) (https://www.fao.org/3/ah834e/ah834e02.pdf).

Recent studies have reported that the global biomass of livestock is double the number of human populations (Salmon et al., 2020). This industry was estimated to have a value of about $1.4 trillion in the 2000's (Steinfeld et al., 2006) and currently the global livestock production value has increased with annual growth of about 17.63% in 2022 (Livestock Industry, Retrieved August 4, 2022). The global utilization of pork, beef, and veal is estimated to have risen by about 30% in the past 15 years, with the import demand projected to 38 million in 2030 (Livestock Industry, Retrieved August 4, 2022). Livestock industry is organized into market chains that create employment for about 1.3 billion people globally, also, it directly supports millions of disadvantaged farmers in the global south (Thornton et al., 2010; Alders et al., 2021). This market also enhances protection of vulnerable communities and contributes significantly to the supply of nutrients, 17% kilocalories and 33% protein globally (Rosegrant et al., 2009; Thornton, 2010; Harwatt et al., 2018; Dong et al., 2022).

2.3 History of African indigenous cattle breeds

Africa is richly endowed with indigenous cattle breeds that are adapted to its prevalent harsh environments. In the past, these breeds were regarded as inferior and having little improvement potential. Indigenous breeds carry diverse traits and differ in terms of size, horns, and coat colors. Each breed carries traits that allow it to adapt to a particular environment, i.e. harsh, semi-desert, savannah, or other environmental conditions, depending on the keeper's production preference (milk, meat, and horns). This diversity represents a living genetic bank that is becoming even more important with emerging diseases, global warming, and market change (Scholtz et al., 2016; Dessie and Mwai, 2019). Currently, the research community

recognizes more than 800 cattle breeds (globally), within which 180 have been identified from sub-Saharan Africa, and of those, 150 are recognized as indigenous cattle breeds (Rege, 1999; Mwai et al., 2015). These breeds were previously divided into two major groups; the taurine and the indicine cattle which are morphologically and genetically distinct and are recognized by their physical appearances (Rege, 1999; Pitt et al., 2018). The earliest African cattle were the humpless *Bos taurus* until migration events involving zebu *Bos indicus* reshaped the architecture of African cattle breeds (**Figure** 2.2) (Bradley et al., 1996; Hanotte et al., 2002). These subspecies were introduced into the African continent through interbreeding and human selection process (Rege and Tawah, 1999; Mwai et al., 2015).

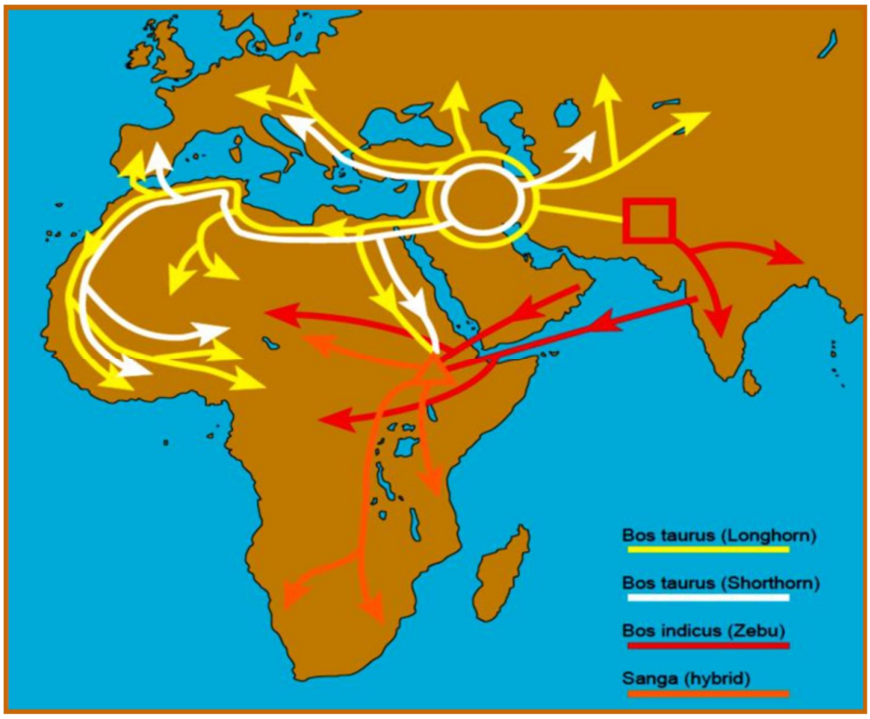

Figure 2.2: Inferred domestication areas and possible migrations that shaped the creation of African cattle breeds (MacHugh, 1997).

Currently the continent is dominated by breeds which are classified into three groups; taurine, indicus, and its derived forms known as *Bos taurus africanus* (Sanga) cattle (**Figure** 2.2) (MacHugh et al*.,* 1997; Mwai et al., 2015). The *Bos taurus indicus* (zebu) is commonly found in arid and semi-arid northern Sahelo Sudan and it is widely distributed in the eastern and the dry parts of West Africa. *Bos taurus taurus* dwells in the humid and sub-humid regions of West Africa, and it is exposed to African tsetse fly and trypanosomes (Rege and Tawah, 1999). The *Bos taurus africanus* (Sanga) is found mostly in arid environments in southern Africa, mainly around the Great Lakes in the western region of central, eastern, and southern Africa (Jahnke,

1988; Rege and Tawah, 1999). The zenga which are the composite of Sanga x zebu types are found in the eastern African region (Rege and Tawah, 1999). The *Bos taurus indicus* possess some unique traits that are well recorded such as; heat-tolerance, low maintenance, and resistance to parasitic agents, and these features enable their physiological structure to be stronger and more protective than that of taurine (Mukasa-Mugerwa, 1989; Scholtz et al., 2016).

The Sanga cattle group includes Nguni, Afrikaner, Sheko, Tuli and Ankole, and they are distinguished by their small cervico-thoracic hump (Rege and Tawah, 1999; Scholtz et al., 2016). Sanga breeds such as Nguni, Tuli, Afrikaner, Tswana, Tonga, and Mashona are known beef breeds found in the southern Africa region (Scholtz et al., 2016). SA is one of the countries which is greatly endowed with several well-adapted indigenous cattle breeds, among which is the Afrikaner, Nguni, Tuli, Bonsmara and Drakensberger. Furthermore, previous findings have highlighted that African breeds are genetically distinct from their European counterparts, e.g. the *Bos taurus* breeds (Angus, Brown swiss and Holstein) (Makina et al., 2014; Scholtz et al., 2016). Other important African indigenous breeds that have been studied include, but not limited to N'Dama which is known for its resistance to parasites, Kenana for its milk production and Ankole for its beef production and its ability to survive under severe conditions (Kim et al., 2017). African breeds are highly adaptable to their respective environments and they are potentially valuable to breeding programs due to their potential to perform in environments with biological stresses (Hanotte and Jianlin, 2005). **Table** 2.1 shows the origin of African indigenous cattle and their current distribution. The description and distribution of the African and SA breeds used in this study is shown in **Figure** 3.1.

Table 2.1: The geographic information of breeds found in Africa including breeds that are part of this study. The information used to distinguish the breeds was generated from (https://www.thecattlesite.com/).

Breed	Subspecies	Region/country of origin	Other distributions
Kenana	*Bos indicus*	East Africa	Sudan
Ankole	*Bos taurus x indicus*	East Africa	Uganda, Ghana, Rwanda, South Africa and Burundi.
Bonsmara	Mixed ancestry (Shorthorn, Hereford and Afrikaner)	South African	South Africa, Brazil, Paraguay, Colombia and Argentina
Afrikaner	*Bos taurus x indicus*	South Africa	Australia, South Africa and other tropical regions.
N'Dama	*Bos taurus*	West Africa	Southern Senegal, Nigeria, Gambia, Mali, Guniea Bissau, Sierra Leone and Ivory.
Nguni	*Bos taurus x indicus*	South Africa	Eswatini, Namibia, Zimbabwe, Botswana, and Angola
Brahman	*Bos indicus*	USA	Argentina, South Africa, Paraguay and Mexico
Tuli	*Bon taurus indicus*	South Africa	Zimbabwe, North America and Australia
Wagyu	*Bos taurus*	Eurasia	USA, Canada, Chili, South Africa

2.3.1. The Kenana cattle breed

Figure 2.3: The picture of Kenana (zebu) cattle with its unique features. Visible unique features are grey colour, large ears, loose skin, and a hump above shoulders (https://www.thecattlesite.com/).

Kenana cattle fall under the zebu *Bos indicus* group, and it is distinguished from other breeds by its hump (humped breed) (Bahbahani et al., 2015). Its name was adopted from the native farmers known as Kenana, who are divided into nomadic and semi-nomadic tribes. The blueprint of their migration patterns is similar to the migration events of humans in Africa (Payne, 1964; https://www.thecattlesite.com). This cattle breed was developed in southern Asia through human selection from non-humped cattle, and soon they were taken to Africa, USA, India, and the Middle East. Presently, it is estimated that 270 million zebu (case for zebu cattle in general) cattle are found in India, followed by Brazil with an estimated 155 million, and the USA carries about 2 million cattle, showing a great spread of this breed and a good population size (Rege, 1999, cited in Epstein, 1971). This breed arrived in Africa around 1500 BC and later a second influx occurred in large numbers around 670 AD (Rege, 1999, cited in Epstein, 1971). Kenana breed is easy to recognize because of its unique characteristics such as; large hump and the long flap of skin "dewlap" located on its chest area. Furthermore, it is distinguished from other breeds by the blue-gray color, present horns, and loose skin (Dessie and Mwai, 2019). Notably, their horns are shorter and never exceed 30-35 cm in size. During maturity (age 5) their body sizes differ depending on the environment and managerial setup.

Their males dominate with a body size of 300 kg to 500 kg compared to females with 250 kg to 350 kg (Saeed et al., 1987). Similar to many other Indicus breeds, zebu is also found in hot, and humid environments, they have evolved to adapt to such areas, and they can resist parasites and diseases that other cattle cannot deal with (Mukasa-Mugerwa, 1989; Taye et al., 2017).

2.3.2. The Ankole cattle

Figure 2.4: Unique features of the Ankole cattle breed, visible unique features are; spotted colour, medium size, long and asymmetrical horns https://www.thecattlesite.com/).

Ankole is an African Sanga-type cattle that occupies the Lake Mobutu to Lake Tanganyika regions in the African east coast (Monda, 2017; Zorloni, 2022, cited in Rahway, 1985). The earliest animals of this kind were introduced to northern Uganda in the 13th and 15th centuries by the native group known as the Hamitic tribes. Their evolutionary events include the inter-breeding of the lateral-horned zebu and the Hamitic Longhorn along the southbound routes. These cattle could have co-migrated with humans until they reached the Zambezi River (Monda, 2017). They have been reported in East African countries such as Tanzania, Uganda, Burundi, Democratic Republic of Congo, and in other African regions (Monda, 2017). Ankole cattle have unique genetic features that enable them to adapt and tolerate drought, heat, and parasites, further, they have an enhanced ability to utilize low-quality indigenous nutrition (Dessie and Mwai, 2019). Interbreeding events have been reported as a factor that played a role in the composition of this breed, e.g., an inter-genetic exchange was observed in Congo between indigenous cattle of the same district which led to the development of “Bashi” and

"Barundi" (a smaller variety of the Ankole with short and fine horns) (Monda, 2017). Many local African communities have used their own name to refer to this breed. In Uganda, they refer to them as Ankole, in Rwanda and Burundi Tutsi people know this breed as 'Watusi' (Monda, 2017). The original Ankole breeds are characterized by features such as the medium-long head, a small sized neck with a deep lip, and the narrow-structured chest. They have a cervico-thoracic hump like bush elephants which are small and barely visible in the cow (Zorloni, 2022, cited in Rahway, 1985). Their color is often red, black, and pied are not uncommon (Zorloni, 2022).

2.3.3. The Afrikaner cattle breed

Figure 2.5: The above picture is showing unique features of the Afrikaner cattle breed, visible unique features are; light tan red colour, medium size, loose skin, with the picture showing the polled type (https://www.thecattlesite.com/).

The Afrikaner cattle breed emerged from the wild-based cattle of the Asian grasslands. Thereafter they migrated to Africa, from Egypt through the Sahara to the tropical region until they arrived in SA. Historical migration processes that occurred resulted in only individuals that were best adapted to particular environments and some of the benefits were adaptability to arid climes where they were exposed to extremely hot temperatures, tropical diseases, and susceptibility to most parasites. The breed was domesticated by Khoi communities even before the arrival of Jan van Riebeeck in 1652 (https://southafrica.co.za/afrikaner-cattle.html) (Labuschagne, 2013). Due to its remarkable features, the research community has named the Afrikaner the "no-nonsense" breed because of its characteristics and performance. This breed carries unique characteristics such as small to medium size, easy calving, ability to round off

on natural grazing hardiness, red coat, and large horns (Scholtz, 2010; Pienaar et al., 2014). They require minimal maintenance and historically they were used for transportation by the Boer community. The other features that make this breed outstanding is its role in crossbreeding where it has been used to develop new synthetic cattle breeds (Scholtz, 2010) such as Afrigus, Bonsmara, Afrisim, SA Bradford, and Sanganer. It is one of the main breeds that represents the gene pool of southern African indigenous cattle (Scholtz, 2010; Pienaar et al., 2014).

2.3.4. The N'Dama cattle breed

Figure 2.6: The above picture is showing unique features of the N'Dama cattle breed, visible unique features are; short legs, sandy color, small sized, thick and deep neck, horns are curved upwards (https://www.roysfarm.com/).

The N'Dama cattle breed is a West African breed that is distinguished by its structure and known as the humpless Longhorn (*Bos taurus longifrons*). Other names include the Boyenca, Boenca and Fauta Longhorn. Research reported that N'Dama could have descended from the first domesticated cattle population of the Humpless Hamitic Longhorn cattle from the location known as the Fertile Crescent sometime around 9000 BC (Payne and Hodges, 1997). They are now populated in many countries around the West and the central African regions e.g. Ghana, Togo, Nigeria, etc., further, they are mostly found in areas that are infested by trypanosomes (tsetse flies) (DAGRIS, 2005). The most prominent feature this breed has is the ability to resist tsetse trypanosomes (Blench et al., 1993; 1998; Kim et al., 2017). These breeds are

characterized by features such as short legs with fine bones, thick neck and they also have a deep broad and fairy back. Other unique aspects include 60 cm horns, medium sized body of about 100 cm height (females) and 120 cm (males) with large and strong head parts. The breed is also known for its peculiarity to resist tick-borne infections, excluding rinderpest (Mattioli et al., 1995; Mattioli et al., 2002). It is well adapted, strives and lives in stressful environmental conditions in West Africa (Mattioli et al., 1995; Mattioli et al., 2002; Kim et al., 2017).

2.3.5. The Nguni cattle breed

Figure 2.7: The above picture is showing unique features of the Nguni cattle breed, visible unique features are; spotted colour, strong legs and hoofs, smaller in size, small sized horns and brush-like tail (https://www.thecattlesite.com/).

The Nguni cattle breed has a complex history with several events of migration. It was introduced into the African continent by nomadic travellers from North Africa (Mwai et al., 2015). Archaeological data (Schoeman, 1989) from SA, showed that the Nguni breed dates to 2000 years ago (Schoeman, 1989; Sanarana et al., 2015). This breed is part of the Sanga subspecies (*Bos taurus africanus*) that were brought through imports from Arabia (Hanotte et al., 2000; Decker et al., 2014). Several forces of evolution (migration, genetic drift, selection, and mutation) influenced the genetic composition of the Nguni cattle breed, and it is now distributed in countries such as Swaziland, Namibia, Zimbabwe, Botswana, and Angola. In addition, the genetic analysis of mitochondrial DNA unearthed the Nguni breed as the composite of the *Bos indicus* and the humpless *Bos taurus* as it was previously reported on the

origion of Sanga breeds (Rege and Tawah, 1999; Horsburgh et al., 2013). Presently, five recognized Nguni ecotypes are found in SA, namely, Venda, Pedi, Bartlow, Shangaan, and Makhathini, these breeds are linked to the geographical area they are found in and mostly named after the tribe/ethnic group farming it (FAO, 2007; Sanarana et al., 2015).

This breed has a small frame size compared to other beef cattle but these traits enable it to thrive in highveld regions of Africa. The Nguni breed is heat tolerant and can handle harsh environmental conditions. They are known for their ability to resist internal and external parasites and they carry a natural immunity to tick-borne diseases (https://www.thecattlesite.com). Some of their well-recorded traits that distinguished them from other indigenous cattle breeds includes their ability to reproduce in harsh conditions, distribution of fat longevity, hard body structure, parasitic resistance, harsh temperature tolerance, quality carcass and for being good foragers (Spickett et al., 1989; Scholtz et al., 2010). The Nguni bulls are medium-sized (weight of about 500 kg-600 kg), cows are small in size (weigh between 300 kg-400 kg), and their horns are round in matured cows and clearly lyre-shaped (Sanarana et al., 2015).

2.3.6 The Bonsmara cattle

Figure 2.8: The above picture is showing unique features of the Bonsmara cattle breed, visible unique features are; red in colour, strong legs and a larger mediam size body showing blend of *Bos indicus* and *Bos taurus*, and de-horned as per breed standard. (https://www.thecattlesite.com/).

The Bonsmara breed is one of the well-documented composite breeds. It is one of the man-made breeds and it was bred at the Mara and Messina Research Station around the year 1937 to 1963 by Prof Jan Bonsma and colleagues (Bonsma, 1980; Bosman et al., 2017). It was developed to improve beef production, targeting traits for growth and adaptation in sub-tropical conditions of SA. The term Bonsmara is the confluence of Prof Bonsma's surname (the leading scientist in its breeding) and the Mara station (a place where the breed was created). The development of this breed was done over a long period and several cross-breeding experiments were tried using around 20 commercial herds across different geographic locations in SA. These attempts gave birth to the best performing crossbred. The structure of cross breeding was formulated as follows; 5/8 Afrikaner and 3/8 Exotic Hereford/Shorthorn (Bonsma, 1980; Bosman et al., 2017). Currently, this breed is the most produced beef cattle in the country, and it has attracted many farmers, including those beyond the SA borders. Its genetics has been exported to other countries such as Namibia and Brazil (Bignardi et al., 2014). Breeds that are part of the stud group are subjected to mandatory performance recordings, and these efforts made it possible for the collection and recording of production traits that can be used in the improvements of the breed (production which includes fertility, growth, and carcass). Further, this also allows genetic assessments possible for breed improvement and it contributed to quantitative scientific studies on the breed performance (Maiwashe et al., 2002; Steyn et al., 2014; Bosman et al., 2017). A well-adapted Bonsmara carries unique characteristics with a strong constitution smooth-coat during summer, good pigmented eyes, sound feet and legs, and udders and hooves (Bonsma, 1980).

2.4 Current genomic research on African indigenous cattle breeds

Characterization of indigenous cattle genomes and identifying regions that contribute to their adaptability is crucial in the development of breeding strategies for enhanced production of milk and meat (Taye et al., 2018). The development of large-scale genetic analysis has introduced strategies for identifying genes and regions associated with phenotypes (Mackay et al., 2009) and created a potential for generating knowledge regarding environmental adaptations in different species (Kim et al., 2017). Recently we have seen the use of genome-wide techniques to investigate the genetic composition and diversity of African cattle (Bahbahani et al., 2015; Kim et al., 2017). Kim et al. (2017) studied 48 genome sequences of five African indigenous populations alongside 53 commercial taurine breeds, and they found the highest genetic diversity among the African zebu/*Bos indicus* and the Sanga cattle. Further, on selection analysis, they discovered regions associated with pathways controlling different

traits including anaemia and feeding behavior on N'Dama cattle breed. In Ankole they discovered genes responsible for coat color and horn growth as well as genes associated with heat tolerance and tick resistance (Kim et al., 2017). Taye et al. (2018) conducted an analysis on East African indicus cattle together with the European and Asian lineages using selection signatures methods XP-EHH and XP-CLR, and identified several pathways and genes that are important for many biological processes such as domestication, behaviour, thermoregulation, feeding, metabolism, tick and parasitic resistance, growth and reproduction as well as immune system response.

Zwane et al. (2019) discovered a total of 1,678,360 novel variants from SA indigenous cattle breeds; Nguni, Drakensberger, and Afrikaner using whole-genome sequence data. The aim of the study was to identify new SNPs and variants, check their distribution in the genome, as well as identifying genes enriched in novel SNPs regions. The study identified regions that discriminates the breeds, selection signatures and genes associated with, among others, coat color and fertility. The study by Nanaei et al. (2020) looked at the genes linked with high milk production in Kenana cattle and their analysis showed several candidate genes involved in milk production, i.e. they identified genes such as *B4GALT1*, *IGFBP-2, ABCG2, GHR* and *RORA* that could be under positive selection for milk production. In 2021, Kim et al. (2021) conducted a comprehensive study to understand the history and the contribution of gene flow to the genomic architecture of African pastoralism. They collected samples from 16 breeds and a total of 172 animals from African indigenous cattle and identified a taurine × indicine cattle gene flow event that is linked to circa 750–1,050 years ago that has shaped the genomic make-up of today's cattle populations. They also identified 16 loci that are involved in environmental adaptations, immune response, heat-tolerance and reproduction. Lastly, they discovered highly divergent locus in African taurine cattle which is linked to trypanotolerance, a trait found in breeds from trypanosomosis-infested areas. Based on their findings they concluded that present-day breeds are the product of admixture between the *taurus* and *indicine* lineages, which is the root of the present success of African pastoralism (Kim et al., 2021). These are not the only studies conducted on African breeds, there are other studies that took advantage of the latest genomics technologies such as Kim et al. (2017b), Tijjani et al. (2019) and Mauki et al. (2022).

2.5 Brief history of genomic markers and availability of genome-wide data

The prospect of genomic research to understand and improve genetic traits in livestock has recently progressed with major technological advances in place (Womack, 2012; Herring et al., 2013; Mukhopadhyay et al., 2020). Back in the 1970s and 1980s, several DNA-based methods were developed, and through their improvements, the concept of genetic markers started. Initially, enzymes were used to cut DNA at a specific position to collect different DNA fragments. Allozymes were the first markers to be applied in livestock DNA sequence studies (Grapes et al., 2004) and they gave rise to DNA analysis methods such as Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPDs), and Amplified Fragmented Length Polymorphism (AFLPs) (Hocquette et al., 2007). Subsequent to these, method of identifying DNA repeats in various units called microsatellites was introduced, and these markers were successfully used in different studies including characterisation of African cattle (Sanarana et al., 2015; Madilindi et al., 2019). The knowledge gathered from the use of microsatellite markers paved a way to the discovery of single-base changes in DNA called single nucleotide polymorphism (SNPs), which became the mostly used genetic markers recently. This advancement took advantage of the development of the Polymerase Chain Reaction (PCR) technologies and revolutionized the way of conducting genetic research (Beuzen et al., 2000; Herring et al., 2013). Currently millions of these markers are being generated using new next-generation sequencing platforms (Pugach et al., 2015).

Single-nucleotide polymorphisms markers are single base changes that occur in large numbers across the genome but less commonly in the coding regions (Toro and Fernandez, 2006; 2011). Currently, there are several methods for assaying SNPs and mostly in low and high-density (SNP chips) (**Table** 2.2). These genotype arrays were developed for analysis such as mapping, admixture, and identity by descent and, also, for analysis of the relationships between genotypes and phenotypes (Hayes, 2009; Zwane et al., 2016). In recent years we have seen the SNP array's applications in population genetics studies and they have helped scientists answer some important biological questions (Edea et al., 2013; 2015; Makina et al., 2014; 2016; Pugach et al., 2015), including unravelling genes of economic importance and complex traits (Makina et al., 2015; Dixit et al., 2021).

Table 2.2: Examples of available commercial bovine SNP arrays and their sizes

Platform	SNP chips	Size
Affymetrix	Axiom Genome wide BOS1	648 875
Neogen-Geneseek	GGP-LD version 2	19721
	GGP-LD version 2	26151
	GGP HD	76 879
	GGP HDi	12 189
	GGP indicus	35 090
	GGP 150k	139 480
	GGP F250	230 000
Illumina	Bovine LD version 1.1	6912
	Bovine HD	777 962
	BovineSNP50v.1[h]	54,001
	BovineSNP50v.2i	54,609

Due to these developments, the cost of generating SNP data has decreased significantly and many species have been sequenced resulting in the discovery of vast amounts of SNPs from the whole-genome which are now accessible in public databases (Lindblad-Toh et al., 2005; Ramos et al., 2009; Dalloul et al., 2010; Wang et al., 2012; Qiu et al., 2012; Jiang et al., 2014; Dong et al., 2012; Huang et al., 2015). The availability and accessibility of these data are bolstering the discovery of mutations responsible for genetic diversity in both simple and complex traits, and scientists are able to study the in-depth characterization of cattle breeds (Stothard et al., 2011; Mukhopadhyay et al., 2020). Currently, several platforms are available for sharing of data and thanks to projects such as the bovine HapMap project (2009), the bovine genome project (2009), and large-scale SNP discovery studies (Kim et al., 2017; Zwane et al., 2019; Kim et al., 2020). Since the sequencing of the Human Genome Project, the sharing of data has become easy and platforms such as National Centre for Biotechnology Information (NCBI), and the Single Nucleotide Polymorphism Database (dbSNP) (http://www.ncbi.nlm.nih. gov/projects/SNP) share their data freely for scientists across the globe (Eggen, 2012). The availability of this data is behind the convergence of biology and technology which gave rise to a field of bioinformatics and computational biology (Esposito et al., 2016).

2.6 Benefits of applying bioinformatics in livestock genomics research

The exponential growth in genomics research, with vast amounts of genotype and sequence data produced from NGS technologies, has positively influenced in-depth genome-wide studies in livestock, and this has also influenced the growth of effective and cutting-edge bioinformatics tools and statistical packages for data analysis (Esposito et al., 2016). Bioinformatics involves processing files through linked steps that are known as pipelines/workflow systems. Normally, these series of steps are done by executable command-based software packages developed for Unix-compatible operating systems (Leipzig, 2017). This allows understanding of genomic information to predict and improve, and can enable agricultural scientists to address challenges relating to production, health and adaptations (Tuggle et al., 2022). It will also allow scientists and farmers to address issues relating to environmental challenges and climate change, as well as the demands of growing populations (Illumina, 2022). Methods of collecting and storing genomic data are important for characterization of new species, meta-analysis, identification of complex traits, enhancing the process of genomic selection and also gene editing for advanced production. These technologies will impact the field of agri-genomics, breeding, and livestock management (Illumina, 2022). Other applications in livestock studies include metagenomics, genome-wide association studies, epigenetics, and the identification of variants in genomes of different species.

African cattle carry unique genetics that have evolved over many years to enable them to adapt and produce in harsh conditions, however, many African cattle have not been fully characterized or studied. Therefore, this study focused on the analysis of whole-genome data of African cattle breeds from different geographical locations (**Figure** 3.1) to study their evolutionary history and relationships, to understand their genetic/genomic diversity, effective population size changes, introgression, gene flow events and selection signatures with annotations of candidate regions. This information will contribute significantly towards whole-genome comparative studies of global cattle populations and will also play an important role in breeding and conservation strategies.

Chapter 3

3. Material and Methods

3.1 Introduction

Sampling of animals from different geographic locations is a good method to understand the genetic diversity and underlying mechanisms of adaptation (Bentley et al., 2017). In this study, whole-genome sequence data were collected from three indigenous SA cattle breeds (Nguni, Afrikaner, and Bonsmara) from different provinces. The data was analysed and compared with previously published data of N'Dama (West Africa), Ankole (East Africa), and Kenana (East/Central Africa) (Kim et al., 2017) to determine their relationships, demographic history and selection signatures analysis. The study was conducted at the Agricultural Research Council-Animal Production (ARC-AP), Pretoria. Samples collection and animals handling were done according to the ethics approval by the ARC-AP (APAEC 2019/12) and the University of Pretoria (UP) (AEC Reference No.: NAS015/2021).

3.2 Sampling and data collection

In this study, data were available in two sets, the primary and secondary data. The primary data was locally generated through sequencing of three SA indigenous cattle breeds Nguni (n=5), Afrikaner (n=5), Bonsmara (n=5), and the secondary sequence data for Ankole (n=5), N'Dama (n=5), Kenana (n=5), and Holstein (n=2), were obtained from the study of Kim et al., (2017), the second out-group (Water Buffalo) was adopted from (Dutta et al., 2020).

3.2.1 Generation of primary data

In total, 15 animals from two SA indigenous and one locally developed beef cattle, Nguni, Afrikaner and Bonsmara (composite-breed) were selected. We chose animals with an estimated weight of at least ±175kg to target the transition period to weaning, which is an ideal period for eased collection of samples in cattle. Blood samples were collected from the animal tail vein into 6 ml vacutainer blood collection tubes and aliquoted into 2ml cryo tubes for storage in -80°C freezer. DNA extraction was done using the Qiagen DNeasy Blood & Tissue Kit (250) (Qiagen, Heinrich Heine University Düsseldorf, Germany) following the prescribed manufactures protocol (which is described in QIAamp DNA Blood Kits (qiagen.com)). Extracted DNA was assessed for quality and purity using Nanodrop 2000c (Thermo Fisher Scientific, Waltham, Massachusetts (USA)) and was verified using Qubit 2.0 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts (USA)). Quantified

DNA was sent to Novogene (Pty) Ltd in United Kingdom (UK) for whole-genome sequencing at 10X coverage using the Hiseq 2500 illumina platform. Raw data was received in fastq format.

3.2.2 Secondary data

Secondary data was downloaded in fastq format from the European Nucleotide Archive (ENA) genomic data bank. To maintain the same data formats and quality, we targeted animals that were sequenced at 10X coverage. The data for (Ankole, N'Dama and Kenana) were accessible under the project PRJNA312138 at National Centre for Biotechnology Information (NCBI) and European Nucleotide Archive (ENA) and can be downloaded using these links; https://www.ncbi.nlm.nih.gov and https://www.ebi.ac.uk/ena/browser/. Additional Water Buffalo were used as an out-group to verify the D-Statistics results observed when Holstein was used as an out-group, this data can be accessed from (Nucleotide Archive (ENA) with study IDs PRJEB3959).

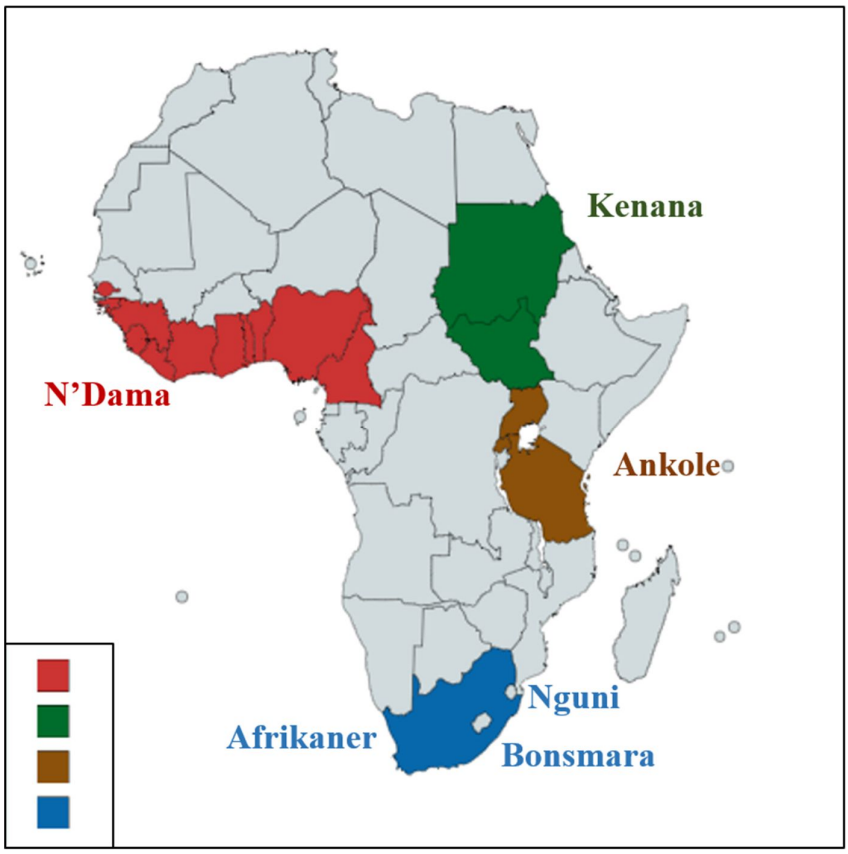

Figure 3.1: The map showing the distribution of the breeds used in this study according to their agro-ecological zones.

Both sample sets were strategically collected to represent different geographical locations (East, West and SA) as shown in **Figure** 3.1.

3.3 Bioinformatics tools and availability of high computing environment

For data analysis, multiple bioinformatics packages were used (**Table** 3.2). This includes sets of latest packages that have been used in recent studies to maintain data quality and current standards. All analysis were run on Lengau server of the South African Centre for High Performance Computing (CHPC). This system uses high-end 24-core cluster nodes that are designed to handle huge amounts of data. Bioinformatics analysis was summarized into several steps; data quality assessment, mapping or aligning, variant calling, and population genomics analysis for understanding relationships and evolutionary processes.

Table 3.1: List of analysis tools and software packages used in the study.

Software/Tools	Function	Source/Authors
BWA mem	Sequence alignment	Li and Durbin (2010; 2013)
SAMtools	Manipulating SAM, BAM and VCF files	Li and Durbin (2010)
Python	To edit and run pre-compiled python scripts	Van Rossum and Drake (2009)
Figtree	Phylogeny construction	Rambaut and Drummond (2009)
ANGSD	Matrix, SFS, Dstats	Korneliussen (2014)
Phylip neighbor	Files conversion	Felsenstein (1993)
Cutadapt	Trimming	Martin (2011)
Gitlib (PSMC)	Demographic history	Li and Durbin (2011)
Plink	Manipulation of Bfiles	Purcell et al (2007)
FastQC	Quality check	Andrews (2010)
Tabix	Manipulating VCF files	Li and Durbin (2010)
Picard	Manipulating SAM, BAM	http://broadinstitute.github.io/picard/
GATK	Variant calling	Poplin et al., (2017)
R	Analysis and visualization	R Core (2013)
Rehh	Selection signatures	Gautier and Vitalis (2016)
Genesis	Visualization	Buchmann and Hazelhurst (2014)

VCFtools	VCF files conversions	Danecek (2011)
Treemix	Inferring migration events	Pickrell and Pritchard (2012)

The workflow followed in this study is shown is **Figure** 3.2, showing in detail adopted packages and population genomic analyses that were applied to the raw data until outputs were generated.

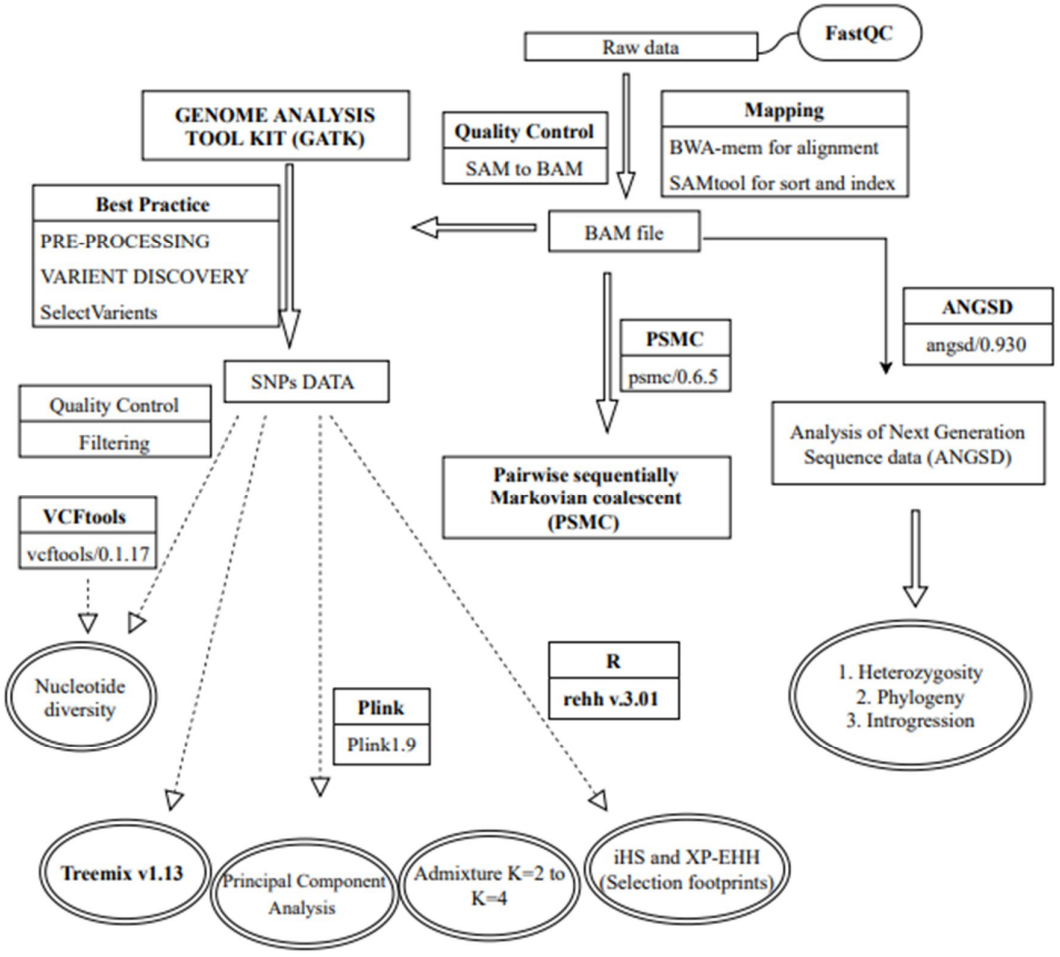

Figure 3.2: The analysis workflow from raw data processing, mapping, variant calling and population genomics analysis. Tools and software packages are highlighted in bold and analysis methods are shown is circles.

3.4 Data quality control, and mapping

Prior to further analysis of the sequence data, FastQC v0.11.9 was used to assess the quality of the raw data (Andrews, 2010). Sequence reads were trimmed to remove adapter sequences, reads shorter than 30 bp, primers poly-A tails, and unwanted sequences using Cutadapt v1.8.1 (Martin 2011). Cattle reads were then merged and mapped against the reference genome ARS-UCD1.2_Btau5.0.1Y (Bellott and Hughes et al., 2014; Rosen and Bickhart et al., 2018) using the Burrow Wheeler Aligner (BWA-mem) (Li and Durbin 2009), Buffalo reads were mapped against water buffalo reference genome UOA_WB_1 assembly and were subjected to standard mapping and file conversion as cattle data (Low et al., 2029). Mapping generated Sequence Alignment Map (SAM) files that were converted to BAM files (a binary version of the SAM file) using SAMtools v1.9 (Li et al., 2009). The binary file (BAM) output was then sorted by coordinates using the SAMtools command line (samtools sort –o sorted). Picard v2.20.3 software was used to add read-group tags and to remove duplicates. Lastly, the sorted BAM file was indexed using SAMtools for downstream analysis. The indexed file was used for variant calling and population genomics analysis in ANGSD.

3.5 Variant calling

Variant calling was done using Genome Analysis Toolkit (GATK) following the GATK best practices (DePristo et al., 2011). Recalibration of bases quality was run to build a model of covariation-based data using known variants ("ARS1.2PlusY_BQSR_v2.vcf.gz" from 1000 bull genome project) sets to produce a calibration table. GATK PrintReads was executed to adjust the base quality scores in the data based on the recalibration table to produce an adjusted BAM file. To establish what had happened to the BAM file, we ran GATK BaseRecalibrator on the recalibrated BAM to generate the "after recalibration" table, which was used to compare the changes made in the recalibration step. To increase the quality of variant calls, we used ARS1.2PlusY_BQSR_v2.vcf.gz, a known variants file of SNPs and INDELs generated from *Bos taurus* and *Bos indicus* Run7 of 1000 bull genome project at tranche 99.9 stringency. This was followed by the creation of Genomic Variant Call Format (GVCF files) using GATK HaplotypeCaller, and finally, we did the GATK DepthOfCoverage to get the data coverage statistics. After generating GVCF files the GVCF consolidation (CombineGVCFs) step was employed, followed by joint genotyping to generate the GenotypeGVCF file for efficient variants partitioning using the SelectVariants. Variants calling was done according to standard recommendation by the 1000 bull genomes project (http://www.1000bullgenomes.com). To

ensure that our variants (SNPs) were of good quality, the data was subjected to hard filtering to retain variants with mapping quality score of ≥ 25 and genotype quality of < 40. Loci with > 2 alleles and within clusters (> 3 SNPs in a 10-bp window) were removed from further analysis (Nanaei et al., 2020).

3.6 Population genomics analysis

3.6.1 Heterozygosity Estimates (Genetic Diversity) and Nucleotide diversity

Genetic diversity based on genome-wide (global) nuclear heterozygosity was estimated by calculating the site frequency spectrum (SFS), which is the distribution of alleles across the genomes as implemented in ANGSD (Korneliussen, 2014). Then SAMtools v1.9 was employed to calculate genotype likelihood using SAMtools method -GL 1, which is employed to estimate heterozygosity and to account for sequencing errors and application of corrections. Since the ancestral genome was unavailable, a reference genome (ARS-UCD1.2_Btau5.0.1Y *) was used as the ancestor sequence. GATK generated BAMfiles were used as input for genotype calling in GATK. Because the BAM files carry raw alignments, nucleotide bases were filtered in ANGSD. Filters; (-remove_bads) was applied to remove bad reads, (-uniqueOnly) removed reads that did not map to a unique position, (-minmapq) reads with quality of less than 30 were removed, (-only_proper_pairs) bases with no mating pairs were removed and (-minQ) bases with a quality score of less than 20 were also discarded. Site allele frequency (SAF) calculations are built on the genotype likelihood, we calculated these estimates using -GL 1 using ANGSD. To calculate the site allele frequency, the Hardy-Weinberg equilibrium (HWE) was assumed, and other parameters were left at default settings, (–nSites= 200, -tolerance (tole) =1e-8. Thereafter, the sub-program realSFS was used to estimate SFS from SAF, then SFS was used to compute individual estimates of heterozygosity. The number of heterozygous sites (the second entry in the SFS) was divided by the total number of sites (first entry + second entry in the SFS) to generate the amount of heterozygous sites for each individual genome. Then genetic diversity estimates were inferred using non-overlapping sliding windows of 100 kb across autosomal regions using VCFtools/0.1.17 (Danecek et al., 2011). The results of these analysis were visualised as population representative boxplot and bar graph in R.

3.6.2 Analysis of genetic relationships and structure

Relationships and admixture were determined using phylogeny, principal component analysis (PCA), and admixture (Purcell, 2007; Korneliussen, 2014).

A. Evolutionary tree

The ANGSD software package was also employed to compute the distance matrix using –doIBS (a parameter for calculating identity by sequence for all pairwise comparisons). IBS is the simple model used for modeling mutations and estimating the allele sharing genotype frequencies for a pair of individuals (Korneliussen, 2014). Parameter –doIBS 2 had two options available for the tree construction (1= using a random base and 2= using the consensus base), in our case we opted for option 2 to construct a distance matrix using a similar DNA structure of nucleotide bases. Then -makeMatrix 1 was employed to print the distance matrix. Parameter -ref was used to denote the reference genome, thereafter the distance matrix was converted to a nexus file which is compatible with the Phylip Neighbour package. The nexus file was reconciled into a phylogenetic tree using the neighbour-joining method (Saitou and Nei, 1987), and lastly, visualization and annotation of the tree were produced by the Figtree software v1.4.4 (Rambaut and Drummond 2009).

B. PCA and Admixture populations relationships analysis

The population structure was further investigated using PCA and Admixture programs available on Plink v1.9 (Purcell et al., 2007). Plink (--pca) was used to extract principal components of the variance-standardized relationships matrix to produce default files; Eigenvectors (plink.eigenvec and plink.eigenval), which were then processed using –eigenvec and –eigenval to produce evec file. The output was then used as input file for visualization in Genesis v0.2.6 (Buchmann and Hazelhurst, 2014). Later, Plink was used to generate the BED (a binary biallelic genotype table) file from the VCF files. This file was used to run Cross-Validation (CV) (4-fold CV), an analysis that produced two files; the Q file which contains individual clusters assignments, and the P file which carries information of the SNPs frequencies. These two files were used to plot the admixture **K=2** to **K=4** using an online plotter POPHELPER Structure Web App v1.0.10 (Francis, 2016).

3.7 Introgression, migration, admixture and gene flow analysis of African breeds using D-statistics and Treemix

3.7.1 Introgression analysis using the D-statistics

To get an overview of introgression events of African cattle breeds, a D-statistics analysis was done, an example is depicted in **Figure** 3.3. This method was employed to estimate the frequency of shared alleles between genomes using a predefined four-taxon phylogenetic tree. We analyzed the differential patterns of derived alleles within two sister's taxa, with the inclusion of the out-groups Holstein and Water Buffalo because of their distant genetic relationships with the sister/sub-sisters (Nguni, Afrikaner, Ankole, Bonsmara, N'Dama, Kenana and Ankole). However, due to the close relationship between Bonsmara and Holstein that could influence signals of gene flow/introgression, we introduced an additional out-outgroup, the Water Buffalo, which is distant related to the studied breeds (Table 4.3). In the example presented in **Figure** 3.3, Afrikaner and Nguni breeds were tested with Kenana, with Holstein used as an out-group. The D-statistics analysis showed allele sharing/introgression between Nguni and Kenana (**Figure** 3.3). This strategy computes either a positive or a negative D-statistic. The positive output represents the ABBA pattern (with A as the ancestral allele and B as the derived allele), showing that there is a gene flow between the individual in position P2 and the ingroup P3 (**Figure** 3.3). The negative D-statistics represent the BABA pattern, this is a sign of gene flow between the individual in position **P2** and the in-group (**P3**) (**Figure** 3.3). This analysis was carried out as in the following papers: Green et al. (2010), Patterson et al. (2012), and Martin et al. (2014).

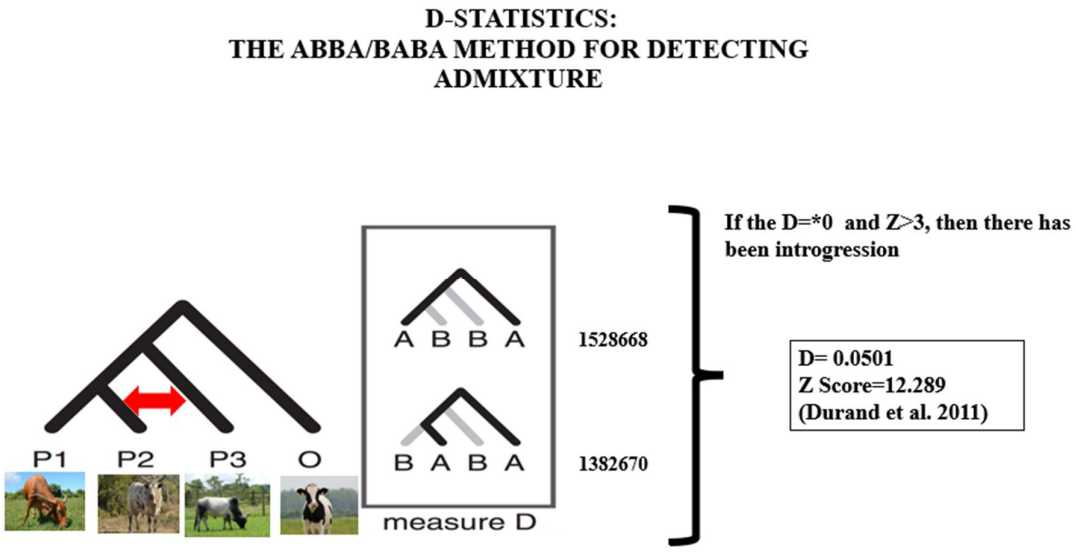

Figure 3.3: This figure shows the ABBA/BABA patterns and positions of individuals (genomes, P1, P2, P3, and O) on the tree topology. Note: The red arrow between P2 and P3 is showing an example of a possible gene flow between P2 and P3 with significant ABBA.

The D-statistics analysis depends on the relationships observed on the tree topology (**Figure** 4.4). There are two parameters available (options) –doAbbababa (do abbababa test 1= random base per site, 2= consensus base). Option 2 was chosen and applied to instruct the program to consider bases that are similar to each other. Then -useLast parameter was applied to give the command to the package to use the last sequence on the BAMlist as the out-group for all analysis. For the first step, only one output file was generated and a BAMlist without an out-group was made. This file was run through an R script. The program uses the black jackknifed procedure to estimate errors by splitting chromosomes into 5Mb chunks. Inter-country tests were designed to construct possible introgression events for the studied breeds. With the consideration of geographic location, we expected to detect excessive signals of gene flow between individuals who are geographically closer to each other, and less or no gene flow between individuals distant from each other. As a special consideration, we also took into account information reported in previous studies when we were designing this analysis (Taye et al., 2018; Zwane et al., 2019; Kim et al., 2021).

3.7.2 Inference of migration/gene flow (Treemix)

Signals of admixture were further assessed using the Treemix v1.13 (Pickrell and Pritchard, 2012) to determine admixture events and population splits over time. The inferential procedure was run for 1 to 4 migration edges and the Holstein was used as an out-group (It should be noted that a more distant out-group can yield more reliable results, however the Holstein still captured informative insights as shown when **Table** 4.2 results were compared to **Table** 4.3). Visualization was done using the R v 4.1.0.

3.8 Effective population size changes in African cattle breeds

To reconstruct the demographic histories of African indigenous breeds (Nguni, Bonsmara, Afrikaner, Ankole, Kenana and N'dama), the Pairwise Sequentially Markovian Coalescent method (PSMC) of Li and Durbin. (2011) was used. This method can infer historical population sizes and the splitting of breeds into different geographical locations. The method was used to infer population size in several species e.g. humans (Li and Durbin, 2011), the woolly mammoth (Palkopoulou et al., 2015) and recently in cattle studies (Wu et al., 2018; Weldenegodguad et al., 2021). This strategy uses the density of heterozygous sites across a

diploid genome of a single individual to infer the distribution of times to the most recent common ancestor (TMRCA) between the two alleles across all chromosomes (Palkopoulou et al., 2015). To eliminate biases from sex chromosomes which tend to either represent one lineage, we created a BAMfile of autosomes from the deduplicated (dedup) BAMfile as in Palkopoulou et al. (2015). All the duplicated regions were removed using a command in SAMtools (Samtools rmdup –S). Calling consensus was done using SAMtools; parameters –Q and –q in mpileup were given to apply the cut-offs of baseQ and mapQ respectively. Other, parameters were applied to perform specific tasks; -v to generate vcf output, (-f) to use the indexed reference genome in fasta format and, (–c) to call consensus sequence from the mpileup using vcfutils.pl script. The minimum and maximum coverage were set at 5 and 30 to allow vcf2fq to filter anything outside that range. Mapping quality was set to 30 and root mean at –Q30. PSMC plotting was done using a mutation rate of 1.1 × 10-8 per generation and an average generation time of 5 years (Kumar and Subramanian, 2002; Murray et al., 2010; MacLeod et al., 2013; Weldenegodguad et al., 2019). However, it was noted that this mutation rate could be uncertain and might be changed in the future when more data and tools are available (Weldenegodguad et al., 2019).

3.9 Selection signatures based on integrated haplotype score (iHS) test

Selection signatures, were identified on pure African indigenous breeds. Bonsmara (composite breed) and Holstein (taurine) were excluded because the aim was to get an overview of selection footprints of Africa indigenous breeds and to lay foundation for future studies that might assess selection signatures at the global level. The integrated haplotype score (iHS) analysis was employed to investigate selection signatures in African indigenous cattle breeds. Beagle was used to statistically estimate phased haplotype data (Browning et al., 2012). The within population positive selection (iHS) was calculated using phased haplotypes to detect ongoing sweeps that could lead to a single haplotype being very long compared to other haplotypes (Voight et al., 2007). The iHS score is based on the statistics of extended haplotype homozygosities (EHH) associated with each allele. The R Package rehh v.3.01 was used to construct iHS, and since there was no available information about the status of the ancestry and derived alleles for each SNPs, iHS was conducted using unpolarized alleles, using a latest option implemented in the rehh package. To cater for missing information, the function was set to “FALSE”, which is an alternative option for studies on domestic animals and non-model organisms (Gautier and Vitalis, 2016). The single-site iHS values were computed across the genome for each breed within non-overlapping windows of 50 kb. Each locus $|iHS|$ score was

computed using the rehh (Gautier and Vitalis, 2017) in R with a |*iHS*| score > 4. The XP-EHH method was then computed to detect alleles that have reached fixation within one population but still polymorphic in another targeting candidate regions that are > 4 (Voight et al., 2007). Candidate regions obtained from the two statistical tests were annotated using default parameters in databases; Ensemble and David and Panther (Thomas et al., 2003; Dennis et al., 2003; Aken et al., 2016).

Chapter 4

4. Results: Evolutionary changes and relationships of African cattle populations

4.1 Introduction

This chapter covers the results obtained from this study, from quality control of the data, mapping, variants statistics and population genomics analysis. The aim of this study was to gain insights into the genomic relationships of African breeds, Afrikaner, Bonsmara, Nguni, Ankole, Kenana and N'Dama, their demographic history, introgression, and identifying selection signatures responsible for their environmental adaptation, using whole-genome sequence data. The population genomics analysis was divided into five major analyses; genomic diversity, introgression and migration/gene flow events, demographic history inference, identification of selection signatures and functional annotations of selection signature regions according to the specific objectives of this study.

4.2 Summary statistics from mapping of the whole-genome data

Whole-genome sequence data for Nguni (5), Afrikaner (5), Bonsmara (5), Ankole (5), Kenana (5), N'Dama (5), and Holstein (2) were generated at 10X coverage using the Illumina platform, the Water Buffalo was generated at 36x coverage from the Illumina platform. **Table** 4.1 shows the summary statistics of the mapped reads. The average mapping coverage was 96.89% across the breeds with most samples mapped at 97% on ARS-UCD1.2_Btau5.0.1Y reference genome using BWA method. The mapping rate was similar to that of Kim et al. (2017) and Kim et al. (2020) who obtained rates of more than 95%, indicating that the quality of the nucleotide bases and coverage were adequate to conduct further analyses. Only two samples from Afrikaner (AFR 20 and 22) had lower mapping coverage of less than 95%. The average fold coverage was 10.63 as expected across all the samples. The average number of called SNPs and InDels per breeds were highest for Kenana (14,152,609; 1,987,935) followed by Ankole (11,738,570; 1,673,009), Nguni (10,690,542; 1,577,835), Afrikaner (9,671,019; 1,461,112), Bonsmara (8,823,934; 1,363,221), N'Dama (8,777,456; 1,355,713) and Holstein (6,479,380; 1,072,710) respectively. There was a direct relationship between the number of mapped reads and the

number of variants detected, the higher the mapped reads the higher the variants detected (**Table** 4.1).

Table 4.1: The summary statistics of the mapped reads and identified SNPs and InDels.

Name of sample	Raw sequences	Mapped reads	Unmapped reads	Reads mapped and paired	Reads duplicated	Indiviadual SNPs	Individual InDels	Percentage of properly paired reads (%)	Coverage
NGI25	220411838	219776458	635380	219601720	46914868	10035504	1464531	97.1	**9.57604**
NGI26	208736228	208325988	410240	208167314	40130485	10604376	1516970	97.0	**9.28472**
NGI27	234287570	233799828	487742	233634130	52940853	10194651	1499572	97.0	**9.96695**
NGI2	264620064	263286331	1333733	263005322	48939649	11383974	1688094	96.9	**11.778**
NGIx	328520860	326571222	1949638	326303806	68381850	11234207	1720009	97.2	**14.149**
BON11	203624592	203260463	364129	203126534	44797341	8552964	1289569	96.9	**8.77729**
BON13	229582672	229179294	403378	229020952	52798505	8660607	1307706	97.3	**9.74337**
BON16	225306504	224872662	433842	224722180	56215949	8644301	1345529	96.6	**9.34645**
BON17	231020660	230499767	520893	230300598	50308496	8832412	1339742	96.5	**9.80498**
BONx	382795996	379073528	3722468	378734366	90529887	9429388	1533561	97.1	**15.4723**
AFR1	240702748	239345914	1356834	239105686	44316114	10210765	1532599	97.2	**10.7215**
AFR20	215554666	215153159	401507	214989942	50806762	9273976	1434609	93.4	**9.11475**
AFR22	208719248	208437392	281856	208294846	51232285	9112008	1374894	94.3	**8.72024**
AFR17	205685502	205254095	431407	205094576	37833341	9598489	1433042	97.1	**9.26589**
AFRx	238407174	237532099	875075	237320080	44046981	10159857	1530414	97.5	**9.11475**
ANK109	270886970	269492223	1394747	268502444	3468428	11840918	1699543	97.6	**9.95913**

ANK111	277542808	275686250	1856558	274373724	1942613	11931535	1705234	97.4	**10.2504**
ANK1	265063462	262545674	2517788	261053832	4844222	11776436	1686937	97.0	**9.61172**
ANK50	309929504	307624905	2304599	305959798	3057240	11865175	1701498	97.2	**11.3706**
ANK73	229107460	227227516	1879944	225986004	2731198	11278787	1571832	97.1	**8.39348**
KEN11	307105878	304606425	2499453	302890768	4377878	14342535	2014654	97.0	**11.207**
KEN13	298725278	296375917	2349361	294703440	5724446	14091012	1974651	96.9	**10.8392**
KEN14	298633066	297111430	1521636	296071678	5482651	14190257	1998729	97.3	**10.8812**
KEN2	298114722	296104233	2010489	294575664	6897220	14102259	1977839	97.2	**10.7976**
KEN4	309243948	306905804	2338144	305462854	5671134	14036983	1973803	97.2	**11.2316**
NDA148	282057630	279686080	2371550	278242452	3244563	8790143	1355647	97.4	**10.3437**
NDA064	253088716	251167842	1920874	250229404	3554736	8739977	1345987	97.6	**9.27586**
NDA158	262549352	259881203	2668149	258720494	2665956	8574811	1320646	97.3	**9.62933**
NDA719	312773758	310268878	2504880	308861156	4972703	8763020	1364862	97.6	**11.4063**
NDA183	296240776	292774156	3466620	290644978	3984951	9019330	1391422	96.9	**10.7888**
HOL1	244997722	244419230	578492	244210726	57452043	6538432	1082751	96.5	**10.3047**
HOL2	241859248	241362578	496670	241110420	48712719	6420327	1062669	96.3	**10.6208**
						Total= 3,32E+08	Total= 49239545	Average= 96.89	**Avg=** 10.36

4.3 Variants statistics

From the 32-sample dataset of the breeds used in this study, a total of 37,482,988 SNPs of high-quality were identified. The number of SNPs identified was high compared to InDels at 4,931,938, multiallelic sites 1,553,700 and multiallelic SNP sites 566,731, indicated in **Figure** 4.1. After filtering, identified SNPs was further used in population genomics analysis.

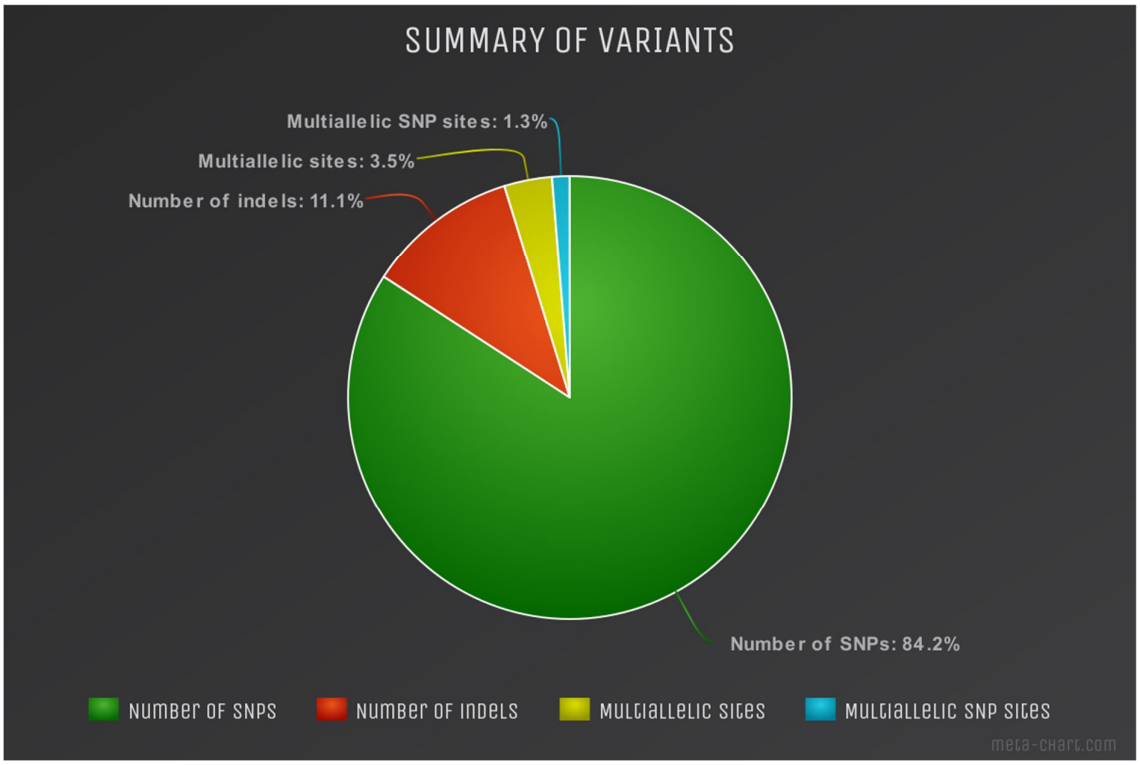

Figure 4.1: Summary statistics of variants identified using GATK best practices, with the highest number of variants identified being SNPs, followed by InDels, multiallelic sites and multiallelic SNPs.

4.4 Heterozygosity estimates and nucleotide diversity

The Site Frequency Spectrum (SFS) and heterozygosity are important in population genomics and they are used to estimate genetic diversity within and between populations (Ferretti et al., 2018). These analyses were done to calculate the genetic diversity estimates of studied populations. Estimates of genome-wide heterozygosity and nucleotide diversity were both significantly higher for Kenana, Ankole, and Nguni compared to Afrikaner, Bonsmara, and N'Dama as shown in **Figure** 4.2. The variation was separated into two halves, the breeds with the highest diversity and the breeds with the least diversity. In order; Kenana had the highest genetic diversity followed by Ankole, Nguni, Afrikaner, Bonsmara and Holstein. The average

genetic diversity observed between the breeds was (Afrikaner 0.0024), (Nguni 0.0028), (Bonsmara 0.00236), (Ankole 0.0032) (Kenana 0.0036), (N'Dama 0.0021) and the out-group (Holstein 0.001534) as indicated in **Figure** 4.2. The average nucleotide diversity (using 100 kb window) is presented in **Figure** 4.3 and it supports estimates generated by heterozygosity estimates across individual's genomes. The *Bos indicus* (Kenana) had the highest genetic diversity, followed by the Sanga group (Nguni, Ankole, Afrikaner) and the composite (Bonsmara). Lastly, the *Bos taurus* had the lowest diversity (N'Dama and Holstein) as in Kim et al., (2021) who observed high diversity in Sanga and zebu breeds.

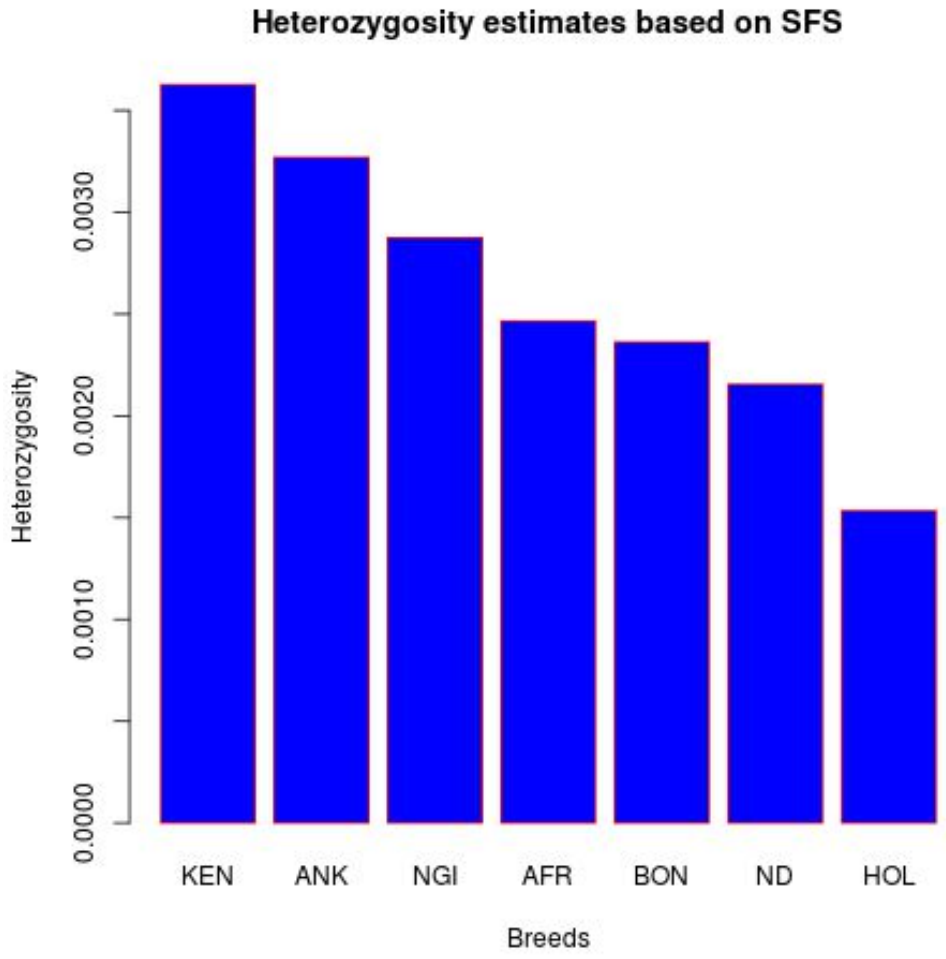

Figure 4.2: Genome-wide heterozygosity estimates in all the populations. Kenana (KEN) cattle had the highest variation, followed by Ankole (ANK), Nguni (NGI), Afrikaner (AFR), Bonsmara (BON), N'Dama (ND) and Holstein (HOL). These estimates reflect the diversity within each population and they may be used to inform breeding and conservation programs.

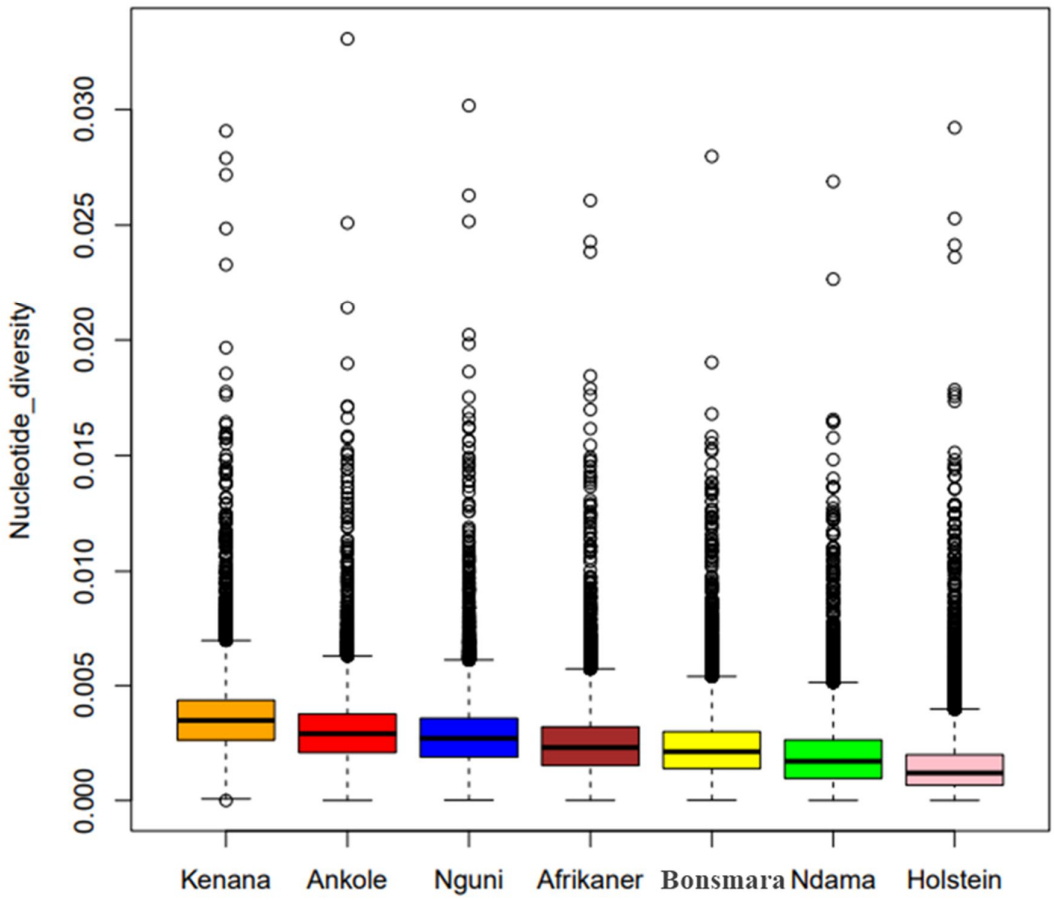

Figure 4.3: Estimations of nucleotide diversity between breeds. These estimates were produced by nucleotide diversity index (Pi) with a non-overlapping window of 100 kb using VCFtools and they are showing variation between the studied populations. Interestingly, Bonsmara (a composite) is showing to have a higher genetic diversity compared to N'Dama as observed in (**Figure** 4.2).

4.5 Genetic relationships and population structure

The phylogenetic tree and the principal component (PCA) supported previous findings on the clustering of the African and European lineages (**Figure** 4.4 and 4.5) (Makina et al., 2014; 2015). Nguni showed expectedly, to have a close relationship with SA breeds (Bonsmara and Afrikaner) and also showed a close relationship with Kenana and the Ankole. **Figure** 4.4 showed three main clusters/groups, 1) that of Kenana, Ankole, and Nguni, 2) for Afrikaner, and 3) the group formed by Bonsmara which was observed to be closely related to the *Bos taurus* breeds (N'Dama (African *taurus*) and Holstein). However, on the PCA (**Figure** 4.5), Ankole and Kenana were not clustered closely to each other, rather, it captured consistent results of SA breeds clustering together. In particular, the Nguni group as shown in **Figure** 4.4,

showed significant within-population genetic differentiation. The link between **Figure** 4.4 and 4.5 is that SA breeds clustered together, with Bonsmara showing to gravitate towards the outgroup (European). The Afrikaner and Nguni cattle showed to have close relationships with Ankole and Kenana and appeared to have a distant relationship with N'Dama. The clustering of Bonsmara is consistent with the findings from previous studies which suggested that this breed has three ancestry lineages, the African taurine, indicine, and European taurine. These results demonstrate the ancestral relationships of the studied breeds.

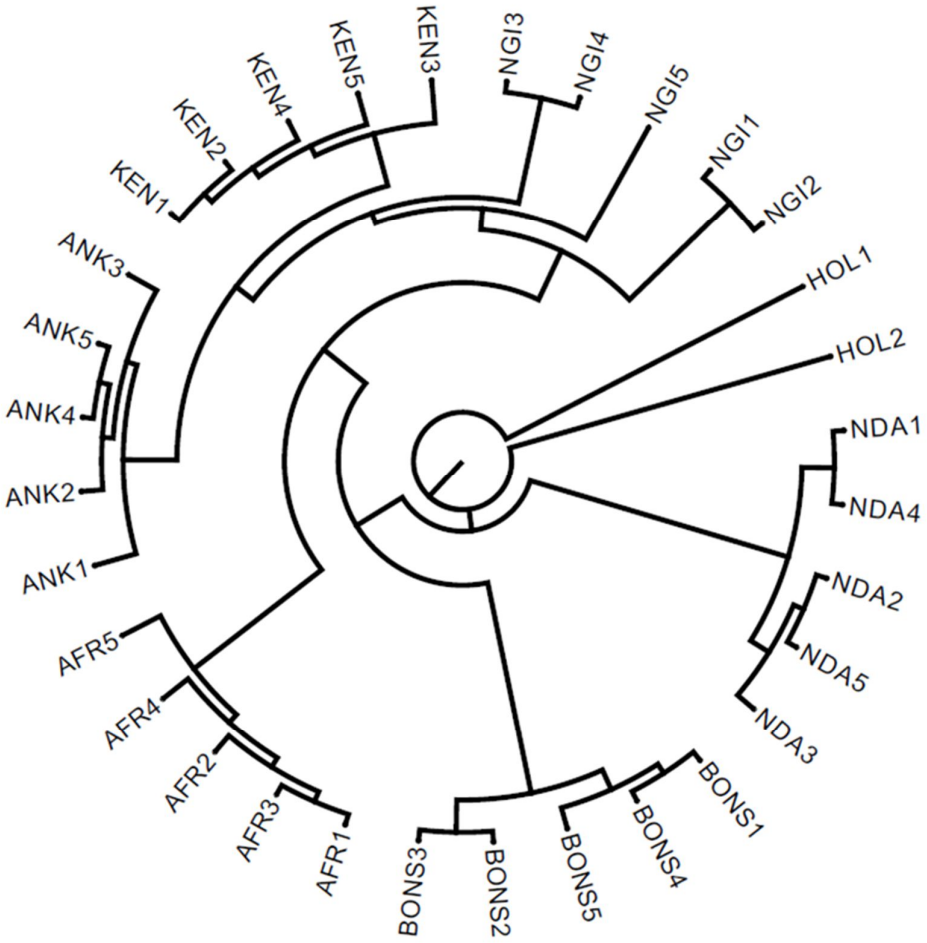

Figure 4.4: Pairwise genetic distances reconciled into a neighbour joining species tree from whole African cattle's genomes. SA breeds are denoted as NGI*, BON* and AFR* and they showed a relationship with other African breeds Ankole (ANK*) and Kenana (KEN*), also Bonsmara (BON) showed a relationship with Holstein (HOL*) and N'Dama (NDA*).

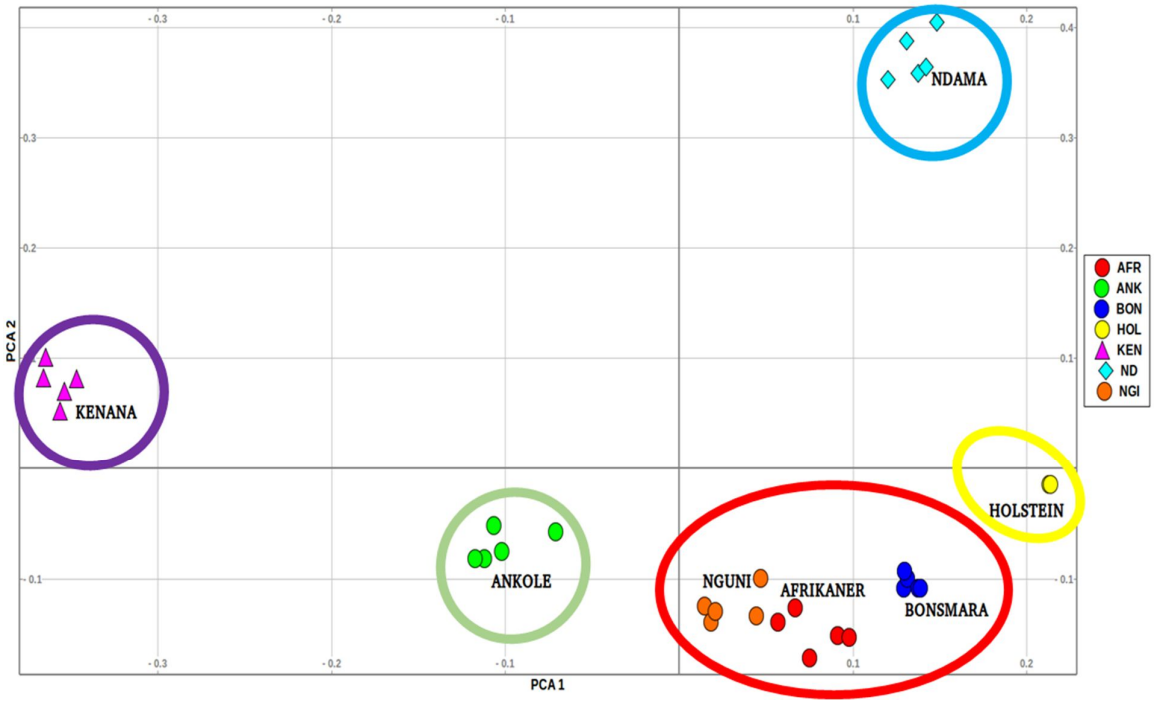

Figure 4.5: Comparative analysis of African indigenous cattle breeds based on the SNP data. SA breeds are denoted by the red circle and their relationships were being compared to other African breeds (Purple; Kenana, Blue; N'Dama, Yellow; Holstein, Green; Ankole).

In order to confirm the PCA results, we constructed the Bayesian clustering algorithm to understand the admixture levels between the studied populations (**Figure** 4.6), and the data was consistent with what was observed from the PCA plots. The informative relationships were observed in **K=3**, which showed the distinctiveness at the geographic level. N'Dama showed to be far from other breeds and this was expected because it was the only sampled West African breed. Bonsmara exhibited its relationship with *Bos taurus* ancestry and the Afrikaner. As expected, all the SA breeds showed signs of relativeness. Sanga breeds showed signs of admixture confirming claims on the classification of Sanga breeds (Rege, 1999; Scholtz et al., 2011; Mwai et al., 2015).

Figure 4.6: The STRUCTURE analysis of studied breeds; the lowest standard error was K=2 with CV error: 0.60915. Although the plot is not showing clearly the ancestral relationships between populations, it managed to capture some relationships, e.g K=3 was informative in terms of the geographic location's relationships.

4.6 Introgression and gene flow analysis (ABBA-BABA Statistics) and Treemix

Looking at the admixture analysis results, we went further to do introgression, gene flow/ migration events analysis by using both Dstat (**Table** 4.2) and Treemix (**Figure** 4.7) respectively. This was done to check the genetic material exchange between the breeds, this analysis have the potential to reveal forces behind the architecture of modern breeds. The inter-regional gene flow was examined using the ABBA-BABA test, guided by the relationships observed on the topology (**Figure** 4.4). The high D-statistics (mostly applied to detect post-divergence gene flow between populations or closely related species) could either indicate differential levels of gene flow or signs of ancestral admixture. When the gene flow between Afrikaner-X-Nguni (SA indigenous breeds) were analysed as sub-sisters against the Ankole, Kenana, and N'Dama, Nguni showed significant levels of gene flow with all the breeds. When

Ankole-X-Kenana (African breeds) were compared, Kenana had significant gene flow with all the other breeds (Nguni, Afrikaner, N'Dama, and Bonsmara). These results were unexpected because as a Sanga breed, Ankole was expected to have significant gene flow with the other Sanga breeds (Afrikaner and Nguni) because of their close ancestry lineage. However, Ankole showed significant genetic exchange with N'Dama when it was paired as sisters with both Nguni and Afrikaner.

Table 4.2: Relationships of inter-regional gene flow/introgression events generated by D-statistics using Holstein as out-group.

Population1	Population2	Population3	Out-group	D-Stat	SE	Z-SCORE
AFR	NGI	KEN	HOL	0.04935329	0.004062787	12.14765
AFR	NGI	ND'	HOL	0.009089576	0.002521326	3.605078
AFR	NGI	ANK	HOL	0.03191823	0.003371067	9.468289
ANK	KEN	NDA	HOL	0.04952499	0.002847004	17.39547
ANK	KEN	NGI	HOL	0.07577715	0.002903927	26.09471
ANK	KEN	AFR	HOL	0.05297526	0.002824412	18.75621
ANK	KEN	BON	HOL	0.03578199	0.002433704	14.70269
ANK	NGI	ND'	HOL	-0.08717984	0.002858506	-30.4984
AFR	ANK	ND'	HOL	0.09551073	0.003087873	30.93091
AFR	NGI	BON	HOL	-0.02978893	0.002537152	-11.74109
ANK	NGI	BON	HOL	-0.164312	0.003528482	-46.56732
ANK	KEN	BON	HOL	0.3816804	0.003648986	104.599
AFR	ANK	BON	HOL	-0.00905599	0.002601872	-3.480567
BON	ND	NGI	HOL	0.111498	0.00308755	36.11212
BON	ND'	AFR	HOL	-0.1412843	0.003360429	-42.04353

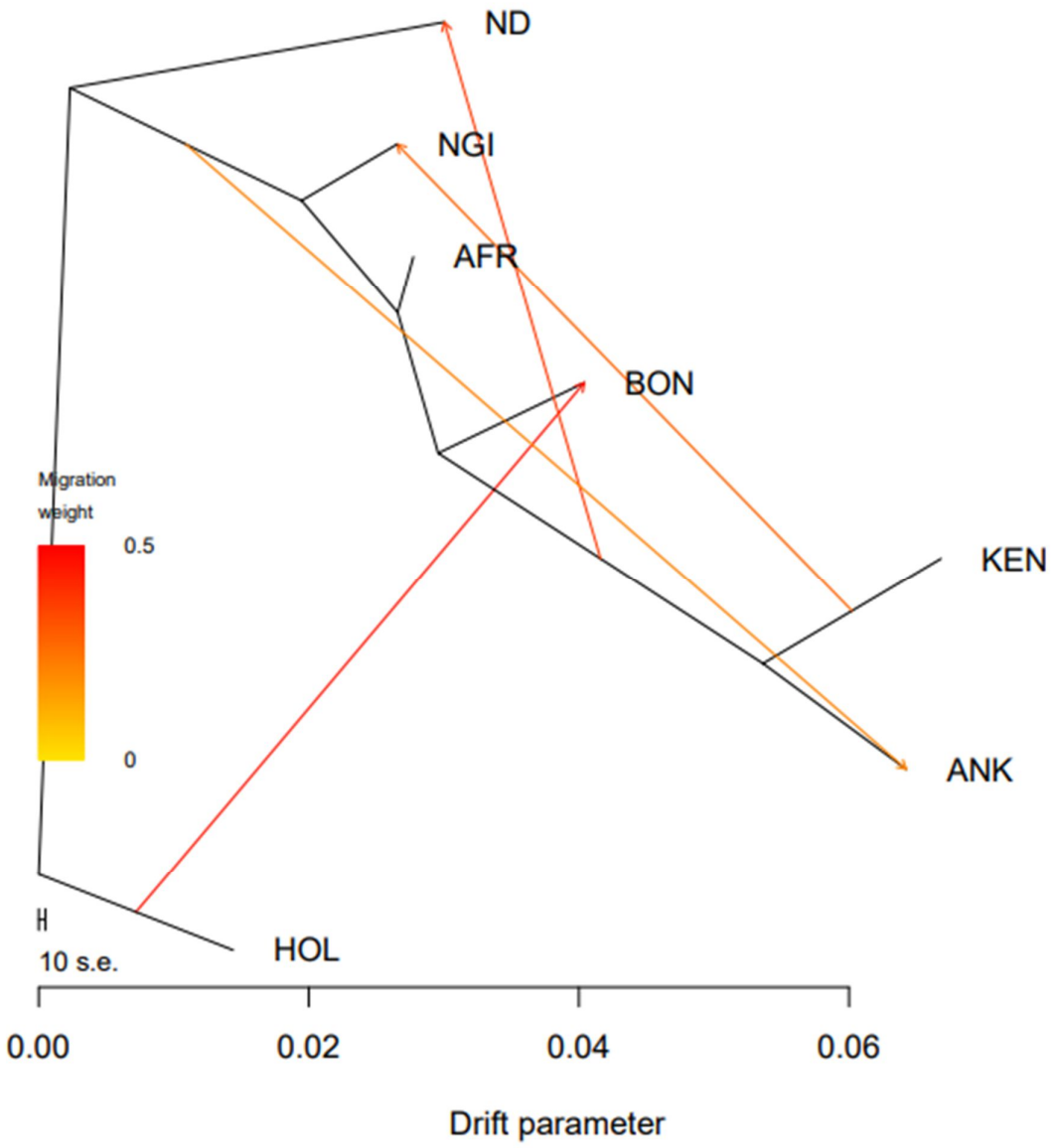

Figure 4.7: Inference of population splits and migration events in different African populations. These are possible migration events that shaped the genomes of the studied populations. This plots presents novel relationships that has not been reported in previous studies e.g. NGI and KEN.

Then, Bonsmara showed significant introgression with Afrikaner in all scenarios they were compared. The results were expected because Afrikaner was used in the formation of Bonsmara. Significant relationships was also observed between Bonsmara and Nguni. Interestingly is the relationship between N'Dama and Ankole with highly significant frequencies (**Table** 4.2). Treemix supported D-stats results, by further highlighting some migration/gene flow events between Holstein-X-Bonsmara, Ankole-X-N'Dama, Kenana-X-Nguni, Ankole-X-N'Dama. The ancestries of Nguni-X-N'Dama-X-Ankole was also detected.

(X sigh is shows detected migration/gene flow events between the breeds, these results can be seen in (**Figure** 4.7 and **Table** 4.2).

4.7 D-statistics follow up-using a more distant out-group

Due to the close relatedness between the ancestry of Bonsmara and Holstein, an alternative out-group was necessary to avoid any possible biases, as in (Moodley et al., 2020). To mitigate these biases a Water Buffalo that is more distant to the rest of studied breeds was used to further understand introgression events. This analysis confirmed some relationships that were observed when Holstein was used as an out-group, the major concordance was the persistent dominance of Nguni and Kenana that showed to have shared genetic components when compared with other breeds (**Table** 4.3; 4.2). However, when we used Water Buffalo as an out-group, there were some differences observed in some breed combinations e.g. When Ankole and Nguni were compared with Bonsmara using Holstein as an out-group, there was an introgression between Ankole and Bonsmara, however when we used the Water Buffalo there was a relationship between Nguni and Bonsmara, an expected results due to their geographic location, the other two combinations that were observed was the introgression between N'DamaXAfrikaner and AnkoleXBonsmara, results that is different from what was observed in **Table** 4.2, showing the effect of using a closely related out-group. In (**Table** 4.3) it can be seen that relationships observed in (**Table** 4.2) could be influenced by the out-group used. Although some relationships did not change, it was observed that the out-group plays a significant role and if it is not chosen cautiously it can cause biases. The results obtained from the two out-groups calls for further analysis that will investigate gene flow or introgression events in African indigenous cattle, these questions were also observed in (Makina et al., 2015), and they will require further assessments of gene flow events.

Table 4.3: Relationships of inter-regional gene flow events generated by D-statistics using Water Buffalo as an out-group.

Population1	Population2	Population3	Out-group	D-Stat	SE	Z-SCORE
AFR	NGI	KEN	BUF	0.008583464	0.001940726	4.422811
AFR	NGI	ND'	BUF	0.04889868	0.004961099	9.85642
AFR	NGI	ANK	BUF	0.02817889	0.00299616	9.405
ANK	KEN	NDA	BUF	0.1826101	0.005296058	34.48037
ANK	KEN	NGI	BUF	0.1614455	0.004413808	36.57737

ANK	KEN	AFR	BUF	0.1763529	0.00478053	36.88982
ANK	KEN	BON	BUF	0.1916325	0.005026423	38.12503
ANK	NGI	ND'	BUF	0.04549223	0.004048602	11.23653
AFR	ANK	ND'	BUF	0.1576777	0.005115402	30.8241
AFR	NGI	BON	BUF	-0.07298508	0.00445908	-16.36774
ANK	NGI	BON	BUF	0.1298295	0.004618526	28.11058
ANK	KEN	BON	BUF	-0.1916325	0.005026423	-38.12503
AFR	ANK	BON	BUF	-0.1975683	0.004901655	-40.30644
BON	ND	NGI	BUF	0.03792016	0.004420451	8.578345
BON	ND'	AFR	BUF	0.06529393	0.004811853	13.5694

4.8 Demographic analysis

The results from PSMC analysis (**Figure** 4.8 and 4.9) indicated that African breeds had two bottlenecks or population declines which affected them in different ways. The first bottleneck occurred around ~1 Million years ago (mya) to 100 thousand years ago (kya) and the second one occurred between 40 kya and 80 kya (**Figure** 4.8 to 4.12). A significant change in population sizes of African breeds was observed between ~2 mya and around the Pleistocene-Halocene transition period (around 11, 000 kya). The Kenana breed had the highest effective population size in all periods, the interglacials (IG) in the Pleistocene and Holocene, the Last Glacial Maximum (LGM), the mid-Pleistocene transition (MPT), and the Plio-Pleistocene transition (PPT) (from the beginning of the trajectory to present time). On the other hand, the N'Dama breed shown to have the lowest effective population size (**Figure** 4.9). Within SA breeds (**Figure** 4.8; 4.10; 4.11 and 4.12) Afrikaner and Nguni had similar trends until around ~60 and 80 kya, a period that seems to have changed the effective population size of many African breeds. This splitting point affected all breeds. A further observed pattern was the huge rise in effective population size of Kenana after the transition period (Pleistocene-Holocene), which behaved differently from the rest of the breeds (**Figure** 4.8; 4.11; 4.12). However, the Ankole breed maintained the same effective population size during this period and showed similar patterns with Nguni than Afrikaner (**Figure** 4.10). The timing and amplitude of effective population sizes (Ne) differed among the breeds (**Figure** 4.8 to 4.12).

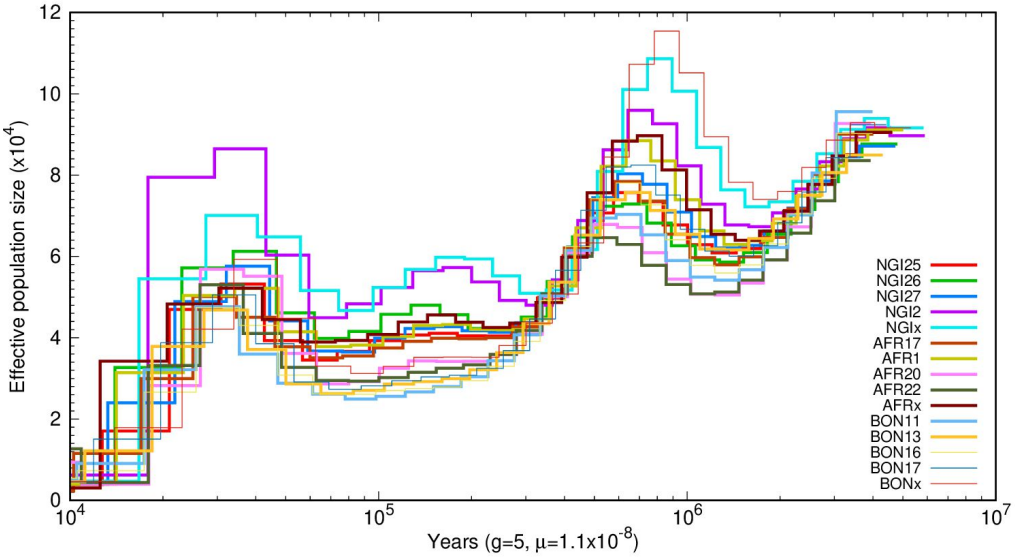

Figure 4.8: Population sizes inferred from SA breeds. The X-axis is the time scaled in (kya), and the Y-axis is representing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$.

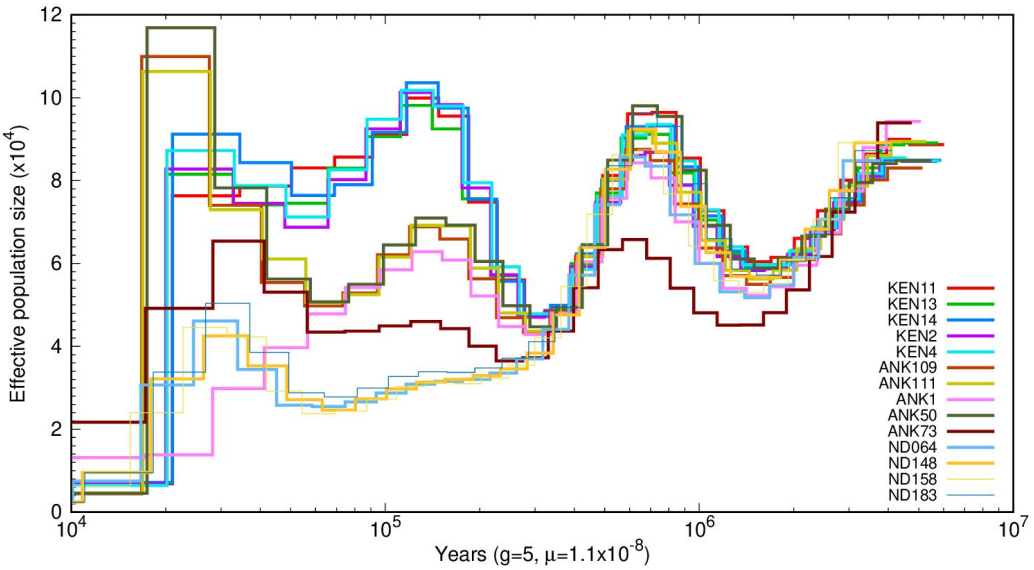

Figure 4.9: Population sizes inferred from West (N'Dama), East (Kenana), and East (Ankole) African breeds. The X-axis is representing the time scaled in (kya), and the Y-axis is showing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$.

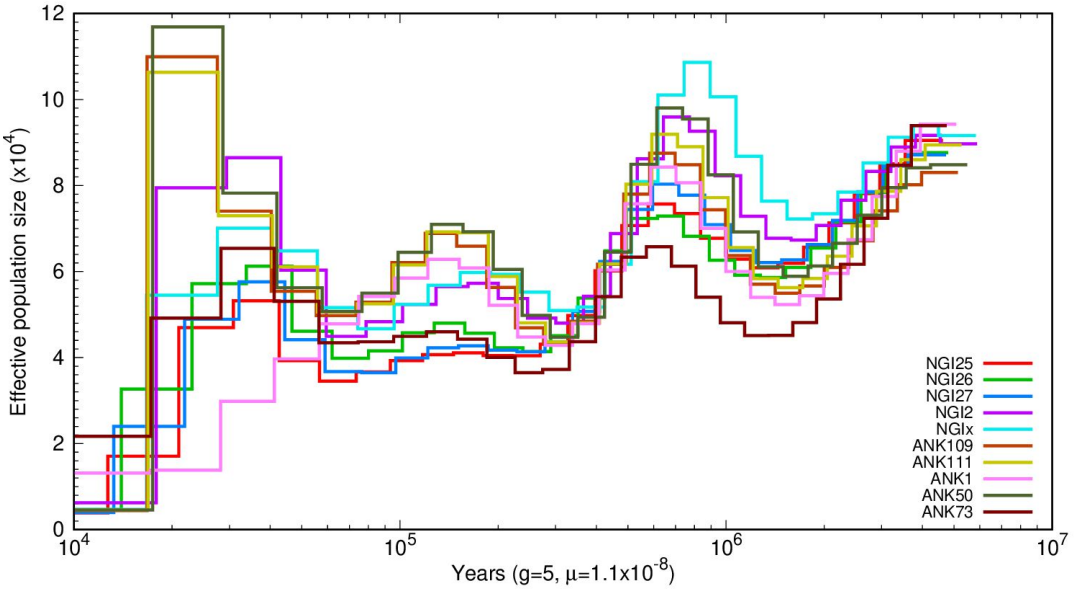

Figure 4.10: Population sizes inferred from SA (Nguni), East (Ankole) African breeds. The X-axis is the time scaled in (kya), and the Y-axis is representing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$. This plot also show that the trajectory is not affected by the number of populations plotted.

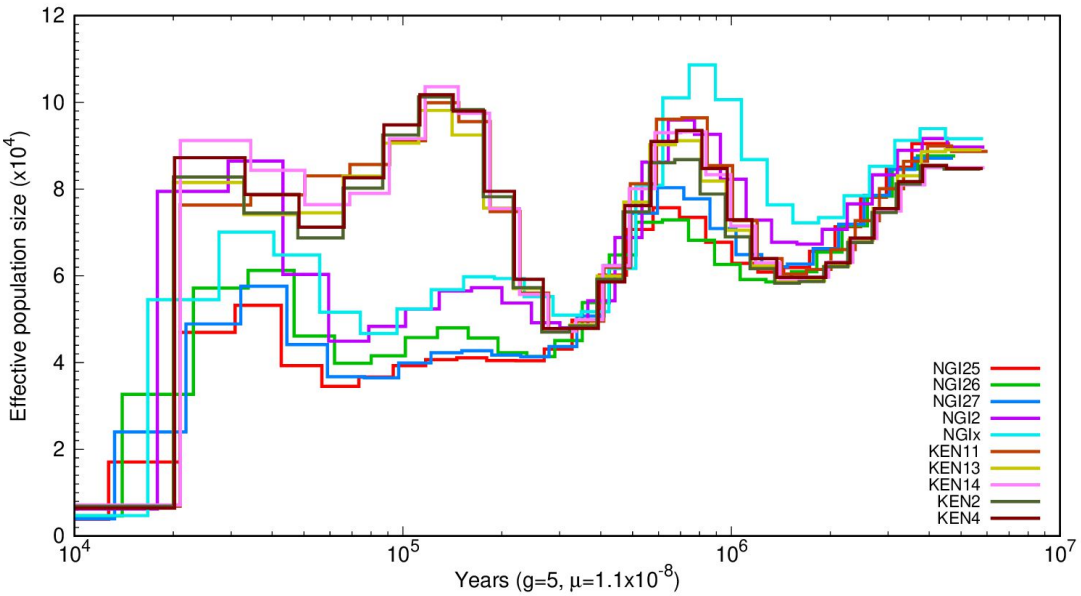

Figure 4.11: Population sizes inferred from, East (Kenana), and SA (Nguni) African breeds. The X-axis the time scaled in thousand years (kya), and the Y-axis is representing the effective population size. g is the generation time in a mutation rate of $1.1x10^{-8}$.

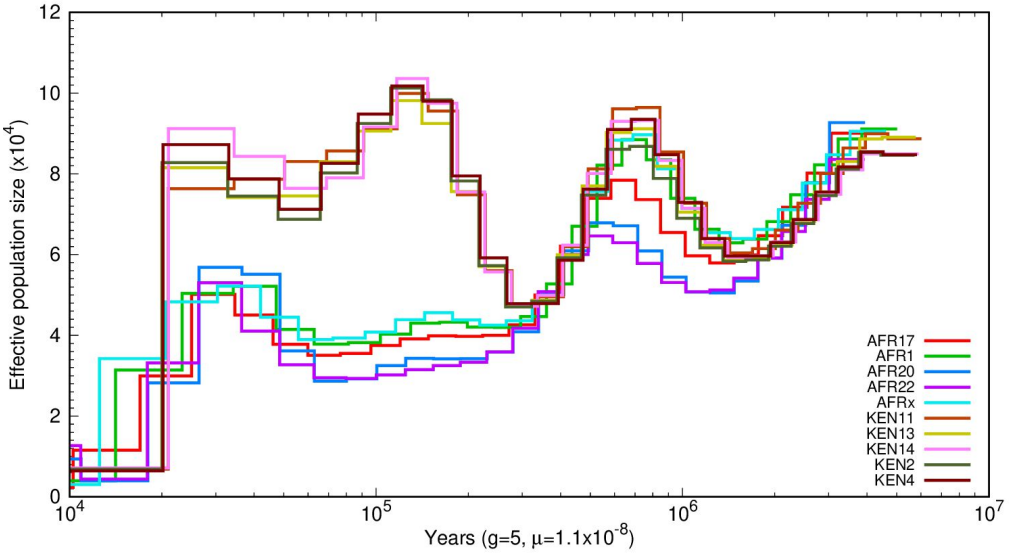

Figure 4.12: Population sizes inferred from East (Kenana), and SA (Afrikaner) African breeds. The X-axis is showing the time scaled in (kya), and the Y-axis is showing changes in effective population size over time. g is the generation time in a mutation rate of $1.1x10^{-8}$.

4.9 Identification of candidate genomic regions under selection

Genomic regions under selection were identified using iHS and XP-EHH statistical methods. Due to recent introduction of Bonsmara and Holstein to the African farming systems, we excluded these breeds from selection footprints analysis. To identify genomic regions harboring signals of positive selection, randomly estimated haplotypes data from Kenana, Ankole, Nguni, Afrikaner and N'Dama were used and the top 1% candidate genes were selected for functional annotations. The distributions of significant and non-significant iHS and XP-EHH P-values are shown in (**Figure** 4.13 to 4.23), and a conservative threshold of 4 (P-value < (0.0001) that determine significant regions of outliers was applied. A total of 1308 candidate regions were detected from the iHS and 1023 from the XP-EHH test. These regions were analysed vs. the Ensembl Cow database to scan for potential genes associated with these regions, and 112 (iHS) and 120 (XP-EHH) candidate genes were captured at 1 MB window in Ensembl. Gene regions were further annotated in DAVID and Panther using default parameters. The top 1% of the iHS and XP-EHH were considered to be candidate genes and some of the genes identified from iHS were associated with growth and development (*CNTN6, LAMA2, DPP6, KCNIP4, KCNMA1, BMP5, CTNND2*), and *(DAPL1)*. Other genes were involved in diseases development or diseases processes such as *EPHB1, KALRN, CHRNA7, HIF-1A, DLC1, KSR1, CNNM2*, and *CYFIP1*. Lastly, we detected genes *(USH2A, ZMPSTE24, CHRNA7, APP, CTNNA3, PTPRQ, CSN3, ABCG2, GHR, RORA, CDH13, and KCNN2)*

harboring important economic traits for milk, fertility and body structures (**Table** 4.4, 4.5, **Figure** 4.9.11). Gene annotation of candidate regions identified by XP-EHH were 34 regions from ANK-AFR, 37 ANK-NGI, 29 KEN-AFR, 49 KEN-NGI, 37 ND-AFR, and 49 ND-NGI (**Table** 4.4) (**Figure** 4.17 to 4.22). Genes such as *NOX4, PDGFD, BANK1, TG, ADAMTS, IGF1R, YAP1, GRB10, TRAPPC9, MAGUK, GRM1, PPFIA1, DLC1, and VDAC1*, with functions for growth, diseases, important roles in pathways and other important biological processes were detected and they were linked with the breed of interest in (**Table** 4.5).

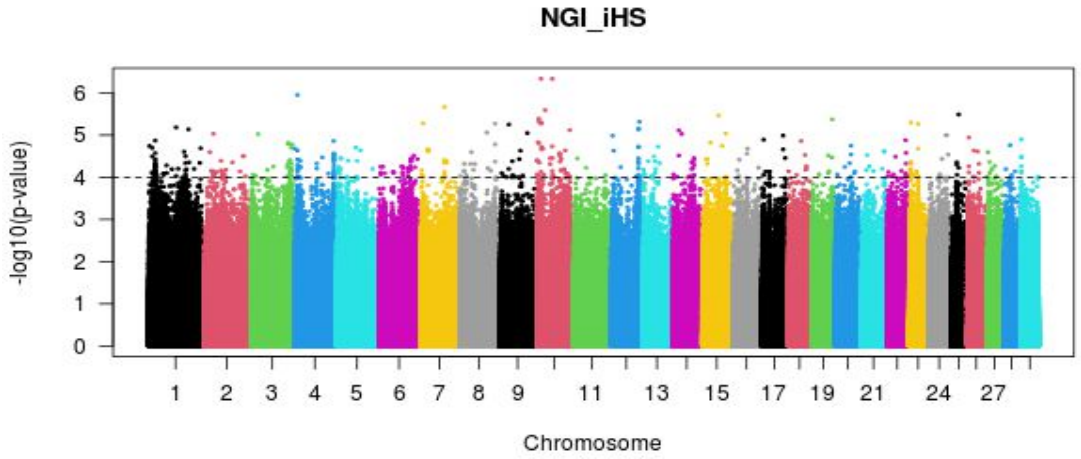

Figure 4.13: The Manhattan plot for the iHS test based on the autosomes of the Nguni cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

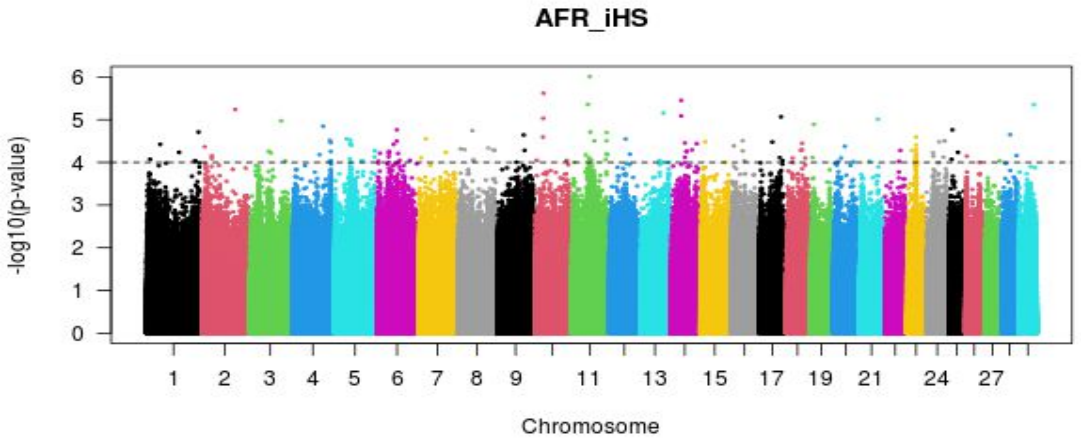

Figure 4.14: The Manhattan plot for the iHS test based on the autosomes of the Afrikaner cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

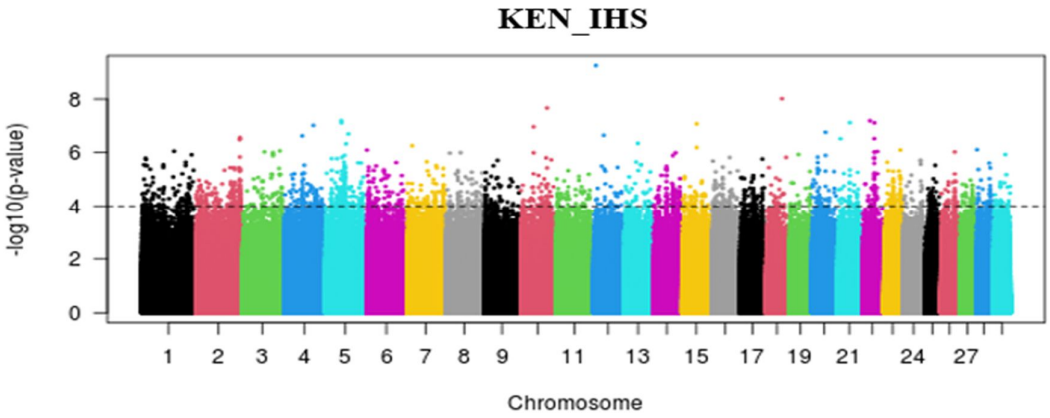

Figure 4.15: The Manhattan plot for the iHS test based on the autosomes of the Kenana cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

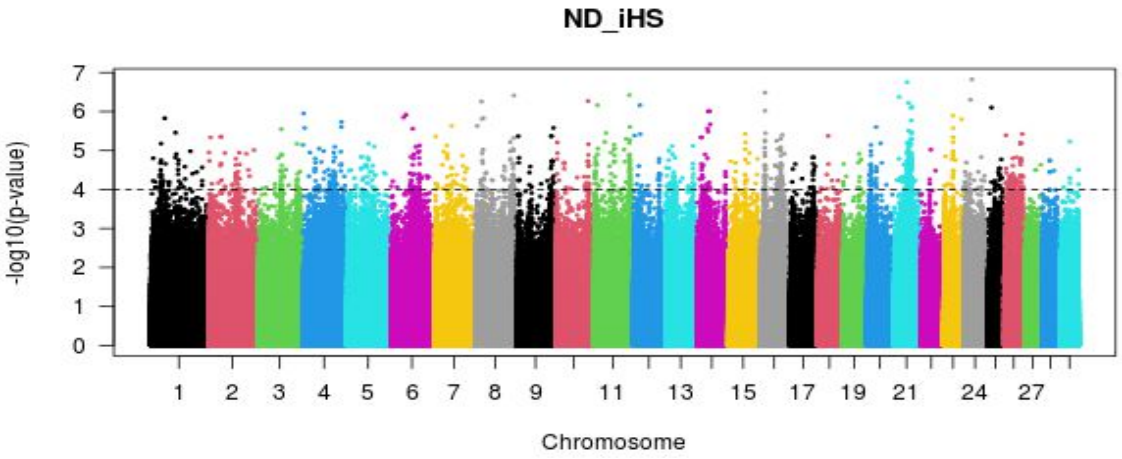

Figure 4.16: The Manhattan plot for the iHS test based on the autosomes of the N'Dama cattle. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

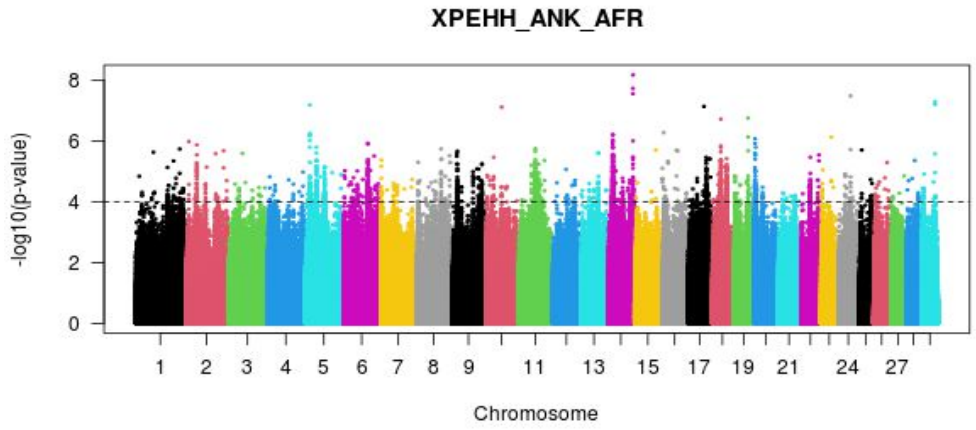

Figure 4.17: The distribution of standardized XP-EHH scores of the Ankole and Afrikaner populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

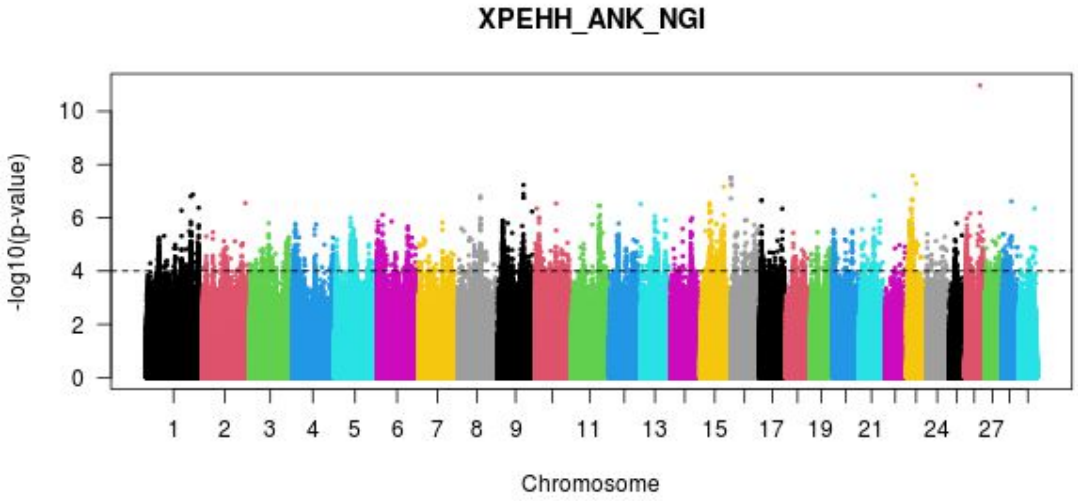

Figure 4.18: The distribution of standardized XP-EHH scores of the Ankole and Nguni populations. The horizontal dotted line is separating significant candidate regions using threshold of (–log10 (p-value) > 4).

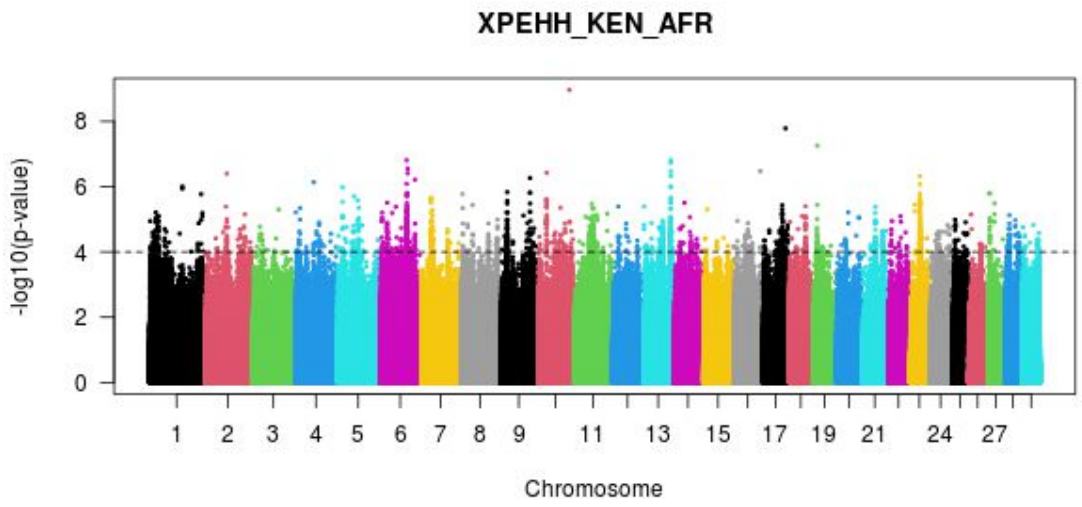

Figure 4.19: The distribution of standardized XP-EHH scores of the Kenana and Afrikaner populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

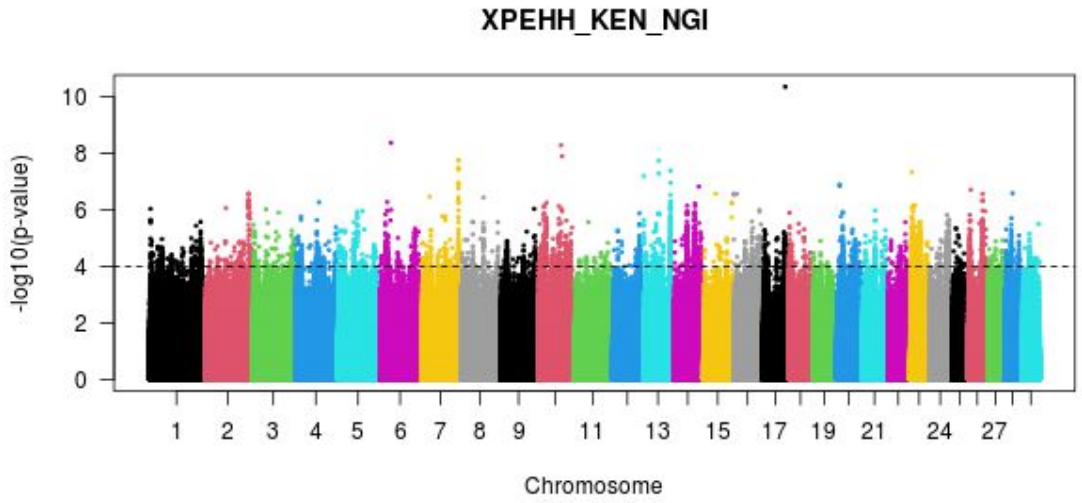

Figure 4.20: The distribution of standardized XP-EHH scores of the Kenana and Nguni populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

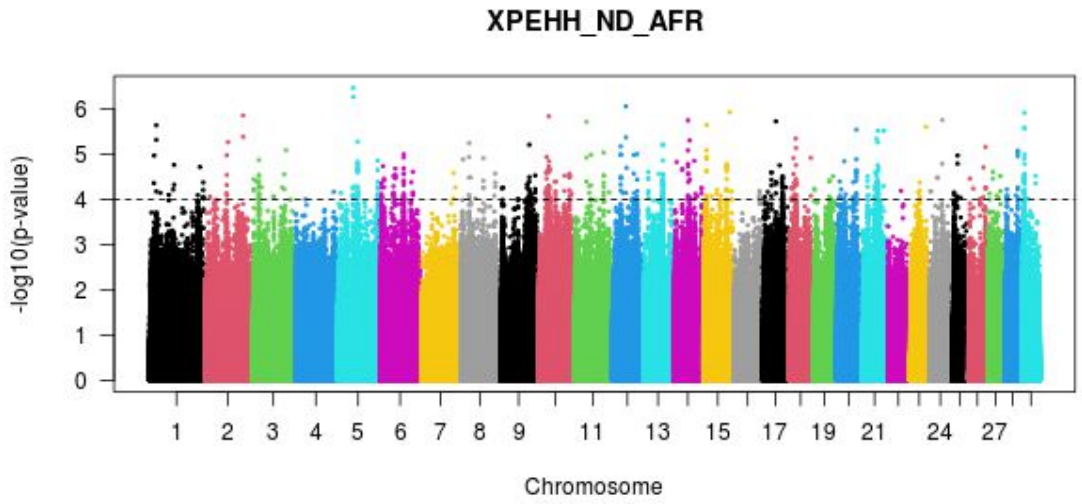

Figure 4.21: The distribution of standardized XP-EHH scores of the N'Dama and Afrikaner populations. The dotted horizontal line is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

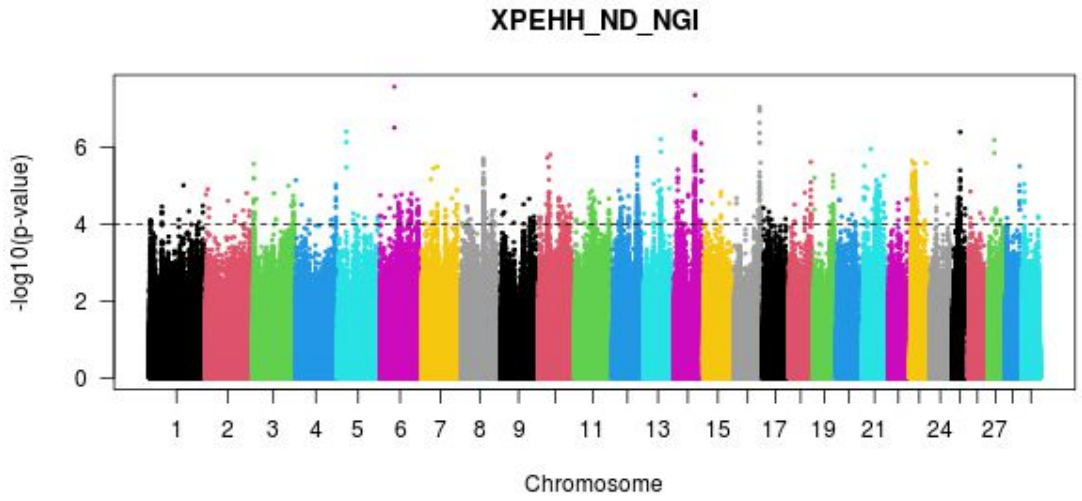

Figure 4.22: The distribution of standardized XP-EHH scores of the N'Dama and Nguni populations. The dotted horizontal lines is separating significant and non-significant candidate regions, on a threshold of (–log10 (p-value) > 4).

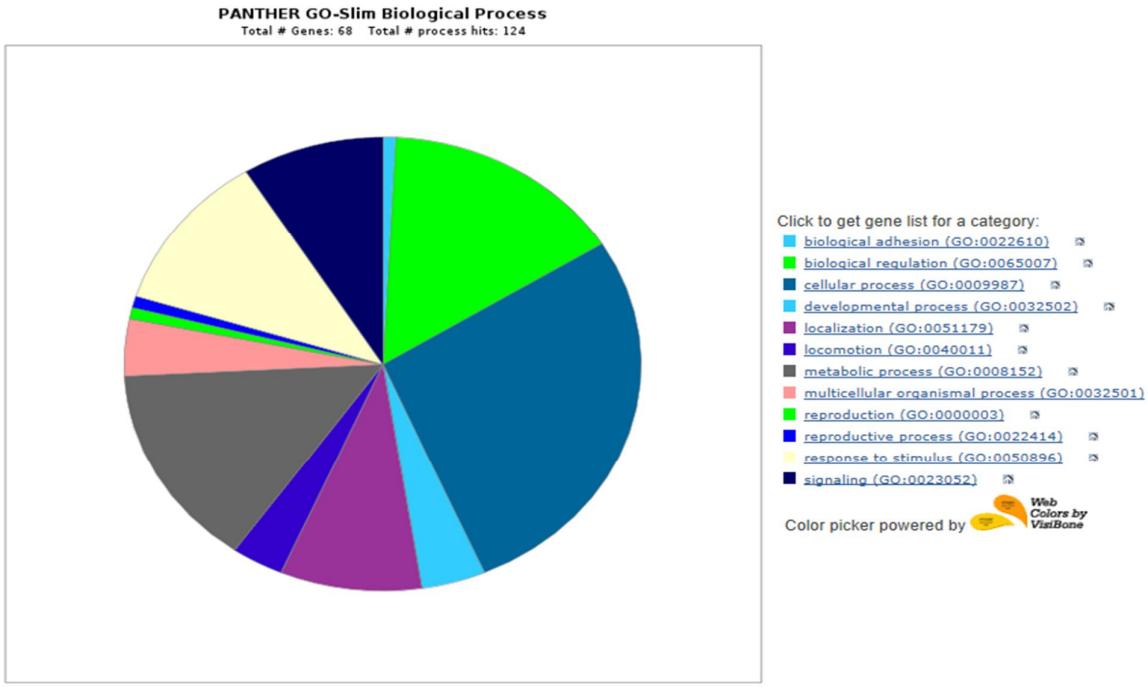

Figure 4.23: A representation of biological processes from the iHS significant genes. (Generated from Panther site).

Table 4.4: Significant candidate regions selected by the two statistical test (iHS). Genes coloured in red are shared between populations.

Gene	Chr	Function	References
		Nguni	
EPHB1	1	Regulation of common cancer cell transforming pathways.	Kampen et al. (2015)
CNTN6	22	Association with neurodevelopmental and autism spectrum disorders.	Repnikova et al. (2020)
KALRN	1	This gene plays a role in several diseases or disorders such as cardiovascular disease, psychiatric disorders, and neurodegeneration.	Mandela et al. (2012)
LAMA2	9	Organization of cells into tissues during embryonic development.	NBCI
DPP6	4	Involved in proteolysis and nervous system development.	Buzanskas et al. (2014)
KCNIP4	6	Involved in heart functions, skeletal muscle growth and immune response processes.	Eydivandi et al. (2021)

USH2A	16	Involved in hearing and vision as a member of the ***USH2*** complex.	NCBI
ZMPSTE24	3	Antiviral defense.	Li et al. (2017)
KCNMA1	28	It is present in the process of paroxysmal dyskinesia and/or epilepsy.	Wang et al. (2017)
BMP5	23	Involved in bone and cartilage development.	Chen et al. (2020)
		Afrikaner	
CHRNA7	21	This gene is mainly expressed in brain and it is considered to be a good candidate gene for the epilepsy phenotypes.	NCBI
APP	1	It is associated with two functions, in humans it is linked with obesity and in pigs with meat tenderness.	Lim et al. (2016)
FANCM	21	Plays a role in eukaryotic DNA-damage detection and it is also involved in the activation of the Fanconi anemia (FA) pathway.	Ito and Nishino, 2021
CTNNA3	29	Associated with features such as body length, weight, eight, and chest circumference in cattle.	Zhao et al. (2022)
CTNND2	20	Associated with brain development and it has been linked with guiding nerve cells to their proper positions for neuronal migrations.	Zhao et al. (2022)
DAPL1	2	Involved in the regulation of testicular steroid system and it can be used to understand symptoms linked with dysregulated endogenous testosterone levels.	Chen et al. (2021)
MAP4K4	11	Associated with the somatic cells score and other different milk traits.	Bhattarai et al. (2017)
AGO2	14	Antiviral	Ludman et al. (2017)
		Ankole	
HIF-1A	13	Associated with a variety of tumors and oncogenic pathways.	Lee et al. (2004)
PTPRQ	5	This gene is involved in the making of meat traits in beef cattle, possibly through the regulation of the *MRF* gene expression.	Robakowska-Hyżorek et al. (2016)
ROBO2	1	Associated with a better prognosis	Ding et al. (2021)
PRKG1	26	Plays a role on milk fatty acids in dairy cattle	Shi et al. (2019)
DLC1	27	Associated with tumorigenesis and development	Xie et al. (2015)

IGF1	5	It is involved in gonadotrophin-releasing hormone during puberty stages, and gonadotrophin-releasing control the process of pubertal development.	Fortes et al. (2013)
PDGFD	15	Involved in embryonic development as well as cell growth, migrations, survival, and chemotaxis.	UniProt
PTK2	14	Associated with dairy production	Wang et al. (2013)
IGF1R	21	Association with economic traits	
KSR1	19	Involved in as the potential therapeutic target for Merlin-deficient tumors.	Zhou et al. (2016)
PLCB4	13	Heat stress	Mohamadipoor et al. (2021)
		Kenana	
EDNRA	17	Involved in gamete transportation.	Ma et al. (2002)
TCP11	23	It is important in sperm function and fertility	Ma et al. (2002)
CSN3	6	Associated milk production	Alim et al. (2014)
ABCG2	6	Associated milk production	Nanaei et al. (2020)
GHR	20	Associated milk production	Nanaei et al. (2020)
RORA	10	Associated milk production	Nanaei et al. (2020)
DGKB	4	Adaptive immunity in mammals	Onzima et al. (2018)
DNAH8	23	Formation of the sperm tail	Weng et al. (2021
		N'Dama	
FBH1	13	Essential for the restoration of normal mitotic progression after decatenation stress.	Laulier et al. (2010)
CNNM2	26	Associated with renal hypomagnesemia, seizures, and developmental delay.	Sharawat, et al. (2021)
SLC22A23	23	Involved in the inflammatory bowel disease (IBD), endometriosis-related infertility	Ekizoglu et al. (2018)
CDH13	18	Fertility traits	Tarekegn et al. (2021)
CPEB3	26	Regulate the translation of dormant and masked mRNA in Xenopus oocytes	Wang et al. (2021)
KCNN2	10	Response to metabolic stress	Olivieri et al. (2016)

CYFIP1	2	Involved in pathological changes in the CNS growth and neuronal connectivity processes.	Pathania et al. (2014)

Table 4.5: List of significant genes from XP-EHH cross population analysis.

Gene	Chr	Function	References
		ANK-AFR	
NOX4	29	It is associated with several cellular functions and it is found in osteoclasts, cells such as melanoma, smooth muscle, hematopoietic stem and endothelial cells. Also, in fibroblasts, keratinocytes, and neurons	Yang et al. (2016)
PDGFD	15	Associated with the regulation of embryonic development, chemotaxis, cell migration, survival and cell proliferation processes.	UniProt
ARRB1	15	Involved in chronic liver diseases	Yang et al. (2015)
BANK1	6	Involved in various autoimmune diseases	Gómez Hernández et al. (2021)
		ANK-NGI	
TG	14	Involved in intramuscular fat content in cattle and it is also used for molecular selection to improve carcass traits in beef cattle	Barendse, 1999
ADAMTS	21	Associated with diseases such tissue and blood disorders, genetic syndromes, osteoarthritis, cancer and Alzheimer's.	Nicholson et al. (2005)
IGF1R	21	Involved in immunomodulation and in bone and muscle.	Ma et al. (2019)
YAP1		Alongside *TAZ* gene, it plays a vital role in granulosa cell proliferation.	Plewes et al. (2019)
WWOX	15	Association with the suppression of tumor	Kośla et al. (2020)
Grb10		It is a role in growth control, cellular proliferation, and insulin signaling processes.	Plasschaert and Bartolomei, (2014)
		KEN-NGI	
TRAPPC9	18	Involved in breast and colon cancer as well as liver diseases.	Mbimba et al. (2019)
MAGUK	14	Tumor suppression	Alday-Parejo et al. (2020)
		NDA-AFR	

GRM1	4	Involved in the maintenance of melanoma cells and has been linked with the treatment of melanoma.	Wangari-Talbot et al. (2012)
		NDA- NGI	
PPFIA1	9	Reported to be an essential promoter of migration and invasion in breast cancer.	Alfarsi et al. (2020)
VDAC1	7	Plays a role in the metabolic communication between cross-talk between the mitochondria (MtDNA) and other cells**.**	Shoshan-Barmatz et al. (2018)

Chapter 5

5. General Discussion and Conclusions

5.1 Introduction

This study applied whole-genome sequence data to generate information on the state of genomic diversity, introgression, gene flow, changes in effective population's sizes and detection of selection signatures of African breeds. The inclusion of SA populations (Nguni, Afrikaner and Bonsmara) in this study unveiled new relationships that has not been reported before. Genomic make-up of cattle breeds may show different historical events including adaptation to different climate conditions, speciation, gene flow and effects of artificial selection (Chen et al., 2021), and also allow identification of important genes associated to these events. Therefore, characterization of these populations is essential for conservation and preservation of these cattle breed resources (Mwai et al., 2015). Because African cattle are exposed to several challenges including disease pressures, poor feeding and extreme climates (Mwai et al., 2015; Dessie and Mwai, 2019), this study was necessary to bridge the knowledge gap by generating new data and information on the genomic diversity and adaptation mechanisms of these breeds using whole-genome sequence data. The data for these breeds were generated through sequencing of animals at 10X coverage using Illumina platform, for both selected SA (locally sequenced) and African (Kim et al., 2017) cattle populations.

5.2 Sequencing, mapping and variants calling

The use of the ARS-UCD1.2_Btau5.0.1Y reference genome assembly, known variants (ARS1.2PlusY_BQSR_v2.vcf.gz) and the best practice methods for mapping and calling variants, generated good and reliable sequence reads with expected coverage of = 10.36 X (**Table** 4.1), and the average mapping rate of (minimum, 93.4 %, maximum 97%). The mapping rate was almost similar to Kim et al. (2020) with (minimum: 91.70%, maximum: 99.91%). Comparing this study to Kim et al., (2017) and Kim et al. (2020) who provided the secondary data, it was observed that mapped reads and variants calls are influenced by the reference genome and variant calls algorisms employed (Talenti et al., 2022). This study adapted the Kim et al., (2020) statistical approach and discovered results that were comparable to other studies such as (Kim et al., 2017; Kim et al., 2020; Mauki et al., 2022). For SNP calling, the GATK best practices and the guideline recommended by the 1000 Bull genomes project were adopted. This was done to maximize reliable mapped reads and SNP identification, as

well as minimizing omissions and calling of biased positive reads and variants. It was observed that some of the existing pipelines such as SAMtools variant caller and the use of old reference genomes such as Btau_4.0 and UMD3.1, fail to detect all variants due to their biases such as favouring one breed over the other (Yi et al., 2014; Hwang et al., 2015; Kim et al., 2017). Therefore caution was taken when the variant calling methods were adopted. SAMtools and GATK have been reported to be more effective when applied to Ion Proton data and Illumina data respectively (Yi et al., 2014; Hwang et al., 2015), and in this study, GATK generated comparable outputs. However, better results could be achieved with the use of the reference assemblies that incorporate the global diversity, and also use of high coverage sequence data to improve the number of mapped reads and variants calls (Talenti et al., 2022).

5.3 Genomic diversity and population structure of African cattle breeds

African cattle breeds such as Kenana, Nguni, Ankole, Afrikaner, Bonsmara and N'Dama carry unique features that will be required in the era of climatic changes (Mwai et al., 2015; Kim et al., 2017; Zwane at al., 2019). The application of the latest genomic technologies to study their level of genetic diversity is important for breed's improvement programs and conservation. Since these breeds originated from either *B. indicus*, *B. taurus* or the hybrid of both, it was important to conduct a comparative study to understand their differences (**Table** 2.1) and it was observed that genetic diversity vary between these ancestral lineages, with B.*indicus* showing higher diversity as it was previously speculated (Kim et al., 2017; Mauki et al., 2022). These breeds are an important resource mainly to smallholder farmers due to their low maintenance and their ability to survive in harsh environmental conditions (Morgado, 2000). Moreover, population structure and genetic diversity information is important for effective breeding (Maciel et al., 2013; Mwai et al., 2015). In this study nucleotide diversity and heterozygosity estimates were done to assess the breed's genetic diversity followed by phylogeny, PCA and admixture analyses which were constructed to elucidate the breeds relatedness and ancestry levels.

Genome-wide heterozygosity and nucleotide diversity revealed high genetic diversity in zebu cattle (Kenana), followed by Sanga breeds Nguni, Ankole, and Afrikaner and low levels of genetic diversity was observed in N'Dama and Holstein. High genetic diversity (**Figure** 4.2; 4.3) observed in zebu and Sanga cattle corroborated findings from previous studies (Makina et al., 2014; Kim et al., 2017; Kim et al., 2021). Reduced genetic diversity of N'Dama was observed previously by Kim et al., (2017) and Kim et al., (2020) and it was attributed to

possible influence of human selection pressures (Makina et al., 2014; Kim et al., 2017; Mauki et al., 2022). This could also be due to the low effective population size observed in this study (**Figure** 4.9), with a signature of high genetic drift that could have caused its effective population to shrink. Additionally, genetic diversity in N'Dama has been linked to past disease challenges it faced in trypanosome-infested region of West Africa (D'Ieteren et al., 1998; Kim et al., 2017). Genetic diversity within SA breeds supported findings by (Makina et al., 2014), however it was different from Lin et al. (2010) and Edea et al. (2015) observations, who reported that *Bos taurus* breeds have more genetic diversity than *Bos indicus*. In this study *Bos taurus indicus* breeds demonstrated high genetic diversity (**Figure** 4.2; 4.3). Therefore, estimates from Lin et al. (2010) and Edea et al. (2015) could be due to SNPs chips used because most SNP data used in the design of the SNP chips were based on Eurasian breeds and African breeds were either partially included or entirely excluded, which causes biases (Pugach et al., 2015). Interestingly, Bonsmara showed stable genetic diversity, despite it being a composite breed. Its diversity was higher than that of N'Dama, indigenous cattle. Thus, Bonsmara could be used as a reference breed for future cross-breeding programs following Bonsma, (1980) recommendations. Based on these results, it is evident that African breeds are diverse and carry unique genetic resources that should be conserved and studied as suggested by (Mwai et al., 2015). It further suggests adoption of latest technologies and the inclusion of more data to clearly construct differences between existing cattle populations.

To characterize the structure/relationships of African cattle populations from different geographic locations, tree topology and the PCA analysis were performed (**Figure** 3.1; 4.4 and 4.5). From SA breeds, Nguni showed to be more related to other African breeds such as Kenana and Ankole with a closer relationship shown in the phylogeny (**Figure** 4.4). This was followed by Afrikaner that showed to be closely related to Kenana and Ankole (**Figure** 4.4), however, in the PCA both Nguni and Afrikaner showed to be distantly related to Kenana and Ankole but closer to Bonsmara and each other as observed by Makina et al., (2016). African breeds Ankole, N'Dama and Kenana showed consistent relationships that were previously reported by (Kim et al., 2017; Kim et al., 2020; Mauki et al., 2022). Notable results from the PCA which is consistent with the phylogeny is the relationships between Sanga breeds Nguni, Ankole and Afrikaner that showed signs of relatedness (**Figure** 4.5). In the PCA N'Dama was distantly related to other breeds, however in the phylogeny it was closely related to breeds that carry *Bos taurus* ancestry, confirming reports that it is a pure African taurine breed. In the admixture K=2 to K=4, Ankole, Kenana and N'dama showed relationships observed in (Kim et al., 2017).

Afrikaner breed showed to have a close genetic relationship with Nguni, confirming observation by Makina et al. (2014) who reported that Nguni cattle shares about 8% of its genetic components with Afrikaner, a relationship that could be due to their co-ancestry and the structure of their migration into the southern Africa (Scholtz et al., 2011). Bonsmara showed genetic relationships with Nguni cattle and this could be due to the genetic components it inherited from Afrikaner, a genetic data that is shared with the ancestor of Nguni. This observation corroborated the findings of Makina et al. (2014) who discovered that Bonsmara shared about 3% of its genetics with Nguni. However, when Afrikaner was analyzed with Bonsmara, the results were different from that of Makina et al. (2014) who discovered that Afrikaner was more related to Nguni. This study showed Bonsmara to be more related to Afrikaner than Nguni. This could be due to large amounts of data used which covered more sites of the genome and as a result, it could give more insights on the relationships than SNP chips could in previous studies. Genome-wide structure analysis in this study corroborated previous findings on the genetic relationships between SA (indigenous and locally-developed breeds) and *Bos taurus* cattle (McKay et al., 2008). Lastly, our results showed a clear difference between *Bos indicus*, African *Bos taurus*, and European *Bos taurus* breeds, and these results agree with several previous studies (BovineHapMap Consortium, 2009; Flori et al., 2019; Verdugo et al., 2019).

5.4 Introgression signatures, admixture, and migrations events

Assessing evidence of allele sharing yields information about population history and to identify pairs of populations that are related (Kim et al., 2017). These statistics are important for further investigating relatedness of the studied breeds. To achieve this objective, D-statistics analysis (Durand et al., 2011) and Treemix (Pickrell and Pritchard, 2012) were conducted. Although each method has its own shortcomings, the use of these methods provided us with enough information to understand possible genetic exchange that occurred between species. D-statistics analysis supported previous findings about the possible occurrence of admixture events that shaped modern breeds and impacted their genetic diversity (McKay et al., 2008; Edea et al., 2015). These results, showed that African cattle populations are products of genetic exchange between *Bos indicus* and *Bos taurus*. Nguni showed a significant relationship with all the other African breeds when it was compared to Afrikaner and Bonsmara, corroborating patterns observed in (**Figure** 4.4) and the PCA (**Figure** 4.5). When African breeds Ankole, Kenana and N'Dama were compared with SA breeds, Kenana revealed high excess of genetic exchange with SA breeds followed by Ankole, this relationship was supported by both D-

statistics and Treemix and it confirms the relationship between Nguni with both Kenana and Ankole (**Table** 4.2; **Figure** 4.7). This gene flow patterns showed the complexity of genetic exchange that occurred in African populations (**Table** 4.2; **Figure** 4.7).

Using the Treemix we captured 4 significant migration events, the one confirming the relationship we observed between Bonsmara and Holstein, supporting reports that Bonsmara carries both the *B. taurus* and the *B. indicus* ancestry (Bonsmara, 1980; Makina et al., 2014). Interestingly, heterogeneity signals were observed from Ankole towards, Nguni and N'Dama ancestry, however, it was more so with N'Dama, an observation that supports what Kim et al. (2017) observed. The other results that could be instrumental in the future analysis are the sign of migration events between Nguni and Kenana, both of which showed a strong relationship in (**Table** 4.2). These results could be used in understanding domestication events reported by previous studies (Pitt et al., 2018; Verdugo et al., 2019). When previous studies looked at parts of the genome, they couldn't capture the full picture of the gene flow that happened within SA breeds, however through the use of whole-genome data, possible events that occurred between SA breeds and other African breeds were captured (Makina et al., 2016). We observed that using Admixture, D-statistics and Treemix can be instrumental in uncovering evolutionary relationships. The Treemix further captured some migration events such as Holstein-X-Bonsmara, Ankole-X-N'Dama, Kenana-X-Nguni, Ankole-X-N'Dama and the relationship between the ancestries of Nguni-X-N'Dama-X-Ankole and some of the migration captured are consistent with the migration routes reported in the study by Pienaar et al. (2014) and from this it was observed that African breeds have the tendency of exchanging genetic material between each other. However, the Afrikaner breed did not show any sign of migration events in the Treemix as in (Makina et al., 2016), and it was concluded that this could be due to the missing samples from the West coast which Afrikaner shares strong signals of introgression with (unsampled or ghost lineages).

To validate the D-statistics results in avoiding biases that comes with using the close related outgroup, we sought of using a more distant breed, Water Buffalo, to check the gene flow between the breeds. Two Water Buffalo whole genomes from (Dutta et al., 2020) were used, and the results supported the dominance of Nguni and Kenana, further it revealed different patterns of Bonsmara (**Table** 4.3). The novel pattern was between NguniXBonsmara, AnkoleXBonsmara, N'DamaXAfrikaner and NguniXN'Dama. This results showed the importance of using a more distant relationships and why caution should be applied when conducting this kind of analysis (Moodley et al., 2020). D-statistics can cause biases if the out-

group is not the real out-group, meaning it is important to gather information on the out-group, its evolutionary history should be understood (Green et al., 2010; Patterson et al., 2012; Martin et al., 2014). Results obtained from the D-statistics with a Water Buffalo showed a good demonstration of how the out-group can influence the fluctuations of alleles sharing between species. The explanation for the biases and false positives can be explained as sign of introgression that is influenced by the ancestral populations structuring which is caused by incomplete lineage sorting (ILS), the contribution by populations that were not sampled, or ghost lineages as demonstrated in (Moodley et al., 2020).

This study attempted for the first time to reconstruct introgression and migration events that could have shaped modern breeds we see today (with a focus on how SA breeds are related to other African breeds) and these results are clearly indicating the need for pangenomes and sampling of more African data in order to gather more insights into the history of African breeds.

5.5 Demographic changes

Changes in climate during the Quaternary period caused an impact on species effective population sizes (*Ne*) and contributed to population evolutions which led to the process of speciation (Nadachowska-Brzyska et al., 2016). The population expansion following a decline in population sizes of all studied populations was observed during the transition period and this is one of the notable periods of the *Ne* changes in most species, an example is the study of Moose organism, and European cattle breeds (Dussex et al., 2020; Weldenegodguad et al., 2021). The period between 20 kya and 80 kya seems to be an interesting time in the history of changes in effective population sizes, including in cattle because even a comprehensive study by Wu et al. (2018), observed a dramatic decline around that time. This significant period could be one of the important times to inform geneticists about the influence humans had on the genomic composition of cattle breeds (Gautier et al., 2002). Interestingly, this is around a period when humans were spreading around the world (Nielsen et al., 2017; Wu et al., 2018). Subsequently, this period was followed by the period of domestication that affected the studied breeds differently. Around ~10 000 kya most breeds showed a decline in effective population sizes (**Figure** 4.8 to 4.12). During this period the breeds seemed to have separated from each other and were possibly exposed to different environmental conditions. The other explanation could be that during this period these breeds were possibly being managed differently (Hayes et al., 2009).

The N'Dama breed had been shown to have the lowest effective population size followed by further declines in recent times (**Figure** 4.9). The fluctuation of the N'Dama population size is consistent with previously observed bottleneck events that have been attributed to the arrival and adaptation of ancestral populations that occupied the tropical sub-humid and humid western African environments (Williamson et al., 1978; Gautier et al., 2009; Kim et al., 2017). Just like other West African breeds, The N'Dama breed was subjected to harsh environmental conditions (pathogens and parasites) (Williamson et al., 1978; Gautier et al., 2009). The Kenana and Ankole underwent significant population increase than any of the African populations, suggesting novel demographic changes they experienced at the Horn of the continent or possible human selection (Decker et al., 2014; Mwai et al., 2015).

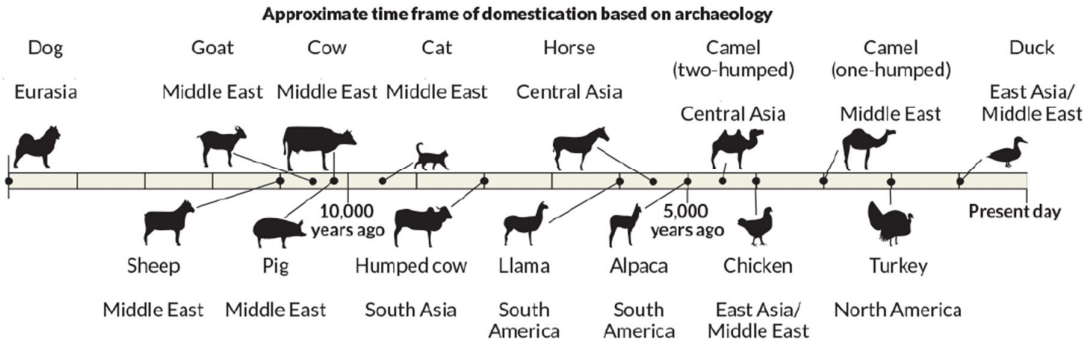

Figure 5.1: This figure is showing different domestication times for different species (including cattle) based on archaeological data (www.sciencenews.org).

Figure 5.1 This figure shows times when animals were being domesticated. The cattle timing supports our observation on the PSMC plots (**Figure** 4.8 to 4.12). Interestingly, SA breeds differed slightly in terms of their effective population sizes, with Nguni having a higher *Ne*, followed by Afrikaner, and then the composite breed Bonsmara. These results are different from Makina et al. (2015b) observations, who reported highest *Ne* for Nguni followed by Bonsmara (using SNP data and LD estimates in determining effective population size). As expected Sanga breeds (Ankole, Nguni, and Afrikaner) displayed similar trends, especially after the domestication period ~10 000 kya and we interpreted this as a possible influence of the management system of that time. For example, Nguni is well distributed across SA, and currently, there are several ecotypes such as Zulu, Shangaan, Pedi, and Venda and this could have contributed to the wide spread of this breed (Sanarana et al., 2015). Although this analysis has some limitations in terms of optimization issues e.g. 1) PSMC results are less accurate for

recent times estimates (i.e. <10ka BP) due to a few coalescence events, 2) low coverage sensitivity and biases, we still managed to capture some insights of historical events as in Weldenegodguad et al. (2021). For example, we managed to capture some insights on changes in effective population size around ~250,000 kya, a period where *Bos taurus* and the *Bos indicus* were reported to have been sharing an aurochs (*Bos primigenius*) as a common ancestor (Pitt et al., 2018). Our results could shed more light on the period where the rise of *Bos taurus* and indicine *Bos indicus* species occurred (Pitt et al., 2018). Despite having a low coverage depth of 10x, there are some similarities in the timing of the population expansion and bottlenecks events between our study and the study by Weldenegodguad et al. (2021), who also conducted this analysis on low coverage data. In future analysis programs such as MSMC & MSMC2 (Schiffels et al., 2020) and Diffusion Approximations for Demographic Inference (dadi) (Gutenkunst et al., 2009) can be used to capture recent demographic changes and help shed light on recent demographic activities.

5.6 Selection signatures

The identification of candidate genomic sites under selection has the potential of revealing genotype and phenotype relationships and an opportunity to better understand the biology of specific traits that are important for adaptation. Moreover they can be used for the development of tools for genomic selection (Moradian et al., 2019). Selection analysis was conducted to get an overview on trends of selection patterns in African indigenous cattle to lay the foundation for more studies. This also allowed the opportunity to understand how SA breeds are adapted to their environments compared to other Sanga breeds such as Ankole. Moreover, it captured the overview on selection patterns in African native breeds. To perform an accurate search for the inter-population signatures of selection in the indigenous breeds, we used the EHH-based methods iHS and XP-EHH (Pickrell et al., 2009; Sabeti et al., 2007). The comparison of breeds from different geographic locations gave us insights into the selective pressures responsible for shaping genomes of African cattle and how it led to their adaptations to harsh environmental conditions (Taye et al., 2018). Despite, focusing on genes that are associated with biological processes, we managed to capture some insightful genes. We observed that most overlapping genes from the two tests were mostly shared between Sanga breeds (**Table** 4.4 and 4.5), e.g. *KCNMA1, IGF1R, PDGFD, and DLC1,* showing consistency in terms of shared relationships, clustering patterns observed from the phylogeny (**Figure** 4.4), Treemix (**Figure** 4.7), and PCA (**Figure** 4.5). This concordance could have originated from the influence of domestication and migration of bovines (Pienaar et al., 2014). Interestingly, the XP-EHH also picked up unique

gene regions that distinguished breeds from each other, and this might be due to their evolutionary history, as observed in (**Figure** 4.8 to 4.12). Some positively selected genes were reported to be associated with reproduction, growth, milk production, meat quality, diseases, and disease resistance. For example, neurodevelopmental and autism spectrum disorders [*CNTN6* (Repnikova et al., 2020)], skeletal muscle growth and immune response [*KCNIP4* (Eydivandi et al., 2021)], meat tenderness [*APP* (Lim et al., 2016)], milk traits [*MAP4K4* (Bhattarai et al., 2017)], fertility [*CDH13* (Tarekegn et al., 2021)], stress [*PLCB4* (Mohamadipoor et al., 2021)] and antiviral [*AGO2* (Ludman et al., 2017). Genes such as *IGF1R, PDGFD, MAP4K4*, *CTNNA3*, and *DAPL1* were shared between breeds and tests (**Figure** 4.13 to 4.23).

Importantly, other genes detected from selection tests were involved in biological processes responsible for resistance to diseases, immune defence, production, and adaptation features, and this indicates a differential pattern of selection between studied breeds as observed from phenotypic depiction (**Figure** 2.3 to 3.8). The unique characteristics of indigenous breeds and their ability to withstand harsh environmental conditions are important. Additionally, the process of identifying genes in this study unveiled genes associated with response to selection pressures (Scarpa et al., 2003; Mwai et al., 2015). The primary defence mechanism of the cattle is the physical structure of the epidermal layer which protects the animal from parasitic attacks (Gautier et al., 2009; Kongsuwan et al., 2010). In line with this, *AGO2*, a gene responsible for response against viruses was identified (Brosseau et al., 2020), the *DGKB* gene, which is responsible for adaptive immunity functions in mammals (Ludman et al., 2017; Onzima et al., 2018), and the *KCNN2* gene response for metabolic stress were identified. These are some of the genes that play a critical role in response and protection of cattle against internal and external invasions (Olivieri et al., 2016).

One of our major focuses was to identify and report economically important genes (*APP, CTNNA3, PTPRQ, CSN3, ABCG2, GHR, RORA*, and *CDH13*) because cattle are important to the African economy and societal development (Kim et al., 2017). The *APP* gene, has been associated with meat tenderness (Lim et al., 2016), *PTPRQ*, in meat production, especially in beef cattle, possibly through its involvement in the regulation of the *MRF* gene expression (Robakowska-Hyżorek et al., 2016). The *CSN3, ABCG2, GHR, RORA* genes are involved in milk production (Alim et al., 2014; Nanaei et al., 2020) and *CDH13* is one of the important genes identified with association to fertility (Tarekegn et al., 2021). Understanding the nature and locations of these genes will play a role in the improvement of these breeds especially now

in the era of gene editing and advanced technologies. Strong positive selection patterns observed here could be the secret behind the resistance and adaptations of African breeds to harsh environments.

5.8 Conclusions and future perspectives

The study assessed genomic relationships of African breeds using next-generation sequencing and discovered that adopting these methods to study our populations carry the potential of generating insights into past events and can potentially shed light on the possibilities of the future. Here for the first-time genomic relationships between Afrikaner, Ankole, Nguni, Bonsmara, N'Dama and Kenana cattle breeds was reported. Population structure, introgression and gene flow analysis revealed that African breeds have complex relationships made up of several gene flow events that possibly created the wealth of diversity existing today. The demographic history analysis showed possible historical events on domestication and opened a set of questions on the role these events have played in shaping the modern cattle. Selection signature analyses unveiled signals of several biological processes that are significantly important and involved in resistance to diseases, immune defence, production, and adaptation. These genomic regions contain potential key genetic variants that may play an important role in adaption of animals to harsh environmental conditions in Africa. Future studies should include samples from other parts of Africa to further understand the nature of native breeds as genetic assets for future livestock breeding programs. These results further highlighted the importance of understanding relationships between breeds, information that is essential for conservation and maximization of production through the use of animals that are adapted to tropical production systems. Lastly, the data generated and all the information gathered in this study can be used as bases for livestock improvement, development of advance livestock genomics tools and also contribute to the overall agricultural industry for the benefit of the African society.

5.7 Limitation of this study

Biases and false-positive are common in this kind of study, therefore more samples and validation strategies might be required to further support our observations. Despite having used robust statistical methods, some of the observations could have been influenced by events that occurred in the past such as natural selection, artificial selection and demographic changes. Future studies may require sampling of both modern and ancient samples. Our sample size was

small (number populations, breeds, coverage) which could have influenced the number of variants and nucleotide bases captured and such biases and underrepresentation of data can create false calls in analysis that require complete data. Furthermore, projects working on reference assemblies that incorporate global diversity and latest variant calls algorisms are underway and using them in the future can help mitigate some biases. Genes identified here could be a result of divergence between two ancient species, and the sampling of many modern populations and comprehensive analysis can help overcome these biases and many others that we could have been missed.

References

Aken, B. L. et al. (2016) "The Ensembl gene annotation system," *Database*. Oxford University Press (OUP). doi: 10.1093/database/baw093.

Alday-Parejo, B. et al. (2020) "MAGI1, a New Potential Tumor Suppressor Gene in Estrogen Receptor Positive Breast Cancer," *Cancers*. MDPI AG. doi: 10.3390/cancers12010223.

Alders, R. G. et al. (2021) "Livestock across the world: diverse animal species with complex roles in human societies and ecosystem services," *Animal Frontiers*. Oxford University Press (OUP). doi: 10.1093/af/vfab047.

Alfarsi, L. H. et al. (2020) "PPFIA1 expression associates with poor response to endocrine treatment in luminal breast cancer," *BMC Cancer*. Springer Science and Business Media LLC. doi: 10.1186/s12885-020-06939-6.

Aliloo, H. et al. (2018) "The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa," *Journal of Dairy Science*. American Dairy Science Association. doi: 10.3168/jds.2018-14621.

Alim, M. A. et al. (2014) "Effect of polymorphisms in the CSN3 (κ-casein) gene on milk production traits in Chinese Holstein Cattle," *Molecular Biology Reports*. Springer Science and Business Media LLC. doi: 10.1007/s11033-014-3648-x.

Alshawi, A. et al. (2019) "Genome Analysis Reveals Genetic Admixture and Signature of Selection for Productivity and Environmental Traits in Iraqi Cattle," *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2019.00609.

Andersson, L. et al. (2015) "Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project," *Genome Biology*. Springer Science and Business Media LLC. doi: 10.1186/s13059-015-0622-4.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Baes, C. F. et al. (2014) "Evaluation of variant identification methods for whole genome sequencing data in dairy cattle," *BMC Genomics*. Springer Science and Business Media LLC. doi: 10.1186/1471-2164-15-948.

Bahbahani, H. et al. (2015) “Signatures of positive selection in East African Shorthorn zebu: A genome-wide single nucleotide polymorphism analysis,” *Scientific Reports*. Springer Science and Business Media LLC. doi: 10.1038/srep11729.

Baichoo, S. et al. (2018) “Developing reproducible bioinformatics analysis workflows for heterogeneous computing environments to support African genomics,” *BMC Bioinformatics*. Springer Science and Business Media LLC. doi: 10.1186/s12859-018-2446-1.

Barendse, W.J., 1999. Assessing lipid metabolism. Int. Pat. Appl. WO99/23248. *World Interrnational Properti Organization*.

Bellott, D. W. et al. (2014) “Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators,” *Nature*. Springer Science and Business Media LLC. doi: 10.1038/nature13206.

Bentley, A. R., Callier, S. and Rotimi, C. N. (2017) “Diversity and inclusion in genomic research: why the uneven progress?,” *Journal of Community Genetics*. Springer Science and Business Media LLC. doi: 10.1007/s12687-017-0316-6.

Beuzen, N. D., Stear, M. J. and Chang, K. C. (2000) “Molecular markers and their use in animal breeding,” *The Veterinary Journal*. Elsevier BV. doi: 10.1053/tvjl.2000.0468.

Bhattarai, D. et al. (2017) “Association of MAP4K4 gene single nucleotide polymorphism with mastitis and milk traits in Chinese Holstein cattle,” *Journal of Dairy Research*. Cambridge University Press (CUP). doi: 10.1017/s0022029916000832.

Bignardi, A. B. et al. (2014) “Models for genetic evaluation of growth of Brazilian Bonsmara cattle,” *Livestock Science*. Elsevier BV. doi: 10.1016/j.livsci.2014.01.031.

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

Bonsma, J.C., (1980). Cross-breeding, breed creation and the genesis of the Bonsmara. Livestock production: A global approach. Cape Town: *Tafelberg Publishers Ltd*, pp.126-3.

Bosman, L. et al. (2017) “Short communication: Population structure of the South African Bonsmara beef breed using high density single nucleotide polymorphism genotypes,” *Livestock Science*. Elsevier BV. doi: 10.1016/j.livsci.2017.01.012.

Bostan, H. and Chiusano, M. (2015) "NexGenEx-Tom: a gene expression platform to investigate the functionalities of the tomato genome," *BMC Plant Biology*. Springer Science and Business Media LLC. doi: 10.1186/s12870-014-0412-2.

Bovine HapMap Consortium (2009) "Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds". *Science*. doi: 10.1126/science.1167936.

Bradley, D. G. et al. (1996) "Mitochondrial diversity and the origins of African and European cattle," *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.93.10.5131.

Brosseau, C. et al. (2020) "Natural variation in the Arabidopsis AGO2 gene is associated with susceptibility to potato virus X," *New Phytologist*. Wiley. doi: 10.1111/nph.16397.

Browning, B. L. et al. (2021) "Fast two-stage phasing of large-scale sequence data," *The American Journal of Human Genetics*. Elsevier BV. doi: 10.1016/j.ajhg.2021.08.005.

Buchmann, R. and Hazelhurst, S., (2014) "Genesis manual". *Johannesburg: University of the Witwatersrand*.

Buzanskas, M. E. et al. (2014) "Genome-Wide Association for Growth Traits in Canchim Beef Cattle," *PLoS ONE*. Edited by S. Zhao. Public Library of Science (PLoS). doi: 10.1371/journal.pone.0094802.

Chen, H. et al. (2021) "DAPL1 is a novel regulator of testosterone production in Leydig cells of mouse testis," *Scientific Reports*. Springer Science and Business Media LLC. doi: 10.1038/s41598-021-97961-6.

Chen, N. et al. (2020) "BGVD: An Integrated Database for Bovine Sequencing Variations and Selective Signatures," *Genomics, Proteomics & Bioinformatics*. Elsevier BV. doi: 10.1016/j.gpb.2019.03.007.

Chen, Q. et al. (2020) "Whole genome analyses revealed genomic difference between European taurine and East Asian taurine," *Journal of Animal Breeding and Genetics*. Wiley. doi: 10.1111/jbg.12501.

Cheruiyot, E. K. et al. (2018) "Signatures of Selection in Admixed Dairy Cattle in Tanzania," *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2018.00607.

Chiusano, M. L. et al. (2008) "ISOL@: an Italian SOLAnaceae genomics resource," *BMC Bioinformatics*. Springer Science and Business Media LLC. doi: 10.1186/1471-2105-9-s2-s7.

D'ieteren, G. D. M. et al. (1998) "Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis," *Revue Scientifique et Technique de l'OIE*. O.I.E (World Organisation for Animal Health). doi: 10.20506/rst.17.1.1088.

Dagris.ilri.cgiar.org. (2005). Welcome to DAGRIS | DAGRIS. [online] Available at: <http://dagris.ilri.cgiar.org> [Accessed 11 July 2022].

Dalloul, R. A. et al. (2010) "Multi-Platform Next-Generation Sequencing of the Domestic Turkey (Meleagris gallopavo): Genome Assembly and Analysis," *PLoS Biology*. Edited by R. J. Roberts. Public Library of Science (PLoS). doi: 10.1371/journal.pbio.1000475.

Danecek, P. et al. (2011) "The variant call format and VCFtools," *Bioinformatics*. Oxford University Press (OUP). doi: 10.1093/bioinformatics/btr330.

Decker, J. E. et al. (2014) "Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle," *PLoS Genetics*. Edited by G. McVean. Public Library of Science (PLoS). doi: 10.1371/journal.pgen.1004254.

Dennis, G., Jr et al. (2003) "DAVID: Database for Annotation, Visualization, and Integrated Discovery," *Genome Biology*. Springer Science and Business Media LLC. doi: 10.1186/gb-2003-4-9-r60.

DePristo, M. A. et al. (2011) "A framework for variation discovery and genotyping using next-generation DNA sequencing data," *Nature Genetics*. Springer Science and Business Media LLC. doi: 10.1038/ng.806.

Dessie, T. and Mwai, A, O. (2019) "The story of cattle in Africa: Why diversity matters". ILRI (aka ILCA and ILRAD).

Ding, C. et al. (2021) "ROBO2 hampers malignant biological behavior and predicts a better prognosis in pancreatic adenocarcinoma," *Scandinavian Journal of Gastroenterology*. Informa UK Limited. doi: 10.1080/00365521.2021.1930144.

Dixit, S. P. et al. (2021) "Genome analyses revealed genetic admixture and selection signatures in Bos indicus," *Scientific Reports*. Springer Science and Business Media LLC. doi: 10.1038/s41598-021-01144-2.

Dong, L. et al. (2022) "Quantification and prediction of enteric methane emissions from Chinese lactating Holstein dairy cows fed diets with different dietary neutral detergent fiber/non-fibrous carbohydrate (NDF/NFC) ratios," *Journal of Integrative Agriculture*. Elsevier BV. doi: 10.1016/s2095-3119(21)63825-x.

Dong, Y. et al. (2012) "Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus)," *Nature Biotechnology*. Springer Science and Business Media LLC. doi: 10.1038/nbt.2478.

Durand, E. Y. et al. (2011) "Testing for Ancient Admixture between Closely Related Populations," *Molecular Biology and Evolution*. Oxford University Press (OUP). doi: 10.1093/molbev/msr048.

Dussex, N. et al. (2020) "Moose genomes reveal past glacial demography and the origin of modern lineages," *BMC Genomics*. Springer Science and Business Media LLC. doi: 10.1186/s12864-020-07208-3.

Dutta, P. et al. (2020) "Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication," *Nature Communications*. Springer Science and Business Media LLC. doi: 10.1038/s41467-020-18550-1.

Edea, Z. 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*. Frontiers Media SA. doi: 10.3389/fgene.2013.00035.

Edea, Z. et al. (2015) "Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds," *Animal. Elsevier BV*. doi: 10.1017/s1751731114002560.

Eggen, A. (2012) "The development and application of genomic selection as a new breeding paradigm," *Animal Frontiers*. Oxford University Press (OUP). doi: 10.2527/af.2011-0027.

Ekizoglu, S. et al. (2018) "Investigation of the SLC22A23 gene in laryngeal squamous cell carcinoma," *BMC Cancer*. Springer Science and Business Media LLC. doi: 10.1186/s12885-018-4381-y.

Epstein, H., (1971). "The origin of the domestic animals of Africa". https://cgspace.cgiar.org/handle/10568/70619.

Esposito, A. et al. (2016) “Bioinformatics for agriculture in the Next-Generation sequencing era,” *Chemical and Biological Technologies in Agriculture*. Springer Science and Business Media LLC. doi: 10.1186/s40538-016-0054-8.

Eydivandi, S. et al. (2021) “Genomic scans for selective sweeps through haplotype homozygosity and allelic fixation in 14 indigenous sheep breeds from Middle East and South Asia,” *Scientific Reports*. Springer Science and Business Media LLC. doi: 10.1038/s41598-021-82625-2.

FAO. (2007). “The state of the world's animal genetic resources for food and agriculture” *Food & Agriculture Org*. http://www.fao.org/3/a1250e/a1250e00.htm.

FAO. (2021). (https://www.fao.org/3/ah834e/ah834e02.pdf).

Felsenstein, J., (1993). “PHYLIP (phylogeny inference package”, *version 3.5 c*. Joseph Felsenstein.

Ferretti, L., Ribeca, P. and Ramos-Onsins, S. E. (2018) “The Site Frequency/Dosage Spectrum of Autopolyploid Populations,” *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2018.00480.

Fisher, R.A. (1930) “The Genetical Theory of Natural Selection”. *Clarendon Press, Oxford*. http://dx.doi.org/10.5962/bhl.title.27468.

Flori, L. et al. (2019) “A genomic map of climate adaptation in Mediterranean cattle breeds,” *Molecular Ecology*. Wiley. doi: 10.1111/mec.15004.

Fortes, M. R. S. et al. (2013) “The IGF1 pathway genes and their association with age of puberty in cattle,” *Animal Genetics*. Wiley. doi: 10.1111/j.1365-2052.2012.02367. x.

Francis, R. M. (2016) “<scp>pophelper</scp>: an R package and web app to analyse and visualize population structure,” *Molecular Ecology Resources*. Wiley. doi: 10.1111/1755-0998.12509.

Garcia, J. F. et al. (2012) “How Bioinformatics Enables Livestock Applied Sciences in the Genomic Era,” *Advances in Bioinformatics and Computational Biology*. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-31927-3_17.

Gautier, A. (2002) “The Evidence for the Earliest Livestock in North Africa: or Adventures with Large Bovids, Ovicaprids, Dogs and Pigs,” *Droughts, Food and Culture*. Kluwer Academic Publishers. doi: 10.1007/0-306-47547-2_12.

Gautier, M. et al. (2009) “A whole genome Bayesian scan for adaptive genetic divergence in West African cattle,” BMC Genomics. Springer Science and Business Media LLC. doi: 10.1186/1471-2164-10-550.

Gautier, M., Klassmann, A. and Vitalis, R. (2016) “<scp>rehh</scp>2.0: a reimplementation of the R package<scp>rehh</scp>to detect positive selection from haplotype structure,” *Molecular Ecology Resources*. Wiley. doi: 10.1111/1755-0998.12634.

Ghosh, M. et al. (2018) “Transformation of animal genomics by next-generation sequencing technologies: a decade of challenges and their impact on genetic architecture,” *Critical Reviews in Biotechnology*. Informa UK Limited. doi: 10.1080/07388551.2018.1451819.

Gómez Hernández, G., Morell, M. and Alarcón-Riquelme, M. E. (2021) “The Role of BANK1 in B Cell Signaling and Disease,” *Cells*. MDPI AG. doi: 10.3390/cells10051184.

Grapes, L. et al. (2004) “Comparing Linkage Disequilibrium-Based Methods for Fine Mapping Quantitative Trait Loci,” *Genetics*. Oxford University Press (OUP). doi: 10.1534/genetics.166.3.1561.

Green, R. E. et al. (2010) “A Draft Sequence of the Neandertal Genome,” *Science*. American Association for the Advancement of Science (AAAS). doi: 10.1126/science.1188021.

Groeneveld, L. F. et al. (2010) “Genetic diversity in farm animals - a review,” *Animal Genetics*. Wiley. doi: 10.1111/j.1365-2052.2010.02038. x.

Gutenkunst, R. N. et al. (2009) “Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data,” *PLoS Genetics*. Edited by G. McVean. Public Library of Science (PLoS). doi: 10.1371/journal.pgen.1000695.

Habier, D., Fernando, R. L. and Dekkers, J. C. M. (2007) “The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values,” *Genetics*. Oxford University Press (OUP). doi: 10.1534/genetics.107.081190.

Hanotte, O. and Jianlin, H., (2005) “Genetic characterization of livestock populations and its use in conservation decision-making”. *The role of biotechnology*.

Hanotte, O. 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*. Wiley. doi: 10.1046/j.1365-294x.2000.00858. x.

Hanotte, O. et al. (2002) “African Pastoralism: Genetic Imprints of Origins and Migrations,” *Science*. *American Association for the Advancement of Science (AAAS)*. doi: 10.1126/science.1069878.

Harwatt, H. (2018) “Including animal to plant protein shifts in climate change mitigation policy: a proposed three-step strategy,” *Climate Policy*. Informa UK Limited. doi: 10.1080/14693062.2018.1528965.

Hayes, B et al. (2012) “1000 bull genomes consortium project”. In *Plant and Animal Meeting*.

Hayes, B. J. et al. (2009) “Accuracy of genomic breeding values in multi-breed dairy cattle populations,” *Genetics Selection Evolution*. Springer Science and Business Media LLC. doi: 10.1186/1297-9686-41-51.

Herring, A.D et al. (2013)” Beef cattle genomics: Promises from the past, looking to the future”. *62nd Annual Florida Beef Cattle Short Course*.

Hocquette, J.-F. et al. (2007) “Recent advances in cattle functional genomics and their application to beef quality,” *Animal*. Elsevier BV. doi: 10.1017/s1751731107658042.

Hoffmann, I. (2010) “Climate change and the characterization, breeding and conservation of animal genetic resources,” *Animal Genetics*. Wiley. doi: 10.1111/j.1365-2052.2010.02043. x.

Horsburgh, K. A. et al. (2013) “The Genetic Diversity of the Nguni Breed of African Cattle (Bos spp.): Complete Mitochondrial Genomes of Haplogroup T1,” *PLoS ONE*. Edited by G. Larson. Public Library of Science (PLoS). doi: 10.1371/journal.pone.0071956.

Huang, J. et al. (2015) “Donkey genome and insight into the imprinting of fast karyotype evolution,” *Scientific Reports*. Springer Science and Business Media LLC. doi: 10.1038/srep14106.

Hughes, A. R. et al. (2008) “Ecological consequences of genetic diversity,” *Ecology Letters*. Wiley. doi: 10.1111/j.1461-0248.2008.01179. x.

Hwang, S. et al. (2015) “Systematic comparison of variant calling pipelines using gold standard personal exome variants,” *Scientific Reports*. Springer Science and Business Media LLC. doi: 10.1038/srep17875.

Illumina, (2022) “Agrigenomics Research: An Overview of Recent Agrigenomics Research Publications Featuring Illumina® Technology”: Agrigenomics Research (illumina.com): (Accessed: 26 January 2022).

Illumina, (2022) “Genome Sequencing in Agriculture and the Environment”. Available at: Plant and Animal Sequencing | Genome sequencing in agriculture (illumina.com) (Accessed: 26 January 2022).

Ito, S. and Nishino, T. (2021) “Structural analysis of the chicken FANCM–MHF complex and its stability,” *Acta Crystallographica Section F Structural Biology Communications*. International Union of Crystallography (IUCr). doi: 10.1107/s2053230x20016003.

Jahnke, H.E et al. (1988) “Livestock production in tropical Africa, with special reference to the tsetse-affected zone”. Livestock production in tsetse-affected areas of Africa, pp.3-21.

Jiang, Y. et al. (2014) “The sheep genome illuminates biology of the rumen and lipid metabolism,” *Science*. American Association for the Advancement of Science (AAAS). doi: 10.1126/science.1252806.

Kampen, K. R. et al. (2015) “EphB1 Suppression in Acute Myelogenous Leukemia: Regulating the DNA Damage Control System,” *Molecular Cancer Research*. American Association for Cancer Research (AACR). doi: 10.1158/1541-7786.mcr-14-0660-t.

Kanzi, A. M. et al. (2020) “Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance,” *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2020.544162.

Kim, J. et al. (2017) “The genome landscape of indigenous African cattle,” Genome Biology. Springer Science and Business Media LLC. doi: 10.1186/s13059-017-1153-y.

Kim, K. et al. (2020) “The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism,” *Nature Genetics*. Springer Science and Business Media LLC. doi: 10.1038/s41588-020-0694-2.

Kim, S.-J. et al. (2017b) “Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N’Dama,” *BMC Genomics*. Springer Science and Business Media LLC. doi: 10.1186/s12864-017-3742-2.

Kongsuwan, K. et al. (2010) “Activation of several key components of the epidermal differentiation pathway in cattle following infestation with the cattle tick, Rhipicephalus (Boophilus) microplus,” *International Journal for Parasitology*. Elsevier BV. doi: 10.1016/j.ijpara.2009.10.013.

Korneliussen, T. S., Albrechtsen, A. and Nielsen, R. (2014) “ANGSD: Analysis of Next Generation Sequencing Data,” *BMC Bioinformatics*. Springer Science and Business Media LLC. doi: 10.1186/s12859-014-0356-4.

Kośla, K., Kałuzińska, Ż. and Bednarek, A. (2020) “The WWOX gene in brain development and pathology,” *Experimental Biology and Medicine*. SAGE Publications. doi: 10.1177/1535370220924618.

Kumar, S. and Subramanian, S. (2002) “Mutation rates in mammalian genomes,” *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.022629899.

Laikre, L et al. (2010) “Neglect of genetic diversity in implementation of the convention of biological diversity”. *Conservation Biology*, Wiley for Society for Conservation Biology. doi: 10.HH/j.1523-1739.2009.0l425.

Laulier, C. et al. (2010) “Mammalian Fbh1 is important to restore normal mitotic progression following decatenation stress,” *DNA Repair*. Elsevier BV. doi: 10.1016/j.dnarep.2010.03.011.

Lee, J.W. et al. (2004) “Hypoxia-inducible factor (HIF-1)α: its protein stability and biological functions,” *Experimental & Molecular Medicine*. Springer Science and Business Media LLC. doi: 10.1038/emm.2004.1.

Leipzig, J. (2016) “A review of bioinformatic pipeline frameworks,” *Briefings in Bioinformatics*. Oxford University Press (OUP). doi: 10.1093/bib/bbw020.

Li, H. and Durbin, R. (2009) “Fast and accurate short read alignment with Burrows-Wheeler transform,” *Bioinformatics*. Oxford University Press (OUP). doi: 10.1093/bioinformatics/btp324.

Li, H. and Durbin, R. (2010) "Fast and accurate long read alignment with Burrows-Wheeler transform," Bioinformatics. Oxford University Press (OUP). doi: 10.1093/bioinformatics/btp4432.

Li, H. and Durbin, R. (2011) "Inference of human population history from individual whole-genome sequences," *Nature*. Springer Science and Business Media LLC. doi: 10.1038/nature10231.

Li, S. et al. (2017) "ZMPSTE24 Is Downstream Effector of Interferon-Induced Transmembrane Antiviral Activity," *DNA and Cell Biology*. Mary Ann Liebert Inc. doi: 10.1089/dna.2017.3791.

Lim, D. et al. (2016) "Genomic Footprints in Selected and Unselected Beef Cattle Breeds in Korea," *PLOS ONE*. Edited by G. A. Brockmann. Public Library of Science (PLoS). doi: 10.1371/journal.pone.0151324.

Lin, B. Z., Sasazaki, S. and Mannen, H. (2010) "Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers," *Animal Science Journal*. Wiley. doi: 10.1111/j.1740-0929.2010.00744. x.

Lindblad-Toh, K. et al. (2005) "Genome sequence, comparative analysis and haplotype structure of the domestic dog," *Nature*. Springer Science and Business Media LLC. doi: 10.1038/nature04338.

Liu, Y. et al. (2009) "Bos taurus genome assembly," *BMC Genomics*. Springer Science and Business Media LLC. doi: 10.1186/1471-2164-10-180.

Livestock Industry. (2022). Reportlinker.com. 2022. [online] Available at: <https://www.reportlinker.com/market-report/Livestock/8/Livestock> [Accessed 11 July 2022].

Low, W.Y et al (2019) "Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity". *Nature Communications*. doi: s41467-020-18550-1

Ludman, M., Burgyán, J. and Fátyol, K. (2017) "Crispr/Cas9 Mediated Inactivation of Argonaute 2 Reveals its Differential Involvement in Antiviral Responses," *Scientific Reports*. Springer Science and Business Media LLC. doi: 10.1038/s41598-017-01050-6.

Ma, Y. (2002) “Molecular characterization of the TCP11 gene which is the human homologue of the mouse gene encoding the receptor of fertilization promoting peptide,” *Molecular Human Reproduction*. Oxford University Press (OUP). doi: 10.1093/molehr/8.1.24.

MacHugh, D. E. et al. (1997) “Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (Bos taurus and Bos indicus),” *Genetics*. Oxford University Press (OUP). doi: 10.1093/genetics/146.3.1071.

Maciel, S. et al. (2014) “The effect of geographical region of birth on the reproductive performance of the Nguni in southern Mozambique,” *South African Journal of Animal Science*. African Journals Online (AJOL). doi: 10.4314/sajas.v43i5.11.

Mackay, T. F. C., Stone, E. A. and Ayroles, J. F. (2009) “The genetics of quantitative traits: challenges and prospects,” *Nature Reviews Genetics*. Springer Science and Business Media LLC. doi: 10.1038/nrg2612.

MacLeod, I. M. et al. (2013) “Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors,” *Molecular Biology and Evolution*. Oxford University Press (OUP). doi: 10.1093/molbev/mst125.

Madella, M. et al. (2014) “Microbotanical Evidence of Domestic Cereals in Africa 7000 Years Ago,” *PLoS ONE*. Edited by X. Yang. Public Library of Science (PLoS). doi: 10.1371/journal.pone.0110177.

Madilindi, M. A. et al. (2019) “Genetic diversity and relationships among three Southern African Nguni cattle populations,” *Tropical Animal Health and Production*. Springer Science and Business Media LLC. doi: 10.1007/s11250-019-02066-y.

Maiwashe, A. N. et al. (2002) “Genetic parameter estimates for body measurements and growth traits in South African Bonsmara cattle,” *Livestock Production Science*. Elsevier BV. doi: 10.1016/s0301-6226(01)00324-4.

Makina, S. O. et al. (2014) “Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel,” *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2014.00333.

Makina, S. O. et al. (2015) “Genome-wide scan for selection signatures in six cattle breeds in South Africa,” *Genetics Selection Evolution*. Springer Science and Business Media LLC. doi: 10.1186/s12711-015-0173-x.

Makina, S. O. et al. (2015b) “Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds,” *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2015.00337.

Makina, S. O. et al. (2016) “Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide,” *Genetics Selection Evolution*. Springer Science and Business Media LLC. doi: 10.1186/s12711-016-0266-1.

Mandela, P. et al. (2012) “Kalrnplays key roles within and outside of the nervous system,” *BMC Neuroscience*. Springer Science and Business Media LLC. doi: 10.1186/1471-2202-13-136.

Mariani, M.A. et al. (2016) “Parallelization of bioinformatics methods using a computational grid”, *South American Journal of Academic Research*: Special Edition.

Martin, M. (2011) “Cutadapt removes adapter sequences from high-throughput sequencing reads,” *EMBnet.journal*. EMBnet Stichting. doi: 10.14806/ej.17.1.200.

Martin, S. H., Davey, J. W. and Jiggins, C. D. (2014) “Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci,” *Molecular Biology and Evolution*. Oxford University Press (OUP). doi: 10.1093/molbev/msu269.

Mattioli, R. C. et al. (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*. Springer Science and Business Media LLC. doi: 10.1007/bf02236320.

Mattioli, R. C. et al. (2002) “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*. Elsevier BV. doi: 10.1016/s0001-706x(00)00063-2.

Mauki, D. H. et al. (2022) “Genome-wide investigations reveal the population structure and selection signatures of Nigerian cattle adaptation in the sub-Saharan tropics,” *BMC Genomics*. Springer Science and Business Media LLC. doi: 10.1186/s12864-022-08512-w.

Mbimba, T. et al. (2018) “TRAPPC9: Novel insights into its trafficking and signaling pathways in health and disease (Review),” *International Journal of Molecular Medicine*. Spandidos Publications. doi: 10.3892/ijmm.2018.3889.

McKay, S. D. et al. (2008) "An assessment of population structure in eight breeds of cattle using a whole genome SNP panel," *BMC Genetics*. Springer Science and Business Media LLC. doi: 10.1186/1471-2156-9-37.

Mohamadipoor Saadatabadi, L. et al. (2021) "Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds," *BMC Veterinary Research*. Springer Science and Business Media LLC. doi: 10.1186/s12917-021-03077-4.

Monda, T., (2017) *"Ankole cattle breed ...the cattle of kings | Celebrating Being Zimbabwean".* [online] *Thepatriot.co.zw.* Available at: <https://www.thepatriot.co.zw/old_posts/ankole-cattle-breed-the-cattle-of-kings/> [Accessed 15 September 2022].

Moradian, H. et al. (2019) "Whole genome detection of recent selection signatures in Sarabi cattle: a unique Iranian taurine breed," Genes & Genomics. Springer Science and Business Media LLC. doi: 10.1007/s13258-019-00888-6.

Morgado, F.P. (2004) "A Pecuária no Centro de Moçambique", *Anais dos Serviços de Veterinária e Industria Animal.* Nova Vega 1ª edição.

Moodley, Y. (2020) "Interspecific gene flow and the evolution of specialization in black and white rhinoceros," *Molecular biology and evolution*. doi: 10.1093/molbev/msaa148

Mukasa-Mugerwa, E. (1989) "A Review of Reproductive Performance of Female Bos indicus (Zebu) Cattle". *Monograph* No. 6, ILCA, Addis Ababa, Ethiopia, 101-104.

Mukhopadhyay, C. S., Kumar, A. and Deb, R. (2020) "Cattle genomics: genome projects, current status, and future applications," *Genomics and Biotechnological Advances in Veterinary, Poultry, and Fisheries*. Elsevier. doi: 10.1016/b978-0-12-816352-8.00001-1.

Murray, C. et al. (2010) "Cattle demographic history modelled from autosomal sequence variation," *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society. doi: 10.1098/rstb.2010.0103.

Mwacharo, J. M. et al. (2006) "The small East African shorthorn zebu cows in Kenya. I: Linear body measurements," *Tropical Animal Health and Production*. Springer Science and Business Media LLC. doi: 10.1007/s11250-006-4266-y.

Mwai, O. et al. (2015) "- Invited Review - African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World," *Asian-Australasian Journal of Animal Sciences*. Asian Australasian Association of Animal Production Societies. doi: 10.5713/ajas.15.0002r.

Nadachowska-Brzyska, K. et al. (2016) "<scp>PSMC</scp>analysis of effective population sizes in molecular ecology and its application to black-and-whiteFicedulaflycatchers," Molecular Ecology. Wiley. doi: 10.1111/mec.13540.

Nanaei, H.A et al. (2020) "Whole-genome resequencing reveals selection signatures associated with milk production traits in African Kenana dairy zebu cattle". *Genomics*. Elsevier BV. doi: 10.1016/j.ygeno.2019.06.002.

Nicholson, A. C. et al. (2005) "Functional evolution of ADAMTS genes: Evidence from analyses of phylogeny and gene organization," *BMC Evolutionary Biology*. Springer Science and Business Media LLC. doi: 10.1186/1471-2148-5-11.

Nielsen, R. et al. (2017) "Tracing the peopling of the world through genomics," *Nature*. Springer Science and Business Media LLC. doi: 10.1038/nature21347.

Notter, D. R. (1999) "The importance of genetic diversity in livestock populations of the future.," *Journal of Animal Science*. Oxford University Press (OUP). doi: 10.2527/1999.77161x.

Ntshepe, L. (2011) "Marketing information needs of smallholder livestock farmers in the Moretele area in the Bojanala Platinum district municipality of the North-West province". MSc Thesis. University of South Africa, Pretoria.

Nyamushamba, G. B. et al. (2016) "Conservation of indigenous cattle genetic resources in Southern Africa's smallholder areas: turning threats into opportunities — A review," *Asian-Australasian Journal of Animal Sciences*. Asian Australasian Association of Animal Production Societies. doi: 10.5713/ajas.16.0024.

Olivieri, B. F. et al. (2016) "Genomic Regions Associated with Feed Efficiency Indicator Traits in an Experimental Nellore Cattle Population," *PLOS ONE*. Edited by R. N. PENA i SUBIRÀ. Public Library of Science (PLoS). doi: 10.1371/journal.pone.0164390.

Onzima, R.B et al. (2018) “Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds”. *Frontiers in Genetics*. Frontiers Media. 10.3389/fgene.2018.0031.

Palkopoulou, E. et al. (2015) “Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth,” *Current Biology*. Elsevier BV. doi: 10.1016/j.cub.2015.04.007.

Pathania, M. et al. (2014) “The autism and schizophrenia associated gene CYFIP1 is critical for the maintenance of dendritic complexity and the stabilization of mature spines,” *Translational Psychiatry*. Springer Science and Business Media LLC. doi: 10.1038/tp.2014.16.

Patterson, N. et al. (2012) “Ancient Admixture in Human History,” *Genetics*. Oxford University Press (OUP). doi: 10.1534/genetics.112.145037.

Payne, W.J.A. (1964). “Origin of domestic cattle in Africa”. *Empire Journal of Experimental Agriculture*, 32(126), p.97.

Payne, W.J.A. and Hodges, J. (1997) “Tropical cattle: origins, breeds and breeding policies”. Blackwell Science Ltd.

Pereira, G.L. et al. (2013) “State of the art of genomic sequencing in livestock”. *Ars Veterinaria*, Jaboticabal, SP, v.29.

Picard, T. (2018). Broad institute, GitHub repository.

Pickrell, J. and Pritchard, J. (2012) “Inference of population splits and mixtures from genome-wide allele frequency data,” *Nature Precedings*. Springer Science and Business Media LLC. doi: 10.1038/npre.2012.6956.1.

Pienaar, L. (2014) “Genetic diversity in the Afrikaner cattle breed*”.*Doctoral dissertation, University of the Free State.

Pitt, D. et al. (2018) “Domestication of cattle: Two or three events?,” *Evolutionary Applications*. Wiley. doi: 10.1111/eva.12674.

Plasschaert, R. N. and Bartolomei, M. S. (2014) “Tissue-specific regulation and function of Grb10 during growth and neuronal commitment,” *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.1411254111.

Plewes, M. R. et al. (2019) "Yes-associated protein 1 is required for proliferation and function of bovine granulosa cells in vitro†," *Biology of Reproduction*. Oxford University Press (OUP). doi: 10.1093/biolre/ioz139.

Poplin, R. et al. (2017) "Scaling accurate genetic variant discovery to tens of thousands of samples." Cold Spring Harbor Laboratory. doi: 10.1101/201178.

Pugach, I. and Stoneking, M. (2015) "Genome-wide insights into the genetic history of human populations," *Investigative Genetics*. Springer Science and Business Media LLC. doi: 10.1186/s13323-015-0024-0.

Purcell, S. et al. (2007) "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses," *The American Journal of Human Genetics*. Elsevier BV. doi: 10.1086/519795.

Qiu, Q. et al. (2012) "The yak genome and adaptation to life at high altitude," *Nature Genetics*. Springer Science and Business Media LLC. doi: 10.1038/ng.2343.

Qwabe, S. et al. (2013) "Short communication: Evaluation of the BovineSNP50 genotyping array in four South African cattle populations," *South African Journal of Animal Science*. African Journals Online (AJOL). doi: 10.4314/sajas.v43i1.7.

Rahway, N.J. (1985) "Genus Bos: Cattle Breeds of the World. MSO-AGVET". Merck Co., Inc. Available online: http://afs. okstate. edu/breeds/cattle/tswana/(accessed on 11 December 2018.

Rambaut, A. (2009) "FigTree. Tree figure drawing tool". *http://tree. bio. ed. ac. uk/software/figtree/*.

Ramos, A. M. 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*. Edited by L. Orban. Public Library of Science (PLoS). doi: 10.1371/journal.pone.0006524.

Rege, J. E. O. and Tawah, C. L. (1999) "The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains," *Animal Genetic Resources Information*. Cambridge University Press (CUP). doi: 10.1017/s1014233900001152.

Repnikova, E. A. et al. (2020) “CNTN6 copy number variations: Uncertain clinical significance in individuals with neurodevelopmental disorders,” *European Journal of Medical Genetics*. Elsevier BV. doi: 10.1016/j.ejmg.2019.02.008.

Rewe, T. O. et al. (2009) “Breeding Indigenous Cattle Genetic Resources for Beef Production in Sub-Saharan Africa,” *Outlook on Agriculture*. SAGE Publications. doi: 10.5367/000000009790422205.

Robakowska-Hyżorek, D. et al. (2016) “Effects of polymorphism in the bovine PTPRQ gene on the expression of MYF6 and MYF5 genes in skeletal muscle and on meat production traits in beef bulls,” *Czech Journal of Animal Science*. Czech Academy of Agricultural Sciences. doi: 10.17221/118/2014-cjas.

Rosegrant, M. W. et al. (2009) “Looking into the future for agriculture and AKST”, *International Assessment of Agricultural Knowledge, Science and Technology for Development (IAASTD)*: Agriculture at a Crossroads, global report. Washington, DC, USA: Island Press. pp.307-376.

Rosen, B.D. et al. (2018) “Modernizing the bovine reference genome assembly”. In *Proceedings of the 11th world congress on genetics applied to livestock production* (pp. 11-16).

Saeed, A.M. (1987) “Characterisation of Kenana cattle at Um Banein”, *Sudan (Vol. 16)*. ILRI (aka ILCA and ILRAD).

Saitou, N. and Nei, M. (1987) “The neighbor-joining method: a new method for reconstructing phylogenetic trees”. *Molecular biology and evolution*.

Salmon, G. R. et al. (2020) “Exploring the landscape of livestock ‘Facts,’” *Global Food Security*. Elsevier BV. doi: 10.1016/j.gfs.2019.100329.

Sanarana, Y. et al. (2015) “Genetic diversity in South African Nguni cattle ecotypes based on microsatellite markers,” *Tropical Animal Health and Production*. Springer Science and Business Media LLC. doi: 10.1007/s11250-015-0962-9.

Scarpa, R. et al. (2003) “Valuing indigenous cattle breeds in Kenya: an empirical comparison of stated and revealed preference value estimates,” *Ecological Economics*. Elsevier BV. doi: 10.1016/s0921-8009(03)00094-6.

Schiffels, S. and Wang, K. (2020) “MSMC and MSMC2: the multiple sequentially markovian coalescent”. In *Statistical population genomics*. Humana, New York, NY.

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

Scholtz, M. M. and Theunissen, A. (2010) “The use of indigenous cattle in terminal cross-breeding to improve beef cattle production in Sub-Saharan Africa,” *Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales*. Cambridge University Press (CUP). doi: 10.1017/s2078633610000676.

Scholtz, M.M. (2010) “Beef Breeding in South Africa”. 2nd ed. Pretoria: Agricultural Research Council.

Scholtz, M.M. (2011) “The Nguni breed of cattle: past, present and future”. *The Nguni cattle Breeder’s Society,* Pretoria, South Africa.

Scholtz, M.M. et al. (2016) “The Afrikaner breed of cattle: South Africa’s heritage for food security”. 1st ed. Pretoria: Agricultural Research Council.

Sharawat, I., Panda, P. and Lourembam, R. (2021) “CNNM2 heterozygous variant presenting as hypomagnesemia and west syndrome: Expanding the spectrum of cnnm2 gene-Related epileptic disorders,” *Annals of Indian Academy of Neurology*. Medknow. doi: 10.4103/aian.aian_1130_20.

Shi, L. et al. (2019) “A post-GWAS confirming effects of PRKG1 gene on milk fatty acids in a Chinese Holstein dairy population,” *BMC Genetics*. Springer Science and Business Media LLC. doi: 10.1186/s12863-019-0755-7.

Shoshan-Barmatz, V., Krelin, Y. and Shteinfer-Kuzmine, A. (2018) “VDAC1 functions in Ca2+ homeostasis and cell life and death in health and disease,” *Cell Calcium*. Elsevier BV. doi: 10.1016/j.ceca.2017.06.007.

Spickett, A.M. et al. (1989) “Resistance of Nguni, Bonsmara and Hereford cattle to ticks in a bushveld region of South Africa”. *University of Pretoria*. repository.up.ac.za

Stanford University. (2010) "Environmental and social impact of the 'livestock revolution'." ScienceDaily. *ScienceDaily*. <www.sciencedaily.com/releases/2010/03/100316101703.htm>.

Steinfeld, H. (2006) “Livestock's long shadow: environmental issues and options”. *Food & Agriculture Org*. Books.google.com

Stothard, P. et al. (2011) “Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery,” *BMC Genomics*. Springer Science and Business Media LLC. doi: 10.1186/1471-2164-12-559.

Talenti, A. et al. (2022) “A cattle graph genome incorporating global breed diversity,” *Nature Communications*. Springer Science and Business Media LLC. doi: 10.1038/s41467-022-28605-0.

Tarekegn, G. M. et al. (2021) “Single-step genome-wide association study uncovers known and novel candidate genomic regions for endocrine and classical fertility traits in Swedish Red and Holstein dairy cows*,” Livestock Science*. Elsevier BV. doi: 10.1016/j.livsci.2021.104731.

Taye, M. (2018) “Exploring the genomes of East African Indicine cattle breeds reveals signature of selection for tropical environmental adaptation traits”. *Cogent Food & Agriculture*. Tandfonline. doi: 10.1080/23311932.2018.1552552.

Taye, M. et al. (2017) “Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef,” *BMC Genetics*. Springer Science and Business Media LLC. doi: 10.1186/s12863-016-0467-1.

Tellam, R. L. et al. (2009) “Unlocking the bovine genome,” *BMC Genomics*. Springer Science and Business Media LLC. doi: 10.1186/1471-2164-10-193.

Telugu, B. P., Park, K. E. and Park, C. H. (2017) “Genome editing and genetic engineering in livestock for advancing agricultural and biomedical applications,” *Mammalian Genome*. Springer Science and Business Media LLC. doi: 10.1007/s00335-017-9709-4.

The Cattle Site. 2022. *Breeds - Beef*. [online] Available at: <https://www.thecattlesite.com/breeds/beef> [Accessed 11 July 2022].

Thornton, P. K. (2010) “Livestock production: recent trends, future prospects,” *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society. doi: 10.1098/rstb.2010.0134.

Thomas, P. D. et al. (2003) “PANTHER: A Library of Protein Families and Subfamilies Indexed by Function,” *Genome Research*. Cold Spring Harbor Laboratory. doi: 10.1101/gr.772403.

Tijjani, A. et al. (2019) "Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypanotolerant West African Muturu Cattle," *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2019.00442.

Toro, M. A. et al. (2011) "Assessing the genetic diversity in small farm animal populations," *Animal*. Elsevier BV. doi: 10.1017/s1751731111000498.

Toro, M.A., Fernandez, J. and Caballero, A. (2006) "Scientific basis for policies in conservation of farm animal genetic resources". *In Proceedings of the 8th World Congress on Genetics Applied to Livestock Production,* Belo Horizonte, Brazil.

Tuggle, C. K. et al. (2022) "The Agricultural Genome to Phenome Initiative (AG2PI): creating a shared vision across crop and livestock research communities," *Genome Biology*. Springer Science and Business Media LLC. doi: 10.1186/s13059-021-02570-1.

Van Rossum, G. & Drake, F.L. (2009) "Python 3 Reference Manual". Scotts Valley, CA: CreateSpace.

Verdugo, M. P. et al. (2019) "Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent," *Science*. American Association for the Advancement of Science (AAAS). doi: 10.1126/science.aav1002.

Voight, B. F. et al. (2007) "Correction: A Map of Recent Positive Selection in the Human Genome," *PLoS Biology*. Public Library of Science (PLoS). doi: 10.1371/journal.pbio.0050147.

Wang, H. et al. (2013) "A Post-GWAS Replication Study Confirming the PTK2 Gene Associated with Milk Production Traits in Chinese Holstein," *PLoS ONE*. Edited by R. Wu. Public Library of Science (PLoS). doi: 10.1371/journal.pone.0083625.

Wang, J. et al. (2017) "KCNMA1 mutation in children with paroxysmal dyskinesia and epilepsy: Case report and literature review," *Translational Science of Rare Diseases*. IOS Press. doi: 10.3233/trd-170018.

Wang, Z. et al. (2012) "Genome sequences of wild and domestic bactrian camels". *Nat Commun **3**, 1202*. https://doi.org/10.1038/ncomms2192.

Wangari-Talbot, J. et al. (2012) "Functional Effects of GRM1 Suppression in Human Melanoma Cells," *Molecular Cancer Research*. American Association for Cancer Research (AACR). doi: 10.1158/1541-7786.mcr-12-0158.

Weldenegodguad, M. (2021) “Genomic characterization of northern Eurasian cattle (Bos taurus) and reindeer (Rangifer tarandus)”. (Doctoral dissertation, Itä-Suomen yliopisto).

Weng, M. et al. (2021) “Mutations in DNAH8 contribute to multiple morphological abnormalities of sperm flagella and male infertility,” *Acta Biochimica et Biophysica Sinica*. China Science Publishing & Media Ltd. doi: 10.1093/abbs/gmab013.

Womack, J. E. (2012) “First steps: bovine genomics in historical perspective,” *Animal Genetics*. Wiley. doi: 10.1111/j.1365-2052.2012.02382.x.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), p.97.

Wu, D.-D. et al. (2018) “Pervasive introgression facilitated domestication and adaptation in the Bos species complex,” *Nature Ecology &amp*; Evolution. Springer Science and Business Media LLC. doi: 10.1038/s41559-018-0562-y.

Xie, C.-R. et al. (2015) “Significance of genetic variants in DLC1 and their association with hepatocellular carcinoma,” *Molecular Medicine Reports*. Spandidos Publications. doi: 10.3892/mmr.2015.3970.

Yang, P., Huang, S. and Xu, A. (2016) “The Oxidative Burst System in Amphioxus,” *Amphioxus Immunity*. Elsevier. doi: 10.1016/b978-0-12-849903-0.00008-7.

Yang, Y. et al. (2015) “β-Arrestin1 enhances hepatocellular carcinogenesis through inflammation-mediated Akt signalling,” *Nature Communications*. Springer Science and Business Media LLC. doi: 10.1038/ncomms8369.

Yi, M. et al. (2014) “Performance comparison of SNP detection tools with illumina exome sequencing data—an assessment using both family pedigree information and sample-matched SNP array data,” *Nucleic Acids Research*. Oxford University Press (OUP). doi: 10.1093/nar/gku392.

Yum, S.-Y. et al. (2018) “Development of genome engineering technologies in cattle: from random to specific,” *Journal of Animal Science and Biotechnology*. Springer Science and Business Media LLC. doi: 10.1186/s40104-018-0232-6.

Zhao, L. et al. (2022) “Expression of ovine CTNNA3 and CAP2 genes and their association with growth traits,” *Gene*. Elsevier BV. doi: 10.1016/j.gene.2021.145949.

Zhou, L. et al. (2016) “The scaffold protein KSR1, a novel therapeutic target for the treatment of Merlin-deficient tumors,” *Oncogene*. Springer Science and Business Media LLC. doi: 10.1038/onc.2015.404.

Zorloni, A. (2022) “Breeds of livestock, Department of Animal Science”. *Breeds of Livestock - Ankole Cattle - Breeds of Livestock, Department of Animal Science*. Available at: http://afs.okstate.edu/breeds/cattle/ankole/ [Accessed June 25, 2022].

Zwane, A. A. et al. (2016) “Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds,” *South African Journal of Animal Science*. African Journals Online (AJOL). doi: 10.4314/sajas.v46i3.10.

Zwane, A. A. et al. (2019) “Genome-Wide SNP Discovery in Indigenous Cattle Breeds of South Africa,” *Frontiers in Genetics*. Frontiers Media SA. doi: 10.3389/fgene.2019.00273.