Selection signatures associated with adaptation in South African Drakensberger, Tuli, and Nguni beef breeds.

by

Maxman Gomo

Supervisor: Prof E. van Marle-Köster

Co-supervisor: Prof C. Visser

Submitted in partial fulfillment of the requirements for the degree

MSc (Agric) Animal Science: Animal Breeding and Genetics

In the Faculty of Natural and Agricultural Sciences

University of Pretoria

22 January 2024

Declaration

I declare that the thesis/dissertation, which I hereby submit for the degree MSc (Agric) Animal Science at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at any other tertiary institution.

Signature…….

Date…….

Acknowledgements

I extend my heartfelt gratitude to the Almighty God for granting me the strength and health to embark on this amazing academic journey. My deepest appreciation goes to my parents for the precious gift of life and their unwavering guidance throughout my formative years.

To Prof E. van Marle-Köster and Prof C. Visser, my esteemed supervisors, I am profoundly thankful for your invaluable guidance and support, unwavering trust, and insightful perspectives. You have not only been mentors but also mothers away from home, offering encouragement during moments when I contemplated giving up on my studies. Your belief in me has been a driving force, and I am truly fortunate to have had your support.

I am deeply grateful for the generous support provided by the University of Pretoria Postgraduate bursary, which has made pursuing my Master's degree possible. I wish to also express gratitude to the South African breed societies for the permission granted to utilize the genotypes in this study.

I also want to express my sincere appreciation to Dr S.F Lashmar, whose expertise in data analysis was instrumental in the success of my study.

To my dear friend Phillipine, your listening ear has provided solace. Your support means a lot to me. Chantelle, my colleague, I appreciate our technical discussions that enriched my academic journey. Dr Kea Ncube, your motivation has been a guiding light, pushing me to strive for excellence.

Each of you has played a unique role in this dissertation, and I am profoundly thankful for your contributions to my academic and personal growth.

Abstract

Climate change has become a major factor influencing beef production systems. Sanga cattle are a unique genetic resource renowned for their adaptability to diverse climatic conditions. The advent of genomic technologies has allowed opportunities to examine these indigenous cattle at the Deoxyribonucleic acid (DNA) level and may provide insight into genome-level variation associated with adaptive traits. This study aimed to identify signatures of selection within and across the South African Drakensberger (DRB), Nguni (NGI), and Tuli (TUL) populations. A total of 1 706 animals, including 1 117 DRB, 377 NGI, and 214 TUL, were genotyped using GeneSeek® Genomic Profiler™ 150K bovine SNP panel. The R (Biscarini *et al.*, 2018) and PLINK v1.90 (Purcell *et al.*, 2007) analysis tools were used to estimate ROH, ROHet and F_{ST} values. A set of 122 632 quality-filtered SNPs were utilized to identify genomic regions under selection based on conserved runs of homozygosity (ROH) and F_{ST}-based differentiation of SNPs. The ROH were calculated for various length categories, and a total of 82 871 ROH were identified across all three breeds (mean±standard deviation ROH/animal: DRB=51.82±21.01; NGI=36.09±12.82; TUL=47.94±15.33), with a mean length of 3.90Mb, 2.31Mb and 3.76Mb respectively. The short ROH segments (ROH <4Mb) were most frequent in all breeds. The highest average F_{ROH} was observed in DRB (0.081±0.046) followed by TUL (0.074±0.031), while the lowest F_{ROH} was found in NGI (0.033±0.024). The estimated mean F_{ST} valued between 0.060 (DRB vs NGI & DRB vs TUL), and 0.040 (NGI vs TUL). To identify within-breed selection signatures, genomic regions with the highest frequency for ROH and runs of heterozygosity (ROHet) and observed in more than 20% and 30%, respectively, of the individuals in the population were considered selection signatures. For across-breed selection signatures, regions with F_{ST} values falling within the top 0.1% of the empirical F_{ST} distribution were considered signatures of positive selection. Annotation of these regions revealed genes which have previously been associated with traits of economic importance such as immunity and adaptation (*FKBP4, CTNNA2, MYC, CYSTM1, SRA1, SD14, WDPCP, DTX1, ELMO3,* and *ADAMTS12),* coat colour (*MCIR, TUBB3*), and reproductive performance (*SPARTA33, TCF35, RPS20, CORIN, TXK, NELL2,* and *TMEM181*). Both within-breed (ROH) and across-breed (F_{ST}) approaches proved to be useful in identifying genomic regions under selection, and this may contribute to the understanding of the genetic architecture underlying the adaptive traits of local Sanga cattle for sustainable beef production in the future.

Keywords: Adaptation, Indigenous breeds, Selection signatures, Candidate genes, Homozygosity

Table of Contents

List of Tables

List of Figures

Abbreviations

AFD – Absolute Allele Difference

ARC-BTP - Agricultural Research Council, Biotechnology Platform

BGP – Beef Genomics Program

BTA – Bovine Autosome

DAFF – Department of Agriculture, Forestry and Fisheries

DCMS – De-correlated Composite of Multiple Signals

DNA – Deoxyribonucleic Acid

DRB – Drakenberger

EHH – Extended Haplotype Heterozygosity

FAO - Food and Agricultural Organization

F_{IS} – Inbreeding Coefficient

F_{PED} – Pedigree-based Inbreeding

F_{ROH} -Runs of Homozygosity based Inbreeding Coefficient

F_{ST} – Fixation Index

GCTA – Genome-wide Complex Trait Analysis

GGP - GeneSeek Genomic Profiler

GO – Gene Ontology

GS – Genomic Selection

GWAS – Genome-wide Association Studies

H_E – Expected Heterozygosity

H_O – Observed Heterozygosity

H_S – Average heterozygosity in the subpopulation

H_T – Average heterozygosity in the metapopulation

HWE – Hardy Weinberg Equilibrium

IBD – Identity by Descent

iHS - integrated Haplotype Score

Kb – Kilobase

Kbp – Kilobase pair

LD – Linkage Disequilibrium

MAF – Minor Allele Frequency

Mb – Mega base pair

NAS – Natural and Agricultural Sciences

NCBI – National Centre for Biotechnology Information

N_e – Effective Population

NGI – Nguni

NGS - Next-generation sequencing

PCA – Principal Component Analysis

PCR - Polymerase chain reaction

QC – Quality Control

QTL – Quantitative Trait Loci

r^2 – Correlation between two loci

ROH – Runs of Homozygosity

ROHet – Runs of Heterozygosity

SA – South Africa

SDG - Sustainable Development Goal

SFS – Single Site Frequency

SNP – Single Nucleotide Polymorphism

TUL – Tuli

UN - United Nations

UP - University of Pretoria

XP-CLR – Cross-Population Composite Likelihood Ratio

XP-EHH – Cross-Population Extended Haplotype Heterozygosity

CHAPTER 1: INTRODUCTION

1.1. Introduction

In Africa, livestock is integral to food security, social, cultural, and economic development (Mapiye *et al.*, 2019). However, the looming threat of climate change is adversely affecting the livestock food supply chain (Godde *et al.*, 2021). Climate change has adverse effects on livestock farming, such as water scarcity, altered precipitation patterns, extreme weather events, and disruptions in agricultural production (Gomez-Zavaglia *et al.*, 2020). Rising temperatures cause heat stress, negatively affecting livestock health, reproduction, and productivity (Lacetera, 2019). Increasing droughts and limited water availability contribute to water scarcity. Extreme weather events, including floods present serious threats to livestock safety. Moreover, disruptions in agricultural production lead to feed and fodder shortages which potentially affect livestock numbers (Bernabucci, 2019; Grossi *et al.*, 2019).

In the face of these challenges, the demand for animal protein is escalating, particularly in developing countries such as South Africa amid projections that over 50% of the world's population growth between 2022 to 2050 is expected to occur in sub-Saharan Africa (United Nations, 2022). In South Africa, where extensive grazing represents up to 83.3% of the total farmland (Pienaar, 2013), livestock farming particularly the beef industry stands out as a prominent player, with 80% of the total cattle population raised for beef production, and the remaining 20% dedicated to dairy (DAFF, 2021). Within the diverse landscape of the South African beef industry, various breeds of cattle play distinctive roles in meeting the nation's protein requirements. According to the report by DAFF (2021), the Sanga cattle types (Drakensberger, Nguni, Tuli, and Afrikaner) account for 29% of the total meat produced, signifying their substantial contribution to the South African beef industry.

In South Africa, beef cattle play a significant role in providing milk, meat, and hides, and, in some areas, serve as a primary source of both draft power and manure for maintaining soil fertility(Scholtz *et al.*, 2008). Burrow (2015), provided a comprehensive definition of adaptation, emphasizing the ability of an animal to thrive, grow, and reproduce in challenging environments. This adaptability is crucial for the sustainability of beef production, as a failure to adapt can result in adverse economic impacts such as decreased yields, animal fatalities, and increased treatment expenses (Prayaga *et al.*, 2006). Adaptation to environmental stressors varies depending on breed types and geographic locations, with cattle needing to respond effectively to various challenges, including ticks (Hanotte *et al.*, 2003), high temperatures and humidity (Mariasegaram *et al.*, 2007), endemic diseases (Marufu *et al.*, 2013), poor-quality feed, appetite and fasting metabolic rate, as well as gastrointestinal pathogens (Prayaga *et al.*, 2006).

South Africa is geographically divided into five distinct and well-defined biomes, each with its unique characteristics further highlighting the importance of adaptation in ensuring cattle's success within their specific ecological contexts (Huntley *et al.*, 1984). The fynbos biome, constituting 5.3% of South African land, is known for its lush shrubs and tufted grass-like plants, primarily located in the Southwest and Southern Cape (Acocks, 1988). The karoo biome, occupying 31.9% of the country's total land, is semi-arid, featuring distinctive succulent plants and hardy shrubs, with coverage extending to areas like the Little Karoo, Namaqualand, Great Karoo, and central and upper Karoo (Acocks, 1988).

The savanna biome comprises vast grasslands characterized by the presence of acacia trees and a diverse range of herbivores and predators. This biome experiences extreme weather variations, from severe cold and snowfall on mountain peaks during winter to high temperatures and intense sunlight in summer (Acocks, 1988). The grassland biome covers 24.1% of the country's land area and is dominated by the grasses and other grass-like plants. It serves as vital grazing land, mainly during rainy season in summer and it also experience dry and frosty winters. South Africa also features the savanna biome, which encompasses 24.2% of the country’s land area and is further divided into arid and moist savanna sub-biomes. Arid savannas feature sparsely vegetated grasslands with widely spaced shrubs, while moist savannas have nutrient-poor, leached soils that can occasionally become waterlogged during the rainy season (Acocks, 1988). Lastly, the forest biome comprises lush, densely wooded areas with towering trees, creating unique habitats for a diverse range of wildlife. Within these diverse environments, various cattle populations have adapted, with notable resilience demonstrated by Sanga cattle, particularly in their ability to thrive across all these biomes (Strydom, 2008).

Research has shown that the adaptability of Sanga types to local environmental conditions and their ability to thrive on natural forage resources contribute to their prominence in the meat production sector (Strydom, 2008; Van Marle-Köster *et al.*, 2021). South African Sanga cattle such as the Drakensberger, Nguni, Tuli, and Afrikaner are indigenous or landrace breeds and are known for their unique adaptive traits that include tolerance to endemic diseases and parasites (both internal and external;(Mapholi, 2015; Mapholi *et al.*, 2022), extreme temperatures (Nyamushamba *et al.*, 2017), changes in feed availability, and low management inputs (Scholtz, 2010). This unique genetic resource can be vital in achieving in achieving sustainable and climate-resilient future for livestock farming.

The development of genomic technology over the past decades created opportunities to study genetic variation and the unique ancestral history of various cattle breeds (Matukumalli *et al.*, 2009). Several genomic diversity studies have been conducted using microsatellites, for example, (Sanarana, 2015; Madilindi *et al.*, 2020) and single nucleotide polymorphism (SNP) markers, (Makina *et al.*, 2014; Zwane *et al.*, 2016; Bosman *et al.*, 2017; Lashmar *et al.*, 2018; Lashmar *et al.*, 2022), which confirmed moderate

levels of genetic diversity in the Afrikaner, Nguni, Tuli and Drakensberger breeds. Despite benchmark studies on the genome-level composition of indigenous cattle, for example Makina *et al.* (2016); Lashmar *et al.* (2018); King *et al.* (2022) and Lashmar *et al.* (2022), the genetic architecture of adaptive traits unique to indigenous populations remains only partially understood.

Natural and artificial selection have been pivotal in shaping the production and morphometric traits of cattle breeds. Through selective breeding, various traits of economic importance such as meat quality, longevity, docility, and adaptation have been improved (Moravčíková *et al.*, 2018). The process of selection refers to how certain genetic traits become more prevalent in a population over time (Brito *et al.*, 2021). Positive selection occurs when a specific allele is favoured and becomes increasingly common, leading to selective sweeps on an animal's genome. This can create either hard, partial, or soft sweeps, depending on how strongly the surrounding neutral sites are affected (Saravanan *et al.*, 2020). Negative selection, on the other hand, occurs when a mutation is disadvantageous, and will eventually be removed from the population (Saravanan *et al.*, 2020; Derks & Steensma, 2021). Genome-wide SNP data harbour helpful information that could be harnessed to manage genetic resources effectively. Genomic technologies allow for the identification of signatures of selection employing various methodologies (Saravanan *et al.*, 2020).

The use of ROH segments in analysing the genomes of farm animals is an important tool for gaining insights into their population history (Fabbri *et al.*, 2021), predicting bottlenecks (Garcia *et al.*, 2023), estimating inbreeding levels (Peripolli *et al.*, 2020), and identifying signatures of selection (Ablondi *et al.*, 2020; Freitas *et al.*, 2021). Such signatures can be identified in South African beef breeds and can be useful in improving traits associated with adaptation, fertility, and meat quality. Detecting selection signatures is critical in characterizing livestock genetic resources and deciphering the genetic factors responsible for variability in traits of economic importance (Yurchenko *et al.*, 2018; Saravanan *et al.*, 2021). They can also be used to identify beneficial mutations and quantitative trait loci (QTLs) that are associated with traits such as disease resistance, immunity and adaptation (Kooverjee *et al.*, 2022).

In South Africa, the Beef Genomics Program (BGP) provided a basis for building reference populations for beef cattle breeds. A total of thirteen SA beef breeds participated in the BGP, providing genotyped animals, which included the three breeds used in this study: Drakensberger, Nguni and Tuli beef breeds. By using SNP genotypes, this study will not only identify runs of homozygosity (ROH) and runs of heterozygosity (ROHet) but also selection signatures in three prominent SA Sanga breeds.

Selection signature studies can help to uncover the underlying genetic mechanisms that drive traits of economic importance, contributing significantly to scientific knowledge and raising awareness about the potential of indigenous breeds.

1.2. Aim and Objectives

Overall aim: To identify selection signatures associated with adaptation in three SA indigenous beef breeds namely, the Drakensberger, Nguni, and Tuli, using genotypes.

To achieve this aim, the set objectives were as follows.

1. To estimate genomic diversity parameters in each breed, including expected and observed heterozygosity (H_E and H_O), minor allele frequency (MAF), linkage disequilibrium (LD), and inbreeding coefficients (F_{IS})
2. To investigate within-population signatures of selection using Runs of Homozygosity (ROH) and Runs of Heterozygosity (ROHet)
3. To identify between breed signatures of selection using the fixation index (F_{ST}).
4. To perform gene annotation for potential genes associated with adaptation traits.

CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

For centuries, humans have played an integral role in shaping livestock populations through artificial selection. From the beginning of domestication, farmers have selected specific traits that meet their needs (Mwai *et al.*, 2015). However, natural selection has also played a part in the evolution of various breeds and ecotypes, where animals have adapted to diverse environments (Moravčíková *et al.*, 2018). To fully understand the evolution and selection processes influencing the development of various breeds, it is essential to explore the genetic and phenotypic differentiation within their genomes. Genomic data provides valuable insights into the selection signatures associated with economically important traits, making it a valuable tool for potential genetic evaluations. The focus of this review is to provide a summary of South African indigenous breeds, followed by a review of relevant literature on signatures of selection and their association with traits of economic importance.

2.2. Origin of African and South African Indigenous breeds

Modern domestic cattle have ancestral origins in two distinct subspecies of wild aurochs found in the Near East (*Bos taurus*) and India (*Bos indicus*) (Zeder, 2008). The domestication of European taurine cattle (*Bos taurus*) can be traced back to approximately 10,000 years ago in the Fertile Crescent, with *Bos primigenius primigenius* as the ancestor (Troy *et al.*, 2001; Zeder, 2008). Roughly 2,000 years later another significant domestication event occurred in Central Asia, where *Bos primigenius namadicus* served as the progenitor of the humped Zebu cattle (*Bos indicus*) (Chen *et al.*, 2018). Due to the advancement of agriculture, the taurine cattle eventually migrated across Africa, Asia, and Europe (Decker *et al.*, 2014). Likewise, indicine cattle accompanied human migrations through South-East Asia and subsequently Africa (Kim *et al.*, 2017). To date, taurine cattle are predominantly found in Europe and Northern Asia, while indicine cattle occupy Southern Asia (Upadhyay *et al.*, 2017). The distinct levels of admixing between taurine and indicine cattle have given rise to African taurine, African zebu, and their hybrid offspring collectively known as Sanga cattle (*Bos taurus africanus*) (Kim *et al.*, 2017). Southern Africa is home to the Sanga type of cattle that are believed to be anatomically and physiologically adapted to diverse environmental conditions within the continent (Hanotte *et al.*, 2002). The Sanga type includes breeds such as the South African Drakensberger, Nguni, and Afrikaner, the Mozambican Landim, Zimbabwean Mashona, Tuli and Nkone, and the Nguni of Zambia (Pienaar *et al.*, 2015; Nyamushamba *et al.*, 2017). Herewith follows a short description of the SA Sanga breeds includeded in this study.

Drakensberger

This breed is characterized by a smooth black coat, medium to large frame, with a mature bull weight of 820kg to 1100kg, mature cows weighing between 550 kg to 720kg, and calves with an average birth weight of 35kg (https://drakensbergers.co.za/English/, 2023). The breed is reported to have high longevity (Bisschoff & Lotriet, 2013), high fertility and milk production, ease of calving (Scholtz, 2010), and good marbling ability and meat quality (Strydom, 2002). Their short glossy blue-black skin makes it difficult for ticks to attach themselves and it also reflect sunlight (Scholtz, 2010). The breed has proven to be an excellent mother line breed in crossbreeding systems (https://drakensbergers.co.za/English/, 2023). Little is known about the genetic composition of this breed, but previous studies have suggested a mixture of European and African taurine and indicine breeds (Makina *et al.*, 2016). Due to their adaptability and consistent performance, they are popular mostly in the Eastern parts of South Africa and are also found in neighboring countries such as Zimbabwe, Namibia, and Zambia (https://drakensbergers.co.za/English/, 2023).

Nguni

The Nguni breed ranks among South Africa's most renowned indigenous cattle breeds, with its name originating from the term “Nguni” which is often used to collectively refer to ethnic groups (Zulu, Xhosa, and Swazi) (Schoeman, 1989). The coats of Nguni cattle are distinguished by their softness and glossy appearance, exhibiting both unicolored and multicolored patterns: white, black, brown, grey, and red (Olson, 1999). The multicolor patterns of Nguni hides provide extensive possibilities for various products, including intact hides for floor and wall coverings, as well as items such as wallets, handbags, sport bags, and footwear, and briefcases (Mapiye *et al.*, 2007). In mature Nguni cows, the cervicothoracic hump is barely noticeable, while it is well-developed and visible in the full-grown bulls. The Nguni cattle is a medium to small breed with the following characteristics: a birth weight of 25kg, a weaning weight of 155kg, a cow mature weight ranging between 320 and 440kg, and a mature bull weight ranging from 500 to 700kg (Bergh *et al.*, 2010).

The cows have small to moderate teat and udder size (Brown, 1959; Zindove & Chimonyo, 2015). They are good mothers (Mapiye *et al.*, 2007), and resistant to ticks and tick-borne diseases (Tada *et al.*, 2013; Mapholi, 2015), resulting in a low mortality rate. Moreover, this breed is known for its temperament and for being excellent foragers, that can graze on steep slopes and in densely vegetated areas. Nguni cattle hold cultural significance, particularly animals with solid black coats, as they play a crucial role in the ceremonial traditions of the Swazi and Zulu communities (Rege & Tawah, 1999). Nguni cattle ecotypes are found not only in South Africa but also in various Sub-Saharan African countries, including eSwatini, Malawi, Namibia, Mozambique, Zambia, and Zimbabwe (Hanotte *et al.*, 1998). Notably, the reach of this

breed extends beyond the African continent, exemplified by their introduction to other countries such as Australia, where the first embryo export took place in 2007 (Roux, 2023).

Tuli

The Tuli breed which originated from Zimbabwe and the Tuli Cattle Breeders Society of South Africa was officially formed on 24 March 1994. The breed is well-suited for extensive beef farming and can thrive on poor-quality grazing while still producing high-quality meat (http://www.tulicattle.co.za/, 2023). The breed is indigenous to Southern Africa, and they possess adaptation attributes such as tolerance to parasites, heat stress, and tolerance to diseases, which allows for their competitive production (Visser, 2023). The animals usually have yellow coats with short straight hair and have a small erect cervicothoracic hump (http://www.tulicattle.co.za/, 2023). Tuli cattle are described by the breed society for their good maternal traits, hardiness, and adaptability to hot and dry conditions, as well as their juicy and tender meat with good marbling (http://www.tulicattle.co.za/, 2023). In addition, they have the advantage of being an early maturing and docile breed and easy to manage (http://www.tulicattle.co.za/, 2023). The breed is naturally polled. (http://www.tulicattle.co.za/, 2023). The geographical distribution of Drakensberger, Nguni, and Tuli stud breeders across South Africa, with their locations estimated in proximity to the nearest towns is shown in **Figure 2.1**

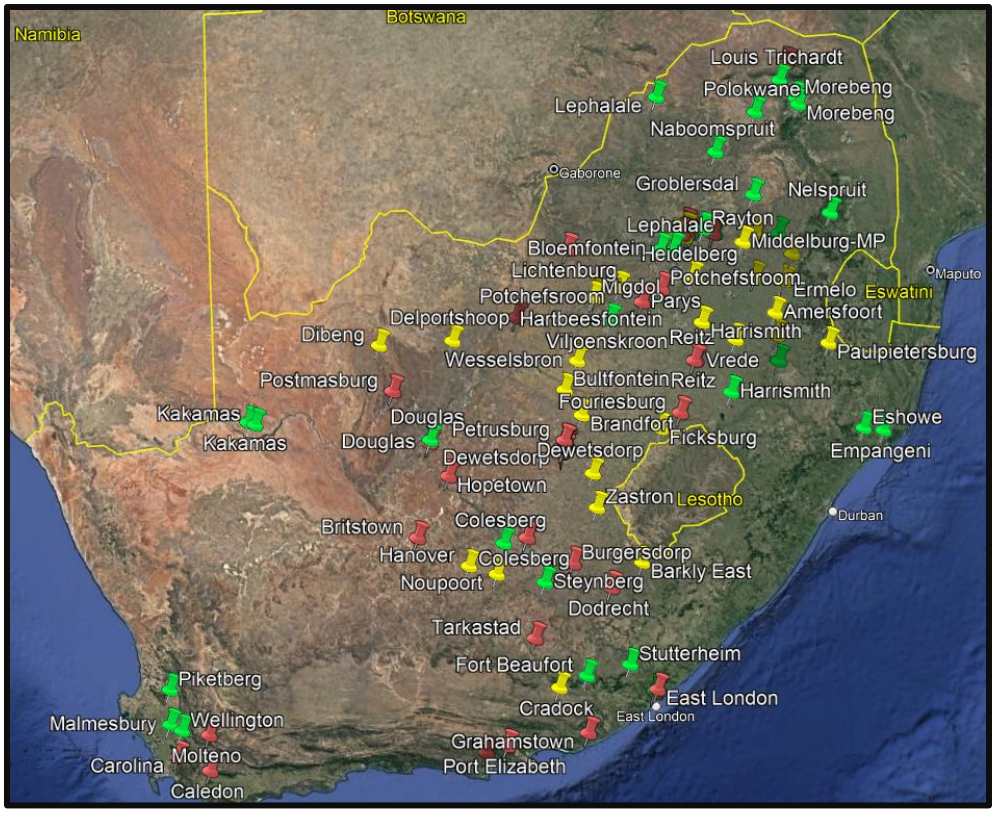


Figure 2.1: Geographical locations of Drakensberger (yellow), Nguni (green), and Tuli (red) stud breeders across South Africa.

All three breeds have established breed societies and actively participate in animal recording with the SA Stud Book**.** Among these populations, the Nguni breed exhibits notably lower participation rates, with just 40% of farmers actively participating with the SA Stud Book despite showing the highest number of registered animals as shown in **Table 2.1.**

Table 2.1: Overview of the breed society establishment, number of breeders and animals registered and/or actively participating in LOGIX beef (SA Stud Book, 2022).

Breed	Breed Society established	N of breeders (Registered)	N of breeders (Participating)	% Participation	Current number of registered animals
Drakensberger	1947	46	46	100	15 006
Nguni	1986	111	44	40	19891
Tuli	1984	34	33	97	9113

N = Number of breeders; % Participation = Percentage participation

Since the inception of the National animal recording schemes in South Africa during the 1960s, breeds such as the Drakensberger, Nguni and Tuli have actively taken part in official animal recording. They contributed by submitting performance and pedigree records for genetic evaluation. Participation rates in the animal recording have varied, ranging from 42% for Nguni to 85% for the Drakensberger breeds (Scholtz, 2010).

2.3. Adaptive traits in cattle under Tropical environmental stressors

Tropical adaptation refers to an animal's ability to thrive and reproduce amidst prevalent environmental stressors in tropical environments (Burrow, 2012). The inability of an animal to adapt has economic implications for production systems, leading to high mortalities and treatment costs and the production of low-quality products (Ayalew *et al.*, 2023). Structural and functional genomic variations play an important role in shaping animal adaptation to diverse agroecological zones (Zwane *et al.*, 2021). Therefore, gaining insight into the genetic basis of adaptive mechanisms in extreme environments is important for effective conservation and genetic improvement, especially when faced with climate change. African cattle breeds are faced with challenging environmental conditions in the form of dry or hot climates (Kim *et al.*, 2017). Natural and artificial selection have played a role in shaping various patterns of genomic variation among these cattle populations, facilitating the development of tropical adaptation features (Mwai *et al*., 2015; Taye *et al.*, 2018). As a result, the identification of signatures of positive selection has become a primary focus in recent genomics studies (Saravanan *et al.*, 2021). African cattle exhibit various adaptive traits, such as disease and parasite resistance, low nutrient requirements for maintenance, thermotolerance, and exceptional leg conformation (Mapholi, 2015; Kim *et al.*, 2017; Edea *et al.*, 2018). These traits, essential for their survival in diverse environments, will be discussed in detail below.

Thermotolerance

Thermotolerance refers to the capacity of an organism to develop resilience to heat shock following prior exposure to heat (Hariyono & Prihandini, 2022). Several studies reported reduced feed intake, growth rate, milk yield, and reproductive efficiency observed under conditions of heat stress (Amamou *et al.*, 2019; Negri *et al.*, 2021; Singh *et al.*, 2022). Livestock primarily use the cellular response as a key pathway to cope with heat stress challenges. Components involved in the heat shock response include Heat Shock Factor (HSF), Heat Shock Element (HSE), and Heat Shock Protein (HSP) (Archana *et al.*, 2017; Hariyono & Prihandini, 2022). The synthesis and discharge of Heat Shock Proteins (HSPs) represent the end product of this cellular response pathway. HSP70, HSP90, and HSP27 are the most frequently studied HSPs in livestock (Belhadj Slimen *et al.*, 2016; Hariyono & Prihandini, 2022). Among these, HSP70 has been identified as the optimal genetic marker for heat stress in livestock. One of the early physiological responses linked to the stress-induced buildup of HSP70 is the development of acquired thermotolerance (Chirico *et*

al., 1988). The cytoprotective role of HSP70 has been confirmed in various organs, such as the intestine, kidney, and embryo of cattle (Bhat *et al.*, 2016). HSP levels surge in response to exposure to various stressors, such as cold, heat, infection, starvation, and hypoxia (Wang *et al.*, 2015).

Tick tolerance

Trypanosomiasis, a tick-borne disease, is one of the most prevalent diseases significantly affecting cattle production in Africa (Perry, 2002). It has symptoms such as anemia and weight loss, potentially leading to fatal consequences and impacting productivity through infertility and calf abortion (Hill *et al.*, 2005). Various cattle breeds of African taurine origin such as Lagune, Muturu, and N'Dama exhibit tolerance to trypanosomiasis and have shown resilience and remain productive compared to indicine or European taurine breeds (Kambal *et al.*, 2023).

In addition, Zebu and Nguni cattle, also display greater tick resistance due to their distinctive morphological features such as coat type, coat colour, skin thickness, and skin secretions (Mapholi, 2015; Shyma *et al.*, 2015; Marima, 2017). Host resistance to ticks is acknowledged as genetically controlled despite the intricate nature of the tick–host interaction. The structural composition of the epidermal layers and skin colour are key defense mechanisms against ectoparasites such as ticks (Mapholi *et al.*, 2014). Notably, keratin genes have previously been linked with the secretion of cytokines, which trigger localized inflammatory responses crucial for tick resistance (Taye *et al.*, 2018). Kim *et al.* (2017) identified *Bovine lymphocyte antigen* (*BOLA*) gene in African cattle, which is also involved in tick resistance. Moreover, Tabor *et al.* (2017) and Marima *et al.* (2020) confirmed that breeds such as Brahman and Nguni exhibit unique immunogenetic responses regulated by both innate immune and adaptive systems.

Adaptation to high altitude

The genetic adaptations of cattle to high altitudes have been the subject of several studies. For example, studies have compared livestock populations from high altitudes (>2500 m above sea level) and low altitudes (<1500 m above sea level) to understand the influence of natural selection on cattle genomes (Wang *et al.*, 2021; Terefe *et al.*, 2022; Kambal *et al.*, 2023). SNP and whole-genome sequencing data were used in these studies to identify genomic regions associated with adaptation to high-altitudes. Using an F_{ST}, comparison between low and highland-dwelling cattle breeds revealed differentiated loci that overlap with genes linked to responses to hypoxia (Terefe *et al.*, 2022). These are proxy indicators for the adaptations of cattle to high-altitude environments and associated challenges, such as hypoxia.

Drought tolerance

African cattle face recurrent droughts and environmental adversities, resulting in resource scarcity, poor forage quality, and increased susceptibility to diseases (Nardone *et al.*, 2010; Tadesse & Dereje, 2018).

Despite facing climate-induced challenges that lead to cattle deaths and diminished productivity, local breeds have developed adaptive strategies, particularly in dealing with poor nutrition (Henry *et al.*, 2018). The selective pressures in tropical environments have prompted African populations to improve digestive efficiency and lower metabolic rates (Taye *et al.*, 2018). Genomic studies have identified candidate genes associated with adaptive responses to feed scarcity, such as *HCRTR1*, *GH1*, *MAP3K5*, and *ZRANB3* (Kim *et al.*, 2017; Edea *et al.*, 2018). These genes regulate feeding behaviour and feed efficiency, contributing to enhanced feed efficiency amid a limited food supply. In extensive production conditions, these genomic footprints could be essential in addressing feed scarcity, particularly in drought-prone regions across the African continent.

2.4. Challenges in recording adaptation traits

The major challenge in the genetic improvement of adaptive traits in cattle farming lies in the difficulties in accurately measuring phenotypes (Burrow, 2015). Heat stress is measured by indicators, such as heart and pulse rates, and rectal temperature (Lemerle & Goddard, 1986), as well as coat type and colour (Ghassemi Nejad *et al.*, 2017). However, indicators such as heart and pulse rates are difficult to measure by farmers, making coat type and color mostly used approaches (Sejian *et al.*, 2022). The easiest way to measure heat stress is by measuring rectal temperature; nevertheless, it is still invasive (Visser & Snyman, 2023). For traits like nematode resistance, the complexity of the molecular mechanisms and the lack of a single accurate indicator further complicate their recording. A more precise way of estimating nematodes resistance require phenotypic indicators such as immunoglobulin levels, fecal egg count, and packed cell volume as prerequisites (McManus *et al.*, 2014). Tick and lice counts face difficulties under extensive farming, influenced by environmental factors like season, rainfall, and temperature (Ramzan *et al.*, 2021). To accurately record reproductive traits such as birth weight is challenging under extensive farming conditions, for instance in case of stillborn calves. For growth traits, traditional selection methods like best linear unbiased prediction (BLUP) are feasible, while disease and heat resistance traits may benefit from direct selection based on genetic markers or genomically enhanced breeding values (GEBV) (Visser & Snyman, 2023). Comprehensive pedigree and performance recording are both essential for accurate EBV and GEBV estimation. For more than four decades, the SA Stud Book has documented traits related to the growth and production performance of Drakensberger, Nguni and Tuli cattle breeds as shown in **Table 2.2**.

Table 2.2: Measurements for the growth traits given as average per breed for male and female animals (SA Stud Book, 2022)

Growth Trait	Average per breed					
	DRB		NGI		TUL	
	Male	Female	Male	Female	Male	Female
Birth weight (kg)	35.7	33.9	26.2	25.0	31.7	30.1
Weaning weight (kg)	211	197	162	149	198	186
12 Month weight (kg)	324	231	228	175	242	220
18 Month weight (kg)	389	310	267	237	340	289
Feed Conversion ratio	6.09	5.87	6.93	-	6.46	
Average Daily Gain (kg)	1.614	1.459	1.178	-	1.367	

Feed Conversion ratio measured in kg/kg; - No records.

Adaptive traits (parameters for heat tolerance, disease, and pathogens resistance) and their heritabilities for South African Sanga cattle are currently unavailable in the literature. Since the three breeds under the current study have access to both genetic evaluations and genotyping services, estimated breeding values (EBV) have been used as valuable tools to assist in selection. The average performance for growth and fertility traits, shown in **Table 2.3** can be used as indicators of adapted animals in South Africa populations.

Table 2.3: A summary of the mean performance in fertility traits among the three indigenous breeds (SA Stud Book, 2022).

Breed	Cow weight at calving (kg)	Cow weight at weaning (kg)	Age at 1st calving (months)	Inter calving period (days)	Days since last calving	Cow calf ratio
DRB	473	493	36.4	436	340	41%
NGI	346	377	32	409	267	42%
TUL	447	456	34.6	406	369	43%

Age at 1st calving = Age at first calving

Despite challenges associated with the high cost of genotyping, implementing genomic selection programs in South Africa is gradually advancing. A considerable number of genotypes have been generated using low to medium density SNP panels resulting in establishment of reference populations. Additionally, imputation has emerged as a strategic approach to alleviate genotyping cost within South African cattle populations (Lashmar *et al.*, 2019). Integrating genotypic data with pedigree records can increase accuracy

in documenting adaptive traits and other traits that are not easy to physically measure (Utsunomiya *et al.*, 2015).

2.5. A brief overview of the development of genomic tools

Recent advances in genomics and bioinformatics have paved the way for comprehensive research on the genetic mechanism underlying traits of economic importance in SA beef cattle. With the focus of existing studies centered on genetic characterization of several South African indigenous cattle breeds in comparison to commercial breeds (Makina *et al.*, 2014; Sanarana, 2015; Van Marle-Köster *et al.*, 2021), there remains a gap in understanding the genetic mechanisms underlying various economically important traits.

The development of genome-wide SNP arrays with a capacity to identify over 100 thousand SNPs has significantly contributed to unraveling the genetic complexities in cattle populations (Matukumalli *et al.*, 2009). Over the decades, companies such as Affymetrix®, Illumina®, and Neogen's GeneSeek®, have developed commercial SNP bead chips that are available for cattle, offering a valuable resource for genetic analysis (Nicolazzi *et al.*, 2015). **Table 2.4** shows a non-comprehensive overview of the currently available commercial bead chips for cattle.

Table 2.4:A non-comprehensive list of commercial single nucleotide polymorphism bovine bead chips adapted from Nicolazzi *et al.,* (2015)

Company	Bead chip	Number of SNPs
Affymetrix®	Axiom® Genome-wide BOS1	648 875
Geneseek®	GeneSeek Dairy Ultra LD v2	7 049
	GGP-LD	
	Version 1 (GGP 9K)	8 610
	Version 2 (GGP 20K)	19 721
	Version 3	26 151
	GGP indicus	35 090
	GGP HD	76 879
	GGP 150K	139 480
Illumina®	Golden-Gate-Bovine-3K	2 900
	Bovine LD	
	Version 1	6 909
	Version 1.1	6 912
	Version 2	7 931
	Bovine SNP50	
	Version 1	54 001
	Version 2	54 609
	Bovine-HD	777 962

GGP = GeneSeek Genomic profiler

Several bead chips listed in **Table 2.4** share a considerable number of SNPs, and some are updated versions of old panels, incorporating additional SNPs. For instance, Illumina's Bovine LD, released in 2011, was designed to replace Golden Gate 3K panel of 2010. The Illumina's Bovine LD included 2,159 shared SNPs with the Golden Gate (Wiggans *et al.*, 2013). The GGP Indicus bead chip is comprised of 35,000 SNPs that are indicine-specific originated from prominent indicine breeds such as Brahman and Nellore, and various composite breeds (Ferraz *et al.*, 2020). In addition, efforts have been made to develop several other lower-density bead chips with the intention of retaining specific SNP subsets shared with higher density chips (Boichard *et al.*, 2012). GGP 150K SNP bead chip, is equipped with 139,480 SNPs with an average inter-SNP distance of 19 kb. This chip incorporates a considerable number of common SNPs from the low density to high density bead chips.

Historically, cattle genetic improvement relied on analyzing performance and pedigree data (based on BLUP), and using microsatellite markers primarily for population genetics (Gutierrez-Reinoso *et al.*, 2021). Microsatellite markers were also valuable for identifying quantitative trait loci (QTL) affecting various economically important cattle traits (Holmberg & Andersson-Eklund, 2006). However, microsatellite marker genotyping is a labour-intensive process, and allele calls tend to be specific to individual laboratories (Guichoux *et al.*, 2011; Sabir *et al.*, 2014). In recent years, there has been a shift towards using single nucleotide polymorphisms (SNPs) for genetic analysis (Cortes *et al.*, 2022). SNPs are cheaper and easier to assay on a per marker basis, and are more densely distributed and abundant compared to microsatellites, with a frequency of approximately one SNP per kb and one SNP per 500 base pairs in humans and cattle, respectively (Heaton *et al.*, 2002; Coates *et al.*, 2009). The availability of high-throughput SNP genotyping platforms allows for extensive scans using large numbers of high density SNP markers (Barabaschi *et al.*, 2016).

In South Africa, the Bovine SNP arrays were investigated by Qwabe *et al.* (2013), and the applicability of the bovine SNP50 bead chip was confirmed in South African cattle. The study further suggested that the SNP array can be used for genomic studies across commonly used cattle breeds in South Africa, such as Angus, Holstein, Nguni, Bonsmara, Drakenberger, and Afrikaner, in both beef and dairy production (Qwabe *et al.*, 2013). In addition, several other studies focusing on Sanga beef cattle have used genotypic data originating from low to high density panels, including Illumina bovine SNP50 (Zwane *et al.*, 2016; Pierce *et al.*, 2018), GGP 80K (King *et al.*, 2022), and GGP 150K (Lashmar *et al.*, 2018; Kooverjee *et al.*, 2022), for different genomic applications. Similarly, during the establishment of the reference population, the BGP generated low-density and high-density SNP genotypes that has been explored in several studies.

2.6. Genetic diversity

Genetic diversity reflects differences among individuals in the same populations and provides the basis for evolutionary and adaptive processes, which may assist in effective breed management (Eusebi *et al.*, 2019). Natural and artificial selection forces have been reported to be the main forces influencing animal genetic diversity (Eusebi *et al.*, 2019), as they impact processes such as segregation, mutation, genetic drift, and selective breeding (Bosse *et al.*, 2019). Measuring and studying genetic diversity is important because variation within or across breeds allows animals to respond to selection and environmental changes (Caballero & Toro, 2000). Genetic diversity dictates the ability of a specific population to react positively to selection; hence allelic variations are vital for long-term survival and ensuring food security in the future (Zhao *et al.*, 2015; Moravčíková *et al.*, 2018).

Ascertainment bias refers to a phenomenon that occurs when SNPs selected for the development of an array are determined through sequencing a limited number of samples or a specific set of breeds (Dokan *et al.*, 2021). This limited sampling can introduce a bias in the selection process, as the chosen SNPs may not adequately represent the genetic variation present in larger and more diverse populations. This bias leads to an overrepresentation of common polymorphic loci while underrepresenting low-frequency polymorphisms (Geibel *et al.*, 2021). Consequently, various genetic diversity parameters, such as linkage disequilibrium and allele frequency distribution can be biased (Lachance & Tishkoff, 2013).

There are several parameters used to explain genetic variation within breeds which include MAF, H_O and H_E, effective population size (Ne), and linkage disequilibrium (LD). These parameters have been widely used to estimate genetic variation since the discovery of microsatellites and SNP markers, population structure, and effective population sizes. H_O can be influenced by ascertainment bias; therefore, it is imperative to compare H_O with H_E, the expected heterozygosity estimated from population allele frequencies (Ajmone-Marsan *et al.*, 2023). Due to occasional non-random mating or sampling bias, H_O is often lower than H_E (Eusebi *et al.*, 2019). Disparities between H_O and H_E indicate considerable deviations from the Wright-Fisher ideal population (e.g., $H_O > H_E$), indicate population admixture or hybridization, whereas $H_O < H_E$ may suggest population subdivision or inbreeding (Ajmone-Marsan *et al.*, 2023).

H_O and H_E have been extensively used to explore the genetic diversity of South African cattle populations. For example, a microsatellite based study on Afrikaner, Bonsmara, Drakensberger, Nguni, and Tuli populations has revealed moderate to high levels of genetic diversity (van der Westhuizen *et al.*, 2020). The mean heterozygosity ranged from 0.57 to 0.74 in nine South African cattle populations investigated using a panel of 11 microsatellite markers. Other studies also reported heterozygosity levels ranging from 0.70 to 0.75 in the Nguni cattle using microsatellite markers (Sanarana, 2015; Madilindi *et al.*, 2020). Other studies used SNP data and found moderate heterozygosity levels for the Afrikaner (0.22 to 0.24)(Zwane *et al.*, 2016), Bonsmara (0.35 to 0.36) (Kooverjee *et al.*, 2022), Drakensberger (0.25 to 0.35) (Lashmar *et al.*, 2018; Van Marle-Köster *et al.*, 2021), and Nguni (0.23 to 0.363)(King *et al.*, 2022) breeds.

Traditionally, inbreeding estimation relied on pedigree information (F_{PED}) alone. However, the advent of genomic technology has ushered in the use of genomic data to estimate inbreeding coefficients. Inbreeding, when it occurs, leads to a reduction in H_O, and a deficiency in H_O may indicate underlying inbreeding. For instance, Van Marle-Köster *et al.*, (2021), assessed inbreeding coefficients based on the deficiency of heterozygotes, revealing coefficients ranging from -0.009 to 0.011 for breeds such as Boran, Bonsmara, Drakensberger, Hereford, Tuli, and Nguni. The introduction of ROH has provided a new approach for inbreeding estimation. In a similar study, F_{ROH} was estimated, reporting inbreeding coefficients ranging from 0.006 to 0.029. Both approaches exhibited lower inbreeding coefficients,

suggesting moderate to higher levels of genetic variability within the population which is important for long-term adaptability (Eusebi *et al.*, 2019).

The non-random association of alleles from different SNPs on the same chromosome is known as linkage disequilibrium (Toro Ospina *et al.*, 2019). LD is commonly observed between SNPs that are physically close to each other, indicating limited recombination between them. However, high LD between distant SNPs can be caused by demographic processes like genetic drift or admixture, as well as selective processes like background selection (Gautason *et al.*, 2021). LD between SNPs on different chromosomes can be an artefact of sampling limited individuals (McVean, 2007). Due to LD, nearby markers are not independent, and LD-based pruning methods are used to select independent markers for genetic analysis (Jemaa *et al.*, 2019). This involves removing SNPs with high LD within sliding windows based on a given threshold. Makina *et al.*, (2015) assessed the LD within four South African breeds (Afrikaner, Nguni, Drakensber and Tuli). The study showed that the Afrikaner cattle exhibited higher levels of LD than other indigenous breeds. High LD values suggest the possibility of bottlenecks in the population, hence LD strength can be used as an estimator of effective population sizes (Makina *et al.*, 2015)

2.7. Selection signatures in cattle breeds

A selection signature refers to a pattern that arises because of selection (both natural and artificial) acting on a specific trait or a combination of traits in a population (Rebelato & Caetano, 2018). It represents evidence or footprints of previous selective pressures that have influenced the genetic composition of individuals within a specific population. These patterns display a localized decrease in genetic variation both downstream and upstream of a mutation due to the rapid fixation of the mutation (Chen *et al.*, 2016). The phenomenon known as a selective sweep occurs when a newly advantageous mutation spreads throughout a population, leading to the elimination or reduction of genetic variation in adjacent neutral sites (Saravanan *et al.*, 2021).

Selection signatures can be created through two main mechanisms: positive selection and purifying selection (Nielsen *et al.*, 2005). Positive selection occurs when a specific genetic variant provides an advantage, leading to its increased frequency in the population over time. Purifying selection, on the other hand, acts to remove deleterious or harmful genetic variants from the population (Saravanan *et al.*, 2020). Selection signatures can be defined in terms of soft and hard selective sweeps, which represent different modes of positive selection (Nielsen *et al.*, 2005). In a soft selective sweep, positive selection acts on existing genetic variation within a population. This means that the beneficial genetic variant already exists in the population at a low frequency, and selection acts to increase its frequency owing to its fitness advantage (Prezeworski *et al.*, 2005). Soft selective sweeps typically result in a diverse range of genetic backgrounds surrounding the selected variant, as multiple offspring carrying different versions of the

variant contribute to its spread (Nielsen *et al.*, 2007). In contrast, a hard selective sweep occurs when positive selection acts on a new mutation that arises in the population. This mutation results in a selective advantage, and as it spreads through the population, it results in reduction of genetic diversity in the surrounding genomic region (Chen *et al.*, 2016). A hard selective sweep results in a more focused selection signature with reduced genetic diversity around the selected variant. Both soft and hard selective sweeps leave distinct genomic patterns that can be detected. These signatures provide insights into the evolutionary history and adaptive changes that have shaped a population's genetic composition.

2.8. Methods of detecting selection signatures

Several studies, including Harris and Meyer (2006), Sabeti *et al.* (2006), and Qanbari and Simianer (2014), have identified factors that must be considered when selecting the most suitable approach for detecting selection signatures using SNP data. In addition, Hohenlohe *et al.* (2010) proposed a decision tree to help identify the most appropriate method for detecting these selective patterns. The decision tree relies on prerequisite factors such as the number of populations being studied, knowledge of substitution class, and the selection mode. Qanbari *et al.* (2014) categorized these methods into two main categories: inter-population statistics and intra-population statistics. The latter involves detecting informative signatures by comparing genomic data within a population using linkage disequilibrium (LD), reduced local variability and site frequency spectrum (SFS) (Weigand & Leese, 2018).

Inter-population statistics detect informative signatures by comparing genomic data between two or more populations or breeds. These methods focus on differentiation between populations therefore, statistics in this group can be divided into (i) single-site differentiation and (ii) haplotype-based differentiation (Bonhomme *et al.*, 2010; Fariello *et al.*, 2013). The commonly used methods are summarized in **Figure 2.2**, followed by a more detailed discussion of the specific intra-population and interpopulation approaches used in this study.

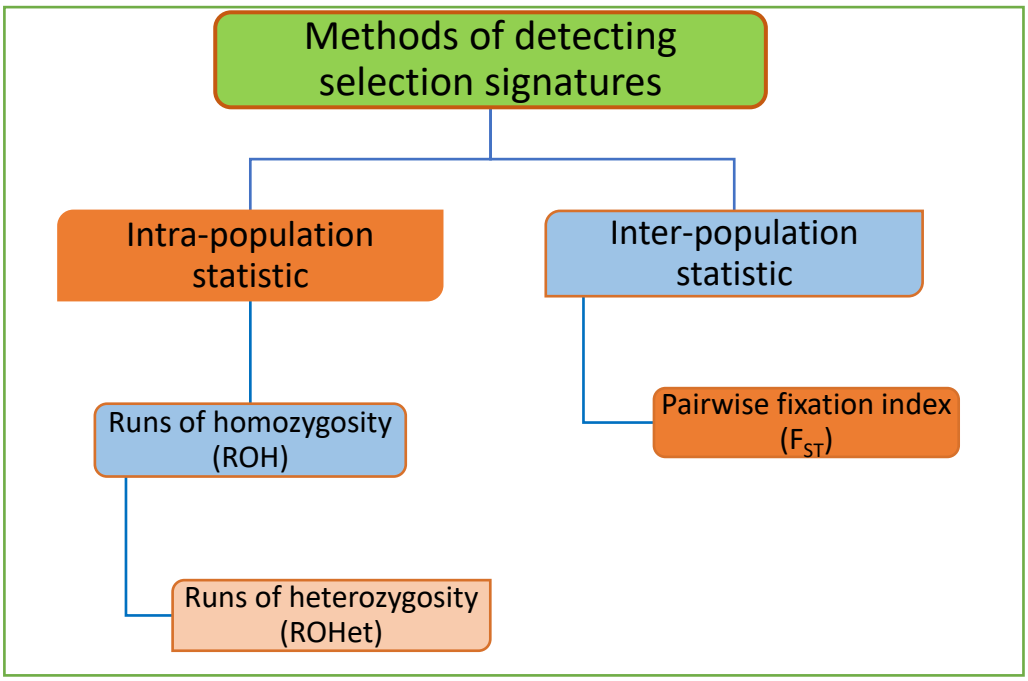


Figure 2.2: A summary of methods to identify selection signatures used in the current study, adapted from Saravanan *et al.,* (2020)

2.9. Intra-population statistics

Genomic regions with a reduced nucleotide diversity or heterozygosity relative to the entire genome within a given population can be identified under reduced local variability using ROH (McQuillan *et al.*, 2008) and ROHet (Biscarini *et al.*, 2020).

2.9.1. Runs of Homozygosity

ROH refer to stretches of homozygous genomic segments that are frequently observed in individuals and populations who share a common ancestor (Gibson *et al.*, 2006). The ROH formation is illustrated in **Figure 2.3** below. According to Almeida *et al.* (2019), ROH segments can serve as a selection footprint, indicating higher levels of homozygosity than the average genome. The analysis of ROH can provide valuable insights into past and recent genetic events within each population, making it a powerful tool for studying demographic evolution, population history, and structure over time (Rebelato & Caetano, 2018). Short ROH segments are typically associated with ancient inbreeding events. As the number of generations increases, the likelihood of recombination events rises, leading to shorter ROH lengths due to the interruption of the formation of long ROH length. (McQuillan *et al.*, 2008).

In farm animals, ROHs are useful in calculating the ROH-based inbreeding coefficients (F_{ROH}) even when pedigree information is unavailable (Ferenčaković *et al.*, 2013). ROH-based inbreeding coefficients

(F_{ROH}) are more accurate than pedigree-based inbreeding coefficients (F_{PED}) and can identify specific genomic regions (selection signatures) with higher precision (Marras *et al.*, 2015; Walsh, 2021). F_{ROH} is a direct measure of the proportion of the genome that exists in a homozygous state due to directional selection or recent inbreeding (Zavarez *et al.*, 2015). It is based on analyzing the actual genetic information rather than relying on pedigree records, which may be incomplete or inaccurate. Their accuracy is explained by the ability of F_{ROH} to provide a higher-resolution assessment of inbreeding by examining the specific genomic regions affected by inbreeding (Marras *et al.*, 2015). Additionally, pedigree-based inbreeding coefficients are typically effective at capturing recent or close inbreeding events but may miss more distant or ancient inbreeding. ROH analysis can detect even relatively ancient ROHs, allowing for a more comprehensive assessment of inbreeding across generations (Ferenčaković *et al.*, 2013). Pedigree-based inbreeding coefficients rely on recorded pedigree information, which may not capture all instances of inbreeding. In many populations, there may be cases of unrecorded or hidden inbreeding due to factors like mating between closely related individuals in small or isolated populations (Szmatoła *et al.*, 2016). ROH analysis can detect these instances by identifying stretches of the genome where an individual has inherited identical copies of genetic material from both parents (Zhao *et al.*, 2015). Furthermore, ROH helps to uncover deleterious variants and reduce the chance of these variants being passed on to the next generation (Biscarini *et al.*, 2020). ROH have high correlation with autozygosity (r~0.7) (McQuillan *et al.*, 2008; Marras *et al.*, 2015), hence can be used to quantify individual autozygosity.

Autozygosity is a genetic phenomenon that arises when two individuals, such as parents, share a common ancestor. In such cases, these individuals pass on chromosomal segments to their offspring that are identical by descent (IBD), resulting in ROH in the offspring's genome (Broman & Weber, 1999). The length, distribution, and frequency of ROH segments in a genome can be influenced by a variety of factors, including mutation rates, linkage disequilibrium, recombination, inbreeding levels, and natural or artificial selection pressures. (Gibson *et al.*, 2006). High-density SNP arrays are a powerful tool for screening the genome for runs of homozygosity (ROH) and identifying IBD haplotypes (Gibson *et al.*, 2006). The presence of ROH provides valuable information about breeding strategies, selective pressures (Purfield *et al.*, 2012), and population bottlenecks (Bosse *et al.*, 2015). ROH analysis has been extensively studied in cattle to estimate inbreeding levels, as evidenced by numerous studies (Sölkner *et al.*, 2010; Purfield *et al.*, 2012; Kim *et al.*, 2013; Curik *et al.*, 2014; Peripolli *et al.*, 2018). Additionally, ROH analysis have been widely used to detect signatures of selection (Yurchenko *et al.*, 2018; Peripolli *et al.*, 2020; Saravanan *et al.*, 2021).

The **Figure 2.3** illustrates the process of ROH formation, with individual A representing the common ancestor of D and E, sire, and dam to offspring F, respectively. The ROH in individual F (depicted in blue),

originated from common ancestor A and was inherited by the parents (D & E) to offspring (individual F). Other colours represent chromosome fragments unrelated to the homologous chromosome of individual A.

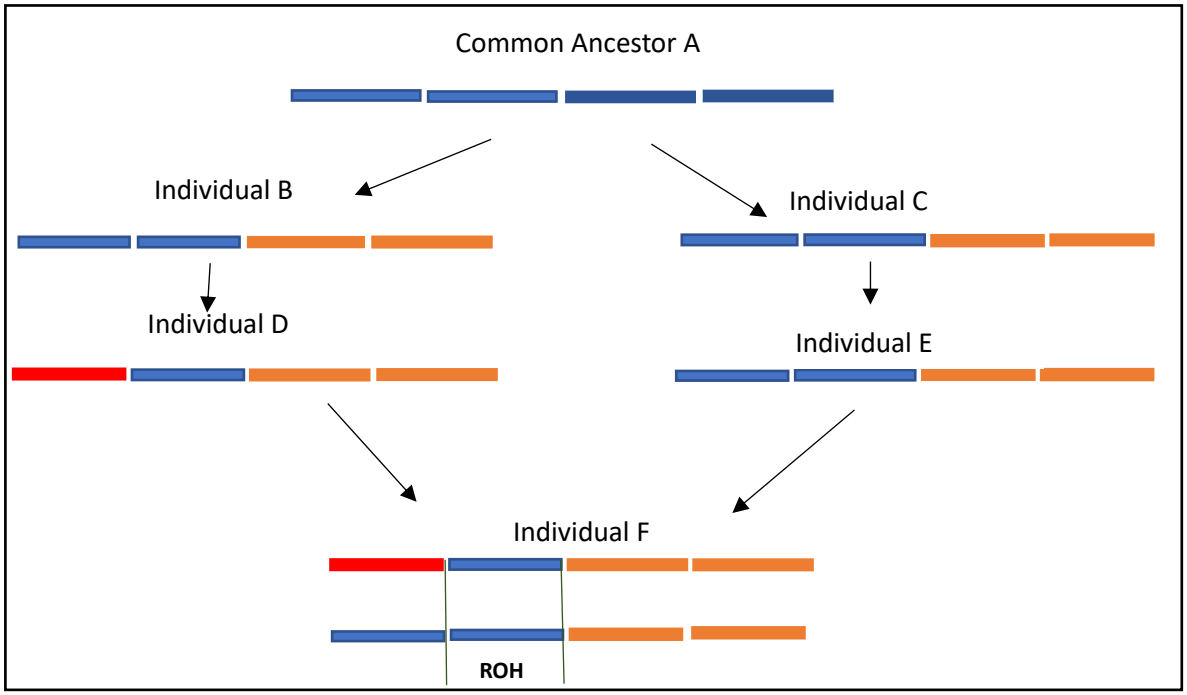


Figure 2.3: The ROH formation in individual F (depicted in blue). This ROH is formed by the combination of homologous segments of the genome inherited from the common ancestor A (Rebelato & Caetano, 2018; Saravanan *et al.,* 2020)

2.9.2. Runs of Heterozygosity (ROHet)

Runs of heterozygosity refer to genomic regions that exhibit high levels of genetic variation, commonly referred to as runs of heterozygosity (ROHet) (Marras *et al.*, 2018). This approach's objective is to pinpoint sections of the genome that show substantial genetic variation. This enables the assessment of population genetic diversity and evolutionary patterns, while also highlighting genomic segments where preserving higher genetic diversity could yield greater advantages (Bizarria dos Santos *et al.*, 2021). Mc Parland *et al.* (2009) suggested that ROHet regions may contain loci that carry essential genes associated with vital functional traits such as heat tolerance, disease resistance, and pest tolerance. This indicated that the analysis of ROHet regions may be essential for identifying critical genetic factors that underlie adaptive traits in populations (Samuels *et al.*, 2016). Therefore, ROHet analysis has the potential to facilitate the study of selection and introgression (Marras *et al.*, 2018). Several studies have employed ROHet analysis to investigate genetic variation in different animal populations, i.e., poultry (Marras *et al.*, 2018; Bizarria dos Santos *et al.*, 2021), sheep (Biscarini *et al.*, 2018; Selli *et al.*, 2021), and cattle (Biscarini *et al.*, 2020;

Lashmar *et al.*, 2022; van Marle-Köster *et al.*, 2022). These studies highlight the utility of ROHet analysis as a powerful tool for studying genetic variation and evolutionary history in diverse animal populations.

2.10. Interpopulation statistics

Wright's fixation index (F_{ST}) is a widely used and powerful statistic in population genetics that allows for the estimation of genetic variability across populations or breeds. F_{ST} measures the differentiation between populations by comparing differences in allele frequencies (Wright, 1949; Nei & Chesser, 1983; Myles *et al.*, 2008). F_{ST} assumes that differences in genetic variants among populations result from variations in geographic selection pressures (Wright, 1949; Oleksyk *et al.*, 2008). Therefore, it is possible to detect selection signatures across multiple loci using the F_{ST} estimate (Akey *et al.*, 2002). The F_{ST} statistic provides an estimate of the degree of differentiation between breeds at any given locus, with values ranging from 0 (no differentiation) to 1 (complete differentiation between populations). High F_{ST} values may suggest positive selection, while low F_{ST} values indicate negative selection between populations (Zhao *et al.*, 2015). Natural selection acts on specific genetic loci, resulting in genomic regions that are highly differentiated between populations due to evolutional forces. This phenomenon can cause a systematic deviation of the F_{ST}, a measure of population differentiation, as described by Akey *et al.* (2002). As a result, genes that contribute to variations in phenotypes across different populations are expected to display significant differences in allele frequencies, as stated by Myles *et al.* (2008).

One of the advantages of using F_{ST} is its ability to perform multi-locus testing, such as SFS or LD-based methods. F_{ST} is SNP-specific, and as such, it can potentially reveal the exact genetic variants that are under selection (Qanbari & Simianer, 2014). However, a primary concern with this approach is that it assumes that all populations under study are derived from the same ancestral group and have the same effective population size. This assumption is often not valid, and as a result, genomic scans based on raw F_{ST} may suffer from false positives and bias, leading to the cryptic structure effect, as described by Price *et al.* (2010). As a solution, Oleksyk *et al.* (2008) proposed using large-population datasets. Another solution is to extend the Lewontin and Krakauer (LK) test, which captures the hierarchical population structure through the kinship matrix (Lewontin & Krakauer, 1973). Additional methods have also been proposed to account for differences in population sizes. For instance, as in the case of the current study, using many SNPs (>500) enhances the efficiency of identifying genetic variation in smaller sample sizes, as suggested by Willing *et al.* (2012).

2.11. Selection signatures for adaptation in African cattle

Nielsen *et al.* (2005) laid the foundation for detecting selection signatures using SNP data. Since then, numerous studies have utilized various methods to detect selection signatures, aiming to unravel genomic regions associated with desirable traits (Cesarani *et al.*, 2018), such as adaptation, production, and

feed efficiency (Yurchenko *et al.*, 2018). While genetic diversity studies typically focus on examining the extent of differentiation among populations, selection signature studies delve deeper to unravel the underlying factors driving this variation (Saravanan *et al.*, 2020).

Several studies have identified selection signatures in various cattle populations. Some of these studies were restricted to analyzing specific chromosomes, for example, studies by Hayes *et al.* (2008) and Prasad *et al.* (2008). A study by Prasad *et al.* (2008) used the EHH method targeting BTA 19 and 29 to identify eight signatures of positive selection in Angus cattle, while Hayes *et al.* (2008) employed the iHS approach and identified candidate regions linked to milk production traits on BTA 6 in Norwegian Red cattle. **Table 2.5** presents candidate genes with verified polymorphisms that are known for their association with adaptive traits in African cattle breeds.

Table 2.5: A summary of potential candidate genes reported for adaptive traits in cattle breeds found in Africa

Trait(s)	Breed(s)	Candidate genes	Reference
Heat tolerance	Bonsmara and Nguni	*HSPH1, HSPB9*	Kooverjee *et al.* (2022)
	Nguni	*HSPB9*	Makina *et al.* (2015)
	Afrikaner, Drakensberger, and Nguni	*KRT24, KRT25, KRT26, KRT27, KRT28 HSPB9*	Zwane *et al.* (2019)
	N'Dama, Ankole, Boran, Ogaden, N´Dama and Kenana	*PPP3CA, PPP2R5E PPP4R3B, IGF-I, HSF5*	Taye *et al.*, (2017)
	N´Dama, Baoulé, Somba, Nadoba, Lagune, Borgou, Sudanese Fulani, Kur	*EDNRB, TRSP1 KRTAP8-1*	Gautier *et al.* (2009)
Coat color	Nguni	*MC1R*	Makina *et al.* (2015)
	Ankole, Boran, N´Dama Kenana, and Ogaden	*SLC45A2, MLPH, RAB17, RAB37, ATRN*	Taye *et al.*, (2017)
Response to scarce feed supply	Bonsmara and Nguni	*DNAJB13*	(Kooverjee *et al.*, 2022)
	Brune de l'Atlas, Guelmoise, Cheurfa, Oulmes	*GH1*	(Ben-Jemaa *et al.*, 2020)
Tick resistance	Afrikaner, Nguni, Drakensberger and Bonsmara	*TNFAIP8L3 SLC25A48*	Makina *et al.* (2015)
	Nguni	*GPR142, PRKG1, LRBA, SMIM12, FER, LINGO2*	Mapholi *et al.* (2016)
	Boran, Ogaden, and Kenana	*BOLA,TNFAIP8L3, SLC25A48, KRT33A, PRG3, SLC45A2, MLPH, MC5R, TGM1, TGM3*	Taye *et al.*, (2018)
Immune response	N'Dama	*STOM, SLC40A1, SBDS, EPB42 and RPS26*	Kim *et al.* (2017)
	Lagunaire, N'Dama, Zebu Bororo, Borgou, Kuri, Lobi	*IRAK2, MLH1, KIF15 and DDX39B*	Goyache *et al.* (2021)
	Afrikaner, Nguni, Drakensberger , Bonsmara	*MTPN, CYM, CDC6, CDK10, TNS4, SLC25A48, NDUFA12, EBFI*	Makina et al. (2015)
Oxidative stress response	Bale, Choke, Semian, Afar, Boran, Ogaden	*CBF2A, SLC23A, PLCB1, SLC26A2, CLCA2*	Terefe *et al.*, (2022)

Several selection signature studies have been conducted with special focus on regions associated with adaptation to various climatic conditions i.e., (Yurchenko *et al.*, 2018; Igoshin *et al.*, 2019). A microsatellites-based study identified the *CXCR4* gene associated with trypanotolerance and immune system development in West African taurine cattle (Dayo *et al.*, 2009). Taye *et al.*, (2018) identified key genes associated with gastrointestinal parasite resistance (*TNFAIP3*, *TNFAIP83*, and *DMBT1*), and *DMBT1* plays a critical role in mucosal defence, cellular immune defense, and epithelial differentiation. The *SLC45A2* and *MATP* genes have been shown to be involved in modulating melanogenesis in melanocytes, influencing morphological traits like skin thickness, hair cells, and coat colour (Taye *et al.*, 2018).

The *MC1R* gene on BTA 18 was identified as differentially selected among Nguni and Holstein cattle, which might suggest the multi-colour skin patterns seen in these breeds (Makina *et al.*, 2015). The same study also identified several candidate genes associated with antigen recognition, a crucial process in immune response. These genes were found in Afrikaner (*MTPN* and *CYM*) and Nguni (*CDK10* and *KCNBI*) populations. Additionally, other genes such as *NDUFA12*, *ALOX15B*, and *ALOX12B* were identified in Bonsmara breed. Other recent studies were conducted using SA local breeds to detect signatures of selection relating to environmental adaptation (Zwane *et al.*, 2021; Kooverjee *et al.*, 2022). In Nguni and Bonsmara cattle breeds, candidate genes associated with heat stress (*DNAJB13, HSPB9, HSP90AB1, HSPH1, and DNAJC5B*) were found to be under positive selection (Kooverjee *et al.*, 2022). The same study also identified *Methyltransferase like 21A* (*METTL21A*) on BTA 2 in the Nguni population directly involved in methylation of a heat shock protein *HSPA1-K561*. The interaction of this gene with heat shock protein demonstrates the contribution of DNA methylation processes in responding to heat stress (Jakobsson *et al.*, 2016).

2.12. Conclusion

Genotypic data enables the exploration of population structure, genomic diversity, and selection signatures within cattle populations. Multiple methods can be employed to identify selection signatures, and for enhanced accuracy, a combination of approaches is often recommended. In this study, ROH and ROHet methods were utilized to detect within-breed selection signatures, while F_{ST} was applied to identify signatures of selection across breeds. By examining the selection signatures in the DRB, NGI, and TUL breeds, the current study has potential to unravel the selection forces that have acted on these populations. These regions can be linked to traits such as heat tolerance, feed efficiency, stature and body size, and DNA methylation. Findings from studies of this nature offer the opportunity to identify causative variants that contribute to the inherent adaptation observed in various South African indigenous cattle.

CHAPTER 3: MATERIALS AND METHODS

3.1. Introduction

The genotypic data used in the current study was provided by SA Stud Book with consent from the Drakensberger, Nguni, and Tuli Cattle Breeders' Society of South Africa. The study was approved by the Ethics Committee in the Faculty of Natural and Agricultural Sciences at the University of Pretoria (NAS235/2022).

3.2. Materials

A total of 1 722 genotypes representing three breeds were available for the study as shown in **Table 3.1.** The animals were originally selected for the establishment of a reference population for these breeds for genomic selection and are representatives of the breeds originating from different regions of South Africa. They were genotyped with the GeneSeek® Genomic Profiler™ 150K bovine SNP panel as part of the Beef Genomic Program (BGP) at the Agricultural Research Council, Biotechnology Platform (ARC-BTP).

Table 3.1: Summary of the number and origin of the genotypes used in the current study.

Breed	Number of genotypes	N of breeders participated
Drakensberger (DRB)	1 125	42
Nguni (NGI)	381	44
Tuli (TUL)	216	38
Total	1722	

N = Number of breeders participated

The GGP 150K BeadChip, included approximately 140 113 SNPs spread across 29 autosomes and sex chromosomes. The genotypic data were obtained in map and ped format and subsequently converted using PLINK into binary files.

3.3. Methods

3.3.1. Quality Control

The markers and samples underwent standard quality control (QC) procedures using the PLINK version 1.9 software package (Purcell *et al.*, 2007). Only autosomal markers mapped to the UMD 3.1.1 bovine reference genome were used (Zorc *et al.*, 2019). In addition, SNPs with duplicated genomic positions were excluded. Individuals with uncalled genotypes (call rate < 90%), markers with low call rates (≤95%) and low minor allele frequencies (MAF <1%) were excluded from the analysis as shown in **Table**

3.2. Following these QC steps, a total of 1 708 animals and 122 632 common SNPs remained for subsequent downstream analyses.

Table 3.2: Animal and marker-based quality control, highlighting the number of animals and markers before quality control, individuals and variants removed, and the number of animals and markers remaining after quality control.

Breed	N before QC	Sample call rate(< 90%)	N after QC	SNPs before QC	SNP call rate (< 95%)	SNPs after QC
DRB	1 125	8	1 117	132 271	9 639	122 632
NGI	381	4	377	132 271	9 202	123 609
TUL	216	2	214	132 271	7 357	124 914
Total	**1 722**	**14**	**1708**			

N=Number of animals; QC=Quality Control; DRB=Drakensberger; NGI=Nguni; TUL=Tuli

3.3.2. Within-breed genetic diversity parameters.

The H_O and H_E, MAF, and LD were estimated using PLINK version 1.9 (Purcell *et al.*, 2007). Specifically, the --het command was used to compute H_O and H_E, while the --freq command was used to compute MAF. PLINK's --r^2 command was executed to estimate the within-breed and across-breed persistence of linkage disequilibrium (LD) on a per-chromosome basis for SNPs. Microsoft Excel (2019) was used to estimate the average H_O and H_E, MAF, and LD respectively. Furthermore, the SNP-by-SNP-based --het command was implemented to derive F_{IS}, based on the discrepancy between the expected and observed numbers of homozygous genotypes.

3.3.3. Population structure

The SNP-based genetic relatedness between individuals was estimated using the Genome-Wide Complex Trait Analysis (GCTA) software v1.24 (Yang *et al.*, 2011). To compute the first three principal components and generate eigenvalues and eigenvectors, a genetic relationship matrix was created using the commands "--make-grm" and the command "--pca 3". The first principal component separated samples into three groups. The first and second principal components were plotted using Microsoft Excel (2019) for visualization. To determine the between-breed relatedness and population sub-structure of the animals, ADMIXTURE v1.23 (Alexander *et al.*, 2009) was used. The most suitable K-value was determined based on the lowest cross-validation error estimate, obtained by including the --cv command while running ADMIXTURE for 4 K-values (2-5). Two output files (.Q and .P) were generated, and model-based plots were subsequently created using GENESIS version 0.2.3 (Buchmann & Hazelhurst, 2014).

3.3.4. Detection of ROH and ROHet

Consistent with the proposal by Meyermans *et al.* (2020), the pruning of minor allele frequency (MAF) was done before the runs of heterozygosity (ROHet) analysis. No MAF pruning was performed before ROH and F_{ST} analysis to avoid the removal of SNPs that are either fixed or highly homozygous (Meyermans *et al.*, 2020). Removal of low-MAF SNPs before ROHet analysis is necessary to avoid underestimation of length of segments. In general, low-MAF SNPs typically exhibit homozygous genotypes, resulting in identification of longer ROHet, and sometimes more but shorter ROHet. Additionally, no LD pruning was performed before either ROH or ROHet analyses. The consecutive-SNP-based detection method of the detectRUNS R-package v 0.96 (Biscarini *et al.*, 2018; Saravanan *et al.*, 2021), was used for the identification of ROH and ROHet. To minimize the risk of spurious ROH detection due to low SNP densities, stringent criteria illustrated in **Table 3.3** were applied, and at least one SNP per 75 kb was allowed (Biscarini *et al.*, 2020). The minimum number of SNPs constituting a segment (l = 52 DRB; l = 54 NGI; l = 50 TUL) was estimated using a formula by Purfield *et al.* (2012):

$$l = \frac{log_e \frac{\alpha}{n_s n_i}}{log_e (1-\overline{het})},$$

where n_s and n_i were the numbers of SNPs and individuals, respectively, α (set to 0.05) represented the proportion of false positives identifications, and het was the mean SNP-wide heterozygosity.

The mean number and length of ROH were calculated per animal and the ROH were grouped into four classes: ROH<4Mb, 4≤ROH<8Mb, 8≤ROH<16Mb, and greater than 16Mb.

To detect runs of heterozygosity (ROHet) in the three beef cattle breeds, the consecutive-SNP-based method was applied using the detectRUNS R-package (Biscarini *et al.*, 2018). The parameters used for calling ROHet segments are summarized in **Table 3.3**. The lengths of ROHet were classified into five classes: ROHet<0.25Mb, 0.25≤ROHet<0.5, 0.5≤ROHet<1, 1≤ROHet≤2, and ROHet>2 Mb (Biscarini *et al.*, 2020). The number, proportion and mean length of ROHet were calculated within each length category per breed.

Due to significant divergence among various African populations, some groups may inadequately capture the diversity of others (Gurdasani *et al.*, 2015). A substantial portion of very rare variants might contribute significantly to African genomes, and considering that the identification of a single SNP depends largely on allele frequency, rare SNPs are more likely to go undiscovered (Clark *et al.*, 2005). Ascertainment bias, which arises during the selection of SNPs for genotyping, can impact the accuracy of selection signature analysis. SNP panels designed based on European breeds may not adequately represent the genetic variation in African breeds. This issue is particularly challenging in studies involving African

populations since the majority of arrays are favor European breeds, leading to an underrepresentation of African diversity (Colonna *et al.*, 2014). The site frequency spectrum (SFS) varies in populations with African ancestry, with a higher proportion of selection candidates occurring at frequencies of 30% or less (Liu *et al.*, 2013; Willemse, 2019). Therefore, lowering thresholds helps mitigate the impact of ascertainment bias, allowing for a more inclusive analysis that considers the unique genetic makeup of African cattle.

To identify within-breed selection signatures using ROH and ROHet approaches, the detectRUNS R-package v 0.96 (Biscarini *et al.*, 2018) was used to identify genomic regions most commonly associated with ROH and ROHet (ROH and ROHet islands). The percentage of occurrence of a SNP in ROH and ROHet was calculated by counting the number of times the SNP appeared in those ROH and ROHet across animals within each population. The function "topRuns" from the detectRUNS package was used to retrieve the most common runs using a threshold value of 0.2 and 0.3 for ROH and ROHet, respectively. This means that a ROH and ROHet have to be present in at least 20% and 30%, respectively, of each population to be included in an ROH and ROHet island (signature of selection).

Table 3.3: Parameters for ROH and ROHet detection

Parameter	ROH	ROHet	References
Minimum length	1000Kb	10Kb	Biscarini *et al.* (2020)
The maximum gap between SNPs	1000Kb	1000Kb	Lashmar *et al.* (2022)
Threshold	0.05	n/a	Lashmar *et al.* (2022)
Maximum opposing run	0	3	Biscarini *et al.* (2020)
SNPs with missing genotypes	2	2	Biscarini *et al.* (2020)

ROH = Runs of Homozygosity; Kb = Kilobase; ROHet = Runs of Heterozygosity; n/a = not applicable

3.4. Between Population Signatures of Selection (F_{ST})

The F_{ST} was used to estimate the genetic differentiation between populations using the formula proposed by Nei (1986):

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Where:

F_{ST} = the reduction in heterozygosity due to the structure of the population,

H_S = Average heterozygosity in the subpopulation

H_T = Average heterozygosity in the metapopulation

The F_{ST} values range from 0 to 1, where 0 indicates no differentiation, and 1 means complete differentiation between populations (Moradi *et al.*, 2012). Higher F_{ST} values show greater allelic frequency

divergence at a locus, indicating a greater selection divergence among the populations under investigation (Maiorano *et al.*, 2018). To compare the three beef cattle populations (DRB vs NGI, DRB vs TUL, and NGI vs TUL), PLINK's --fst command was used (Caivio-Nasner *et al.*, 2021) to determine genomic regions that exhibit increasing differentiation. Negative F_{ST} values were set to 0 since they do not have any biological interpretation (Akey *et al.*, 2002). Pairwise F_{ST} values of SNPs were plotted relative to their position within each autosome. The top 0.1% F_{ST} values were used to represent selection signatures (significant SNPs related to genes affecting adaptive and economically important traits) according to Kijas *et al.* (2012), Zhao *et al.* (2015), and Saravanan *et al.* (2021). Prior to annotation, windows of 250kb downstream and upstream of the significant SNPs were investigated to verify overlapping gene segments (Maiorano *et al.*, 2018).

3.4.1. Annotation of genomic regions

After identifying significant regions under selection, annotation was done using the National Centre for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) based on the UMD3.1.1 bovine genome assembly (https://www.ncbi.nlm.nih.gov/assembly/228231). Furthermore, functional annotations such as Gene Ontology (GO) keywords (biological process, cellular component, and molecular function) and pathways were assigned to genes via ShinyGO 0.77 (http://bioinformatics.sdstate.edu/go/) and the PANTHER version 16 software (http://www.pantherdb.org) (Mi *et al.*, 2021).

CHAPTER 4: RESULTS

4.1. Introduction

This study aimed to identify selection signatures within and between the South African Drakensberger (DRB), Nguni (NGI), and Tuli (TUL) populations using genome-wide SNP data. The genotypes were analyzed to quantify the genomic regions under selection based on conserved runs of homozygosity (ROH), runs of heterozygosity (ROHet), and F_{ST}-based differentiation of SNPs.

4.2. Genetic diversity parameters

In **Table 4.1,** the within-population genetic diversity parameters for the three breeds under study were summarized. The Nguni breed displayed the lowest H_E value of 0.320±0.001, while the Drakensberger breed exhibited the highest H_E value of 0.347±0.005. Similarly, the Drakensberger breed had the highest observed H_O value of 0.342±0.016, whereas the Nguni breed had the lowest H_O value of 0.320±0.015. The mean MAF varied between 0.238±0.152 in the Nguni breed and 0.262±0.144 in the Drakensberger breed. Among the three breeds, the Drakensberger breed exhibited the highest and the only positive inbreeding coefficient (F_{IS} = 0.014±0.034).

Table 4.1: Summary statistics (mean±standard deviation) of the within-population genetic diversity parameters across three cattle breeds in South Africa

Breed	N	H_E	H_O	MAF	LD	F_{IS}
DRB	1 117	0.347±0.005	0.342±0.016	0.262±0.144	0.4485	0.014±0.034
NGI	377	0.320±0.001	0.320±0.015	0.238±0.152	0.4606	-0.0006±0.048
TUL	214	0.332±0.0003	0.334±0.017	0.246±0.148	0.4608	-0.0079±0.053

MAF = Minor Allele Frequency; F_{IS} = Inbreeding coefficient; N = Number of animals; LD = linkage disequilibrium

4.3. Runs of Homozygosity (ROH) and F_{ROH} Distribution

A total of 82 871 ROH were identified across all three breeds, with the highest number of ROH (59 232) observed in DRB, while the lowest (10 710) was detected in TUL (**Table 4.2**). The mean number of ROH per animal was highest in the DRB population (mean±standard deviation: 51.82±21.01 with a range of 2-255) and lowest in NGI (mean±standard deviation: 36.09±12.84 with a range of 2-122). The mean ROH length was highest in DRB (3.86±4.40Mb), and lowest in NGI (2.31±2.83Mb). The average genome-wide F_{ROH} values were highest in the DRB , followed by TUL, and NGI populations.

Table 4.2: Total number of ROH per breed, FROH, mean number of ROH per individual and mean ROH length per population.

Breed	Total nROH	Mean $F_{ROH±SD}$	Mean±SD $nROH_{ind}$	Maximum ROH length (Mb)	$MeanROH_{length}$ (Mb)
DRB	59232	0.081±0.046	51.82±21.01	71.9	3.86
NGI	12929	0.033±0.024	36.09±12.84	65.5	2.31
TUL	10710	0.074±0.031	47.94±15.36	53.8	3.74
Total	**82 871**				

SD = Standard deviation; nROH = number of ROH; $nROH_{ind}$ = number of ROH per individual

The longest ROH were observed on BTA 6 in both DRB and TUL, with a length of 71.9 Mb and 53.8 Mb harbouring 4 296 and 3 230 SNPs, respectively. In the NGI population, the longest ROH (65.5 Mb) was identified on BTA 20 harbouring 2 963 SNPs.

The mean ROH length per category varied across populations (**Table 4.3**). The DRB breed exhibited the highest mean ROH length (1.94±0.78 Mb) in the shortest length category (ROH<4Mb), while the TUL population had the highest mean ROH length (5.65±1.14 Mb) in the category of 4≤ROH<8Mb. In the NGI population, the mean ROH length was highest at 11.21±2.18 Mb in the 8≤ROH<16Mb category and 23.6±7.37 Mb in the >16Mb category.

Table 4.3: Summary statistics of runs of homozygosity (ROH) and mean FROH identified in different length categories for the Drakensberger (DRB), Nguni (NGI), and Tuli (TUL) cattle populations.

ROH Categories	DRB			NGI			TUL		
	nROH	$MeanROH_{length}$ ±SD (Mb)	Mean F_{ROH}	nROH	$MeanROH_{length}$ ±SD (Mb)	Mean F_{ROH}	nROH	$MeanROH_{length}$ ±SD (Mb)	Mean F_{ROH}
ROH<4Mb	42711	1.94±0.78	0.055	11751	1.68±0.59	0.018	8011	1.84±0.74	0.049
4≤ROH<8Mb	10104	5.63±1.12	0.038	650	5.56±1.14	0.020	1493	5.65±1.14	0.035
8≤ROH<16Mb	4882	10.97±2.17	0.024	365	11.21±2.18	0.018	897	11.00±2.17	0.021
>16Mb	1535	23.01±7.44	0.020	163	23.6±7.37	0.021	309	22.62±6.52	0.019

nROH=number of runs of homozygosity; SD=standard deviation; ROH_{length}=length of runs of homozygosity

F_{ROH} = ROH-based Inbreeding coefficient

In the ROH<4Mb and 4≤ROH<8Mb categories, the DRB exhibited the highest mean F_{ROH} values (F_{ROH} = 0.055 and 0.038, respectively). In the ROH>16Mb category, a higher F_{ROH} value (F_{ROH} = 0.021) was observed in the NGI, while TUL had the lowest (F_{ROH} = 0.019). The highest F_{ROH} values were observed

for ROH shorter than 4 Mb (ROH<4Mb = 0.055 and 0.49) for DRB and TUL, respectively. Conversely, the NGI had higher mean values for F_{ROH} calculated based on longer ROH(ROH>16Mb = 0.021).

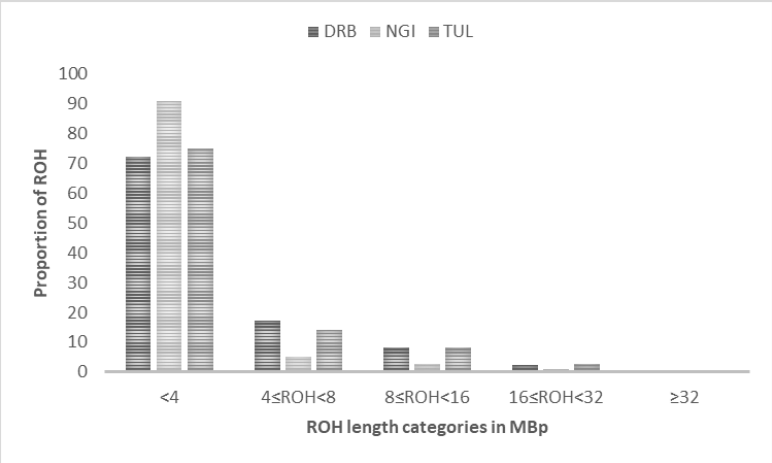


Figure 4.1: The proportion of runs of homozygosity within four ROH length categories for the three populations.

The highest proportion of ROH across all breeds was observed within the shortest length category (ROH<4Mb) with proportions ranging from 71.45% (DRB) breed to 91.33% (NGI) breed (**Figure 4.1**).

4.4. Runs of Heterozygosity (ROHet)

A total of 134 633 runs of heterozygosity (ROHet) were identified across all three cattle populations. Notably, the DRB had by far the highest number of ROHet with 97 162, followed by NGI with 21 103, and the TUL with the lowest count of 16 368. The average length of ROHet detected across all autosomes was comparable in all categories among the breeds. The longest ROHet was observed on BTA10 (2.84 Mb) for the DRB, and BTA7 for both NGI (1.60 Mb) and TUL (1.48 Mb), respectively. There was a higher occurrence of short ROHet fragments (ROHet<0.25Mb) compared to other size categories and the number of ROHet decreased as the length of ROHet increased. Furthermore, only one ROHet exceeding 2 Mb was observed in DRB on BTA10. The number and the average length of ROHet for the three cattle populations were grouped into length categories as shown in **Table 4.4.**

Table 4.4: The number and mean length of runs of heterozygosity (ROHet) identified for DRB, NGI, and TUL within different length categories.

Category (Mb)	DRB		NGI		TUL	
	nROHet	ROHetlength±SD	nROHet	ROHetlength±SD	nROHet	ROHetlength±SD
ROH<0.25	76 802	0.14±0.07	16 915	0.13±0.07	13 138	0.13±0.07
0.25≤ROHet<0.5	19 655	0.31±0.05	4 053	0.32±0.06	3 132	0.31±0.05
0.5≤ROHet<1	632	0.69±0.16	97	0.65±0.15	79	0.66±0.17
1≤ROHet≤2	72	1.23±0.09	38	1.40±0.08	19	1.21±0.09
ROHet>2	1	2.84±0.00	-	-	-	-
Total	**97 162**		**21 103**		**16 368**	

nROHet=Number of Runs of heterozygosity; SD=Standard Deviation.

It was observed that the highest and lowest number of ROHet were on BTA25 (7 279) and BTA 21 (906) in DRB, on BTA25 (1 494) and BTA 18 (189) in NGI, and BTA 26 (1 343) and BTA 22 (112) in TUL respectively

4.5. Population structure analyses

4.5.1. Principal Component Analysis

The genetic relatedness between the individuals from three different populations is shown in **Figure 4.2**, where PCA1 accounted for 5.4% and PCA2 explained 1.7% of the total variation among the three cattle populations. The three populations were separated into three different clusters.

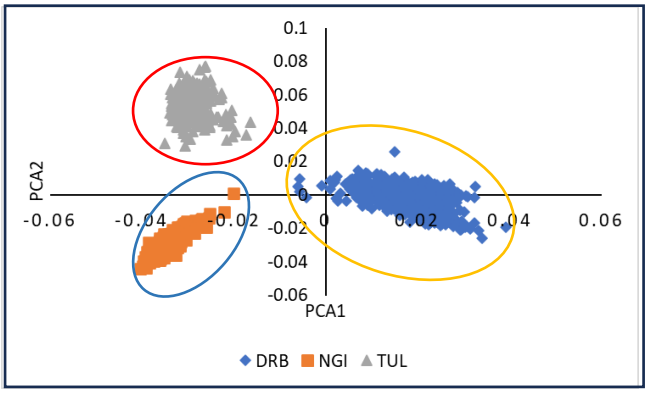


Figure 4.2: The genetic relatedness of the three SA beef cattle populations plotted against PCA1 (5.4%) and PCA2 (1.7%).

4.5.2. Admixture

The ideal number of ancestral populations (K) determined by the lowest cross-validation error (CV=0.52032) was 5. The cross-validation score for each K value ranging from 1 to 7 was plotted (showing a reduction in cross-validation error values) in **Figure 4.3**.

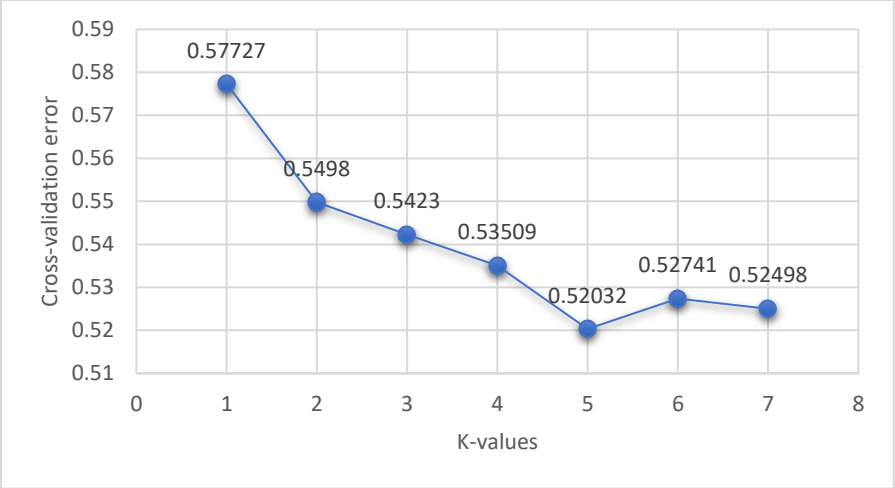


Figure 4.3: A cross-validation plot, showing the estimated cross-validation error rate for different K values (K=1-7)

The proportion of individuals in each of the breeds inferred by the ADMIXTURE was presented in **Table 4.5**. with an ideal K = 5.

Table 4.5: The ancestral population proportions for each individual

Predefined populations	**Inferred clusters**				
	1	2	3	4	5
DRB	0.29839	**0.39010**	0.24451	0.03663	0.03037
NGI	0.01563	0.03816	0.01211	**0.89196**	0.04213
TUL	0.01644	0.04417	0.01452	0.09900	**0.82590**

The model-based admixture plots (**Figure 4.4**), depict the genetic relatedness of the three cattle populations based on their ancestral origins. The ideal K value grouped the populations into three different clusters, with some level of heterogeneity observed from K = 4 and 5.

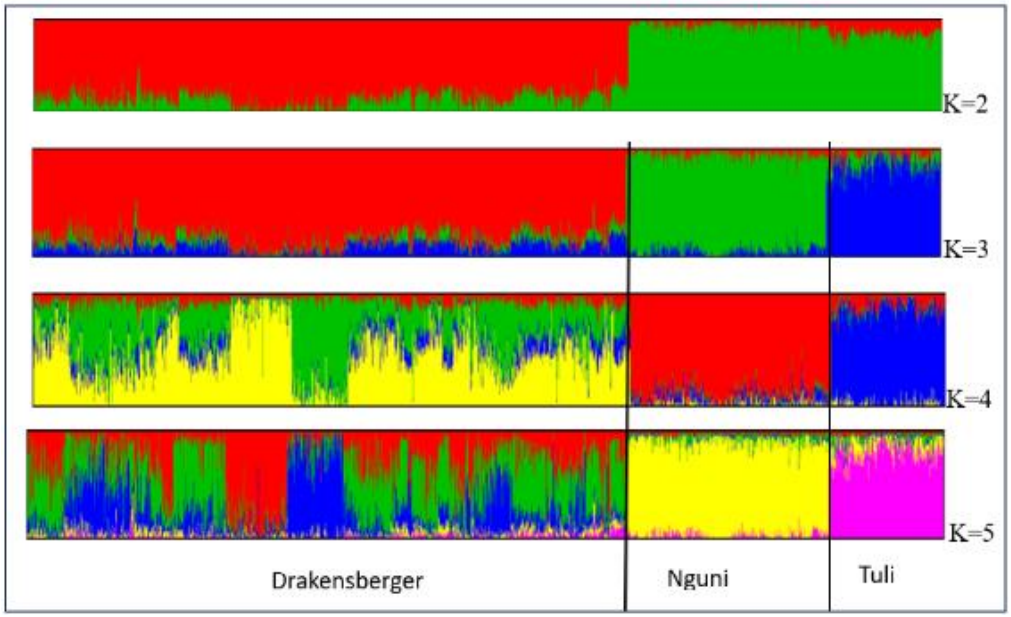
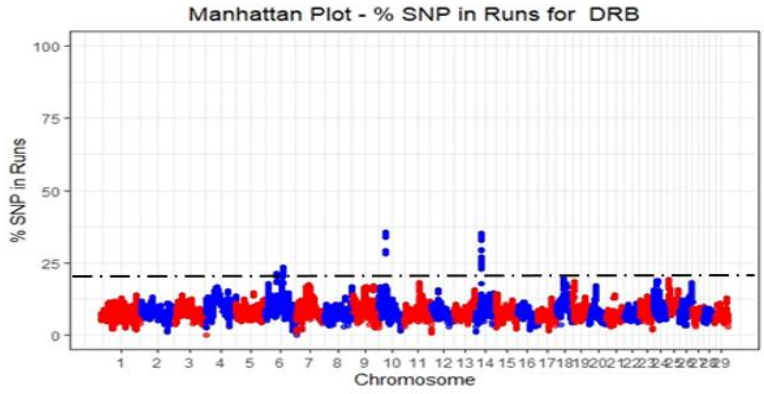


Figure 4.4: Admixture plots showing ancestral populations proportions for each individual at K=5

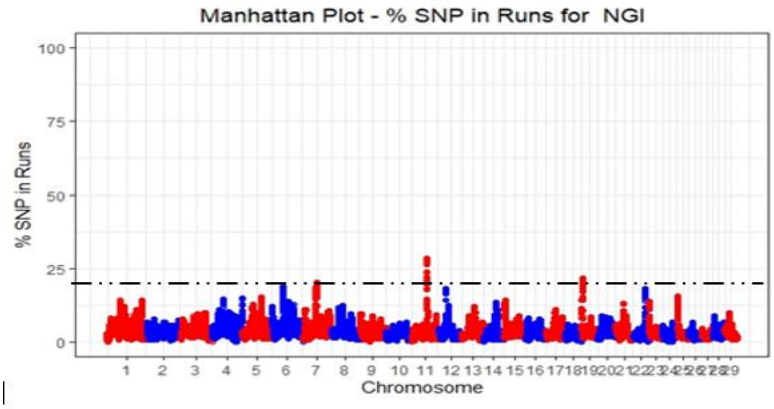
4.6. Selection signatures within populations

4.6.1. Runs of homozygosity (ROH) approach

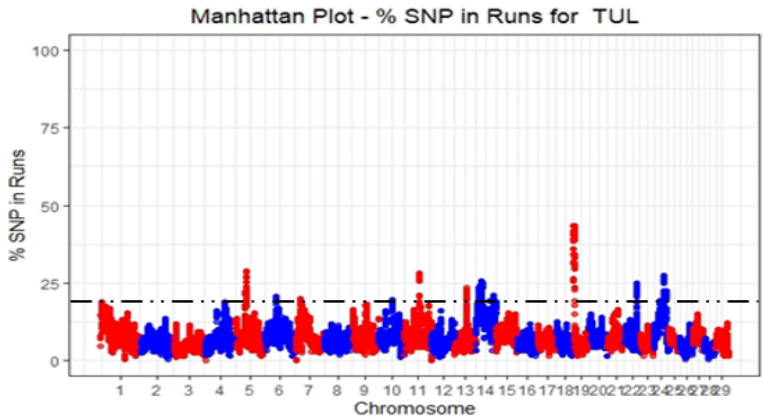
Genomic regions characterized by the highest frequency of ROH occurrence are referred to as ROH islands. Regions that occur in more than 20% of the individuals within the population are considered common regions and can also harbour genes associated with traits of economic importance in cattle production. A total of 43 genomic regions (DRB = 5; NGI = 5; and TUL = 33) were identified as ROH islands in the three populations. These regions were located on BTA6, 10 and 14 in DRB, BTA 7, 11, and 19 in NGI, and BTA 4, 5, 6, 7, 10, 14, 19, 22, and 24 in TUL. Among these regions, only 5 did not contain any coding sequences, specifically on BTA 11 and 19 in NGI. The subsequent annotation of these candidate regions identified 250 (DRB = 29; NGI = 12; & TUL = 208) candidate genes linked to several biological processes. Manhattan plots illustrating the thresholds for all three populations are shown in **Figures 4.5 A, B**, and **C**.



A



B



C

Figure 4.5: A, B and C. Manhattan plots for ROH islands distribution across the autosomes in Drakensberger, Nguni, and Tuli Populations. The dotted line represents a threshold of 20% in DRB, NGI, and TUL respectively.

A total of 250 genes enriched in 14 biological processes were found in 43 candidate regions across all populations. Only genes associated with adaptation were presented in **Table 4.6**. Some of these genes were previously related to body size and stature (*FAM110B, UBXN2B, CYP7A1, SDCBP, NSMAF, TOX, PLAG1, LYN, TMEM68, RPS20, MOS, CHCHD7*, and *KCNIP4),* reproduction traits (*RPS20, CORIN, TXK, SPIRE1,* and *PLAG1*), immunity and adaptation traits (*CTNNA2, MYC, CYSTM1, CD14, WDPCP, and FKBP4).* Genes which were shared by two or three populations were highlighted in bold. Only candidate genes associated with adaptation were listed in **Table 4.6**, and a full list of other genes identified in the candidate regions were listed in Addendum.

Table 4.6: Functional enrichment analysis for genes identified within ROH candidate regions.

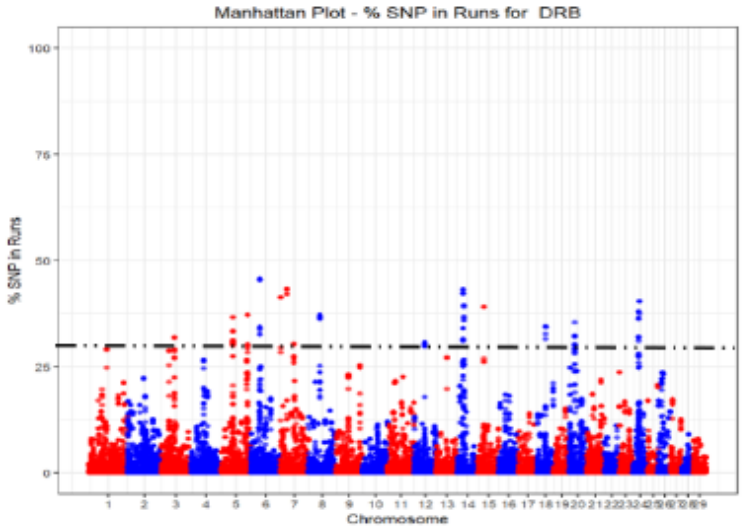
BREED	Gene ID	Ensembl Gene ID	BTA	Position (Mb)	Description
DRB	*SLIT2*	ENSBTAG00000005108	6	39.79	slit guidance ligand 2
	KCNIP4	ENSBTAG00000047743	6	40.25	potassium voltage-gated channel interacting protein 4
	TXK	ENSBTAG00000005055	6	66.75	TXK tyrosine kinase
	CORIN	ENSBTAG00000002199	6	66.27	corin, serine peptidase
	UBXN2B	ENSBTAG00000009138	14	24.58	UBX domain protein 2B
	CYP7A1	ENSBTAG00000005287	14	24.66	cytochrome P450, family 7, subfamily A, polypeptide 1
	SDCBP	ENSBTAG00000019910	14	24.72	syndecan binding protein
	TOX	ENSBTAG00000004954	14	24.94	thymocyte selection associated high mobility group box
TUL	*LYZ2*	ENSBTAG00000026088	5	44.36	lysozyme C-2
	LYZ1	ENSBTAG00000046511	5	44.39	lysozyme (renal amyloidosis)
	LYZ3	ENSBTAG00000046628	5	44.42	lysozyme 3
	LYZ	ENSBTAG00000026779	5	44.50	lysozyme
	RGS20	ENSBTAG00000003454	14	21.90	regulator of G protein signaling 20
	TMEM68	ENSBTAG00000005893	14	23.03	transmembrane protein 68
	LYN	ENSBTAG00000020034	14	23.13	LYN proto-onco, Src family tyrosine kinase
	MOS	ENSBTAG00000019145	14	23.29	MOS proto-onco, serine/threonine kinase
	PLAG1	ENSBTAG00000004022	14	23.33	PLAG1 zinc finger
	CHCHD7	ENSBTAG00000049910	14	23.37	coiled-coil-helix-coiled-coil-helix domain containing 7
	FAM110B	ENSBTAG00000050550	14	24.36	family with sequence similarity 110 member B
	UBXN2B	ENSBTAG00000009138	14	24.58	UBX domain protein 2B
	CYP7A1	ENSBTAG00000005287	14	24.66	cytochrome P450, family 7, subfamily A, polypeptide 1
	SDCBP	ENSBTAG00000019910	14	24.72	syndecan binding protein
	NSMAF	ENSBTAG00000008958	14	24.76	neutral sphingomyelinase activation associated factor
	TOX	ENSBTAG00000004954	14	24.94	thymocyte selection associated high mobility group box
	KIF2B	ENSBTAG00000017345	19	3.77	kinesin family member 2B
	SPIRE1	ENSBTAG00000010542	24	42.88	spire type actin nucleation factor 1
	CA8	ENSBTAG00000017529	14	25.95	carbonic anhydrase 8
	CACNA2D3	ENSBTAG00000013117	22	45.92	calcium voltage-gated channel auxiliary subunit alpha2delta 3
	LYPLA1	ENSBTAG00000004243	14	21.99	lysophospholipase 1
	XKR4	ENSBTAG00000044050	14	22.64	XK related 4
	TGS1	ENSBTAG00000005898	14	23.07	trimethylguanosine synthase 1
	ERC2	ENSBTAG00000010029	22	44.54	ELKS/RAB6-interacting/CAST family member 2
	MC5R	ENSBTAG00000009143	24	43.54	melanocortin 5 receptor

Bold = Genes observed in more than one population; Mb = Megabase pairs; DRB = Drakensberger; NGI = Nguni; TUL = Tuli

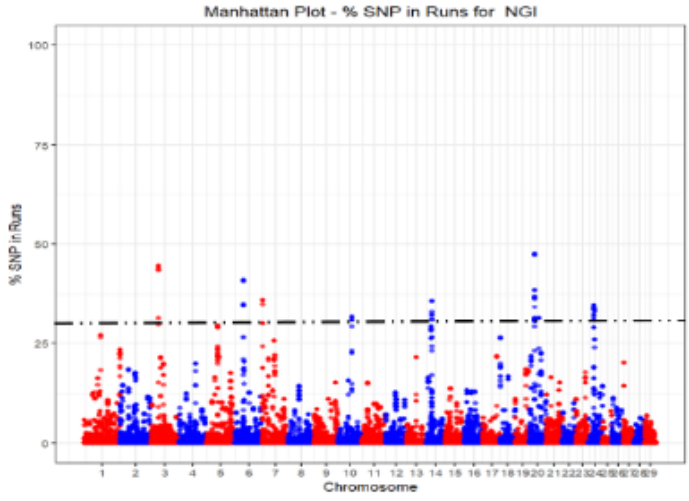
4.6.2. Runs of heterozygosity (ROHet) approach

Regions of the genome characterized by a high frequency of runs of heterozygosity (ROHet) are referred to as "ROHet islands." These common regions, found in over 30% of individuals within each population, are of significance and could potentially harbour genes linked to economically important traits

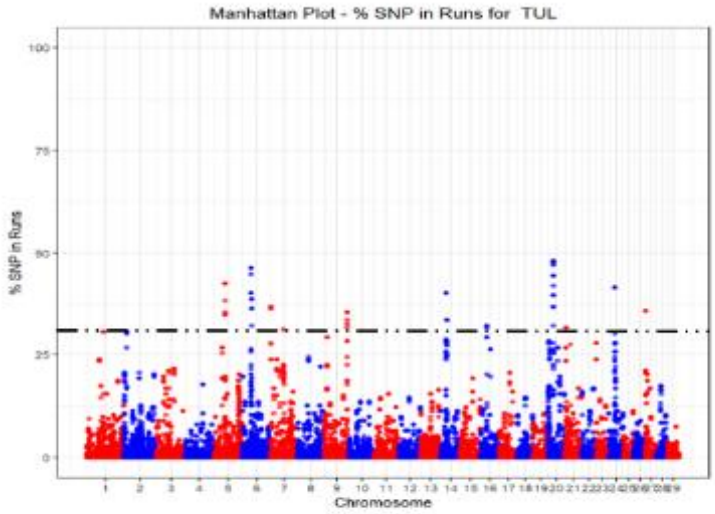
in beef production. The present study identified 143 genomic regions as ROH islands across the three populations, with 54 in DRB, 35 in NGI, and 54 in the TUL populations, respectively. Annotation of these candidate regions yielded 148 candidate genes, where 49 were found in the DRB, 39 in NGI, and 60 in TUL. For a visual representation of the thresholds for all three populations, the results were plotted in Figures **4.6 A, B,** and **C**.



A



B



C

Figure 4.6: A, B, and C Manhattan plots for ROHet islands distribution across the autosomes in Drakensberger, Nguni, and Tuli Populations. The dotted line represents a threshold of 30% in DRB, NGI, and TUL respectively.

A total of 148 genes enriched in 14 biological processes were found in these 143 candidate regions. Only genes relevant to this study were included in **Table 4.7** below. Some of these genes were related to reproduction traits (*NELL2, TMEM181*) and immunity and adaptation traits (*DXT1, ELMO3,* and *ADAMTS12*). Some genes particularly those highlighted in bold were common in two or three populations. Only candidate genes associated with adaptation were listed in **Table 4.7**, and a full list of other genes identified in the candidate regions were listed in Addendum.

Table 4.7: List of candidate genes identified within genomic regions considered to be selection signatures using the ROHet approach in all three beef breeds.

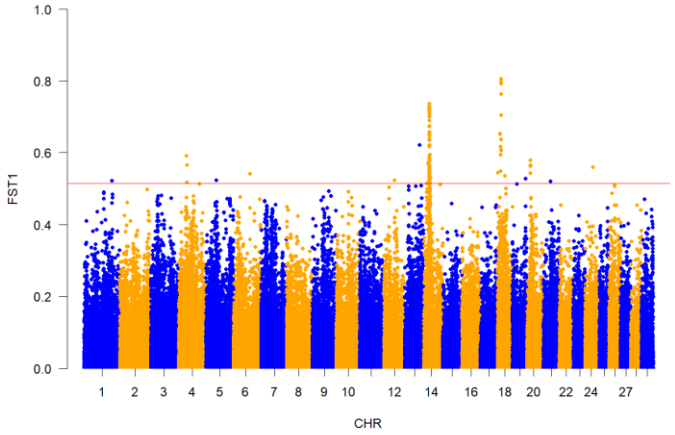
BREED	Gene ID	Ensembl Gene ID	BTA	Position (Mb)	Description
DRB	*PRDM15*	ENSBTAG00000021253	1	142.07	*PR/SET domain 15*
	LAMTOR5	ENSBTAG00000014970	3	33.03	*late endosomal/lysosomal adaptor, MAPK and MTOR activator 5*
	CRACR2A	ENSBTAG00000018940	5	106.15	*calcium release activated channel regulator 2A*
	PRMT8	ENSBTAG00000009447	5	106.32	*protein arginine methyltransferase 8*
	FKBP4	ENSBTAG00000007605	5	106.95	*FKBP prolyl isomerase 4*
	FAM184B	ENSBTAG00000005932	6	37.18	*family with sequence similarity 184 member B*
	WDPCP	ENSBTAG00000005151	11	61.72	*WD repeat containing planar cell polarity effector*
	LYN	ENSBTAG00000020034	14	23.13	*LYN proto-onco, Src family tyrosine kinase*
	FAM110B	ENSBTAG00000050550	14	24.37	*family with sequence similarity 110 member B*
	CHD7	ENSBTAG00000021841	14	26.36	*chromodomain helicase DNA binding protein 7*
	E2F4	ENSBTAG00000012063	18	34.83	*E2F transcription factor 4*
	ELMO3	ENSBTAG00000001286	18	34.83	*engulfment and cell motility 3*
	FHOD1	ENSBTAG00000032427	18	34.86	*formin homology 2 domain containing 1*
	ATP6V0D1	ENSBTAG00000014553	18	35.05	*ATPase H+ transporting V0 subunit d1*
	AGRP	ENSBTAG00000014556	18	35.09	*agouti related neuropeptide*
NGI	*R3HDM2*	ENSBTAG00000018361	5	56.08	*R3H domain containing 2*
	STAC3	ENSBTAG00000018358	5	56.19	*SH3 and cysteine rich domain 3*
	NDUFA4L2	ENSBTAG00000031503	5	56.21	*NDUFA4 mitochondrial complex associated like 2*
	FAM184B	ENSBTAG00000005933	6	37.18	*family with sequence similarity 184 member B*
	CYSTM1	ENSBTAG00000016596	7	51.36	*cysteine rich transmembrane module containing 1*
	SRA1	ENSBTAG00000001449	7	51.72	*steroid receptor RNA activator 1*
	CD14	ENSBTAG00000015032	7	51.76	*CD14 molecule*
	SDCBP	ENSBTAG00000019910	14	24.73	*syndecan binding protein*
	NSMAF	ENSBTAG00000008958	14	24.77	*neutral sphingomyelinase activation associated factor*
	DTX1	ENSBTAG00000016738	17	61.09	*deltex E3 ubiquitin ligase 1*
	ADAMTS12	ENSBTAG00000012558	20	39.87	*ADAM metallopeptidase with thrombospondin type 1 motif 12*
TUL	*NELL2*	ENSBTAG00000032183	5	35.48	*neural EGFL like 2*
	FAM184B	ENSBTAG00000005934	6	37.18	*family with sequence similarity 184 member B*
	CYSTM1	ENSBTAG00000016595	7	51.36	*cysteine rich transmembrane module containing 1*
	HBEGF	ENSBTAG00000021766	7	51.53	*heparin binding EGF like growth factor*
	TMEM181	ENSBTAG00000010838	9	95.03	*transmembrane protein 181*
	FAM110B	ENSBTAG00000050551	14	24.37	*family with sequence similarity 110 member B*
	CLVS1	ENSBTAG00000043978	14	26.85	*clavesin 1*
	ADAMTS12	ENSBTAG00000012559	20	39.87	*ADAM metallopeptidase with thrombospondin type 1 motif 12*

Bold = Genes observed in more than one population; Mb = Megabase pairs; DRB = Drakensberger; NGI = Nguni; TUL = Tuli

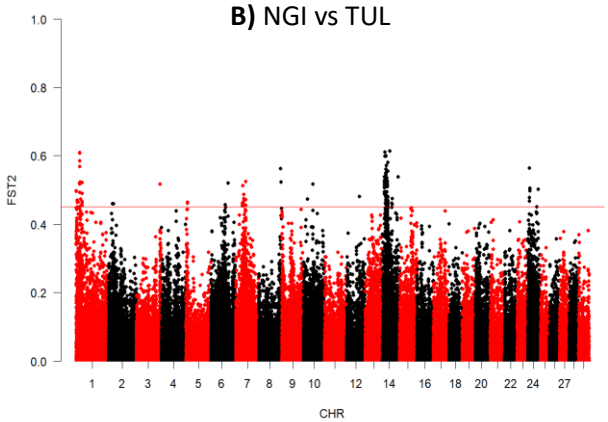
4.6.3. Wright's Fixation Index approach

The study detected signals of divergent selection by assessing F_{ST} values between cattle populations. The regions with F_{ST} values falling within the top 0.1% of the empirical F_{ST} distribution were considered signatures of positive selection. Using this approach, a total of 357 genomic regions undergoing divergent selection were identified. The genome-wide distribution of selection signatures within the 3 breed pairs was visualized by plotting the F_{ST} values against genomic positions for each SNP as shown in **Figure 4.7 A, B & C** below.

A) DRB vs NGI



B) NGI vs TUL



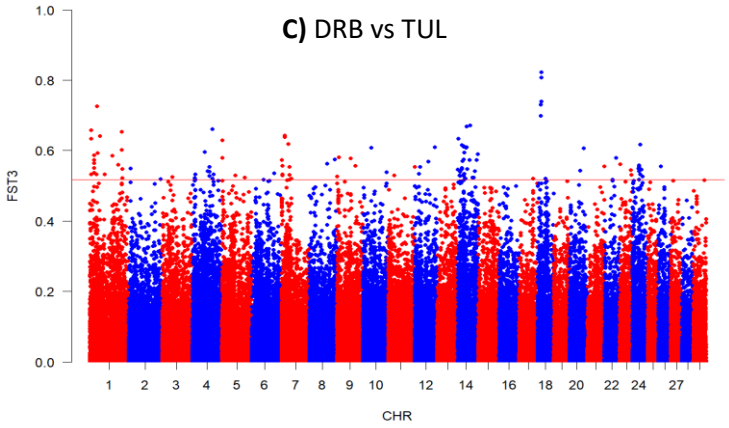


Figure 4.7: Manhattan plots for the distribution of top 0.1 FST values within the autosomes for three breed combinations: A) DRB vs NGI, B) NGI vs TUL, and C) DRB vs TUL

The analysis found 499 genes associated with these divergently selected regions. Subsequent functional analyses of these candidate genes unveiled their potential roles in various traits of economic importance across the three cattle breeds. These included genes linked to immunity and adaptation traits (*LYZ1, LYZ2, LYZ3, NMU, LMAN2, CDK10,* and *CBFA2T3*), and coat colour (e.g., *TUBB3* and *MC1R*) in these indigenous cattle breeds. Some genes, particularly those highlighted in bold (**Table 4.8**), were found to be common in two or three populations. Table 4.8 exclusively presents candidate genes associated with adaptation, while a comprehensive list of other genes identified in the candidate regions was provided in the Addendum.

Table 4.8: A selected list of candidate genes identified within genomic regions revealed as selection signatures using F_{ST} values.

Breed	Gene ID	Ensembl Gene ID	BTA	Position (Mb)	Description
DRBvsNGI	*LYZ2*	ENSBTAG00000026088	5	44.37	*lysozyme C-2*
	LYZ1	ENSBTAG00000046511	5	44.39	*lysozyme (renal amyloidosis)*
	LYZ3	ENSBTAG00000046628	5	44.42	*lysozyme 3*
	LYZ	ENSBTAG00000026779	5	44.51	*lysozyme*
	UBXN2B	ENSBTAG00000009138	14	24.59	*UBX domain protein 2B*
	CYP7A1	ENSBTAG00000005287	14	24.66	*cytochrome P450, family 7, subfamily A, polypeptide 1*
	SDCBP	ENSBTAG00000019910	14	24.73	*syndecan binding protein*
	TOX	ENSBTAG00000004954	14	24.95	*thymocyte selection associated high mobility group box*
	CHD7	ENSBTAG00000021841	14	26.36	*chromodomain helicase DNA binding protein 7*
	CLVS1	ENSBTAG00000043978	14	26.85	*clavesin 1*

	CBFA2T3	ENSBTAG00000010927	18	14.05	*CBFA2/RUNX1 partner transcriptional co-repressor 3*
	SPG7	ENSBTAG00000012041	18	14.46	*SPG7 matrix AAA peptidase subunit, paraplegin*
	CPNE7	ENSBTAG00000012215	18	14.50	*copine 7*
	CDK10	ENSBTAG00000033333	18	14.57	*cyclin dependent kinase 10*
	FANCA	ENSBTAG00000001906	18	14.61	*FA complementation group A*
	SPIRE2	ENSBTAG00000040356	18	14.65	*spire type actin nucleation factor 2*
	MC1R	ENSBTAG00000023731	18	14.71	*melanocortin 1 receptor*
	TUBB3	ENSBTAG00000023730	18	14.71	*tubulin beta 3 class III*
	MYLK3	ENSBTAG00000014818	18	15.03	*myosin light chain kinase 3*
DRBvsTUL	*MTERF2*	ENSBTAG00000010144	5	70.25	*mitochondrial transcription termination factor 2*
	RGS20	ENSBTAG00000003454	14	21.90	*regulator of G protein signaling 20*
	LYPLA1	ENSBTAG00000004243	14	21.99	*lysophospholipase 1*
	LYN	ENSBTAG00000020034	14	23.13	*LYN proto-onco, Src family tyrosine kinase*
	MOS	ENSBTAG00000019145	14	23.30	*MOS proto-onco, serine/threonine kinase*
	TOX	ENSBTAG00000004955	14	24.95	*thymocyte selection associated high mobility group box*
	CHD7	ENSBTAG00000021842	14	26.36	*chromodomain helicase DNA binding protein 7*
	CLVS1	ENSBTAG00000043979	14	26.85	*clavesin 1*
	PREX2	ENSBTAG00000022169	14	31.98	*phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2*
	NCOA2	ENSBTAG00000020312	14	33.84	*nuclear receptor coactivator 2*
	XKR9	ENSBTAG00000048138	14	34.36	*XK related 9*
	LY96	ENSBTAG00000008864	14	37.24	*lymphocyte antigen 96*
	CBFA2T3	ENSBTAG00000010928	18	14.05	*CBFA2/RUNX1 partner transcriptional co-repressor 3*
	SPG7	ENSBTAG00000012042	18	14.46	*SPG7 matrix AAA peptidase subunit, paraplegin*
	CPNE7	ENSBTAG00000012216	18	14.50	*copine 7*
	CDK10	ENSBTAG00000033334	18	14.57	*cyclin dependent kinase 10*
	FANCA	ENSBTAG00000001907	18	14.61	*FA complementation group A*
	SPIRE2	ENSBTAG00000040357	18	14.65	*spire type actin nucleation factor 2*
	MC1R	ENSBTAG00000023732	18	14.71	*melanocortin 1 receptor*
	TUBB3	ENSBTAG00000023731	18	14.71	*tubulin beta 3 class III*
NGIvsTUL	*KCNE1*	ENSBTAG00000001150	1	1.04	*potassium voltage-gated channel subfamily E regulatory subunit 1*
	HOXD1	ENSBTAG00000015840	2	20.72	*homeobox D1*
	HOXD3	ENSBTAG00000004835	2	20.74	*homeobox D3*
	HOXD4	ENSBTAG00000039581	2	20.76	*homeobox D4*
	HOXD8	ENSBTAG00000049845	2	20.78	*homeobox D8*
	HOXD9	ENSBTAG00000016033	2	20.79	*homeobox D9*
	HOXD10	ENSBTAG00000016030	2	20.79	*homeobox D10*
	HOXD11	ENSBTAG00000033330	2	20.80	*homeobox D11*
	HOXD12	ENSBTAG00000004314	2	20.81	*homeobox D12*
	NMU	ENSBTAG00000044161	6	71.03	*neuromedin U*

LMAN2	ENSBTAG00000008034	7	38.83	*lectin, mannose binding 2*
RGS20	ENSBTAG00000003455	14	21.90	*regulator of G protein signaling 20*
LYPLA1	ENSBTAG00000004244	14	21.99	*lysophospholipase 1*
XKR4	ENSBTAG00000044050	14	22.64	*XK related 4*
UBXN2B	ENSBTAG00000009139	14	24.59	*UBX domain protein 2B*
CYP7A1	ENSBTAG00000005288	14	24.66	*cytochrome P450, family 7, subfamily A, polypeptide 1*
TOX	ENSBTAG00000004956	14	24.95	*thymocyte selection associated high mobility group box*
CHD7	ENSBTAG00000021843	14	26.36	*chromodomain helicase DNA binding protein 7*
CLVS1	ENSBTAG00000043980	14	26.85	*clavesin 1*
CDH7	ENSBTAG00000014488	24	10.78	*cadherin 7*

Bold = Genes observed in more than one population; Mb = Megabase pairs; DRB = Drakensberger; NGI = Nguni; TUL = Tuli

4.7. Common/shared genomic regions

The study has also identified shared candidate genes within genomic regions identified by all three methods as selection signatures. For instance, the *FAM184B* gene, present in all three breeds, was identified through ROH analysis on BTA 6 at 37.18 Mb. Similarly, a genomic segment on BTA 14, spanning from 23.13 Mb to 26.85 Mb, has been identified as a common region across all three methods, harbouring genes such as *LYN*, *SDCBP*, *CHD7*, and *CLVS1*. Furthermore, numerous other genes were found to be common to both the ROH and F_{ST} approaches on BTA 14. These genes include *LYPA1*, *TCEA1*, *MYC*, *FBXO32*, *ATAD2*, *ZHX1*, *FAM83A*, *HAS2*, *CEBPD*, *SPIDR*, *MCM4*, *RGS20*, and *TOX*.

CHAPTER 5: DISCUSSION

5.1. Introduction

The study used genotypes generated using the bovine 150K SNP array to detect selection signatures in populations of three South African indigenous breeds. The Bovine 150K SNP chip has proven to be an effective tool in numerous studies for the identification of ROH, ROHet and/or selection signatures in diverse cattle breeds (Cheruiyot *et al.*, 2018; Lashmar *et al.*, 2018; Caivio-Nasner *et al.*, 2021; Van Marle-Köster *et al.*, 2021; Kooverjee *et al.*, 2022). The primary objectives of this study were to analyze the genotypes of the Drakensberger (DRB), Nguni (NGI), and Tuli (TUL) breeds to quantify within and between population genetic variation and identify potential selection signatures within the population. This study employed conserved runs of homozygosity and runs of heterozygosity to identify selection signatures within populations. Additionally, Wright's Fixation Index (F_{ST}) was used to detect signatures of selection between populations. Furthermore, candidate genes associated with selection were annotated within the identified regions of interest.

Investigating the genetic variation between populations presented certain challenges due to varying sample sizes across the three breeds. To overcome this limitation, the populations were initially analyzed individually to maximize the information obtained from each breed. Following quality control of individual populations, a total of 122 632 common SNPs to all three populations were used for a comparative analysis.

Little to no research has been conducted on these three breeds collectively to investigate selection signatures using ROH, ROHet and F_{ST} approaches. Furthermore, there is a scarcity of similar studies on South African cattle breeds that have explored runs of homozygosity and runs of heterozygosity as a method of detecting signatures of selection. As a result, the availability of local studies to compare with the outcomes of this research is limited. Studies that have included related parameters include, Makina *et al.,* (2015), Lashmar *et al.,* (2018); Lashmar *et al.,* (2022); Van Marle-Köster *et al.,* (2021); King *et al.,* (2022), and Kooverjee *et al.,* (2022). The current study is the first comprehensive analysis to use runs of homozygosity and runs of heterozygosity to identify selection signatures in South African cattle breeds.

5.2. Genomic diversity parameters

Genomic tools have been used to study the genetic diversity of SA beef cattle populations. Studies based on SNP arrays indicated moderate heterozygosity levels for the Drakensberger (0.250 - 0.360), Nguni (0.280 - 0.338), and Tuli (0.23 - 0.34) (Makina *et al.*, 2014; Zwane *et al.*, 2016; Lashmar *et al.*, 2018; Van Marle-Köster *et al.*, 2021; King *et al.*, 2022; Kooverjee *et al.*, 2022). Likewise, the results of this study indicated a moderate level of genetic diversity in all three populations. The level of expected heterozygosity in the DRB population aligned with those reported by Lashmar *et al.* (2018) (H_E = 0.347). However, it was slightly lower in DRB than the results reported by Van Marle-Köster *et al.* (2021) (H_E = 0.364), possibly

due to a larger sample size used in the current study. Notably, higher levels of expected heterozygosity were observed compared to studies by Makina *et al.* (2014) and Zwane *et al.* (2016) in DRB (H_E = 0.25 and 0.30) and NGI (H_E = 0.28 and 0.23) respectively, likely attributed to larger sample sizes and the use of higher-density SNP array in the current study. The H_E results of the current study were comparable with those of other studies on indigenous cattle breeds of Africa. For instance, King *et al.* (2021) reported mean H_E of 0.303 and 0.298 for Mozambican Tete and Landim breeds, while for Ethiopian Gofa, Nuer and Sheko breeds, the values were 0.302, 0.313 and 0.309, respectively (Meseret *et al.*, 2020). Additionally, Tanzanian Sukuma, Tarime and Maasai cattle populations showed mean H_E values of 0.396, 0.395 and 0.390, respectively (Msalya *et al.*, 2017)

Among the studied populations, DRB cattle showed the highest observed heterozygosity (H_O = 0.342), while NGI had the lowest (H_O = 0.320). The H_O in the DRB aligned with the results reported by Van Marle-Köster *et al.* (2021), where DRB (0.360) displayed the highest H_O and NGI (H_O = 0.338) had the lowest. Slightly lower H_O values (H_O = 0.315) in the Nguni population reported by King *et al.* (2022) may be attributed to the use of a lower SNP density (80K SNP array) and a smaller sample size compared to this present study (H_O = 0.320). In the NGI population, the H_E was equal to the H_O (H_O = H_E = 0.320), suggesting random mating (Mburu & Hanotte, 2005). In addition, this could be due to the highly purposeful sampling where founder animals genotyped for BGP were selected based on their EBVs, hence only preferred animals were selected (SA Nguni Breeders' Society, 2016), potentially resulting higher chance of relatedness.

The slightly higher average observed (DRB: 0.342, TUL: 0.334) than expected heterozygosity (DRB: 0.347, TUL: 0.332) can be attributed to the lack of relatedness among animals within these populations. The higher diversity values observed in the DRB compared with the other two breeds, likely reflects its larger population size. Among the populations involved in the current study, the DRB (14 071) has the highest number of animals in production in the country, followed by NGI (8 525) and TUL (6 484) as reported by the SA Stud Book (2022).

Negative F_{IS} values observed in NGI (-0.0006) and TUL (-0.0079) may suggest random sampling error or controlled mating of unrelated individuals which diminished the likelihood of inbreeding (Frankham *et al.*, 2002). Despite small discrepancies between H_E and H_O across all populations, sufficient diversity exists both within and between these breeds, supporting their potential for development. Moreover, it has been noted that ascertainment bias can impact genetic diversity parameters such as H_E which rely on allele frequency (Clark *et al.*, 2005). The absence of rare SNPs during genotyping of non-discovery breeds may lead to an inflation in the heterozygosity of the polymorphic SNPs, while

simultaneously causing an underestimation of heterozygosity across the entire set of SNPs due to missing SNPs (Clark *et al.,* 2005).

The average MAF from the current study was lower for NGI than previously reported by (Sanarana *et al.*, 2021), but comparable for NGI (0.233) and TUL (0.242) breeds (King *et al.*, 2022). Furthermore, this study, like other previous studies by King *et al.* (2022) and Lashmar *et al.* (2018), highlights the presence of ascertainment bias in South African indigenous beef breeds as indicated by low MAF values (DRB = 0.262; NGI = 0.238 and TUL = 0.246). This bias could result from the exclusion of these breeds during the development of commercial SNP genotyping panels that favour European *Bos taurus* populations. The differences in levels of MAF in this study could be attributed to the varying proportions of European taurine ancestry within each population (Makina *et al.*, 2016).

5.3. Runs of homozygosity

In the study by Rocha *et al.* (2023), it was reported that there is an interplay between the sample size, the parameters set to define an ROH and the resulting number of identified ROHs. This observation aligns with the results of the current study, where the DRB (59 232) exhibited the highest number of ROH, followed by NGI (12 929) and TUL (10 710) in decreasing order of population sizes.

A high frequency of ROH < 4Mb (range: 0.715 for DRB to 0.913 for NGI) compared to other categories was detected in all three populations. According to Moravčíková *et al.* (2018), the distribution of ROH is correlated with the number of generations stemming from the common ancestor, and the length thereof has a negative correlation with the time of co-ancestry (Mastrangelo *et al.*, 2017). This occurs due to recombination events in each generation that break long homozygous segments into smaller haploblocks as the number of generations increases. As a result, the high frequency of short ROH (ROH< 4Mb) may reflect a strong ancient relationship within a population (Peripolli *et al.*, 2018) or recent admixture leading to the breakdown of longer ROH (Liu *et al.*, 2021). The results of this study align with results from various international studies on cattle, such as those reported by Marras *et al.* (2015), Szmatoła *et al.* (2016), and Peripolli *et al.* (2018). These studies consistently observed an abundance of shorter ROH. However, it is worth noting that there are discrepancies in the criteria used to define ROH, making it challenging to compare ROH studies across different research projects. This lack of a standardized consensus for defining thresholds across studies has also been pointed out by Howrigan *et al.* (2011). Additionally, it is important to highlight the report by Ferenčaković *et al.* (2013), who previously indicated that medium SNP arrays tend to overestimate the number of shorter segments (ROH<4Mb) due to their limited sensitivity to detecting these shorter segments.

The ROH lengths of 10Mb, 5Mb, and 2.5Mb would be associated with 5, 10, and 20 generations, respectively, according to the estimation by Howrigan *et al.* (2011). The ROH can allow for the estimation

of the number of generations involved in the formation of these runs which would be impossible using pedigree data, particularly in cases where pedigree recording is a major concern. The results of this study are similar to those reported by Van Marle-Köster *et al.* (2021), Lashmar *et al.* (2022), and King *et al.* (2022) (albeit with slightly higher values) for the proportions of ROH observed in different length categories for these indigenous breeds. The TUL breed (0.290) showed the highest proportion of ROH > 16Mb relative to the NGI and DRB. This suggests that recent inbreeding events may have occurred more frequently in this population. Despite the current low proportions of ROH (ROH >16 Mb) ranging from 0.101 in NGI to 0.249 in TUL breeds, routine monitoring is needed when developing genomic selection pipelines for these breeds. It is also crucial to emphasize that the lower counts of long ROH in this study could be attributed to the stringent parameters not permitting any heterozygous calls within a ROH region. This precaution was taken to prevent the potential overestimation of the presence of long ROH segments (Peripolli *et al.*, 2018).

The ROHs longer than 16Mb were estimated to have been formed less than three generations ago whereas those less than 8Mb were formed six generations ago (Cardoso *et al.*, 2020; Mastrangelo *et al.*, 2020). The lower abundance of ROH found in the TUL populations may suggest that a larger effective population size was preserved over generations (King *et al.*, 2022). It is also worth noting that animals with identical homozygous genome lengths in the current study (the DRB and NGI: 251.54Mb) displayed different numbers of ROH. The variation in ROH among animals with the same genome length has previously been associated with distinct genetic distances from a common ancestor by Mészáros *et al.* (2015). Normally, when examining animals sharing the same homozygous genome length, it can be deduced that those with fewer ROH possess a greater proportion of longer genetic segments, indicating a closer genetic relationship with the common ancestor compared to those with a higher ROH count. However, the results of the current study reported the opposite because the NGI displayed fewer ROH counts compared to DRB and had the least proportion of long ROH counts.

Since Sanga cattle result from the interbreeding of taurine and indicine breeds, it is expected that their genetic makeup could introduce disruptions in homozygous genome stretches (Purfield *et al.*, 2012). Therefore, using ROH as a tool for trait association in these breeds might be beneficial. For example, in the present study, the presence of shared homozygous segments in up to 37.11%, 28.39%, and 44.39% of the sampled DRB, NGI, and TUL populations, respectively, may indicate the potential fixation of specific segments due to selection pressures.

The genomic inbreeding coefficient derived from runs of homozygosity (F_{ROH}) in different length categories is a highly reliable metric for assessing autozygosity, providing insights into both recent and ancient inbreeding patterns (Ferenčaković *et al.*, 2013). In the ROH<4Mb category, higher F_{ROH} values,

such as DRB (F_{ROH} = 0.055) and TUL (F_{ROH} = 0.049), indicate ancient inbreeding (Hulsegge *et al.*, 2022). Conversely, NGI displayed the highest F_{ROH} values in the ROH>16Mb category suggesting a potential occurrence of recent inbreeding (Mastrangelo *et al.*, 2016), albeit with slightly lower values. Despite slightly lower F_{ROH} values in the ROH>16Mb category for all breeds, it is crucial to closely monitor mating patterns to prevent further recent inbreeding. The F_{ROH} values align with the distribution of ROH in the DRB and TUL populations, showing decreasing F_{ROH} values with increasing ROH length.

5.4. Runs of heterozygosity

Although ROHet have been linked to specific regions, they remain relatively unexplored compared to ROH (Tsartsianidou *et al.*, 2021). These regions have been associated with loci that prevent the detrimental effects of continuous homozygosity, favouring the advantages of heterozygotes. The benefits thereof were reported in immune-related genes, as well as in traits related to productivity and reproduction (Chen *et al.*, 2022; Ruan *et al.*, 2022). Typically, ROHet are concentrated in genomic regions associated with disease resistance, where increased diversity can aid populations in addressing potential health challenges, especially when facing novel threats (Sanglard *et al.*, 2021).

To date, only a few ROHet studies have been conducted in livestock species, exhibiting varying number of ROHet per individual ranging from 9.9 for cattle (Biscarini *et al.*, 2020), 28.3 for sheep (Tsartsianidou *et al.*, 2021), and 52.2 for horses (Bizarria dos Santos *et al.*, 2021). The mean number of ROHet detected per animal in the current study was (mean±standard deviation: 86.81±11.73 in the DRB population, 55.98±10.00 in the NGI population, and 76.49±12.33 in the TUL population). In contrast to the results reported by Biscarini *et al.* (2020) and Szmatoła *et al.* (2023) where the number of ROH identified (3 332 and 113 177) were higher than ROHet (1471 and 81 924), respectively, the current study displayed the opposite trend. These disparities in findings could potentially be attributed to variations in parameter adjustments used for ROHet detection or the significant genetic diversity inherent within the populations under investigation in the current study.

Of particular significance is the observation that the DRB population exhibited the highest prevalence of ROHet, constituting over 72.17% of the total ROHet identified in the present study. These results align with expectations, as the DRB breed has previously been characterized as an admixed breed with a genetic composition comprising 46% European, 38% African taurine, and 15% indicine ancestry (Makina *et al.*, 2016). This genetic makeup likely contributes to the higher observed heterozygosity within this population. Furthermore, the substantial disparity in ROHet prevalence between the DRB population and the other two breeds (NGI and TUL) may be attributed to differences in sample sizes across these populations.

Unlike several previous studies, such as those conducted by Biscarini *et al.,* (2020), Lashmar *et al.*, (2022), and van Marle Koster *et al.,* (2022), which often reported higher proportions for regions within the ROHet category 0.5 - 1 Mb, this study yielded different results. This study showed that ROHet≤0.25 Mb exhibited the highest mean proportions across all populations. This observation provides evidence of selection favouring heterozygotes, potentially attributable to the admixture present within Sanga cattle populations.

5.5. Population structure analyses

The PCA and model-based admixture results consistently demonstrated a clear differentiation among the three Sanga cattle breeds: DRB, NGI, and TUL, as evident in their distinct separation along PCA1 and PCA2. This differentiation is primarily attributed to variations in the genetic composition of these populations. When PCA1 and PCA2 were plotted, the distinct clustering observed between NGI and TUL aligns with King *et al.* (2022), where these two breeds were notably distant from each other. The DRB exhibited a notable clustering into three distinct groups (K = 1, 2 & 3), a phenomenon explained by its genetic heritage derived from three ancestral origins: African *Bos taurus*, European *Bos Taurus*, and *Bos indicus* (Makina *et al.,* 2016). Sanga cattle, in general, are predominantly of taurine lineage with different ratios of African and European ancestry, as previously reported by Makina *et al.* (2016). Moreover, the three sanga breeds, share comparable evolutionary history as they originated from the same ancestral groups (*Bos taurus* and *Bos indicus)* (Ramsay, 2010) and later went through geographical separation due to tribal ownership (Makina *et al.*, 2016).

The distinctive clustering observed in the DRB can be attributed to its historical geographic isolation in the Drakensberg Mountain area over an extended period (https://drakensbergers.co.za/English/, 2023). Furthermore, directional selection for black coat colour likely played a role in maintaining breed integrity within the DRB population (https://drakensbergers.co.za/English/, 2023). The results of the current study indicated that, of the three populations studied, the DRB exhibited the highest level of admixture, while the NGI population displayed the lowest. Higher admixture levels in the DRB agree with the historical narrative surrounding the origin of this breed, characterized by an unclear and uncertain origin (Scholtz, 2010). Furthermore, the genetic differentiation, particularly in the TUL breed which originated from Zimbabwe, may have hindered gene flow among these populations due to geographical locations.

5.6. Selection signatures

Ascertainment bias has previously been reported as a major issue within genetic investigations involving African breeds stemming from their exclusion in the development of the current genotyping arrays designed predominantly for European breeds (Falchi *et al.*, 2023). This has resulted in the under-representation of rare variants in the African samples (non-European breeds) and can be attributed to low

thresholds used to detect selection signatures in the current study. Other studies have also reiterated this concern, highlighting a limited set of polymorphic SNPs in the African breeds when compared to European breeds (Randhawa *et al.*, 2016; Zwane *et al.*, 2019)

5.6.1. Selection signatures within populations

Using the ROH-based approach, this study identified selection signatures in 43 genomic regions, which collectively contained 249 genes. Specifically, the TUL breed contributed to 33 of these regions, while the DRB breed contributed to 5, and the other 5 regions were from the NGI, contributing to 208, 29, and 12 genes, respectively. Using the ROHet approach, 143 genomic regions were identified as selection signatures across the three populations, with 54 in DRB, 35 in NGI, and 54 in TUL populations. Upon annotation of these candidate regions, a total of 148 candidate genes were identified, with 49 in DRB, 39 in NGI, and 60 in the TUL population. While many of these gene associations were initially studied in model animals such as mice and humans rather than in cattle, their functions are generally the same due to the similarities between human and cattle DNA (Costilla *et al.,* 2020). The results of this study were compared with previous studies conducted both in cattle and other livestock species to ensure and validate the reliability of the findings.

Genes related to immunity and adaptation

In the DRB breed, the *FKBP4* gene, located on BTA 5: 106.95 Mb was identified. This gene was previously associated with thermotolerance in Gir cattle, where it plays a role in heat shock protein binding (Saravanan *et al.*, 2021). The present study also identified *CTNNA2*, which has been linked to nervous system development in previous studies (Amaral *et al.*, 2020). This study also identified the *CYSTM1* gene in both NGI and TUL cattle, located on BTA 7: 51.36Mb. This gene has previously been linked to QTLs responsible for immune response and gestation length in cattle (Fang *et al.*, 2019). In addition, the *SRA1* gene found in NGI cattle on BTA 7, has been part of selective sweeps associated with cold adaptation in Western Finncattle (Weldenegodguad *et al.*, 2019; Huang *et al.*, 2023). The *CD14* gene, selected in NGI cattle on BTA 7:51.76 Mb, is crucial for innate immunity, offering defense against a wide range of pathogens (Pal *et al.*, 2011). It plays several roles in the immune response to diseases such as glomerulonephritis (Yoon *et al.*, 2003), mastitis (Lee *et al.*, 2003), and treponemiasis (Schröder *et al.*, 2000). In the DRB breed, the *WDPCP* gene, located on BTA 11: 61.72 Mb, has previously been associated with inflammatory response mechanisms to infections (de Las Heras-Saldana *et al.*, 2019). Its role in collective cell movement and cilia formation suggests involvement in disease resistance mechanisms in Nelore cattle (Afonso *et al.*, 2020). The identification of genes associated with innate immunity in the selective sweeps could indicate physiological adaptations, potentially leading to diseases resistance in these breeds.

Using the ROHet approach, several ROHet islands were identified, for instance, an island was observed on BTA 17: 61.09 Mb, in NGI cattle which harboured the *DTX1* gene, associated with negative regulation of lymphocyte activation suggesting its roles in immune systems (Silva *et al.*, 2022). The *ELMO3* gene located on BTA 18 in DRB has previously been associated with phagocytosis and cell migration (engulfment and cell motility) in postpartum dairy cows (Cheng *et al.*, 2015). The same gene was previously reported as a candidate gene for phagocytosis in NGI cattle by van Marle-Köster *et al.* (2022). The *ADAMTS12* gene is also known for its participation in inflammatory responses as well as the regulation of the hepatocyte growth factor (HGF) receptor signaling pathway, which is crucial for processes like epithelial cell proliferation, mobility, morphogenesis, and angiogenesis (Nakamura & Mizuno, 2010). In pigs, the *ADAMTS12* gene has been associated with adaptation to high altitudes suggesting its potential as an adaptive trait in NGI and TUL cattle (Ai *et al.*, 2013).

Genes related to body size and stature

The genomic region spanning from 24.2Mb to 25.6Mb on BTA 14 was found to be in an ROH candidate region shared between both the DRB and TUL breeds. Within this region, several candidate genes were identified, including *FAM110B, UBXN2B, CYP7A1, SDCBP, NSMAF,* and *TOX.* Additionally, another gene *KCNIP4* on BTA 6 was also found to be common between these two populations. There were no common regions (ROH islands) and genes identified between the NGI population and any other population. Genes such as *FAM110B and UBXN2B* were reported as candidate genes for feed efficiency in the South African Nguni and Bonsmara (Kooverjee *et al.*, 2022) as well as in Vrindavan cattle of India (Singh *et al.*, 2020). The selection of genes such as *FAM110B* and *UBXN2B* in these breeds could be explained by their potential to enhance feed efficiency, which is vital for reducing production costs and improving the sustainability of cattle farming.

Furthermore, another homozygous region identified on BTA 14: 21.9 - 23.7Mb in the TUL breed was previously linked to a QTL involved in weaning weight in Brangus breed (Weng *et al.,* 2016). This region harboured candidate genes such as *PLAG1, LYN,* and *CHCHD7*. *LYN*, a protein kinase which has been associated with various traits including body stature, weight and average daily gain in Nellore cattle (Utsunomiya *et al.*, 2013). Genes such as *LYN, CHCHD7, RPS20, MOS*, *PLAG1,* and *TMEM68* have been previously associated with carcass-related traits (Grigoletto *et al.*, 2020), feed intake and growth in cattle (Taye *et al.*, 2018). Similarly, the *PLAG1* gene has been reported as a candidate gene responsible for body stature, weight, and height in cattle (Utsunomiya *et al.*, 2013; Zhong *et al.*, 2019). Additionally, *SLIT2* was also identified in the current study on BTA 6: 40.1- 41.2 Mb in DRB. This genes was previously associated with the regulation of bone weight in Chinese Simmental beef cattle (Niu *et al.*, 2021). In addition, *SLIT2* gene has been reported to be associated with skeletal type traits in Limousine and Angus cattle (Doyle *et*

al., 2020). The choice of this gene in the DRB may be associated with the requirement for strong and robust bones, which is relevant as the DRB predominates the mountainous areas (Drakensberg mountains) where grazing is challenging and physically demanding.

The ROHet island on BTA 3, contained a gene *LAMTOR5* which was identified in both DRB and NGI, has been linked to cell growth in response to growth factor (Gaudet *et al.,* 2011). This gene has also been previously associated with weight gain adjusted for 345 days in Braford and Hereford cattle (Ribeiro *et al.*, 2021). Considering that the DRB and NGI cattle are characterized as small to medium frame animals, selecting for genes that influence weight gain would contribute to the selection of animals with desired body size and weight to meet the market requirements.

Genes associated with reproduction

In animal breeding, selection for high growth rates often leads to an increased incidence of dystocia in cattle, which can compromise herd reproductive performances. This study has identified genes influencing fetal growth, mature body weight of the dam, and other factors contributing to dystocia (Zaborski *et al.,* 2016). Among the identified genes in the TUL, two genes, *RPS20* and *PLAG1,* have been reported to play direct or indirect roles in factors associated with dystocia. *PLAG1,* for instance, has been associated with female reproductive traits such as age at first *corpus luteum*, age at puberty in tropical beef cattle (Fortes *et al.*, 2018) and birth weight in Nellore cattle (Utsunomiya *et al.*, 2013), both of which are pertinent to dystocia. The gene has also previously been reported to affect the beginning of puberty in Nellore heifers (Mota *et al.*, 2020). Furthermore, *RPS20* has been previously associated with fetal growth, which is one of the key determinants of dystocia. By selecting these genes, it becomes possible to enhance calving ease, manage birth weight, and regulate fetal growth, reducing the incidence of dystocia in TUL.

Moreover, this study has also identified genes previously linked to fertility such as *CORIN*, located on BTA 6: 66.3 - 66.9 Mb. *CORIN* is known to promote the invasion of the trophoblast in the uterus during pregnancy (Cui *et al.*, 2012). Additionally, *SPIRE1*, identified in the TUL breed on BTA 24: 40.3-44.9 Mb has been associated with spermatogenesis (Wen *et al.*, 2018), and has been implicated in oocyte meiotic nuclear division in horses (Laseca *et al.*, 2022). The *TXK* (*TXK tyrosine kinase*) gene in DRB, located on BTA 6: 66.76 Mb, is associated with the regulation of processes such as epididymal maturation, motility, and capacitation (Abril-Parreño *et al.*, 2023). Additionally, this gene has been reported to be beneficial in pre-freeze semen gross mortality (semen quality traits in Holstein-Friesian bulls) (Abril-Parreño *et al.*, 2023). The selection of *TXK* in the DRB breed could explain the aim to improve reproductive processes and semen quality. A region on BTA 14 containing the *XKR4* gene was found, and this gene was previously associated with traits such as heifer pregnancy, prolactin level, scrotal circumference, and rump fat thickness in Angus and Brahman (Paim *et al.*, 2020). This study also identified genes such as *SLIT2* and

KCNP4 on BTA 6: 40.1- 41.2 Mb in DRB, these genes have been previously associated with female fertility in Nordic Red cattle (Höglund *et al.*, 2015).

Two noteworthy genes were identified using the ROHet approach, *NELL2* and *TMEM181*. The *NELL2* gene was identified in the TUL, located on BTA 5: 35.48 Mb, and was previously associated with fertilization and embryo development (Sigdel *et al.*, 2020). Furthermore, the *TMEM181* gene was identified in TUL, which was previously associated with reproduction-related traits such as sperm concentration (Sun *et al.*, 2023). The selection of these alleles in the current study explains their relevance in potentially improving male and female reproductive performance in these breeds.

5.7. Selection signatures across the population

Selection signatures based on F_{ST}

Genes associated with reproduction traits

This study identified *MTERF2* located on BTA 5: 70.25 Mb in the DRBvsTUL pair and was previously linked to reproductive traits in Nellore cattle (Oliveira Júnior *et al.*, 2019). This gene has also been previously reported as a selection signal in Nguni and Bonsmara crossbred cattle by Bhika Kooverjee *et al.* (2022). While the *TOX* and *NCOA2* genes were associated with puberty onset in Brahman cattle (Fortes *et al.*, 2011), *NCOA2* has also been linked to reproductive traits such as early pregnancy, age at first calving and days to first calving in Nellore cattle (de Camargo *et al.*, 2015). This study also identified the *MYLK3* gene on BTA 18: 15.03 Mb, this gene has been previously associated with calving ease (reproductive performance) and oxytocin signaling pathways in the Mozambican and South African Nguni and Tuli cattle (King *et al.*, 2022). In this study, *SPIRE2* and *FANCA* genes were identified on BTA 18, has previously been related to reproductive traits in indigenous Chinese pigs (Zhang *et al.*, 2020), and the *SPIRE2* gene had also been associated with fecundity traits in goats (Wang *et al.*, 2022). The results of this study indicated that these genes are under positive selection, potentially due to their involvement in long-term beef selection breeding of these breeds. This study also identified *CPNE7* and *SPG7* genes which were previously identified in a group selection pattern related to reproduction, and the development of the nervous system in Chinese local cattle (Xu *et al.*, 2019).

Genes related to immunity and adaptation

The current study identified three lysozyme genes, *LYZ1*, *LYZ2*, and *LYZ3* on BTA 5: 44.37 - 44.5 Mb, which possess bacteriolytic properties and play diverse roles ranging from digestion to immune response (Wu *et al.*, 2012). The overexpression of these genes has been reported to help in sustaining mucosal inflammation of the intestines, which in turn contributes to the development of host resistance to

intestinal worms (Li *et al.*, 2015). These genes have previously been associated with the defense mechanisms of the mammary gland, contributing to mastitis resistance in cattle (Sahoo *et al.*, 2012).

Heat shock protein (HSP) serves as a cellular and tissue defense mechanism and is expressed in high volumes during heat shock (Archana *et al.*, 2017). If overexpressed, it protects the animal against hyperthermia during heat stroke which signifies the pivortal role of HSP in cytoprotection (Dangi *et al.*, 2017). The current study identified the *LMAN2* gene on BTA7 in the NGIvsTUL pair, which was previously associated with functions of heat shock protein binding (GO:0031072) in beef cattle (Ghebrewold, 2018). This gene responds to heat shock through intracellular and extracellular signals which activate the expression of Heat Shock Protein such as HSP27 (Shibata, 2014), HSP70 (Bhat *et al.*, 2016) and HSP90 in cattle (Archana *et al.*, 2017).

Another gene, *NMU* located on BTA 6: 71.03 Mb in the NGIvsTUL pair, was previously associated with the regulation of stress responses and thermoregulation in cattle (Yayou *et al.*, 2009). The *CDK10* gene located on BTA 18: 14.57 Mb, was identified as a candidate gene in this study and has been previously associated with immunity in Nguni cattle (Makina *et al.*, 2015). Its selection could be due to its role in antigen recognition crucial for immune response. *CBFA2T3* has been identified on BTA 18: 14.05 Mb in the current study and was previously under positive selection in Iraqi cattle (Alshawi *et al.,* 2019), and plays a role in the innate immune response, particularly in response to mammary gland inflammation (Li *et al.*, 2020). *Lymphocyte Antigen 96* (*LY96*), a gene which has been linked with a crucial role of detecting lipopolysaccharide, serving as pattern recognition receptor, and positive regulation of phagocytosis was found on BTA 14: 37.24 Mb in DRBvsTUL pair (Dou *et al.*, 2013). This gene plays a crucial role in the early identification of pathogens and the activation of immune signalling pathways that then engage the adaptive immune response (Dixon *et al.,* 2013). Consequently, the identification of this gene in the current study implies that the DRB and TUL breeds likely possess a more effective mechanism for recognizing and responding to pathogens.

5.9.3 Genes Affecting Coat Color

Mammalian skin and hair serve essential roles in providing physical protection and regulating body temperature (Jian *et al.,* 2014). Certain characteristics of the coat colour in cattle play a role in enhancing conductive and convective heat loss while minimizing the absorption of solar radiation, making them valuable phenotypic markers for heat tolerance (Lenis-Sanin *et al.,* 2016). Cattle adapted to tropical environments, exhibit light-coloured, sleek, and shiny hair coats, which effectively reflect a higher proportion of incident solar radiation, thereby reducing heat load (Hansen, 2004). In the current study genes such as *TUBB3* and *MC1R* which are associated with coat colour were identified on BTA 18 at 14.07 Mb. These genes encode proteins that regulate melanogenesis, playing pivotal roles in pigmentation and

inflammation processes (Taye *et al.*, 2018; Goud *et al.*, 2020). Notably, polymorphisms in the *MC1R* gene region have previously been linked to variations in coat colour, such as black or red, in Angus cattle (Klungland *et al.,* 1995) and Holstein-Frisian cattle (Zhao *et al.,* 2015). Drakensberger cattle are distinguished by their solid black coat colour (https://drakensbergers.co.za/English/, 2023), while Nguni cattle exhibit a range of coat colours, including black, brown, and red as primary hues, with variations stemming from these base colours (Olson, 1999). These diverse coat colours hold significant cultural and breed-specific importance, especially among the Nguni ethnic community, influencing their selection preferences (Oosthuizen, 1996). In addition, the Drakensberger and Nguni cattle, being indigenous to specific regions, likely experienced evolutionary pressures favouring traits such as light-coloured, sleek, and shiny hair coats that contribute to effective heat dissipation in their respective tropical environments. Consequently, these phenotypic markers linked to coat colour become advantageous in their agroecological zones, leading to a directional selection of genes associated with these adaptive traits.

5.8. Conclusion and recommendations

This study used 122,632 high-quality SNPs to investigate selection signatures in three South African indigenous cattle breeds: the Drakensberger (DRB), Nguni (NGI), and Tuli (TUL). The genotypes were analyzed to quantify genetic variation, assess the genetic structure of the populations, estimate levels of inbreeding, and identify potential selection signatures within a population (using conserved runs of homozygosity and runs of heterozygosity) and between populations (using Wright's Fixation Index). The analysis of genetic diversity revealed a moderate level of diversity across all three populations, with slight variations in expected heterozygosity. The differentiation among the populations was evident in the population structure analysis, reflecting their distinct genetic compositions and historical origins. Furthermore, measures of genomic inbreeding suggested low levels of inbreeding in all three populations, indicating effective breeding practices. Runs of homozygosity (ROH) analysis indicated varying numbers of ROH segments and their lengths, providing insights into population history and recent admixture. The prevalence of shorter ROH segments was associated with potential selection signatures and ancient inbreeding. The presence of runs of heterozygosity (ROHet) was also explored, revealing associations with traits related to productivity and disease resistance. The Principal Component Analysis and model-based admixture demonstrated clear differentiation among the three Sanga cattle breeds. Selection signatures were identified based on ROH and ROHet, highlighting candidate genes associated with traits of economic importance. Overall, this study provides valuable insights into the genetic diversity, inbreeding, and selection signatures within South African indigenous cattle breeds.

In the face of climate change and a rapid increase in protein demand, prioritizing the breeding of resilient animals is imperative. African cattle breeds, dispersed across the continent, embody a rich spectrum of unique genetic resources with immense potential for studies on productivity, adaptation, and disease resistance. These breeds shaped by natural and artificial selection forces in the African context have become resilient to environmental stressors and can survive through various challenges. Exploring the genetic footprints of selection forces could provide insights and assist to identify genomic regions associated with adaptation and traits of economic importance. Fewer candidate genes identified in African cattle based on European breed reference assays may be attributed to ascertainment bias. This increases the likelihood of obtaining false positive results in the designated candidate regions, highlighting the importance of assembling the African cattle reference genome. This can mitigate potential biases, enhance the accuracy of genomic analyses, and gain more comprehensive understanding of the unique genetic make-up of African cattle populations. False positives in these studies are not uncommon, despite the stringent parameters employed to identify selection signatures. Therefore, it is imperative to replicate similar studies with larger sample sizes using high density genomic data (e.g., Whole-genome sequence data).

References

Ablondi, M., Dadousis, C., Vasini, M., Eriksson, S., Mikko, S., & Sabbioni, A. (2020). Genetic Diversity and Signatures of Selection in a Native Italian Horse Breed Based on SNP Data. *Animals*, *10*(6), 1005. https://doi.org/10.3390/ani10061005

Abril-Parreño, L., Carthy, T. R., Keogh, K., Štiavnická, M., O'Meara, C., Lonergan, P., Kenny, D. A., & Fair, S. (2023). Genome-wide association study reveals candidate markers related to field fertility and semen quality traits in Holstein-Friesian bulls. *Animal*, *17*(6), 100841.

Acocks, J. P. H. (1988). *Veld types of South Africa*.

Afonso, J., Fortes, M. R. S., Reverter, A., Diniz, W. J. d. S., Cesar, A. S. M., Lima, A. O. d., Petrini, J., de Souza, M. M., Coutinho, L. L., & Mourão, G. B. (2020). Genetic regulators of mineral amount in Nelore cattle muscle predicted by a new co-expression and regulatory impact factor approach. *Scientific reports*, *10*(1), 8436.

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

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome research*, *12*(12), 1805-1814.

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

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

Amamou, H., Beckers, Y., Mahouachi, M., & Hammami, H. (2019). Thermotolerance indicators related to production and physiological responses to heat stress of holstein cows. *Journal of thermal biology*, *82*, 90-98.

Amaral, A. J., Pavão, A. L., & Gama, L. T. (2020). Genomic tools for the conservation and genetic improvement of a highly fragmented breed—The Ramo Grande cattle from the Azores. *Animals*, *10*(6), 1089.

Archana, P., Aleena, J., Pragna, P., Vidya, M., Niyas, A., Bagath, M., Krishnan, G., Manimaran, A., Beena, V., & Kurien, E. (2017). Role of heat shock proteins in livestock adaptation to heat stress. *J. Dairy Vet. Anim. Res*, *5*(1), 00127.

Ayalew, W., Wu, X.-y., Tarekegn, G. M., Min, C., Liang, C.-n., Tessema, T. S., & Ping, Y. (2023). Signatures of positive selection for local adaptation of African Native Cattle populations: a review. *Journal of integrative agriculture*.

Barabaschi, D., Tondelli, A., Desiderio, F., Volante, A., Vaccino, P., Valè, G., & Cattivelli, L. (2016). Next generation breeding. *Plant Science*, *242*, 3-13.

Belhadj Slimen, I., Najar, T., Ghram, A., & Abdrrabba, M. (2016). Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review. *Journal of animal physiology and animal nutrition*, *100*(3), 401-412.

Ben-Jemaa, S., Mastrangelo, S., Lee, S.-H., Lee, J. H., & Boussaha, M. (2020). Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle. *Scientific reports*, *10*(1), 19466.

Bergh, L., Gerhard, R., Scholtz, M., & Mamabolo, M. (2010). Introduction to the information on beef and dual purpose breeds in South Africa. *Beef breeding in South Africa*(Ed. 2), 150-158.

Bernabucci, U. (2019). Climate change: impact on livestock and how can we adapt. *Animal frontiers: the review magazine of animal agriculture*, *9*(1), 3.

Bhat, S., Kumar, P., Kashyap, N., Deshmukh, B., Dige, M. S., Bhushan, B., Chauhan, A., Kumar, A., & Singh, G. (2016). Effect of heat shock protein 70 polymorphism on thermotolerance in Tharparkar cattle. *Veterinary World*, *9*(2), 113.

Bhika Kooverjee, B., Soma, P., Neser, F., van der Nest, M., & Scholtz, M. (2022). Copy number variation analysis in Nguni and Bonsmara crossbred cattle. Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges,

Biscarini, F., Cozzi, P., Gaspa, G., & Marras, G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes.

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G., & Ciampolini, R. (2020). Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in Maremmana semi-feral cattle using pedigree and genomic data. *Animals*, *10*(12), 2285.

Bisschoff, C., & Lotriet, R. (2013). The Drakensberger as competitive breed of cattle in the South African beef industry. 19th International Farm Management Congress,

Bizarria dos Santos, W., Pimenta Schettini, G., Fonseca, M. G., Pereira, G. L., Loyola Chardulo, L. A., Rodrigues Machado Neto, O., Baldassini, W. A., Nunes de Oliveira, H., & Abdallah Curi, R. (2021). Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. *Journal of Animal Breeding and Genetics*, *138*(2), 161-173.

Boichard, D., Chung, H., Dassonneville, R., David, X., Eggen, A., Fritz, S., Gietzen, K. J., Hayes, B. J., Lawley, C. T., & Sonstegard, T. S. (2012). Design of a bovine low-density SNP array optimized for imputation. *PloS one*, *7*(3), e34130.

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

Bosman, L., Van Marle-Köster, E., Van der Westhuizen, R., Visser, C., & Berry, D. (2017). Short communication: Population structure of the South African Bonsmara beef breed using high density single nucleotide polymorphism genotypes. *Livest. Sci*, *197*, 102.

Bosse, M., Megens, H.-J., Madsen, O., Crooijmans, R. P., Ryder, O. A., Austerlitz, F., Groenen, M. A., & de Cara, M. A. R. (2015). Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome research*, *25*(7), 970-981.

Bosse, M., Megens, H. J., Derks, M. F., de Cara, Á. M., & Groenen, M. A. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evolutionary applications*, *12*(1), 6-17.

Brito, L., Bédère, N., Douhard, F., Oliveira, H., Arnal, M., Peñagaricano, F., Schinckel, A., Baes, C. F., & Miglior, F. (2021). Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. *Animal*, *15*, 100292.

Broman, K. W., & Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *The American Journal of Human Genetics*, *65*(6), 1493-1500.

Brown, D. (1959). The Nguni breed of cattle. A descriptive review. *Empire Journal of Experimental Agriculture*, *27*, 276-290.

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

Burrow, H. (2012). Importance of adaptation and genotype× environment interactions in tropical beef breeding systems. *Animal*, *6*(5), 729-740.

Burrow, H. M. (2015). Genetic aspects of cattle adaptation in the tropics. In *The genetics of cattle* (pp. 571-597). CABI Wallingford UK.

Caballero, A., & Toro, M. A. (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetics research*, *75*(3), 331-343.

Caivio-Nasner, S., López-Herrera, A., González-Herrera, L. G., & Rincón, J. C. (2021). Diversity analysis, runs of homozygosity and genomic inbreeding reveal recent selection in Blanco Orejinegro cattle. *Journal of Animal Breeding and Genetics*, *138*(5), 613-627.

Cardoso, D. F., Fernandes Junior, G. A., Scalez, D. C. B., Alves, A. A. C., Magalhães, A. F. B., Bresolin, T., Ventura, R. V., Li, C., de Sena Oliveira, M. C., & Porto-Neto, L. R. (2020). Uncovering sub-structure and genomic profiles in across-countries subpopulations of Angus cattle. *Scientific reports*, *10*(1), 8770.

Cesarani, A., Sorbolini, S., Criscione, A., Bordonaro, S., Pulina, G., Battacone, G., Marletta, D., Gaspa, G., & Macciotta, N. (2018). Genome-wide variability and selection signatures in Italian island cattle breeds. *Animal Genetics*, *49*(5), 371-383.

Chen, M., Pan, D., Ren, H., Fu, J., Li, J., Su, G., Wang, A., Jiang, L., Zhang, Q., & Liu, J.-F. (2016). Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genetics Selection Evolution*, *48*(1), 1-12.

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., Hu, S., Huang, S., Zhang, H., & Zheng, Z. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications*, *9*(1), 2337.

Chen, Z., Zhang, Z., Wang, Z., Zhang, Z., Wang, Q., & Pan, Y. (2022). Heterozygosity and homozygosity regions affect reproductive success and the loss of reproduction: A case study with litter traits in pigs. *Computational and Structural Biotechnology Journal*, *20*, 4060-4071.

Cheng, Z., Oguejiofor, C. F., Swangchan-Uthai, T., Carr, S., & Wathes, D. C. (2015). Relationships between circulating urea concentrations and endometrial function in postpartum dairy cows. *Animals*, *5*(3), 748-773.

Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., & Mujibi, F. D. (2018). Signatures of selection in admixed dairy cattle in Tanzania. *Frontiers in Genetics*, *9*, 607.

Chirico, W. J., Waters, M. G., & Blobel, G. (1988). 70K heat shock related proteins stimulate protein translocation into microsomes. *Nature*, *332*(6167), 805-810.

Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H., & Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome research*, *15*(11), 1496-1502.

Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D., & Lewis, L. C. (2009). Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. *Journal of Heredity*, *100*(5), 556-564.

Cortes, O., Cañon, J., & Gama, L. T. (2022). Applications of microsatellites and single nucleotide polymorphisms for the genetic characterization of cattle and small ruminants: An overview. *Ruminants*, *2*(4), 456-470.

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science*, *166*, 26-34.

DAFF (2021) Department of Agriculture, Forestry and Fisheries. A profile of the South African beef market value chain. http://webapps1.daff.gov.za/AmisAdmin/upload/Beef%20Market%20Value%20Chain%20Profile%202021.pdf. 3-22.

Dangi, S. S., Bharati, J., Samad, H. A., Bhure, S. K., Singh, G., Maurya, V. P., Sarkar, M., & Kumar, P. (2017). Expression dynamics of heat shock proteins (HSP) in livestock under thermal stress. *Heat shock proteins in veterinary medicine and sciences*, 37-79.

de Camargo, G. M., Costa, R. B., de Albuquerque, L. G., Regitano, L. C., Baldi, F., & Tonhati, H. (2015). Polymorphisms in TOX and NCOA2 genes and their associations with reproductive traits in cattle. *Reproduction, Fertility and Development*, *27*(3), 523-528.

de Las Heras-Saldana, S., Clark, S. A., Duijvesteijn, N., Gondro, C., van der Werf, J. H., & Chen, Y. (2019). Combining information from genome-wide association and multi-tissue gene expression studies

to elucidate factors underlying genetic variation for residual feed intake in Australian Angus cattle. *BMC genomics*, *20*(1), 1-16.

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., & Praharani, L. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS genetics*, *10*(3), e1004254.

Derks, M. F., & Steensma, M. (2021). balancing selection for deleterious alleles in livestock. *Frontiers in Genetics*, *12*, 761728.

Dokan, K., Kawamura, S., & Teshima, K. M. (2021). Effects of single nucleotide polymorphism ascertainment on population structure inferences. *G3*, *11*(9), jkab128.

Dou, T., Fu, M., Wang, Y., Zhao, Y., Wang, Z., Bian, Z., & Zhou, Y. (2013). Signatures of positive selection in LY96 gene in vertebrates. *Journal of biosciences*, *38*, 899-904.

Doyle, J. L., Berry, D. P., Veerkamp, R. F., Carthy, T. R., Walsh, S. W., Evans, R. D., & Purfield, D. C. (2020). Genomic regions associated with skeletal type traits in beef and dairy cattle are common to regions associated with carcass traits, feed intake and calving difficulty. *Frontiers in Genetics*, *11*, 20.

Edea, Z., Dadi, H., Dessie, T., Uzzaman, M., Rothschild, M. F., Kim, E. S., Sonstegard, T., & Kim, K. S. (2018). Genome-wide scan reveals divergent selection among taurine and zebu cattle populations from different regions. *Animal Genetics*, *49*(6), 550-563.

Eusebi, P. G., Martinez, A., & Cortes, O. (2019). Genomic tools for effective conservation of livestock breed diversity. *Diversity*, *12*(1), 8.

Fabbri, M. C., Dadousis, C., Tiezzi, F., Maltecca, C., Lozada-Soto, E., Biffani, S., & Bozzi, R. (2021). Genetic diversity and population history of eight Italian beef cattle breeds using measures of autozygosity. *PloS one*, *16*(10), e0248087.

Falchi, L., Cesarani, A., Mastrangelo, S., Senczuk, G., Portolano, B., Pilla, F., & Macciotta, N. P. (2023). Analysis of runs of homozygosity of cattle living in different climate zones. *Journal of Animal Science*, *101*, skad061.

Fang, L., Jiang, J., Li, B., Zhou, Y., Freebern, E., Vanraden, P. M., Cole, J. B., Liu, G. E., & Ma, L. (2019). Genetic and epigenetic architecture of paternal origin contribute to gestation length in cattle. *Communications biology*, *2*(1), 100.

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

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

Ferraz, J., Wu, X.-L., Li, H., Xu, J., Ferretti, R., Simpson, B., Walker, J., Silva, L., Garcia, J., & Tait, R. (2020). Development and evaluation of a low-density single-nucleotide polymorphism chip specific to Bos indicus cattle. *Animal Production Science*, *60*(15), 1769-1776.

Fortes, M., Reverter, A., Nagaraj, S., Zhang, Y., Jonsson, N., Barris, W., Lehnert, S., Boe-Hansen, G., & Hawken, R. (2011). A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. *Journal of Animal Science*, *89*(6), 1669-1683.

Fortes, M. R. S., Enculescu, C., Porto Neto, L. R., Lehnert, S. A., McCulloch, R., & Hayes, B. (2018). Candidate mutations used to aid the prediction of genetic merit for female reproductive traits in tropical beef cattle. *Revista Brasileira de Zootecnia*, *47*.

Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). *Introduction to conservation genetics*. Cambridge university press.

Freitas, P. H., Wang, Y., Yan, P., Oliveira, H. R., Schenkel, F. S., Zhang, Y., Xu, Q., & Brito, L. F. (2021). Genetic diversity and signatures of selection for thermal stress in cattle and other two bos species adapted to divergent climatic conditions. *Frontiers in Genetics*, *12*, 604823.

Garcia, A. O., Otto, P. I., Glatzl Junior, L. A., Rocha, R. d. F. B., Dos Santos, M. G., de Oliveira, D. A., da Silva, M. V. G. B., Panetto, J. C. d. C., Machado, M. A., & Verneque, R. d. S. (2023). Pedigree reconstruction and population structure using SNP markers in Gir cattle. *Journal of Applied Genetics*, 1-12.

Gautason, E., Schönherz, A., Sahana, G., & Guldbrandtsen, B. (2021). Genomic inbreeding and selection signatures in the local dairy breed Icelandic Cattle. *Animal Genetics*, *52*(3), 251-262.

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

Geibel, J., Reimer, C., Weigend, S., Weigend, A., Pook, T., & Simianer, H. (2021). How array design creates SNP ascertainment bias. *PloS one*, *16*(3), e0245178.

Ghassemi Nejad, J., Kim, B. W., Lee, B. H., & Sung, K. I. (2017). Coat and hair color: hair cortisol and serotonin levels in lactating Holstein cows under heat stress conditions. *Animal Science Journal*, *88*(1), 190-194.

Ghebrewold, R. (2018). Genome-wide Association Study for the Relationship Between Temperature and Feed Intake in Beef Cattle.

Gibson, J., Morton, N. E., & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human molecular genetics*, *15*(5), 789-795.

Godde, C. M., Mason-D'Croz, D., Mayberry, D., Thornton, P. K., & Herrero, M. (2021). Impacts of climate change on the livestock food supply chain; a review of the evidence. *Global food security*, *28*, 100488.

Gomez-Zavaglia, A., Mejuto, J. C., & Simal-Gandara, J. (2020). Mitigation of emerging implications of climate change on food production systems. *Food Research International*, *134*, 109256.

Goud, T. S., Upadhyay, R. C., Onteru, S. K., Pichili, V. B. R., & Chadipiralla, K. (2020). Identification and sequence characterization of melanocortin 1 receptor gene (MC1R) in Bos indicus versus (Bos taurus X Bos indicus). *Animal Biotechnology*, *31*(4), 283-294.

Goyache, F., Pérez-Pardal, L., Fernández, I., Traoré, A., Menéndez-Arias, N. A., & Álvarez, I. (2021). Ancient autozygous segments subject to positive selection suggest adaptive immune responses in West African cattle. *Gene*, *803*, 145899.

Grigoletto, L., Ferraz, J. B., Oliveira, H. R., Eler, J. P., Bussiman, F. O., Abreu Silva, B. C., Baldi, F., & Brito, L. F. (2020). Genetic architecture of carcass and meat quality traits in Montana Tropical® composite beef cattle. *Frontiers in Genetics*, *11*, 123.

Grossi, G., Goglio, P., Vitali, A., & Williams, A. G. (2019). Livestock and climate change: impact of livestock on climate and mitigation strategies. *Animal frontiers*, *9*(1), 69-76.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., & Salin, F. (2011). Current trends in microsatellite genotyping. *Molecular ecology resources*, *11*(4), 591-611.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M. O., & Choudhury, A. (2015). The African genome variation project shapes medical genetics in Africa. *Nature*, *517*(7534), 327-332.

Gutierrez-Reinoso, M. A., Aponte, P. M., & Garcia-Herreros, M. (2021). Genomic analysis, progress and future perspectives in dairy cattle selection: A review. *Animals*, *11*(3), 599.

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

Hanotte, O., Okomo, M., Bradley, D., Verjee, Y., Ochieng, A., Teale, A., & Rege, J. (1998). Geographical distribution and frequency of taurine Bos taurus and zebu B. indicus Y chromosome haplotypes amongst sub-Saharan African cattle breeds. Proc. 4th Global Conf. on Conservation Domestic Animal Genetic Resources, Nepal, 17–21 Aug,

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., Sugimoto, Y., Kemp, S., Gibson, J., & Korol, A. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proceedings of the National Academy of Sciences*, *100*(13), 7443-7448.

Hariyono, D. N. H., & Prihandini, P. W. (2022). Association of selected gene polymorphisms with thermotolerance traits in cattle–A review. *Animal Bioscience*, *35*(11), 1635.

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

Hayes, B. J., Lien, S., Nilsen, H., Olsen, H. G., Berg, P., MacEachern, S., Potter, S., & Meuwissen, T. (2008). The origin of selection signatures on bovine chromosome 6. *Animal Genetics*, *39*(2), 105-111.

Heaton, M. P., Harhay, G. P., Bennett, G. L., Stone, R. T., Grosse, W. M., Casas, E., Keele, J. W., Smith, T. P., Chitko-McKown, C. G., & Laegreid, W. W. (2002). Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. *Mammalian Genome*, *13*, 272-281.

Henry, B., Eckard, R., & Beauchemin, K. (2018). Adaptation of ruminant livestock production systems to climate changes. *Animal*, *12*(s2), s445-s456.

Hill, E. W., O'Gorman, G. M., Agaba, M., Gibson, J. P., Hanotte, O., Kemp, S. J., Naessens, J., Coussens, P. M., & MacHugh, D. E. (2005). Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics. *Veterinary Immunology and Immunopathology*, *105*(3-4), 247-258.

Höglund, J. K., Buitenhuis, B., Guldbrandtsen, B., Lund, M. S., & Sahana, G. (2015). Genome-wide association study for female fertility in Nordic Red cattle. *BMC genetics*, *16*, 1-11.

Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International journal of plant sciences*, *171*(9), 1059-1071.

Holmberg, M., & Andersson-Eklund, L. (2006). Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. *Journal of dairy science*, *89*(9), 3664-3671.

Howrigan, D. P., Simonson, M. A., & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC genomics*, *12*(1), 1-15.

http://www.tulicattle.co.za/. (2023). *Tuli Cattle - The Intelligent Choice*. http://www.tulicattle.co.za/. Retrieved 20 March 2023 from

https://drakensbergers.co.za/English/. (2023). *ORIGIN AND HISTORY*. https://drakensbergers.co.za/English/origin-an-history/. Retrieved 20 March 2023 from

Huang, N., Zhao, L., Wang, J., Jiang, Q., Ju, Z., Wang, X., Yang, C., Gao, Y., Wei, X., & Zhang, Y. (2023). Signatures of selection in indigenous Chinese cattle genomes reveal adaptive genes and genetic variations to cold climate. *Journal of Animal Science*, *101*, skad006.

Hulsegge, I., Oldenbroek, K., Bouwman, A., Veerkamp, R., & Windig, J. (2022). Selection and drift: A comparison between historic and recent Dutch Friesian cattle and recent Holstein Friesian using WGS data. *Animals*, *12*(3), 329.

Huntley, B., De Booysen, P., & Tainton, N. (1984). Ecological effects of fire in South African ecosystems. *Ecological studies*, *48*, 1-18.

Igoshin, A. V., Yurchenko, A. A., Belonogova, N. M., Petrovsky, D. V., Aitnazarov, R. B., Soloshenko, V. A., Yudin, N. S., & Larkin, D. M. (2019). Genome-wide association study and scan for signatures of

selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. *BMC genetics*, *20*(1), 5-14.

Jemaa, S. B., Thamri, N., Mnara, S., Rebours, E., Rocha, D., & Boussaha, M. (2019). Linkage disequilibrium and past effective population size in native Tunisian cattle. *Genetics and Molecular Biology*, *42*, 52-61.

Kambal, S., Tijjani, A., Ibrahim, S. A., Ahmed, M. K. A., Mwacharo, J. M., & Hanotte, O. (2023). Candidate signatures of positive selection for environmental adaptation in indigenous African cattle: A review. *Animal Genetics*.

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., & Gietzen, K. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS biology*, *10*(2), e1001258.

Kim, E.-S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., Liu, G., Da, Y., & Sonstegard, T. S. (2013). Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PloS one*, *8*(11), e80813.

Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., Agaba, M., Kim, K., Kwak, W., & Sung, S. (2017). The genome landscape of indigenous African cattle. *Genome biology*, *18*, 1-14.

King, F., Visser, C., & Banga, C. (2022). Genetic characterization of Mozambican Nguni cattle and their relationship with indigenous populations of South Africa. *Livestock Science*, *264*, 105044.

King, F. J., Banga, C. B., & Visser, C. (2021). Genetic diversity and population structure of three native cattle populations in Mozambique. *Tropical animal health and production*, *53*, 1-7.

Kooverjee, B. B., Soma, P., Van Der Nest, M. A., Scholtz, M. M., & Neser, F. W. (2022). Selection Signatures in South African Nguni and Bonsmara Cattle Populations Reveal Genes Relating to Environmental Adaptation. *Frontiers in Genetics*, *13*.

Lacetera, N. (2019). Impact of climate change on animal health and welfare. *Animal frontiers*, *9*(1), 26-31.

Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays*, *35*(9), 780-786.

Laseca, N., Demyda-Peyrás, S., Valera, M., Ramón, M., Escribano, B., Perdomo-González, D. I., & Molina, A. (2022). A genome-wide association study of mare fertility in the Pura Raza Español horse. *Animal*, *16*(3), 100476.

Lashmar, S., Muchadeyi, F., & Visser, C. (2019). Genotype imputation as a cost-saving genomic strategy for South African Sanga cattle: A review. *South African Journal of Animal Science*, *49*(2), 262-280.

Lashmar, S., Visser, C., van Marle-Köster, E., & Muchadeyi, F. C. (2018). Genomic diversity and autozygosity within the SA Drakensberger beef cattle breed. *Livestock Science*, *212*, 111-119.

Lashmar, S. F., Visser, C., Okpeku, M., Muchadeyi, F. C., Mapholi, N. O., & van Marle-Köster, E. (2022). A within-and across-country assessment of the genomic diversity and autozygosity of South African and eSwatini Nguni cattle. *Tropical animal health and production*, *54*(6), 1-7.

Lee, J.-W., Paape, M. J., Elsasser, T. H., & Zhao, X. (2003). Recombinant soluble CD14 reduces severity of intramammary infection by Escherichia coli. *Infection and immunity*, *71*(7), 4034-4039.

Lemerle, C., & Goddard, M. (1986). Assessment of heat stress in dairy cattle in Papua New Guinea. *Tropical animal health and production*, *18*, 232-242.

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

Li, R., Li, C., Chen, H., Li, R., Chong, Q., Xiao, H., & Chen, S. (2020). Genome-wide scan of selection signatures in Dehong humped cattle for heat tolerance and disease resistance. *Animal Genetics*, *51*(2), 292-299.

Li, R. W., Wu, S., Li, C.-J., Li, W., & Schroeder, S. G. (2015). Splice variants and regulatory networks associated with host resistance to the intestinal worm Cooperia oncophora in cattle. *Veterinary Parasitology*, *211*(3-4), 241-250.

Liu, D., Chen, Z., Zhao, W., Guo, L., Sun, H., Zhu, K., Liu, G., Shen, X., Zhao, X., & Wang, Q. (2021). Genome-wide selection signatures detection in Shanghai Holstein cattle population identified genes related to adaption, health and reproduction traits. *BMC genomics*, *22*(1), 1-19.

Liu, X., Ong, R. T.-H., Pillai, E. N., Elzein, A. M., Small, K. S., Clark, T. G., Kwiatkowski, D. P., & Teo, Y.-Y. (2013). Detecting and characterizing genomic signatures of positive selection in global populations. *The American Journal of Human Genetics*, *92*(6), 866-881.

Madilindi, M. A., Banga, C. B., Bhebhe, E., Sanarana, Y. P., Nxumalo, K. S., Taela, M. G., Magagula, B. S., & Mapholi, N. O. (2020). Genetic diversity and relationships among three Southern African Nguni cattle populations. *Tropical animal health and production*, *52*(2), 753-762.

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

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

Makina, S. O., Muchadeyi, F. C., Van Marle-Köster, E., Taylor, J. F., Makgahlela, M. L., & Maiwashe, A. (2015). Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetics Selection Evolution*, *47*(1), 47-92. https://doi.org/10.1186/s12711-015-0173-x

Makina, S. O., Whitacre, L. K., Decker, J. E., Taylor, J. F., MacNeil, M. D., Scholtz, M. M., van Marle-Köster, E., Muchadeyi, F. C., Makgahlela, M. L., & Maiwashe, A. (2016). Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide. *Genetics Selection Evolution*, *48*(1), 1-7.

Mapholi, N. O. (2015). *Exploring genetic architecture of tick resistance in South African Nguni cattle* Stellenbosch: Stellenbosch University].

Mapholi, N. O., Banga, C., Dzama, K., Matika, O., Riggio, V., Nyangiwe, N., & Maiwashe, A. (2022). Prevalence and tick loads in Nguni cattle reared in different environmental conditions across four provinces of South Africa. *Veterinary World*, *15*(8).

Mapholi, N. O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S., MacNeil, M., Banga, C., Taylor, J., & Dzama, K. (2016). Genome-wide association study of tick resistance in South African Nguni cattle. *Ticks and tick-borne diseases*, *7*(3), 487-497.

Mapholi, N. O., Marufu, M. C., Maiwashe, A., Banga, C. B., Muchenje, V., MacNeil, M. D., Chimonyo, M., & Dzama, K. (2014). Towards a genomics approach to tick (Acari: Ixodidae) control in cattle: A review. *Ticks and tick-borne diseases*, *5*(5), 475-483.

Mapiye, C., Chikwanha, O. C., Chimonyo, M., & Dzama, K. (2019). Strategies for sustainable use of indigenous cattle genetic resources in Southern Africa. *Diversity*, *11*(11), 214.

Mapiye, C., Chimonyo, M., Muchenje, V., Dzama, K., Marufu, M. C., & Raats, J. G. (2007). Potential for value-addition of Nguni cattle products in the communal areas of South Africa: a review. *African Journal of Agricultural Research*, *2*(10), 488-495.

Mariasegaram, M., Chase Jr, C., Chaparro, J., Olson, T., Brenneman, R., & Niedz, R. (2007). The slick hair coat locus maps to chromosome 20 in Senepol-derived cattle. *Animal Genetics*, *38*(1), 54-59.

Marima, J., Nel, C., Marufu, M. C., Jonsson, N., Dube, B., & Dzama, K. (2020). A genetic and immunological comparison of tick-resistance in beef cattle following artificial infestation with Rhipicephalus ticks. *Experimental and applied acarology*, *80*, 569-590.

Marima, J. K. (2017). *Gene expression profiles associated with beef cattle resistance to Rhipicephalus ticks* Stellenbosch: Stellenbosch University].

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J. L., & Macciotta, N. P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal Genetics*, *46*(2), 110-121.

Marras, G., Wood, B., Makanjuola, B., Malchiodi, F., Peeters, K., Van As, P., Baes, C., Biscarini, F., & Turkeys, H. (2018). Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (Meleagris gallopavo) population. Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand,

Marufu, M., Chimonyo, M., Mans, B., & Dzama, K. (2013). Cutaneous hypersensitivity responses to Rhipicephalus tick larval antigens in pre-sensitized cattle. *Ticks and tick-borne diseases*, *4*(4), 311-316.

Mastrangelo, S., Tolone, M., Ben Jemaa, S., Sottile, G., Di Gerlando, R., Cortés, O., Senczuk, G., Portolano, B., Pilla, F., & Ciani, E. (2020). Refining the genetic structure and relationships of European cattle breeds through meta-analysis of worldwide genomic SNP data, focusing on Italian cattle. *Scientific reports*, *10*(1), 14522.

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M., & Portolano, B. (2016). Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal*, *10*(5), 746-754.

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

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P., & Sonstegard, T. S. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PloS one*, *4*(4), e5350.

Mburu, D., & Hanotte, O. (2005). A practical approach to microsatellite genotyping with special reference to livestock population genetics. *ILRI, Nairobi, Kenya*.

Mc Parland, S., Kearney, F., & Berry, D. P. (2009). Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genetics Selection Evolution*, *41*(1), 1-8.

McManus, C., do Prado Paim, T., de Melo, C. B., Brasil, B. S., & Paiva, S. R. (2014). Selection methods for resistance to and tolerance of helminths in livestock. *Parasite*, *21*.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., & Tenesa, A. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, *83*(3), 359-372.

Meseret, S., Mekonnen, Y. A., Brenig, B., Schütz, E., Hanotte, O., Gültas, M., & Schmitt, A. O. (2020). Genetic diversity and population structure of six ethiopian cattle breeds from different geographical regions using high density single nucleotide polymorphisms. *Livestock Science*, *234*, 103979.

Mészáros, G., Boison, S. A., Pérez O'Brien, A. M., Ferenčaković, M., Curik, I., Da Silva, M. V. B., Utsunomiya, Y. T., Garcia, J. F., & Sölkner, J. (2015). Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. *Frontiers in Genetics*, *6*, 173.

Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC genomics*, *21*(1), 1-14.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T., & Thomas, P. D. (2021). PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic acids research*, *49*(D1), D394-D403.

Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., & McEwan, J. C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC genetics*, *13*(1), 1-15.

Moravčíková, N., Simčič, M., Mészáros, G., Sölkner, J., Kukučková, V., Vlček, M., Trakovická, A., Kadlečík, O., & Kasarda, R. (2018). Genomic response to natural selection within alpine cattle breeds. *Czech journal of animal science*, *63*(4), 136-143.

Mota, L. F., Lopes, F. B., Fernandes Júnior, G. A., Rosa, G. J., Magalhães, A. F., Carvalheiro, R., & Albuquerque, L. G. (2020). Genome-wide scan highlights the role of candidate genes on phenotypic plasticity for age at first calving in Nellore heifers. *Scientific reports*, *10*(1), 6481.

Msalya, G., Kim, E.-S., Laisser, E. L., Kipanyula, M. J., Karimuribo, E. D., Kusiluka, L. J., Chenyambuga, S. W., & Rothschild, M. F. (2017). Determination of genetic structure and signatures of selection in three strains of Tanzania Shorthorn Zebu, Boran and Friesian cattle by genome-wide SNP analyses. *PloS one*, *12*(1), e0171088.

Mwai, O., Hanotte, O., Kwon, Y. J., & Cho, S. (2015). African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. *Asian-Australas J Anim Sci*, *28*(7), 911-921. https://doi.org/10.5713/ajas.15.0002R

Myles, S., Davison, D., Barrett, J., Stoneking, M., & Timpson, N. (2008). Worldwide population differentiation at disease-associated SNPs. *BMC medical genomics*, *1*(1), 1-10.

Nakamura, T., & Mizuno, S. (2010). The discovery of hepatocyte growth factor (HGF) and its significance for cell biology, life sciences and clinical medicine. *Proceedings of the Japan Academy, Series B*, *86*(6), 588-610.

Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M. S., & Bernabucci, U. (2010). Effects of climate changes on animal production and sustainability of livestock systems. *Livestock Science*, *130*(1-3), 57-69.

Negri, R., Aguilar, I., Feltes, G. L., & Cobuci, J. A. (2021). Selection for test-day milk yield and thermotolerance in Brazilian Holstein cattle. *Animals*, *11*(1), 128.

Nei, M. (1986). Definition and estimation of fixation indices. *Evolution*, *40*(3), 643-645.

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of human genetics*, *47*(3), 253-259.

Nicolazzi, E. L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., Pirani, A., Soans, C., Brew, F., & Jorjani, H. (2015). SNPchiMp v. 3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC genomics*, *16*(1), 1-6.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nature reviews genetics*, *8*(11), 857-868.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome research*, *15*(11), 1566-1575.

Niu, Q., Zhang, T., Xu, L., Wang, T., Wang, Z., Zhu, B., Gao, X., Chen, Y., Zhang, L., & Gao, H. (2021). Identification of candidate variants associated with bone weight using whole genome sequence in beef cattle. *Frontiers in Genetics*, *12*, 750746.

Nyamushamba, G., Mapiye, C., Tada, O., Halimani, T., & Muchenje, V. (2017). Conservation of indigenous cattle genetic resources in Southern Africa's smallholder areas: turning threats into opportunities—A review. *Asian-Australasian Journal of Animal Sciences*, *30*(5), 603.

Oleksyk, T. K., Zhao, K., De La Vega, F. M., Gilbert, D. A., O'Brien, S. J., & Smith, M. W. (2008). Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. *PloS one*, *3*(3), e1712.

Oliveira Júnior, G. A., Santos, D. J., Cesar, A. S., Boison, S. A., Ventura, R. V., Perez, B. C., Garcia, J. F., Ferraz, J. B. S., & Garrick, D. J. (2019). Fine mapping of genomic regions associated with female fertility in Nellore beef cattle based on sequence variants from segregating sires. *Journal of Animal Science and Biotechnology*, *10*(1), 1-13.

Olson, T. (1999). Genetics of colour variation. *The genetics of cattle*, *1*, 33-53.
Oosthuizen, M. P. (1996). *Uchibidolo: the abundant herds: a descriptive study of the Sanga-Nguni cattle of the Zulu people, with special reference to colour-pattern terminology and naming-practice*
Paim, T. d. P., Hay, E. H. A., Wilson, C., Thomas, M. G., Kuehn, L. A., Paiva, S. R., McManus, C., & Blackburn, H. (2020). Genomic breed composition of selection signatures in Brangus beef cattle. *Frontiers in Genetics*, *11*, 710.
Pal, A., Sharma, A., Bhattacharya, T., Chatterjee, P., & Chakravarty, A. (2011). Molecular characterization and SNP detection of CD14 gene of crossbred cattle. *Molecular biology international*, *2011*.
Peripolli, E., Reimer, C., Ha, N.-T., Geibel, J., Machado, M. A., Panetto, J. C. d. C., do Egito, A. A., Baldi, F., Simianer, H., & da Silva, M. V. G. B. (2020). Genome-wide detection of signatures of selection in indicine and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data. *BMC genomics*, *21*(1), 1-16.
Peripolli, E., Stafuzza, N. B., Munari, D. P., Lima, A. L. F., Irgang, R., Machado, M. A., Panetto, J. C. d. C., Ventura, R. V., Baldi, F., & da Silva, M. V. G. B. (2018). Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. *BMC genomics*, *19*(1), 1-13.
Perry, B. D. (2002). *Investing in animal health research to alleviate poverty*. ILRI (aka ILCA and ILRAD).
Pienaar, L., Neser, F., Grobler, J., Scholtz, M., & MacNeil, M. (2015). Pedigree analysis of the Afrikaner cattle breed. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales*, *57*, 51-56.
Pienaar, P. L. (2013). *Typology of smallholder farming in South Africa's former homelands: Towards an appropriate classification system* Stellenbosch: Stellenbosch University].
Pierce, M. D., Dzama, K., & Muchadeyi, F. C. (2018). Genetic diversity of seven cattle breeds inferred using copy number variations. *Frontiers in Genetics*, *9*, 163.
Prasad, A., Schnabel, R., McKay, S., Murdoch, B., Stothard, P., Kolbehdari, D., Wang, Z., Taylor, J., & Moore, S. (2008). Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal Genetics*, *39*(6), 597-605.
Prayaga, K., Barendse, W., & Burrow, H. (2006). Genetics of tropical adaptation. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006,
Prezeworski, M., Coop, G., & Wall, J. D. (2005). The signature of positive selection on standing genetic variation. *Evolution*, *59*(11), 2312-2323.
Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature reviews genetics*, *11*(7), 459-463.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.
Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC genetics*, *13*(1), 1-11.
Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., Nielsen, R., & Simianer, H. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS genetics*, *10*(2), e1004148.
Qanbari, S., & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock Science*, *166*, 133-143.
Qwabe, S. O., Maiwashe, A., & Muchadeyi, F. (2013). Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. *South African Journal of Animal Science*, *43*(1), 64-67.

Ramsay, K. (2010). Adaptive traits of Sanga cattle: Their importance in meeting the challenges associated with climate change in the tropics and sub tropics. *Advances in Animal Biosciences*, *1*(2), 381-382.

Ramzan, M., Murtaza, G., Sattar, A., Munawar, N., Ullah, A., Ejaz, A., Ayaz, F., Anwar, S., Jameel, K., & Kamran, F. (2021). Techniques for managing ticks and tick-borne diseases under changing climate; A review. *Egyptian Academic Journal of Biological Sciences, B. Zoology*, *13*(1), 117-128.

Randhawa, I. A., Khatkar, M. S., Thomson, P. C., & Raadsma, H. W. (2016). A meta-assembly of selection signatures in cattle. *PloS one*, *11*(4), e0153013.

Rebelato, A. B., & Caetano, A. R. (2018). Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. *Pesquisa Agropecuária Brasileira*, *53*, 975-984.

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

Ribeiro, V. M. P., Gouveia, G. C., de Moraes, M. M., de Araújo, A. E. M., Raidan, F. S. S., de Souza Fonseca, P. A., Cardoso, E. P., da Silva, M. V. G. B., & Toral, F. L. B. (2021). Genes underlying genetic correlation between growth, reproductive and parasite burden traits in beef cattle. *Livestock Science*, *244*, 104332.

Rocha, R. d. F. B., Garcia, A. O., Otto, P. I., da Silva, M. V. B., Martins, M. F., Machado, M. A., Panetto, J. C. d. C., & Guimarães, S. E. F. (2023). Runs of homozygosity and signatures of selection for number of oocytes and embryos in the Gir Indicine cattle. *Mammalian Genome*, 1-15.

Roux, G. (2023). *Genesis Home of Nguni Cattle Australia*. Retrieved 4 November from

Ruan, D., Yang, J., Zhuang, Z., Ding, R., Huang, J., Quan, J., Gu, T., Hong, L., Zheng, E., & Li, Z. (2022). Assessment of heterozygosity and genome-wide analysis of heterozygosity regions in two duroc pig populations. *Frontiers in Genetics*, *12*, 812456.

SA Stud Book (2022). *SA Stud Book Ayrshire and related types annual report 2021–2022*. Pretoria, South Africa.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T., Altshuler, D., & Lander, E. (2006). Positive natural selection in the human lineage. *Science*, *312*(5780), 1614-1620.

Sabir, J., Mutwakil, M., El-Hanafy, A., Al-Hejin, A., Sadek, M. A., Abou-Alsoud, M., Qureshi, M., Saini, K., & Ahmed, M. (2014). Applying molecular tools for improving livestock performance: From DNA markers to next generation sequencing technologies. *Journal of Food, Agriculture and Environment*, *12*(2), 541-553.

Sahoo, N., Kumar, P., Bhushan, B., Bhattacharya, T., Dayal, S., & Sahoo, M. (2012). Lysozyme in livestock: a guide to selection for disease resistance: a review.

Samuels, D. C., Wang, J., Ye, F., He, J., Levinson, R. T., Sheng, Q., Zhao, S., Capra, J. A., Shyr, Y., & Zheng, W. (2016). Heterozygosity ratio, a robust global genomic measure of autozygosity and its association with height and disease risk. *Genetics*, *204*(3), 893-904.

Sanarana, Y., Maake, M., Tada, O., Muchadeyi, F., & Banga, C. (2021). An analysis of genetic diversity and population structure of South African smallholder dairy cattle herds using SNP markers. *Sustainable Animal Production and Health*, 275.

Sanarana, Y. P. (2015). *Genetic characterization of South African Nguni cattle ecotypes using microsatellite markers* University of Pretoria].

Sanglard, L. P., Huang, Y., Gray, K. A., Linhares, D. C., Dekkers, J. C., Niederwerder, M. C., Fernando, R. L., & Serão, N. V. (2021). Further host-genomic characterization of total antibody response to PRRSV vaccination and its relationship with reproductive performance in commercial sows: genome-wide haplotype and zygosity analyses. *Genetics Selection Evolution*, *53*, 1-17.

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

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

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

Scholtz, M. (2010). *Beef breeding in South Africa*. Agricultural Research Council.

Scholtz, M., Bester, J., Mamabolo, J., & Ramsay, K. (2008). Results of the national cattle survey undertaken in South Africa, with emphasis on beef. *Appl. Anim. Husb. Rural Dev*, *1*, 1-9.

Schröder, N. W., Opitz, B., Lamping, N., Michelsen, K. S., Zähringer, U., Göbel, U. B., & Schumann, R. R. (2000). Involvement of lipopolysaccharide binding protein, CD14, and Toll-like receptors in the initiation of innate immune responses by Treponema glycolipids. *The Journal of Immunology*, *165*(5), 2683-2693.

Sejian, V., Shashank, C. G., Silpa, M. V., Madhusoodan, A. P., Devaraj, C., & Koenig, S. (2022). Non-Invasive Methods of Quantifying Heat Stress Response in Farm Animals with Special Reference to Dairy Cattle. *Atmosphere*, *13*(10), 1642.

Selli, A., Ventura, R. V., Fonseca, P. A., Buzanskas, M. E., Andrietta, L. T., Balieiro, J. C., & Brito, L. F. (2021). Detection and visualization of heterozygosity-rich regions and runs of homozygosity in worldwide sheep populations. *Animals*, *11*(9), 2696.

Shibata, M. (2014). Influence of housing density and grazing on heat shock protein 27 expression in skeletal muscle of beef cattle. *Journal of Fisheries & Livestock Production*, *2*(02).

Shyma, K., Gupta, J. P., & Singh, V. (2015). Breeding strategies for tick resistance in tropical cattle: a sustainable approach for tick control. *Journal of Parasitic Diseases*, *39*, 1-6.

Sigdel, A., Liu, L., Abdollahi-Arpanahi, R., Aguilar, I., & Peñagaricano, F. (2020). Genetic dissection of reproductive performance of dairy cows under heat stress. *Animal Genetics*, *51*(4), 511-520.

Silva, M., Verardo, L., Machado, M., Panetto, J. d. C., Carolino, I., & Carolino, N. (2022). Candidate genes for disease, reproduction and meat quality traits in Portuguese native breeds. Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges,

Singh, A., Mehrotra, A., Gondro, C., Romero, A. R. d. S., Pandey, A. K., Karthikeyan, A., Bashir, A., Mishra, B., Dutt, T., & Kumar, A. (2020). Signatures of selection in composite Vrindavani cattle of India. *Frontiers in Genetics*, *11*, 589496.

Singh, G., Samad, H. A., Karthiga, K., Priyanka, K., Sarma, L., Chouhan, V. S., & Maurya, V. (2022). Comparative Assessment of Thermo-Tolerance of Crossbred and Indigenous Cattle Breeds. In *Climate Change and Livestock Production: Recent Advances and Future Perspectives* (pp. 73-81). Springer.

Sölkner, J., Ferencakovic, M., Gredler, B., & Curik, I. (2010). Genomic metrics of individual autozygosity, applied to a cattle population. Proceedings of the 61st Annual Meeting of the European Association for Animal Production,

Strydom, P. (2002). Factors influencing growth, carcass cattle yield and carcass composition of cattle. *Feedlot management. Ed. Leeuw, KJ. Agricultural Research Council, Animal Production Institute, Irene*, 40.

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

Sun, T., Pei, S., Liu, Y., Hanif, Q., Xu, H., Chen, N., Lei, C., & Yue, X. (2023). Whole genome sequencing of simmental cattle for SNP and CNV discovery. *BMC genomics*, *24*(1), 179.

Szmatoła, T., Gurgul, A., Jasielczuk, I., & Ropka-Molik, K. (2023). Comprehensive analysis of runs of homozygosity and heterozygosity of Holstein cattle on the basis of medium and high density SNP panels and large population sample. *Annals of Animal Science*.

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

Tabor, A. E., Ali, A., Rehman, G., Rocha Garcia, G., Zangirolamo, A. F., Malardo, T., & Jonsson, N. N. (2017). Cattle tick Rhipicephalus microplus-host interface: a review of resistant and susceptible host responses. *Frontiers in cellular and infection microbiology*, *7*, 506.

Tada, O., Muchenje, V., & Dzama, K. (2013). Preferential traits for breeding Nguni cattle in low-input in-situ conservation production systems. *Springerplus*, *2*, 1-7.

Tadesse, G., & Dereje, M. (2018). Impact of Climate Change o Africa-Review. *Climate Change*, *4*(15), 299-313.

Taye, M., Yoon, J., Dessie, T., Cho, S., Oh, S. J., Lee, H.-K., & Kim, H. (2018). Deciphering signature of selection affecting beef quality traits in Angus cattle. *Genes & genomics*, *40*, 63-75.

Terefe, E., Belay, G., Han, J., Hanotte, O., & Tijjani, A. (2022). Genomic adaptation of Ethiopian indigenous cattle to high altitude. *Frontiers in Genetics*, *13*, 960234.

Toro Ospina, A. M., Maiorano, A. M., Curi, R. A., Pereira, G. L., Zerlotti-Mercadante, M. E., dos Santos Goncalves Cyrillo, J. N., Aspilcueta-Borquis, R. R., & de V. Silva, J. A. I. (2019). Linkage disequilibrium and effective population size in Gir cattle selected for yearling weight. *Reproduction in Domestic Animals*, *54*(12), 1524-1531.

Troy, C. S., MacHugh, D. E., Bailey, J. F., Magee, D. A., Loftus, R. T., Cunningham, P., Chamberlain, A. T., Sykes, B. C., & Bradley, D. G. (2001). Genetic evidence for Near-Eastern origins of European cattle. *Nature*, *410*(6832), 1088-1091.

Tsartsianidou, V., Sánchez-Molano, E., Kapsona, V. V., Basdagianni, Z., Chatziplis, D., Arsenos, G., Triantafyllidis, A., & Banos, G. (2021). A comprehensive genome-wide scan detects genomic regions related to local adaptation and climate resilience in Mediterranean domestic sheep. *Genetics Selection Evolution*, *53*(1), 1-17.

Upadhyay, M., Chen, W., Lenstra, J. A., Goderie, C., MacHugh, D., Park, S., Magee, D. A., Matassino, D., Ciani, F., & Megens, H. (2017). Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. *Heredity*, *118*(2), 169-176.

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

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

van der Westhuizen, L., MacNeil, M. D., Scholtz, M. M., Neser, F. W., Makgahlela, M. L., & van Wyk, J. B. (2020). Genetic variability and relationships in nine South African cattle breeds using microsatellite markers. *Tropical animal health and production*, *52*, 177-184.

Van Marle-Köster, Lashmar, S. F., Retief, A., & Visser, C. (2021). Whole-Genome SNP Characterisation Provides Insight for Sustainable Use of Local South African Livestock Populations. *Frontiers in Genetics*, *12*, 714194.

van Marle-Köster, E., Lashmar, S., Okpeku, M., Muchadeyi, F., Mapholi, N., & Visser, C. (2022). Assessing genomic diversity and heterozygosity-rich regions of transboundary Nguni cattle of South Africa

and eSwatini. Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges,

Visser, C. (2023). Breeding for Adaptability. *Fertility, Profitable Veld Cattle Journal 2023*, 14-15.

Visser, C., & Snyman, M. A. (2023). Incorporating new technologies in breeding plans for South African goats in harsh environments. *Animal frontiers*, *13*(5), 53-59.

Walsh, J. B. (2021). Genomic selection signatures and animal breeding. In (Vol. 138, pp. 1-3): Wiley Online Library.

Wang, J.-J., Li, Z.-D., Zheng, L.-Q., Zhang, T., Shen, W., & Lei, C.-Z. (2022). Genome-wide detection of selective signals for fecundity traits in goats (Capra hircus). *Gene*, *818*, 146221.

Wang, S.-H., Cheng, C.-Y., Tang, P.-C., Chen, C.-F., Chen, H.-H., Lee, Y.-P., & Huang, S.-Y. (2015). Acute heat stress induces differential gene expressions in the testes of a broiler-type strain of Taiwan country chickens. *PloS one*, *10*(5), e0125816.

Wang, X., Ju, Z., Jiang, Q., Zhong, J., Liu, C., Wang, J., Hoff, J. L., Schnabel, R. D., Zhao, H., & Gao, Y. (2021). Introgression, admixture, and selection facilitate genetic adaptation to high-altitude environments in cattle. *Genomics*, *113*(3), 1491-1503.

Weigand, H., & Leese, F. (2018). Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society*, *184*(2), 528-583.

Weldenegodguad, M., Popov, R., Pokharel, K., Ammosov, I., Ming, Y., Ivanova, Z., & Kantanen, J. (2019). Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. *Frontiers in Genetics*, *9*, 728.

Wen, Q., Li, N., Xiao, X., Lui, W.-y., Chu, D. S., Wong, C. K., Lian, Q., Ge, R., Lee, W. M., & Silvestrini, B. (2018). Actin nucleator Spire 1 is a regulator of ectoplasmic specialization in the testis. *Cell Death & Disease*, *9*(2), 208.

Wiggans, G., Cooper, T., Van Tassell, C., Sonstegard, T., & Simpson, E. (2013). Characteristics and use of the Illumina BovineLD and GeneSeek Genomic Profiler low-density bead chips for genomic evaluation. *Journal of dairy science*, *96*(2), 1258-1263.

Willemse, M. (2019). *The Effect of Ascertainment Bias on Detecting Signatures of Selection*

Willing, E.-M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers.

Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, *15*(1), 323-354.

Wu, S., Li, R. W., Li, W., Beshah, E., Dawson, H. D., & Urban Jr, J. F. (2012). Worm burden-dependent disruption of the porcine colon microbiota by Trichuris suis infection. *PloS one*, *7*(4), e35470.

Xu, L., Yang, L., Zhu, B., Zhang, W., Wang, Z., Chen, Y., Zhang, L., Gao, X., Gao, H., & Liu, G. E. (2019). Genome-wide scan reveals genetic divergence and diverse adaptive selection in Chinese local cattle. *BMC genomics*, *20*(1), 1-12.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76-82.

Yayou, K., Kitagawa, S., Ito, S., Kasuya, E., & Sutoh, M. (2009). Effects of intracerebroventricular administration of neuromedin U or neuromedin S in steers. *General and comparative endocrinology*, *163*(3), 324-328.

Yoon, H., Shin, J., Yang, S., Chae, D., Kim, H., Lee, D., Kim, H., Kim, S., Lee, J., & Kim, Y. (2003). Association of the CD14 gene–159C polymorphism with progression of IgA nephropathy. *Journal of medical genetics*, *40*(2), 104-108.

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

Zavarez, L. B., Utsunomiya, Y. T., Carmo, A. S., Neves, H. H., Carvalheiro, R., Ferenčaković, M., Pérez O'Brien, A. M., Curik, I., Cole, J. B., & Van Tassell, C. P. (2015). Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. *Frontiers in Genetics*, *6*, 5.

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences*, *105*(33), 11597-11604.

Zhang, W., Yang, M., Zhou, M., Wang, Y., Wu, X., Zhang, X., Ding, Y., Zhao, G., Yin, Z., & Wang, C. (2020). Identification of signatures of selection by whole-genome resequencing of a Chinese native pig. *Frontiers in Genetics*, *11*, 566255.

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

Zhong, J.-L., Xu, J.-W., Wang, J., Wen, Y.-f., Niu, H., Zheng, L., He, H., Peng, K., He, P., & Shi, S.-Y. (2019). A novel SNP of PLAG1 gene and its association with growth traits in Chinese cattle. *Gene*, *689*, 166-171.

Zindove, T., & Chimonyo, M. (2015). Are calving interval, abortions, incidence of stillbirths and pre-weaning losses in Nguni cows associated with linear type traits? *Animal reproduction science*, *160*, 49-56.

Zorc, M., Ogorevc, J., & Dovc, P. (2019). The new bovine reference genome assembly provides new insight into genomic organization of the bovine major histocompatibility complex. *Journal of Central European Agriculture*, *20*(4), 1111-1115.

Zwane, A. A., Maiwashe, A., Makgahlela, M., Choudhury, A., Taylor, J., & van Marle-Köster, E. (2016). Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds. *South African Journal of Animal Science*, *46*(3), 302-312.

Zwane, A. A., Nxumalo, K., Makgahlela, M., Van Marle-Koster, E., & Maiwashe, N. (2021). Gene-set enrichment analysis of selective sweeps reveals phenotypic traits in Nguni cattle. *South African Journal of Animal Science*, *51*(6), 761-777.

Zwane, A. A., Schnabel, R. D., Hoff, J., Choudhury, A., Makgahlela, M. L., Maiwashe, A., Van Marle-Koster, E., & Taylor, J. F. (2019). Genome-wide SNP discovery in indigenous cattle breeds of South Africa. *Frontiers in Genetics*, *10*, 273.

Addendum

Table 5.1: Position and number of SNPs involved in the genomic regions identified as signatures of selection using ROH in DRB, NGI, and TUL breeds.

Breed	Start_SNP	End_SNP	BTA	nSNP	from	To
DRB	BovineHD1000008919	BovineHD1000031294	10	60	27170804	28426603
	BovineHD1400007031	BovineHD1400007419	14	341	24243733	25636017
	BovineHD0600010984	Hapmap24229-BTC-037356	6	53	40141642	41211754
	BovineHD0600018302	BovineHD0600018446	6	23	66336540	66858247
	BovineHD0600018463	BovineHD0600018707	6	44	66900340	67771797
NGI	BovineHD1100016485	ARS-BFGL-NGS-15285	11	65	56735630	58241212
	BTA-33625-no-rs	ARS-BFGL-BAC-11748	11	6	58306286	58442169
	BovineHD1900000626	BovineHD1900000656	19	5	2665924	2782726

	BovineHD1900000668	BovineHD1900000776	19	31	2847032	3357822
	BovineHD0700015166	ARS-BFGL-NGS-96012	7	16	52499466	53101552
TUL	ARS-USDA-AGIL-chr10-27190494-000180	ARS-USDA-AGIL-chr10-27217129-000181	10	4	27190494	27217129
	BovineHD1000009004	BTB-00916903	10	29	27443844	28070182
	BovineHD1000016367	BovineHD1000016380	10	5	55273043	55329485
	BovineHD1100016326	BovineHD1100016674	11	73	55904351	57695824
	BovineHD1300013616	BovineHD1300014104	13	83	46522038	48360800
	BovineHD1400003475	BovineHD4100011184	14	21	11951308	12401118
	BovineHD1400003609	BovineHD4100011226	14	41	12430302	13251085
	BovineHD1400004422	Hapmap26696-BTA-147023	14	5	15760627	15836610
	ARS-BFGL-NGS-33755	UA-IFASA-5528	14	21	16185315	16677121
	BovineHD1400024413	Hapmap54398-rs29020900	14	228	16823272	21870878
	ARS-BFGL-NGS-3198	BovineHD1400006888	14	263	21933950	23735038
	BovineHD1400006889	BovineHD1400007295	14	347	23737521	25164603
	BovineHD1400007296	ARS-BFGL-NGS-119937	14	493	25173600	27777207
	ARS-BFGL-BAC-24232	BovineHD1400015402	14	3	54848245	54902309
	BovineHD1400020220	Hapmap50328-BTA-35439	14	32	72183245	72827899
	BovineHD1900000003	BovineHD1900000913	19	171	90671	3880072
	BovineHD2200012971	BovineHD2200013330	22	64	44811911	46323159
	BovineHD2400010639	Hapmap30511-BTA-129561	24	50	38916982	39964041
	Hapmap50828-BTA-94125	BovineHD2400012305	24	202	40342820	44856048
	BovineHD2400013741	BovineHD2400014236	24	65	49113246	50799830
	ARS-BFGL-NGS-75548	ARS-BFGL-NGS-16805	4	52	76072824	77150031
	BTB-01812364	BovineHD0500012204	5	38	41666974	42701427
	BovineHD0500012208	ARS-BFGL-NGS-7850	5	2	42732874	42775818
	BTB-00226490	BTB-00226490	5	1	43034529	43034529
	BovineHD0500012334	BovineHD0500012336	5	3	43099238	43111315
	BTB-00226702	BovineHD0500012390	5	2	43275588	43282776
	BovineHD0500012414	BovineHD4100003632	5	7	43355001	43490630
	BovineHD0500012484	BovineHD0500013171	5	100	43629951	45753061
	BTB-00250110	Hapmap24229-BTC-037356	6	48	40236966	41211754
	BovineHD0700004782	BovineHD0700004804	7	5	17112594	17221471
	BovineHD0700005009	BovineHD0700005017	7	2	17808655	17856462
	ARS-BFGL-NGS-70183	BovineHD0700005074	7	9	17913294	18089438
	ARS-BFGL-NGS-59838	BovineHD0700005454	7	10	19461567	19599003

nSNP = Number of SNPs; BTA = *Bos Taurus* Autosome

Table 5.2: Position and number of SNPs involved in the genomic regions considered to be signatures of selection using ROHet in DRB, NGI, and TUL

Breed	Start_SNP	End_SNP	BTA	nSNP	from	to
DRB	BovineHD0100020811	BovineHD0100020849	1	11	72714365	72767001
	BTB-01955138	BovineHD0100040777	1	7	1.42E+08	142181172
	BovineHD1100008050	Hapmap23233-BTA-88497	11	10	27030068	27147078
	BTA-88624-no-rs	ARS-BFGL-NGS-68010	11	10	29691021	29774437
	BovineHD1100017557	BovineHD1100017598	11	10	61969147	62046972
	BovineHD1200011752	BovineHD1200011835	12	10	41635295	41970238
	BovineHD1300012084	Hapmap58859-rs29023610	13	10	41724999	41833607
	BovineHD1400003630	BovineHD1400003639	14	3	12508525	12543734
	BovineHD1400006736	BovineHD1400006757	14	18	23240328	23283727
	BovineHD1400007003	BovineHD1400007014	14	11	24158787	24193383
	BovineHD1400007072	BovineHD1400007085	14	14	24384496	24418370
	BovineHD1400007623	BovineHD1400007632	14	10	26474157	26491080
	BovineHD1400007650	BovineHD1400007661	14	11	26574416	26593712
	BovineHD1500006315	BovineHD1500006330	15	13	24211069	24224422
	ARS-BFGL-NGS-93226	BovineHD1800010638	18	12	34817149	35105083
	BovineHD1800019071	BovineHD1800019076	18	6	65493523	65503397
	BovineHD0200018907	BovineHD0200018979	2	11	65404949	65551073
	BovineHD2000001308	ARS-BFGL-NGS-20128	20	10	4057908	4073020
	BovineHD2000006630	Hapmap39254-BTA-119282	20	12	22151938	22165165
	BovineHD2000006643	BovineHD2000006668	20	25	22166679	22213427
	BovineHD2000006671	BovineHD2000006681	20	10	22220534	22234812
	BovineHD2000008971	UA-IFASA-9183	20	11	30865802	31195125
	BovineHD2100016564	BovineHD2100016572	21	8	57625740	57636658
	BovineHD2300015442	ARS-BFGL-BAC-29485	23	13	337522	582242
	BovineHD2400007287	BovineHD2400007307	24	17	26752313	26787812
	BovineHD2400007362	BovineHD2400007379	24	18	26915455	26957084
	BovineHD2400007513	BovineHD4100016534	24	12	27585144	27608874
	ARS-BFGL-NGS-36624	BovineHD2500011713	25	10	41497446	41585097
	BovineHD2600003529	ARS-BFGL-NGS-89510	26	13	13696293	13936262
	BovineHD2600004956	ARS-BFGL-NGS-113339	26	10	19080910	19237604
	ARS-BFGL-NGS-38953	BovineHD0300010302	3	18	32988739	33030685
	BovineHD0300015825	BovineHD0300015838	3	11	52417930	52439195
	BovineHD0300015944	BovineHD0300015958	3	13	52819645	52853050
	BTB-00180546	Hapmap44277-BTA-70372	4	12	48431383	48726071
	BovineHD0500014005	BovineHD0500014014	5	10	48521028	48563017
	BovineHD0500014029	BovineHD0500014069	5	30	48647967	48833337
	BovineHD0500014092	BovineHD0500014102	5	11	48925240	48960291

	BovineHD0500030458	BovineHD0500030473	5	14	1.06E+08	106196654
	BovineHD0500030502	BovineHD0500030513	5	12	1.06E+08	106391372
	BovineHD0500030671	BovineHD0500030681	5	10	1.07E+08	106826774
	BovineHD0500030736	BovineHD0500030762	5	27	1.07E+08	106987104
	BovineHD0600010339	BovineHD0600010340	6	2	37075974	37078344
	BovineHD0600010365	BovineHD0600010379	6	14	37236550	37268940
	BTA-22850-no-rs	BovineHD0600010549	6	11	37983812	38003903
	BovineHD0600010587	BovineHD0600010601	6	13	38232631	38264991
	BovineHD0600010802	BovineHD0600010809	6	10	39246418	39259257
	BovineHD0700000124	BovineHD0700000135	7	11	680008	713300
	BovineHD0700006561	BovineHD0700006574	7	11	23912976	23959904
	BovineHD0700015001	BovineHD0700015166	7	18	51695898	52499466
	BovineHD0700016127	ARS-BFGL-NGS-118213	7	8	55871886	56055399
	BovineHD0800008634	BovineHD0800008681	8	10	28546561	28707852
	BovineHD0800013766	BovineHD0800013893	8	18	46057409	46392802
	BovineHD0900013468	BTB-00391496	9	9	48845335	48975316
	BovineHD0900026959	BovineHD0900026979	9	9	95123948	95178303
NGI	BovineHD0100020815	BovineHD0100020849	1	10	72721380	72767001
	ARS-BFGL-NGS-27443	BovineHD0100045967	1	9	1.57E+08	157209697
	BovineHD1000017747	BovineHD1000017828	10	14	61240034	61661874
	BovineHD1300012084	Hapmap58859-rs29023610	13	10	41724999	41833607
	BovineHD1400006737	BovineHD1400006757	14	17	23244975	23283727
	BTB-01530778	BovineHD1400007122	14	14	24482969	24519904
	BovineHD1400007191	BovineHD1400007208	14	13	24758139	24821431
	BovineHD1700017382	BovineHD1700017424	17	10	60959691	61176858
	BovineHD1800000649	BovineHD1800000665	18	5	2369944	2459027
	BovineHD2000001636	BovineHD2000001645	20	10	5188404	5200939
	BovineHD2000006624	Hapmap39254-BTA-119282	20	17	22139478	22165165
	BovineHD2000006650	BovineHD2000006668	20	18	22177605	22213427
	BovineHD2000006671	BovineHD2000006681	20	10	22220534	22234812
	BovineHD2000011491	BovineHD2000011567	20	12	40180680	40340545
	BovineHD2000014091	BovineHD2000014116	20	6	50876323	50982008
	BovineHD2400007287	BovineHD2400007307	24	18	26752313	26787812
	BovineHD2400007362	BovineHD2400007379	24	17	26915455	26957084
	BovineHD2400007385	BovineHD2400007394	24	10	26993580	27033953
	BovineHD2400007513	BovineHD2400007522	24	11	27585144	27607600
	BovineHD2700000235	BovineHD2700000241	27	2	769259	780853
	ARS-BFGL-NGS-38953	BovineHD0300010302	3	18	32988739	33030685
	BTA-119272-no-rs	Hapmap53369-rs29012431	3	8	41572438	41673315
	BovineHD0500014033	BovineHD0500014047	5	11	48665878	48722739
	BovineHD0500014049	BovineHD0500014069	5	16	48727320	48833337
	BovineHD0500014073	BovineHD0500014086	5	10	48853528	48902876
	BovineHD0500014092	BovineHD0500014111	5	18	48925240	49004461
	BovineHD0500015970	BovineHD0500015981	5	10	56171855	56211341

	BovineHD0600010365	BovineHD0600010379	6	16	37236550	37268940
	BovineHD0600010587	BovineHD0600010601	6	13	38232631	38264991
	BovineHD0600010803	BovineHD4100004606	6	10	39250536	39261628
	BovineHD0700000124	BovineHD0700033469	7	10	680008	707281
	BovineHD0700000140	BovineHD4100005669	7	10	723567	759575
	BovineHD0700006561	BovineHD0700006574	7	11	23912976	23959904
	BTA-79078-no-rs	BovineHD0700014967	7	11	51127488	51520623
	BovineHD0700015001	BovineHD0700015018	7	1	51695898	51772534
TUL	BTA-110282-no-rs	BovineHD0100015120	1	10	53582877	53653911
	BovineHD0100020815	BovineHD0100020849	1	10	72721380	72767001
	BovineHD1400006736	BovineHD1400006747	14	11	23240328	23262932
	BovineHD1400007006	BovineHD1400007015	14	11	24167861	24194599
	BovineHD1400007072	BovineHD1400007085	14	14	24384496	24418370
	BovineHD1400007621	BovineHD1400007632	14	11	26471856	26491080
	BovineHD4100011349	BovineHD4100011353	14	10	26856813	26882634
	BovineHD1400007734	BovineHD1400007740	14	10	26887021	26905915
	BovineHD1600007209	BovineHD1600007243	16	14	25935379	26045099
	Hapmap40674-BTA-121055	ARS-BFGL-NGS-84156	16	11	42625201	42941953
	BovineHD1700011695	Hapmap28103-BTA-154966	17	7	42120875	42206854
	BovineHD0200000815	BovineHD0200000841	2	8	3030698	3140659
	BovineHD0200003380	Hapmap44546-BTA-48673	2	10	12046675	12138830
	BovineHD0200018979	BovineHD0200018979	2	1	65551073	65551073
	BovineHD0200037007	BovineHD0200037016	2	8	1.27E+08	127526531
	BovineHD2000001666	BovineHD2000001679	20	14	5332823	5362947
	BovineHD2000001744	BovineHD2000001753	20	9	5664976	5679628
	BovineHD2000001786	BovineHD2000001797	20	10	5742243	5752697
	BovineHD2000006624	Hapmap39254-BTA-119282	20	17	22139478	22165165
	BovineHD2000006643	BovineHD2000006643	20	1	22166679	22166679
	BovineHD2000006650	BovineHD2000006683	20	32	22177605	22246591
	BovineHD2000011491	BovineHD2000011567	20	12	40180680	40340545
	BovineHD2100000457	BovineHD2100021047	21	14	3075947	3371330
	BovineHD2100006617	BovineHD2100006660	21	10	22543536	22641070
	BovineHD2300015442	ARS-BFGL-BAC-29485	23	13	337522	582242
	BovineHD2300000103	BovineHD2300000153	23	12	736378	936645
	BovineHD2400007169	BovineHD2400007178	24	10	26337388	26356410
	BovineHD2400007287	BovineHD2400007307	24	18	26752313	26787812
	BovineHD2400007362	BovineHD2400007379	24	17	26915455	26957084
	BovineHD2400007513	BovineHD4100016534	24	12	27585144	27608874
	BovineHD2700000235	BovineHD2700000285	27	13	769259	897201
	BovineHD2700002266	BovineHD2700002289	27	8	7202691	7311537
	BovineHD0300015944	BovineHD0300015953	3	11	52819645	52844880
	BovineHD0300021558	BTA-107738-no-rs	3	10	74072243	74315105
	BovineHD0500010200	BovineHD0500010213	5	4	35568704	35624882

BovineHD0500014056	BovineHD0500014069	5	12	48772978	48833337
BovineHD0500014073	BovineHD0500014086	5	10	48853528	48902876
BovineHD0500014092	BovineHD0500014102	5	11	48925240	48960291
BovineHD0500030746	BovineHD0500030756	5	9	1.07E+08	106981208
BovineHD0600010338	BovineHD0600010340	6	3	37075413	37078344
BovineHD0600010365	BovineHD0600010381	6	18	37236550	37279162
BovineHD0600010587	BovineHD0600010600	6	12	38232631	38264184
BovineHD0600010674	BovineHD0600010679	6	10	38592222	38605879
BovineHD0600010802	BovineHD4100004606	6	11	39246418	39261628
BovineHD0700000122	BovineHD0700000137	7	15	656981	715134
BovineHD0700006372	BovineHD0700006383	7	13	23264972	23286745
BovineHD0700006504	BovineHD0700006513	7	10	23736205	23752203
BovineHD0700010950	BovineHD0700010987	7	10	38002844	38136792
BovineHD0700014947	Hapmap41088-BTA-78936	7	21	51427741	52419683
BovineHD0700016127	ARS-BFGL-NGS-118213	7	8	55871886	56055399
ARS-BFGL-NGS-28838	Hapmap31524-BTA-163487	8	11	46020875	46154811
BovineHD0800028832	BovineHD0800028870	8	10	97599268	97748292
BovineHD0900002717	BovineHD0900002730	9	13	10949242	10970517
BTA-24108-no-rs	BovineHD0900026986	9	17	95041591	95199303

nSNP = Number of SNPs; BTA = *Bos Taurus* Autosome

Table 5.3: A comprehensive list of genes identified in the genomic regions considered to be signatures of selection using ROH in DRB, NGI, and TUL

BREED	Gene ID	Ensembl Gene ID	Chr	Position (Mbp)	Description
TUL	*IGFBP3*	ENSBTAG00000003994	4	76.12	insulin like growth factor binding protein 3
TUL	*IGFBP1*	ENSBTAG00000046768	4	76.13	insulin like growth factor binding protein 1
TUL	*ADCY1*	ENSBTAG00000009520	4	76.17	adenylate cyclase 1
TUL	*RAMP3*	ENSBTAG00000020704	4	76.44	receptor activity modifying protein 3
TUL	*TBRG4*	ENSBTAG00000010504	4	76.52	transforming growth factor beta regulator 4
TUL	*CCM2*	ENSBTAG00000008090	4	76.55	CCM2 scaffold protein
TUL	*MYO1G*	ENSBTAG00000006377	4	76.62	myosin IG
TUL	*PURB*	ENSBTAG00000052945	4	76.67	purine rich element binding protein B
TUL	*H2AZ2*	ENSBTAG00000016975	4	76.69	H2A.Z variant histone 2
TUL	*PPIA*	ENSBTAG00000012003	4	76.71	peptidylprolyl isomerase A
TUL	*ZMIZ2*	ENSBTAG00000011997	4	76.73	zinc finger MIZ-type containing 2
TUL	*OGDH*	ENSBTAG00000006029	4	76.76	oxoglutarate dehydrogenase
TUL	*DDX56*	ENSBTAG00000010602	4	76.87	DEAD-box helicase 56
TUL	*NPC1L1*	ENSBTAG00000044146	4	76.89	NPC1 like intracellular cholesterol transporter 1
TUL	*NUDCD3*	ENSBTAG00000006325	4	76.92	NudC domain containing 3
TUL	*CAMK2B*	ENSBTAG00000012653	4	77.03	calcium/calmodulin dependent protein kinase II beta
TUL	*YKT6*	ENSBTAG00000000274	4	77.12	YKT6 v-SNARE homolog
TUL	*KIF21A*	ENSBTAG00000004832	5	41.73	kinesin family member 21A
TUL	*CPNE8*	ENSBTAG00000020914	5	42.19	copine 8 [Source:VGNC Symbol;Acc:VGNC:27664]
TUL	*PTPRR*	ENSBTAG00000015311	5	42.61	protein tyrosine phosphatase receptor type R
TUL	*KCNMB4*	ENSBTAG00000003749	5	43.04	potassium calcium-activated channel subfamily M regulatory beta subunit 4
TUL	*MYRFL*	ENSBTAG00000014829	5	43.52	myelin regulatory factor like
TUL	*RAB3IP*	ENSBTAG00000031950	5	43.66	RAB3A interacting protein
TUL	*BEST3*	ENSBTAG00000002931	5	43.76	bestrophin 3
TUL	*LRRC10*	ENSBTAG00000019157	5	43.86	leucine rich repeat containing 10
TUL	*CCT2*	ENSBTAG00000019156	5	43.87	chaperonin containing TCP1 subunit 2
TUL	*FRS2*	ENSBTAG00000019155	5	43.89	fibroblast growth factor receptor substrate 2
TUL	*YEATS4*	ENSBTAG00000002256	5	44.09	YEATS domain containing 4
TUL	*LYZ2*	ENSBTAG00000026088	5	44.37	lysozyme C-2
TUL	*LYZ1*	ENSBTAG00000046511	5	44.39	lysozyme (renal amyloidosis)
TUL	*LYZ3*	ENSBTAG00000046628	5	44.42	lysozyme 3
TUL	*LYZ*	ENSBTAG00000026779	5	44.51	lysozyme
TUL	*CPSF6*	ENSBTAG00000007323	5	44.56	cleavage and polyadenylation specific factor 6
TUL	*CPM*	ENSBTAG00000013496	5	44.88	carboxypeptidase M
TUL	*MDM2*	ENSBTAG00000010422	5	44.98	MDM2 proto-onco
TUL	*NUP107*	ENSBTAG00000006911	5	45.08	nucleoporin 107
TUL	*RAP1B*	ENSBTAG00000008967	5	45.16	RAP1B, member of RAS onco family
TUL	*MDM1*	ENSBTAG00000019738	5	45.45	Mdm1 nuclear protein

TUL	*IL22*	ENSBTAG00000015407	5	45.52	interleukin 22
TUL	*IL26*	ENSBTAG00000052655	5	45.55	interleukin 26
TUL	*IFNG*	ENSBTAG00000012529	5	45.62	interferon gamma
DRB	*SLIT2*	ENSBTAG00000005108	6	39.78	slit guidance ligand 2
DRB	*PACRGL*	ENSBTAG00000033351	6	40.23	parkin coregulated like
DRB	*KCNIP4*	ENSBTAG00000047743	6	40.25	potassium voltage-gated channel interacting protein 4
DRB	*CORIN*	ENSBTAG00000002199	6	66.27	corin, serine peptidase
DRB	*NFXL1*	ENSBTAG00000002201	6	66.59	nuclear transcription factor, X-box binding like 1
DRB	*CNGA1*	ENSBTAG00000002205	6	66.67	cyclic nucleotide gated channel subunit alpha 1
DRB	*NIPAL1*	ENSBTAG00000009423	6	66.70	NIPA like domain containing 1
DRB	*TXK*	ENSBTAG00000005055	6	66.76	TXK tyrosine kinase
DRB	*TEC*	ENSBTAG00000005062	6	66.82	tec protein tyrosine kinase
DRB	*SLAIN2*	ENSBTAG00000021963	6	67.05	SLAIN motif family member 2
DRB	*SLC10A4*	ENSBTAG00000004888	6	67.17	solute carrier family 10 member 4
DRB	*ZAR1*	ENSBTAG00000004886	6	67.17	zygote arrest 1
DRB	*FRYL*	ENSBTAG00000000137	6	67.18	FRY like transcription coactivator
DRB	*OCIAD1*	ENSBTAG00000010611	6	67.50	OCIA domain containing 1
DRB	*OCIAD2*	ENSBTAG00000001839	6	67.53	OCIA domain containing 2
DRB	*CWH43*	ENSBTAG00000021347	6	67.63	cell wall biosis 43 C-terminal homolog
DRB	*DCUN1D4*	ENSBTAG00000001600	6	67.71	defective in cullin neddylation 1 domain containing 4
TUL	*ZNF414*	ENSBTAG00000007660	7	17.11	zinc finger protein 414
TUL	*MYO1F*	ENSBTAG00000007661	7	17.12	myosin IF
TUL	*ADAMTS10*	ENSBTAG00000014207	7	17.16	ADAM metallopeptidase with thrombospondin type 1 motif 10
TUL	*C3*	ENSBTAG00000017280	7	17.77	complement C3
TUL	*TNFSF14*	ENSBTAG00000012223	7	17.84	TNF superfamily member 14
TUL	*TNFSF9*	ENSBTAG00000046266	7	17.96	TNF superfamily member 9
TUL	*TUBB4A*	ENSBTAG00000021013	7	17.98	tubulin beta 4A class IVa
TUL	*CRB3*	ENSBTAG00000051530	7	18.01	crumbs cell polarity complex component 3
TUL	*SLC25A23*	ENSBTAG00000003491	7	18.01	solute carrier family 25 member 23
TUL	*SLC25A41*	ENSBTAG00000020017	7	18.03	solute carrier family 25 member 41
TUL	*KHSRP*	ENSBTAG00000021018	7	18.04	KH-type splicing regulatory protein
TUL	*GTF2F1*	ENSBTAG00000021016	7	18.06	ral transcription factor IIF subunit 1
TUL	*PSPN*	ENSBTAG00000031794	7	18.07	persephin
TUL	*CLPP*	ENSBTAG00000014712	7	18.08	caseinolytic mitochondrial matrix peptidase proteolytic subunit
TUL	*DPP9*	ENSBTAG00000021134	7	19.43	dipeptidyl peptidase 9
TUL	*MYDGF*	ENSBTAG00000018655	7	19.47	myeloid derived growth factor
TUL	*TNFAIP8L1*	ENSBTAG00000037765	7	19.48	TNF alpha induced protein 8 like 1
TUL	*SEMA6B*	ENSBTAG00000031658	7	19.57	semaphorin 6B
NGI	*PCDHGC3*	ENSBTAG00000017349	7	52.47	protocadherin gamma subfamily B, 4
NGI	*PCDHGC4*	ENSBTAG00000054197	7	52.57	protocadherin gamma subfamily C, 4
NGI	*HDAC3*	ENSBTAG00000017360	7	52.72	histone deacetylase 3
NGI	*RELL2*	ENSBTAG00000017371	7	52.73	RELT like 2
NGI	*FCHSD1*	ENSBTAG00000049248	7	52.73	FCH and double SH3 domains 1

NGI	*ARAP3*	ENSBTAG00000015938	7	52.75	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3
NGI	*PCDH1*	ENSBTAG00000012518	7	52.93	protocadherin 1
NGI	*DELE1*	ENSBTAG00000008435	7	52.99	DAP3 binding cell death enhancer 1
NGI	*PCDH12*	ENSBTAG00000008437	7	53.01	protocadherin 12
NGI	*RNF14*	ENSBTAG00000005249	7	53.04	ring finger protein 14
NGI	*GNPDA1*	ENSBTAG00000007865	7	53.07	glucosamine-6-phosphate deaminase 1
NGI	*NDFIP1*	ENSBTAG00000047747	7	53.07	Nedd4 family interacting protein 1
DRB	*OR4H12*	ENSBTAG00000037411	10	27.17	olfactory receptor family 4 subfamily H member 12
TUL	*OR4N2C*	ENSBTAG00000055225	10	27.20	olfactory receptor family 4 subfamily N member 2C
TUL	*OR4N4*	ENSBTAG00000054809	10	27.21	olfactory receptor family 4 subfamily N member 4
TUL	*OR4L1*	ENSBTAG00000048863	10	27.46	olfactory receptor family 4 subfamily L member 1
TUL	*OR4N5*	ENSBTAG00000048339	10	27.49	olfactory receptor family 4 subfamily N member 5
DRB	*LPCAT4*	ENSBTAG00000020040	10	28.28	lysophosphatidylcholine acyltransferase 4
DRB	*NUTM1*	ENSBTAG00000014948	10	28.29	NUT midline carcinoma family member 1
DRB	*NOP10*	ENSBTAG00000025632	10	28.30	NOP10 ribonucleoprotein
DRB	*SLC12A6*	ENSBTAG00000016236	10	28.30	solute carrier family 12 member 6
DRB	*EMC4*	ENSBTAG00000006416	10	28.39	ER membrane protein complex subunit 4
TUL	*CTNNA2*	ENSBTAG00000031669	11	54.80	catenin alpha 2
TUL	*REG3G*	ENSBTAG00000007135	11	56.61	rerating islet-derived 3 gamma
TUL	*DIP2C*	ENSBTAG00000006531	13	46.53	disco interacting C
TUL	*ZMYND11*	ENSBTAG00000002578	13	46.84	zinc finger MYND-type containing 11
TUL	*PRNP*	ENSBTAG00000048903	13	47.04	prion protein [Source:NCBI gene
TUL	*PRND*	ENSBTAG00000011010	13	47.09	prion like protein doppel
TUL	*RASSF2*	ENSBTAG00000017346	13	47.13	Ras association domain family member 2
TUL	*TMEM230*	ENSBTAG00000006063	13	47.42	transmembrane protein 230
TUL	*PCNA*	ENSBTAG00000006065	13	47.43	proliferating cell nuclear antigen
TUL	*CDS2*	ENSBTAG00000006066	13	47.44	CDP-diacylglycerol synthase 2
TUL	*PROKR2*	ENSBTAG00000015872	13	47.55	prokineticin receptor 2
TUL	*GPCPD1*	ENSBTAG00000008293	13	47.72	glycerophosphocholine phosphodiesterase 1
TUL	*SHLD1*	ENSBTAG00000014048	13	47.92	shieldin complex subunit 1
TUL	*CHGB*	ENSBTAG00000011782	13	48.07	chromogranin B
TUL	*TRMT6*	ENSBTAG00000001314	13	48.11	tRNA methyltransferase 6
TUL	*MCM8*	ENSBTAG00000014623	13	48.12	minichromosome maintenance 8 homologous recombination repair factor
TUL	*CRLS1*	ENSBTAG00000000065	13	48.19	cardiolipin synthase 1
TUL	*LRRN4*	ENSBTAG00000000066	13	48.21	leucine rich repeat neuronal 4
TUL	*FERMT1*	ENSBTAG00000020465	13	48.25	fermitin family member 1
TUL	*MYC*	ENSBTAG00000008409	14	12.65	MYC proto-onco, bHLH transcription factor
TUL	*ANXA13*	ENSBTAG00000048822	14	16.14	annexin A13
TUL	*KLHL38*	ENSBTAG00000016754	14	16.23	kelch like family member 38
TUL	*FBXO32*	ENSBTAG00000016194	14	16.33	F-box protein 3
TUL	*NTAQ1*	ENSBTAG00000009916	14	16.42	N-terminal glutamine amidase 1
TUL	*ATAD2*	ENSBTAG00000051782	14	16.45	ATPase family AAA domain containing 2
TUL	*ZHX1*	ENSBTAG00000015334	14	16.53	zinc fingers and homeoboxes 1

TUL	*C14H8orf76*	ENSBTAG00000015319	14	16.57	chromosome 14 C8orf76 homolog
TUL	*FAM83A*	ENSBTAG00000005877	14	16.59	family with sequence similarity 83 member A
TUL	*TBC1D31*	ENSBTAG00000021156	14	16.63	TBC1 domain family member 31
TUL	*HAS2*	ENSBTAG00000019892	14	18.09	hyaluronan synthase 2
TUL	*CEBPD*	ENSBTAG00000046307	14	19.13	CCAAT enhancer binding protein delta
TUL	*SPIDR*	ENSBTAG00000044106	14	19.13	scaffold protein involved in DNA repair
TUL	*H3-5*	ENSBTAG00000017016	14	19.42	H3 histone, family 3C
TUL	*PRKDC*	ENSBTAG00000017019	14	19.42	protein kinase, DNA-activated, catalytic subunit
TUL	*MCM4*	ENSBTAG00000017021	14	19.55	minichromosome maintenance complex component 4
TUL	*UBE2V2*	ENSBTAG00000023218	14	19.59	ubiquitin conjugating enzyme E2 V2
TUL	*SNAI2*	ENSBTAG00000013227	14	19.94	snail family transcriptional repressor 2
TUL	*PPDPFL*	ENSBTAG00000038286	14	19.98	pancreatic progenitor cell differentiation and proliferation factor like
TUL	*SNTG1*	ENSBTAG00000002448	14	20.40	syntrophin gamma 1
TUL	*PCMTD1*	ENSBTAG00000017492	14	21.01	protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 1
TUL	*ST18*	ENSBTAG00000005560	14	21.16	ST18 C2H2C-type zinc finger transcription factor
TUL	*RB1CC1*	ENSBTAG00000000878	14	21.49	RB1 inducible coiled-coil 1
TUL	*NPBWR1*	ENSBTAG00000016159	14	21.61	neuropeptides B and W receptor 1
TUL	*OPRK1*	ENSBTAG00000000914	14	21.71	opioid receptor kappa 1
TUL	*RGS20*	ENSBTAG00000003454	14	21.90	regulator of G protein signaling 20
TUL	*TCEA1*	ENSBTAG00000003460	14	21.96	transcription elongation factor A1
TUL	*LYPLA1*	ENSBTAG00000004243	14	21.99	lysophospholipase 1
TUL	*SOX17*	ENSBTAG00000005748	14	22.23	SRY-box transcription factor 17
TUL	*RP1*	ENSBTAG00000011203	14	22.34	RP1 axonemal microtubule associated
TUL	*XKR4*	ENSBTAG00000044050	14	22.64	XK related 4
TUL	*TMEM68*	ENSBTAG00000005893	14	23.03	transmembrane protein 68
TUL	*TGS1*	ENSBTAG00000005898	14	23.07	trimethylguanosine synthase 1
TUL	*LYN*	ENSBTAG00000020034	14	23.13	LYN proto-onco, Src family tyrosine kinase
TUL	*MOS*	ENSBTAG00000019145	14	23.30	MOS proto-onco, serine/threonine kinase
TUL	*PLAG1*	ENSBTAG00000004022	14	23.33	PLAG1 zinc finger
TUL	*CHCHD7*	ENSBTAG00000049910	14	23.38	coiled-coil-helix-coiled-coil-helix domain containing 7
TUL	*SDR16C5*	ENSBTAG00000018570	14	23.43	short chain dehydrogenase/reductase family 16C member 5
TUL	*SDR16C6*	ENSBTAG00000040321	14	23.48	short chain dehydrogenase/reductase family 16C, member 6
TUL	*PENK*	ENSBTAG00000004924	14	23.54	proenkephalin
TUL	*BPNT2*	ENSBTAG00000015637	14	23.87	3'(2'), 5'-bisphosphate nucleotidase 2
DRB,TUL	*FAM110B*	ENSBTAG00000050550	14	24.37	family with sequence similarity 110 member B
DRB,TUL	*UBXN2B*	ENSBTAG00000009138	14	24.59	UBX domain protein 2B
DRB,TUL	*CYP7A1*	ENSBTAG00000005287	14	24.66	cytochrome P450, family 7, subfamily A, polypeptide 1
DRB,TUL	*SDCBP*	ENSBTAG00000019910	14	24.73	syndecan binding protein
DRB,TUL	*NSMAF*	ENSBTAG00000008958	14	24.77	neutral sphingomyelinase activation associated factor
DRB,TUL	*TOX*	ENSBTAG00000004954	14	24.95	thymocyte selection associated high mobility group box
TUL	*CA8*	ENSBTAG00000017529	14	25.96	carbonic anhydrase 8
TUL	*CHD7*	ENSBTAG00000021841	14	26.36	chromodomain helicase DNA binding protein 7

TUL	*CLVS1*	ENSBTAG00000043978	14	26.85	clavesin 1
TUL	*ASPH*	ENSBTAG00000026283	14	27.02	aspartate beta-hydroxylase
TUL	*NKAIN3*	ENSBTAG00000054400	14	27.74	sodium/potassium transporting ATPase interacting 3
TUL	*EBAG9*	ENSBTAG00000037571	14	54.84	estrogen receptor binding site associated antigen 9
TUL	*PKHD1L1*	ENSBTAG00000026247	14	54.88	PKHD1 like 1
TUL	*RUNX1T1*	ENSBTAG00000017339	14	72.30	RUNX1 partner transcriptional co-repressor 1 [Source:
TUL	*OR4C1F*	ENSBTAG00000052049	19	0.12	olfactory receptor family 4 subfamily C member 1F
TUL	*OR4C1E*	ENSBTAG00000054316	19	0.18	olfactory receptor family 4 subfamily C member 1E
TUL	*KIF2B*	ENSBTAG00000017345	19	3.78	kinesin family member 2B
TUL	*ERC2*	ENSBTAG00000010029	22	44.54	ELKS/RAB6-interacting/CAST family member 2
TUL	*WNT5A*	ENSBTAG00000020221	22	45.54	Wnt family member 5A
TUL	*CACNA2D3*	ENSBTAG00000013117	22	45.92	calcium voltage-gated channel auxiliary subunit alpha2delta 3
TUL	*LRTM1*	ENSBTAG00000013124	22	46.05	leucine rich repeats and transmembrane domains 1
TUL	*EPB41L3*	ENSBTAG00000019251	24	38.88	erythrocyte membrane protein band 4.1 like 3
TUL	*TMEM200C*	ENSBTAG00000046396	24	39.31	transmembrane protein 200C
TUL	*ARHGAP28*	ENSBTAG00000019036	24	39.83	Rho GTPase activating protein 28
TUL	*LAMA1*	ENSBTAG00000018160	24	39.95	laminin subunit alpha 1
TUL	*PTPRM*	ENSBTAG00000024015	24	40.33	protein tyrosine phosphatase receptor type M
TUL	*RAB12*	ENSBTAG00000010356	24	41.08	RAB12, member RAS onco family
TUL	*MTCL1*	ENSBTAG00000009459	24	41.14	microtubule crosslinking factor 1
TUL	*NDUFV2*	ENSBTAG00000004871	24	41.42	NADH:ubiquinone oxidoreductase core subunit V2
TUL	*ANKRD12*	ENSBTAG00000002755	24	41.44	ankyrin repeat domain 12
TUL	*TWSG1*	ENSBTAG00000001805	24	41.55	twisted gastrulation BMP signaling modulator 1
TUL	*RALBP1*	ENSBTAG00000021250	24	41.62	ralA binding protein 1
TUL	*TXNDC2*	ENSBTAG00000017276	24	41.87	thioredoxin domain containing 2
TUL	*APCDD1*	ENSBTAG00000005154	24	42.13	APC down-regulated 1
TUL	*NAPG*	ENSBTAG00000005158	24	42.18	NSF attachment protein gamma
TUL	*PIEZO2*	ENSBTAG00000011171	24	42.25	piezo type mechanosensitive ion channel component 2
TUL	*GNAL*	ENSBTAG00000002125	24	42.65	G protein subunit alpha L
TUL	*MPPE1*	ENSBTAG00000002128	24	42.72	metallophosphoesterase 1
TUL	*IMPA2*	ENSBTAG00000043951	24	42.75	inositol monophosphatase 2
TUL	*CIDEA*	ENSBTAG00000046547	24	42.78	cell death inducing DFFA like effector A
TUL	*TUBB6*	ENSBTAG00000046337	24	42.80	tubulin beta 6 class V
TUL	*AFG3L2*	ENSBTAG00000011250	24	42.81	AFG3 like matrix AAA peptidase subunit 2
TUL	*PRELID3A*	ENSBTAG00000017983	24	42.85	PRELI domain containing 3A
TUL	*SPIRE1*	ENSBTAG00000010542	24	42.88	spire type actin nucleation factor 1
TUL	*CEP76*	ENSBTAG00000010547	24	43.06	centrosomal protein 76
TUL	*PSMG2*	ENSBTAG00000010552	24	43.07	proteasome assembly chaperone 2
TUL	*PTPN2*	ENSBTAG00000010563	24	43.10	protein tyrosine phosphatase non-receptor type 2
TUL	*SEH1L*	ENSBTAG00000010792	24	43.20	SEH1 like nucleoporin
TUL	*CEP192*	ENSBTAG00000013380	24	43.22	centrosomal protein 192
TUL	*LDLRAD4*	ENSBTAG00000009139	24	43.31	low density lipoprotein receptor class A domain containing 4
TUL	*FAM210A*	ENSBTAG00000009141	24	43.49	family with sequence similarity 210 member A

TUL	*RNMT*	ENSBTAG00000009142	24	43.50	RNA guanine-7 methyltransferase
TUL	*MC5R*	ENSBTAG00000009143	24	43.54	melanocortin 5 receptor
TUL	*SETBP1*	ENSBTAG00000018088	24	44.28	SET binding protein 1
TUL	*DYM*	ENSBTAG00000000024	24	48.72	dymeclin
TUL	*LIPG*	ENSBTAG00000048916	24	49.21	lipase G, endothelial type
TUL	*ACAA2*	ENSBTAG00000002863	24	49.44	acetyl-CoA acyltransferase 2
TUL	*MYO5B*	ENSBTAG00000019455	24	49.49	myosin VB
TUL	*MBD1*	ENSBTAG00000003801	24	49.86	methyl-CpG binding domain protein 1
TUL	*CXXC1*	ENSBTAG00000021884	24	49.95	CXXC finger protein 1
TUL	*SKA1*	ENSBTAG00000018216	24	49.96	spindle and kinetochore associated complex subunit 1
TUL	*MAPK4*	ENSBTAG00000009907	24	50.11	mitogen-activated protein kinase 4
TUL	*MRO*	ENSBTAG00000047635	24	50.32	maestro
TUL	*ME2*	ENSBTAG00000016269	24	50.40	malic enzyme 2
TUL	*SMAD4*	ENSBTAG00000006919	24	50.52	SMAD family member 4
TUL	*MEX3C*	ENSBTAG00000014886	24	50.65	mex-3 RNA binding family member C

Table 5.4: A comprehensive list of genes identified in the genomic regions considered to be signatures of selection using ROHet in DRB, NGI, and TUL

Symbol	Ensembl Gene ID	Chr	Position (Mbp)	Description
TRAT1	ENSBTAG00000049108	1	53.578753	T cell receptor associated transmembrane adaptor 1
PRDM15	ENSBTAG00000021253	1	142.065605	PR/SET domain 15
C2CD2	ENSBTAG00000021259	1	142.14261	C2 calcium dependent domain containing 2
EFHB	ENSBTAG00000013806	1	157.125001	EF-hand domain family member B
RAB5A	ENSBTAG00000046385	1	157.198501	RAB5A, member RAS onco family
ACTR3	ENSBTAG00000003401	2	65.522494	actin related protein 3
LDLRAP1	ENSBTAG00000001050	2	127.482749	low density lipoprotein receptor adaptor protein 1
LAMTOR5	ENSBTAG00000014970	3	33.0268	late endosomal/lysosomal adaptor, MAPK and MTOR activator 5
ZRANB2	ENSBTAG00000002627	3	74.146776	zinc finger RANBP2-type containing 2
PTGER3	ENSBTAG00000019230	3	74.185666	prostaglandin E receptor 3
COG5	ENSBTAG00000005526	4	48.291963	component of oligomeric golgi complex 5
GPR22	ENSBTAG00000046917	4	48.508186	G protein-coupled receptor 22
DUS4L	ENSBTAG00000021849	4	48.572836	dihydrouridine synthase 4 like
SLC26A4	ENSBTAG00000010062	4	48.677577	solute carrier family 26 member 4
NELL2	ENSBTAG00000032183	5	35.475234	neural EGFL like 2
LEMD3	ENSBTAG00000039435	5	48.54286	LEM domain containing 3
WIF1	ENSBTAG00000014758	5	48.687578	WNT inhibitory factor 1
TBC1D30	ENSBTAG00000011352	5	48.952181	TBC1 domain family member 30
R3HDM2	ENSBTAG00000018361	5	56.081334	R3H domain containing 2
STAC3	ENSBTAG00000018358	5	56.193684	SH3 and cysteine rich domain 3
NDUFA4L2	ENSBTAG00000031503	5	56.205859	NDUFA4 mitochondrial complex associated like 2
SHMT2	ENSBTAG00000031500	5	56.208141	serine hydroxymethyltransferase 2
CRACR2A	ENSBTAG00000018940	5	106.146546	calcium release activated channel regulator 2A
PRMT8	ENSBTAG00000009447	5	106.319032	protein arginine methyltransferase 8
TEAD4	ENSBTAG00000019788	5	106.773101	TEA domain transcription factor 4
FKBP4	ENSBTAG00000007605	5	106.946944	FKBP prolyl isomerase 4
FAM184B	ENSBTAG00000005932	6	37.181075	family with sequence similarity 184 member B
CNOT6	ENSBTAG00000017362	7	0.58731	CCR4-NOT transcription complex subunit 6
LYRM7	ENSBTAG00000010961	7	23.24408	LYR motif containing 7
HINT1	ENSBTAG00000010959	7	23.271324	histidine triad nucleotide binding protein 1
MARCHF3	ENSBTAG00000006797	7	26.947768	membrane associated ring-CH-type finger 3
RNF44	ENSBTAG00000017748	7	37.999787	ring finger protein 44
GPRIN1	ENSBTAG00000044035	7	38.070323	G protein regulated inducer of neurite outgrowth 1
SNCB	ENSBTAG00000009803	7	38.091557	synuclein beta
EIF4E1B	ENSBTAG00000003103	7	38.111499	eukaryotic translation initiation factor 4E family member 1B
TSPAN17	ENSBTAG00000017451	7	38.116057	tetraspanin 17
NRG2	ENSBTAG00000010689	7	51.03192	neuregulin 2
PURA	ENSBTAG00000008881	7	51.298957	purine rich element binding protein A
IGIP	ENSBTAG00000025226	7	51.314286	IgA inducing protein

CYSTM1	ENSBTAG00000016595	7	51.364569	cysteine rich transmembrane module containing 1
PFDN1	ENSBTAG00000051779	7	51.435093	prefoldin subunit 1
HBEGF	ENSBTAG00000021766	7	51.5336	heparin binding EGF like growth factor
SLC4A9	ENSBTAG00000021775	7	51.558951	solute carrier family 4 member 9
EIF4EBP3	ENSBTAG00000052310	7	51.722071	eukaryotic translation initiation factor 4E binding protein 3
SRA1	ENSBTAG00000001449	7	51.724953	steroid receptor RNA activator 1
APBB3	ENSBTAG00000001450	7	51.731683	amyloid beta protein binding family B member 3
SLC35A4	ENSBTAG00000030584	7	51.738224	solute carrier family 35 member A4
CD14	ENSBTAG00000015032	7	51.762895	CD14 molecule
NDUFA2	ENSBTAG00000015041	7	51.778423	NADH:ubiquinone oxidoreductase subunit A2
IK	ENSBTAG00000015034	7	51.780969	IK cytokine
WDR55	ENSBTAG00000015042	7	51.795499	WD repeat domain 55
DND1	ENSBTAG00000018845	7	51.800738	DND microRNA-mediated repression inhibitor 1
HARS1	ENSBTAG00000018847	7	51.803863	histidyl-tRNA synthetase 1
HARS2	ENSBTAG00000025212	7	51.817094	histidyl-tRNA synthetase 2, mitochondrial
ZMAT2	ENSBTAG00000005441	7	51.825588	zinc finger matrin-type 2
PCDHA2	ENSBTAG00000046190	7	51.894458	protocadherin alpha 2
PCDHA3	ENSBTAG00000046717	7	51.902109	protocadherin alpha 3
PCDHA13	ENSBTAG00000030227	7	51.98079	protocadherin alpha 13
PCDHAC2	ENSBTAG00000053546	7	52.009933	protocadherin alpha subfamily C, 2
PCDHB1	ENSBTAG00000037885	7	52.169913	protocadherin beta 1
TMEM181	ENSBTAG00000010838	9	95.033934	transmembrane protein 181
SYTL3	ENSBTAG00000010334	9	95.072282	synaptotagmin like 3
EZR	ENSBTAG00000010347	9	95.15627	ezrin
SHC4	ENSBTAG00000022583	10	61.2841	SHC adaptor protein 4
EID1	ENSBTAG00000027764	10	61.370363	EP300 interacting inhibitor of differentiation 1
CEP152	ENSBTAG00000011661	10	61.43723	centrosomal protein 152
CAMKMT	ENSBTAG00000032519	11	26.771728	calmodulin-lysine N-methyltransferase
WDPCP	ENSBTAG00000005151	11	61.723956	WD repeat containing planar cell polarity effector
MDH1	ENSBTAG00000019295	11	62.022816	malate dehydrogenase 1
LYN	ENSBTAG00000020034	14	23.134995	LYN proto-onco, Src family tyrosine kinase
FAM110B	ENSBTAG00000050550	14	24.365744	family with sequence similarity 110 member B
SDCBP	ENSBTAG00000019910	14	24.728895	syndecan binding protein
NSMAF	ENSBTAG00000008958	14	24.765312	neutral sphingomyelinase activation associated factor
CHD7	ENSBTAG00000021841	14	26.361178	chromodomain helicase DNA binding protein 7
CLVS1	ENSBTAG00000043978	14	26.854867	clavesin 1
HHIPL2	ENSBTAG00000017561	16	26.010406	HHIP like 2
CASZ1	ENSBTAG00000019818	16	42.83857	castor zinc finger 1
LHX5	ENSBTAG00000003107	17	60.981921	LIM homeobox 5
PLBD2	ENSBTAG00000019040	17	61.041318	phospholipase B domain containing 2
DTX1	ENSBTAG00000016738	17	61.08571	deltex E3 ubiquitin ligase 1
RASAL1	ENSBTAG00000016739	17	61.125494	RAS protein activator like 1
CFAP73	ENSBTAG00000011924	17	61.167481	cilia and flagella associated protein 73
ZNRF1	ENSBTAG00000034689	18	2.363234	zinc and ring finger 1

LDHD	ENSBTAG00000006779	18	2.453442	lactate dehydrogenase D
EXOC3L1	ENSBTAG00000012035	18	34.820643	exocyst complex component 3 like 1
E2F4	ENSBTAG00000012063	18	34.827984	E2F transcription factor 4
ELMO3	ENSBTAG00000001286	18	34.834852	engulfment and cell motility 3
TMEM208	ENSBTAG00000002327	18	34.855548	transmembrane protein 208
FHOD1	ENSBTAG00000032427	18	34.857829	formin homology 2 domain containing 1
SLC9A5	ENSBTAG00000039190	18	34.875817	solute carrier family 9 member A5
PLEKHG4	ENSBTAG00000021570	18	34.900315	pleckstrin homology and RhoGEF domain containing G4
KCTD19	ENSBTAG00000032396	18	34.909015	potassium channel tetramerization domain containing 19
LRRC36	ENSBTAG00000001838	18	34.936769	leucine rich repeat containing 36
TPPP3	ENSBTAG00000019822	18	34.997818	tubulin polymerization promoting protein family member 3
ZDHHC1	ENSBTAG00000038134	18	35.006309	zinc finger DHHC-type containing 1
HSD11B2	ENSBTAG00000005685	18	35.040304	hydroxysteroid 11-beta dehydrogenase 2
ATP6V0D1	ENSBTAG00000014553	18	35.046554	ATPase H+ transporting V0 subunit d1
AGRP	ENSBTAG00000014556	18	35.086439	agouti related neuropeptide
ZNF329	ENSBTAG00000003465	18	65.485031	zinc finger protein 329
ZNF135	ENSBTAG00000003462	18	65.499669	zinc finger protein 135
UBTD2	ENSBTAG00000034659	20	3.997984	ubiquitin domain containing 2
CPEB4	ENSBTAG00000009995	20	5.606065	cytoplasmic polyadenylation element binding protein 4
NSG2	ENSBTAG00000000564	20	5.750258	neuronal vesicle trafficking associated 2
NNT	ENSBTAG00000011885	20	31.154175	nicotinamide nucleotide transhydrogenase
ADAMTS12	ENSBTAG00000012558	20	39.873127	ADAM metallopeptidase with thrombospondin type 1 motif 12
TARS1	ENSBTAG00000014261	20	40.309676	threonyl-tRNA synthetase 1
SLC28A1	ENSBTAG00000020250	21	22.524664	solute carrier family 28 member 1
ITPK1	ENSBTAG00000009845	21	57.589349	inositol-tetrakisphosphate 1-kinase
KHDRBS2	ENSBTAG00000043990	23	0.270838	KH RNA binding domain containing, signal transduction associated 2
CPEB3	ENSBTAG00000015450	26	13.52854	cytoplasmic polyadenylation element binding protein 3
IDE	ENSBTAG00000019759	26	13.881642	insulin degrading enzyme
CRTAC1	ENSBTAG00000008102	26	19.019143	cartilage acidic protein 1

Table 5.5: A comprehensive list of genes identified in the genomic regions considered to be signatures of selection using F_{ST} in DRB, NGI, and TU*L*

Symbol	Ensembl Gene ID	Chr	Position (Mbp)	Description
RCAN1	ENSBTAG00000020035	1	0.882081	regulator of calcineurin 1
KCNE1	ENSBTAG00000001150	1	1.040523	potassium voltage-gated channel subfamily E regulatory subunit 1
KCNE2	ENSBTAG00000026260	1	1.123943	potassium voltage-gated channel subfamily E regulatory subunit 2
TIAM1	ENSBTAG00000017839	1	4.192104	TIAM Rac1 associated GEF 1
APP	ENSBTAG00000017753	1	10.231296	amyloid beta protein
GABPA	ENSBTAG00000019043	1	10.656634	GA binding protein transcription factor subunit alpha
ATP5PF	ENSBTAG00000000605	1	10.690694	ATP synthase peripheral stalk subunit F6
JAM2	ENSBTAG00000000603	1	10.702387	junctional adhesion molecule 2
USP25	ENSBTAG00000019314	1	21.167188	ubiquitin specific peptidase 25

NRIP1	ENSBTAG00000047293	1	22.189815	nuclear receptor interacting protein 1
GBE1	ENSBTAG00000036262	1	29.107326	1,4-alpha-glucan branching enzyme 1
MTX2	ENSBTAG00000006693	2	20.568451	metaxin 2
HOXD1	ENSBTAG00000015840	2	20.724272	homeobox D1
HOXD3	ENSBTAG00000004835	2	20.740712	homeobox D3
HOXD4	ENSBTAG00000039581	2	20.759992	homeobox D4
HOXD8	ENSBTAG00000049845	2	20.781821	homeobox D8
HOXD9	ENSBTAG00000016033	2	20.789261	homeobox D9
HOXD10	ENSBTAG00000016030	2	20.793704	homeobox D10
HOXD11	ENSBTAG00000033330	2	20.803593	homeobox D11
HOXD12	ENSBTAG00000004314	2	20.812814	homeobox D12
LNPK	ENSBTAG00000018061	2	20.906685	lunapark, ER junction formation factor
PDK1	ENSBTAG00000016836	2	24.02665	pyruvate dehydrogenase kinase 1
ITGA6	ENSBTAG00000017266	2	24.091577	integrin subunit alpha 6
DLX2	ENSBTAG00000005741	2	24.448619	distal-less homeobox 2
ACKR3	ENSBTAG00000018424	3	116.025262	atypical chemokine receptor 3
ZDHHC17	ENSBTAG00000021756	5	6.174722	zinc finger DHHC-type palmitoyltransferase 17
CSRP2	ENSBTAG00000013406	5	6.281008	cysteine and glycine rich protein 2
E2F7	ENSBTAG00000016931	5	6.443192	E2F transcription factor 7
LYZ2	ENSBTAG00000026088	5	44.365581	lysozyme C-2
LYZ1	ENSBTAG00000046511	5	44.392339	lysozyme (renal amyloidosis)
LYZ3	ENSBTAG00000046628	5	44.423778	lysozyme 3
LYZ	ENSBTAG00000026779	5	44.506989	lysozyme
MTERF2	ENSBTAG00000010144	5	70.252258	mitochondrial transcription termination factor 2
SRD5A3	ENSBTAG00000014913	6	70.803119	steroid 5 alpha-reductase 3
TMEM165	ENSBTAG00000001269	6	70.838726	transmembrane protein 165
CLOCK	ENSBTAG00000044044	6	70.874084	clock circadian regulator
NMU	ENSBTAG00000044161	6	71.03301	neuromedin U
EXOC1L	ENSBTAG00000045602	6	71.23075	exocyst complex component 1 like
CRACD	ENSBTAG00000040398	6	71.604475	capping protein inhibiting regulator of actin dynamics
AASDH	ENSBTAG00000020583	6	71.742769	aminoadipate-semialdehyde dehydrogenase
PPAT	ENSBTAG00000010571	6	71.782614	phosphoribosyl pyrophosphate amidotransferase
PAICS	ENSBTAG00000010577	6	71.796672	phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase
SRP72	ENSBTAG00000010593	6	71.846136	signal recognition particle 72
HSTN	ENSBTAG00000048250	6	85.459411	histatherin
CSN3	ENSBTAG00000039787	6	85.645854	casein kappa
CABS1	ENSBTAG00000019849	6	85.734578	calcium binding protein, spermatid associated 1
PRR16	ENSBTAG00000050473	7	33.203692	proline rich 16
MXD3	ENSBTAG00000005010	7	38.779127	MAX dimerization protein 3
LMAN2	ENSBTAG00000008034	7	38.829326	lectin, mannose binding 2
RGS14	ENSBTAG00000008497	7	38.857589	regulator of G protein signaling 14
F12	ENSBTAG00000018872	7	38.898276	coagulation factor XII
DBN1	ENSBTAG00000011400	7	38.954347	drebrin 1
DDX41	ENSBTAG00000015187	7	39.010919	DEAD-box helicase 41

TMED9	ENSBTAG00000001394	7	39.117501	transmembrane emp24 protein transport domain containing 9
B4GALT7	ENSBTAG00000001767	7	39.124716	beta-1,4-galactosyltransferase 7
N4BP3	ENSBTAG00000006368	7	39.216407	NEDD4 binding protein 3
RMND5B	ENSBTAG00000006371	7	39.225401	required for meiotic nuclear division 5 homolog B
NHP2	ENSBTAG00000006374	7	39.244051	NHP2 ribonucleoprotein
ZNF354A	ENSBTAG00000006428	7	39.797175	zinc finger protein 354A
PROP1	ENSBTAG00000007359	7	39.848312	PROP paired-like homeobox 1
OR2Y1	ENSBTAG00000038629	7	39.950418	olfactory receptor family 2 subfamily Y member 1
MGAT1	ENSBTAG00000001546	7	39.979972	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase
OR2AJ9	ENSBTAG00000040033	7	41.681668	olfactory receptor family 2 subfamily AJ member 9
OR2L2B	ENSBTAG00000051864	7	41.773074	olfactory receptor family 2 subfamily L member 2B
OR2L2C	ENSBTAG00000049479	7	41.791167	olfactory receptor family 2 subfamily L member 2C
OR2L3C	ENSBTAG00000052782	7	41.810762	olfactory receptor family 2 subfamily L member 3C
OR2L13	ENSBTAG00000013691	7	41.827752	olfactory receptor family 2 subfamily L member 13
OR2T22	ENSBTAG00000048025	7	41.859337	olfactory receptor family 2 subfamily T member 22
OR2M4	ENSBTAG00000030722	7	41.950264	olfactory receptor family 2 subfamily M member 4
OR2M10	ENSBTAG00000019925	7	41.963265	olfactory receptor family 2 subfamily M member 10
OR2T6	ENSBTAG00000045644	7	42.05442	olfactory receptor family 2 subfamily T member 6
OR2T1	ENSBTAG00000024653	7	42.074026	olfactory receptor family 2 subfamily T member 1
OR2T27	ENSBTAG00000030719	7	42.104424	olfactory receptor family 2 subfamily T member 27
OR2T2	ENSBTAG00000034743	7	42.117972	olfactory receptor family 2 subfamily T member 2
SPOCK1	ENSBTAG00000003502	7	48.367043	SPARC (osteonectin), cwcv and kazal like domains proteoglycan 1
ASTN2	ENSBTAG00000025667	8	105.642638	astrotactin 2
TRIM32	ENSBTAG00000017155	8	105.87612	tripartite motif containing 32
NEO1	ENSBTAG00000004990	10	19.888399	neogenin 1
HCN4	ENSBTAG00000017466	10	20.147117	hyperpolarization activated cyclic nucleotide gated potassium channel 4
NPTN	ENSBTAG00000008218	10	20.37335	neuroplastin
CD276	ENSBTAG00000019734	10	20.490779	CD276 molecule
INSYN1	ENSBTAG00000025803	10	20.546893	inhibitory synaptic factor 1
CA12	ENSBTAG00000019390	10	46.649404	carbonic anhydrase 12
APH1B	ENSBTAG00000010641	10	46.654444	aph-1 homolog B, gamma-secretase subunit
RAB8B	ENSBTAG00000011146	10	46.7805	RAB8B, member RAS onco family
GRHL1	ENSBTAG00000007485	11	87.587451	grainyhead like transcription factor 1
BLCAP	ENSBTAG00000003209	13	66.461377	BLCAP apoptosis inducing factor
NNAT	ENSBTAG00000003212	13	66.465154	neuronatin
CTNNBL1	ENSBTAG00000007658	13	66.604147	catenin beta like 1
VSTM2L	ENSBTAG00000004750	13	66.79939	V-set and transmembrane domain containing 2 like
TTI1	ENSBTAG00000004100	13	66.868177	TELO2 interacting protein 1
CYRIB	ENSBTAG00000020801	14	10.69473	CYFIP related Rac1 interactor B
GSDMC	ENSBTAG00000017478	14	10.798273	gasdermin C
MYC	ENSBTAG00000008409	14	12.653173	MYC proto-onco, bHLH transcription factor
TRIB1	ENSBTAG00000023179	14	14.77905	tribbles pseudokinase 1
NSMCE2	ENSBTAG00000009394	14	14.849669	NSE2 (MMS21) homolog, SMC5-SMC6 complex SUMO ligase

SQLE	ENSBTAG00000005498	14	15.144664	squalene epoxidase
ZNF572	ENSBTAG00000009072	14	15.178243	zinc finger protein 572
MTSS1	ENSBTAG00000020407	14	15.383762	MTSS I-BAR domain containing 1
NDUFB9	ENSBTAG00000020405	14	15.553533	NADH:ubiquinone oxidoreductase subunit B9
RNF139	ENSBTAG00000020399	14	15.593322	ring finger protein 139
TMEM65	ENSBTAG00000022114	14	15.675525	transmembrane protein 65
FER1L6	ENSBTAG00000013537	14	15.865228	fer-1 like family member 6
FAM91A1	ENSBTAG00000005025	14	16.074868	family with sequence similarity 91 member A1
FBXO32	ENSBTAG00000016194	14	16.328949	F-box protein 32
ATAD2	ENSBTAG00000051782	14	16.454611	ATPase family AAA domain containing 2
ZHX1	ENSBTAG00000015334	14	16.534435	zinc fingers and homeoboxes 1
FAM83A	ENSBTAG00000005877	14	16.588055	family with sequence similarity 83 member A
DERL1	ENSBTAG00000020693	14	16.714564	derlin 1
ZHX2	ENSBTAG00000026309	14	16.772152	zinc fingers and homeoboxes 2
HAS2	ENSBTAG00000019892	14	18.093101	hyaluronan synthase 2
CEBPD	ENSBTAG00000046307	14	19.128179	CCAAT enhancer binding protein delta
SPIDR	ENSBTAG00000044106	14	19.129765	scaffold protein involved in DNA repair
PRKDC	ENSBTAG00000017019	14	19.424288	protein kinase, DNA-activated, catalytic subunit
MCM4	ENSBTAG00000017021	14	19.551601	minichromosome maintenance complex component 4
PPDPFL	ENSBTAG00000038286	14	19.981881	pancreatic progenitor cell differentiation and proliferation factor like
ST18	ENSBTAG00000005560	14	21.160085	ST18 C2H2C-type zinc finger transcription factor
RB1CC1	ENSBTAG00000000878	14	21.493003	RB1 inducible coiled-coil 1
ATP6V1H	ENSBTAG00000003450	14	21.85375	ATPase H+ transporting V1 subunit H
RGS20	ENSBTAG00000003454	14	21.903677	regulator of G protein signaling 20
TCEA1	ENSBTAG00000003460	14	21.961033	transcription elongation factor A1
LYPLA1	ENSBTAG00000004243	14	21.994932	lysophospholipase 1
RP1	ENSBTAG00000011203	14	22.335447	RP1 axonemal microtubule associated
XKR4	ENSBTAG00000044050	14	22.640221	XK related 4
LYN	ENSBTAG00000020034	14	23.134995	LYN proto-onco, Src family tyrosine kinase
MOS	ENSBTAG00000019145	14	23.299177	MOS proto-onco, serine/threonine kinase
UBXN2B	ENSBTAG00000009138	14	24.587138	UBX domain protein 2B
CYP7A1	ENSBTAG00000005287	14	24.664833	cytochrome P450, family 7, subfamily A, polypeptide 1
SDCBP	ENSBTAG00000019910	14	24.728895	syndecan binding protein
TOX	ENSBTAG00000004954	14	24.946881	thymocyte selection associated high mobility group box
RAB2A	ENSBTAG00000000948	14	26.181071	RAB2A, member RAS onco family
CHD7	ENSBTAG00000021841	14	26.361178	chromodomain helicase DNA binding protein 7
CLVS1	ENSBTAG00000043978	14	26.854867	clavesin 1
ASPH	ENSBTAG00000026283	14	27.01606	aspartate beta-hydroxylase
NKAIN3	ENSBTAG00000054400	14	27.737786	sodium/potassium transporting ATPase interacting 3
TTPA	ENSBTAG00000022471	14	28.074682	alpha tocopherol transfer protein
YTHDF3	ENSBTAG00000018557	14	28.141155	YTH N6-methyladenosine RNA binding protein 3
CYP7B1	ENSBTAG00000001299	14	29.199622	cytochrome P450 family 7 subfamily B member 1
PREX2	ENSBTAG00000022169	14	31.976746	phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2
SULF1	ENSBTAG00000004720	14	33.382025	sulfatase 1

SLCO5A1	ENSBTAG00000032881	14	33.489213	solute carrier organic anion transporter family member 5A1
PRDM14	ENSBTAG00000000060	14	33.786654	PR/SET domain 14
NCOA2	ENSBTAG00000020312	14	33.837784	nuclear receptor coactivator 2
XKR9	ENSBTAG00000048138	14	34.364687	XK related 9
EYA1	ENSBTAG00000011298	14	34.786252	EYA transcriptional coactivator and phosphatase 1
MSC	ENSBTAG00000005404	14	35.439046	musculin
TRPA1	ENSBTAG00000002062	14	35.611321	transient receptor potential cation channel subfamily A member 1
KCNB2	ENSBTAG00000040496	14	35.905091	potassium voltage-gated channel subfamily B member 2
TERF1	ENSBTAG00000032982	14	36.382236	telomeric repeat binding factor 1
RDH10	ENSBTAG00000020143	14	36.62845	retinol dehydrogenase 10
TMEM70	ENSBTAG00000012920	14	37.225602	transmembrane protein 70
LY96	ENSBTAG00000008864	14	37.240839	lymphocyte antigen 96
JPH1	ENSBTAG00000008842	14	37.502881	junctophilin 1
GDAP1	ENSBTAG00000012608	14	37.626644	ganglioside induced differentiation associated protein 1
CRISPLD1	ENSBTAG00000004411	14	38.295064	cysteine rich secretory protein LCCL domain containing 1
HNF4G	ENSBTAG00000008280	14	38.801752	hepatocyte nuclear factor 4 gamma
EIF3H	ENSBTAG00000014032	14	47.572111	eukaryotic translation initiation factor 3 subunit H
TRPS1	ENSBTAG00000017694	14	48.629593	transcriptional repressor GATA binding 1
UBR5	ENSBTAG00000021209	14	61.987835	ubiquitin protein ligase E3 component n-recognin 5
RRM2B	ENSBTAG00000021208	14	62.133632	ribonucleotide reductase regulatory TP53 inducible subunit M2B
NCALD	ENSBTAG00000026963	14	62.384855	neurocalcin delta
GRHL2	ENSBTAG00000004518	14	62.721044	grainyhead like transcription factor 2
YWHAZ	ENSBTAG00000000236	14	63.393855	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta
PABPC1	ENSBTAG00000046358	14	63.625556	poly(A) binding protein cytoplasmic 1
SNX31	ENSBTAG00000011271	14	63.70707	sorting nexin 31
RNF19A	ENSBTAG00000017833	14	64.047857	ring finger protein 19A, RBR E3 ubiquitin protein ligase
SPAG1	ENSBTAG00000032544	14	64.174147	sperm associated antigen 1
FBXO43	ENSBTAG00000019795	14	64.227507	F-box protein 43
RGS22	ENSBTAG00000019793	14	64.289853	regulator of G protein signaling 22
OSR2	ENSBTAG00000013213	14	65.332265	odd-skipped related transciption factor 2
KCNS2	ENSBTAG00000007768	14	65.828236	potassium voltage-gated channel modifier subfamily S member 2
NIPAL2	ENSBTAG00000018206	14	65.96051	NIPA like domain containing 2
POP1	ENSBTAG00000012596	14	66.086238	POP1 homolog, ribonuclease P/MRP subunit
MTDH	ENSBTAG00000003098	14	66.507284	metadherin
TSPYL5	ENSBTAG00000011572	14	66.852944	TSPY like 5
CPQ	ENSBTAG00000011908	14	66.989157	carboxypeptidase Q
PTDSS1	ENSBTAG00000013901	14	67.893306	phosphatidylserine synthase 1
GDF6	ENSBTAG00000007687	14	68.058424	growth differentiation factor 6
NDUFAF6	ENSBTAG00000007570	14	69.265371	NADH:ubiquinone oxidoreductase complex assembly factor 6
TP53INP1	ENSBTAG00000055158	14	69.40234	tumor protein p53 inducible nuclear protein 1
CCNE2	ENSBTAG00000004906	14	69.448183	cyclin E2
DPY19L4	ENSBTAG00000013621	14	69.526133	dpy-19 like 4
VIRMA	ENSBTAG00000003789	14	69.781691	vir like m6A methyltransferase associated
RAD54B	ENSBTAG00000000632	14	69.85833	RAD54 homolog B

GEM	ENSBTAG00000007596	14	70.066731	GTP binding protein overexpressed in skeletal muscle
CDH17	ENSBTAG00000021964	14	70.104355	cadherin 17
DEPTOR	ENSBTAG00000015341	14	81.286703	DEP domain containing MTOR interacting protein
MTBP	ENSBTAG00000006770	14	81.787034	MDM2 binding protein
PIEZO1	ENSBTAG00000020944	18	13.938926	piezo type mechanosensitive ion channel component 1
CDT1	ENSBTAG00000000638	18	14.008412	chromatin licensing and DNA replication factor 1
APRT	ENSBTAG00000000639	18	14.013051	adenine phosphoribosyltransferase
TRAPPC2L	ENSBTAG00000019184	18	14.032588	trafficking protein particle complex 2 like
CBFA2T3	ENSBTAG00000010927	18	14.051588	CBFA2/RUNX1 partner transcriptional co-repressor 3
ACSF3	ENSBTAG00000015968	18	14.203972	acyl-CoA synthetase family member 3
CDH15	ENSBTAG00000011530	18	14.257853	cadherin 15
SLC22A31	ENSBTAG00000011533	18	14.277507	solute carrier family 22 member 31
ANKRD11	ENSBTAG00000016006	18	14.327315	ankyrin repeat domain 11
SPG7	ENSBTAG00000012041	18	14.463306	SPG7 matrix AAA peptidase subunit, paraplegin
RPL13	ENSBTAG00000012044	18	14.490263	ribosomal protein L13
CPNE7	ENSBTAG00000012215	18	14.500473	copine 7
DPEP1	ENSBTAG00000033326	18	14.525082	dipeptidase 1
CHMP1A	ENSBTAG00000033331	18	14.550491	charged multivesicular body protein 1A
CDK10	ENSBTAG00000033333	18	14.568489	cyclin dependent kinase 10
ZNF276	ENSBTAG00000001904	18	14.593024	zinc finger protein 276
FANCA	ENSBTAG00000001906	18	14.606881	FA complementation group A
SPIRE2	ENSBTAG00000040356	18	14.652352	spire type actin nucleation factor 2
TCF25	ENSBTAG00000012102	18	14.681557	transcription factor 25
MC1R	ENSBTAG00000023731	18	14.705093	melanocortin 1 receptor
TUBB3	ENSBTAG00000023730	18	14.70894	tubulin beta 3 class III
DEF8	ENSBTAG00000039014	18	14.722687	differentially expressed in FDCP 8 homolog
GAS8	ENSBTAG00000007096	18	14.772666	growth arrest specific 8
ORC6	ENSBTAG00000001872	18	15.013053	origin recognition complex subunit 6
MYLK3	ENSBTAG00000014818	18	15.033527	myosin light chain kinase 3
TTYH2	ENSBTAG00000011007	19	57.151008	tweety family member 2
RNF180	ENSBTAG00000043992	20	15.106459	ring finger protein 180
HTR1A	ENSBTAG00000040439	20	15.587584	5-hydroxytryptamine receptor 1A
CDH7	ENSBTAG00000014488	24	10.779992	cadherin 7