SNP-based genetic diversity and population structure of South African beef cattle breeds

By

Taylor Tonkin

Dissertation submitted in fulfilment of the requirements for the degree

With specialization in Animal Breeding and Genetics

In the Faculty of Natural and Agricultural Sciences
University of Pretoria

Pretoria

November 2021

SNP-based genetic diversity and population structure of South African beef cattle breeds

By

Taylor Tonkin

Supervisor: Prof Carina Visser

Co-supervisor: Prof Este van Marle-Köster

Department: Animal Science

Degree: MSc (Agric) Animal Breeding and Genetics

Declaration

I, Taylor May Tonkin, 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…08/11/2021……..

Acknowledgements

Firstly, I would like to thank the Lord for providing me with the ability, strength and perseverance to finish this degree and for Your love and grace. Thank You for providing me with a passion for animal science and making my dreams possible. Nothing is impossible with God (Luke 1:37).

I would like to thank my supervisor, Prof Carina Visser and co-supervisor, Prof Este van Marle-Köster for their pearls of wisdom and patience. Thank you for imparting your knowledge and for your guidance and support. Thank you for always being willing to answer any question and for helping me become a better scientist. I hold you both in high esteem and I am forever grateful to you both.

To my incredible parents, thank you for all of the emotional and financial support – not just for my MSc but for my BSc (Agric) as well. Thank you both for being my shoulder to cry on and people to celebrate all the success with. Thank you for pushing me to keep going when times were tough. Thank you for being my pillars of strength and encouragement when you were in South Africa and when you made your move to the UAE. I could not have done any of this without you!

Thank you to my twin sister for your continuous support and encouragement. Thank you for always being willing to read my drafts and assist in fixing my language.

To my husband, thank you for being my rock and prayer warrior. Thank you for always believing in me and encouraging me to do the same. Thank you for your boundless love, support and encouragement for everything I set out to achieve. Thank you for standing by my side in everything.

Thank you to the SA Studbook Association and the breed societies for the use of the genotypic data and for their continuous work to better animal breeding and genetics in South Africa. Thank you to the Beef Genomics Project for the funding of this MSc and for paving a way for genomic selection in South African beef cattle.

Lastly, I would like to thank Anel Retief and Dr Simon Lashmar for always being willing to lend a helping hand and for their assistance in understanding some programs used in my MSc.

Abstract

Evaluation of genetic diversity in cattle populations is important to understand breed structure and for maintaining diversity. The aim of this study was to assess SNP based genomic diversity and population structure of eight South African beef cattle breeds. In this study 2110 beef cattle genotypes, generated within the BGP, representing eight South African beef cattle breeds (Bonsmara, Beefmaster, Boran, Charolais, Hereford, Drakensberger, Nguni and Tuli) were included for analyses. Genotypic data were generated using the GGP 150k SNP array for all registered, genotyped animals participating in Logix Beef Recording. PLINK was used to estimate genetic diversity parameters within populations and biological types. GCTA and ADMIXTURE were used for population structure analysis. SNeP was used to estimate effective population size for all populations. Results indicated limited loss of heterozygosity for the Beefmaster, Boran, Drakensberg and Hereford breeds (H_e>H_o) and no loss of heterozygosity for Bonsmara, Charolais and Nguni breeds. Results further indicated no loss in genetic diversity for all eight populations. The eight populations were grouped into specific biological types namely indigenous (Drakensberger, Nguni and Tuli), composite (Beefmaster and Bonsmara) and exotic types (Boran, Charolais and Hereford) and population diversity parameters were estimated. Analysis of the percentage of low-MAF SNP (MAF<0.05) was done and revealed the indigenous breed group had the highest percentage of low-MAF SNP across the genome. Inbreeding estimates based on F_{IS} and F_{ROH} indicated limited inbreeding across the populations (F_{IS}:-0.009 to 0.016 and F_{ROH} : 0.001-0.005). ROH analysis revealed that any inbreeding was due to ancient inbreeding, based upon the high number of ROH segments that had a length of between 0 and 3.9 MB. Effective population size (N_e) analysis showed a decline in the N_e for all eight breeds. A principal component analysis (PCA) and admixture plot identified eight distinctive breeds with some admixture present. The indigenous populations clustered together due to common ancestry and time divergence. The study concluded that they SA beef populations in this study exhibit moderate to high levels of genetic diversity with low inbreeding.

Table of contents

List of Figures

Chapter 1: Introduction and Literature Review

Chapter 3: Results

List of Tables

Chapter 1: Introduction and Literature Review

Chapter 2: Methods and Materials

Chapter 3: Results

List of Abbreviations

AFR	Afrikaner
AI	Artificial insemination
ANG	Angus
ARC	Agricultural Research Council
BGP	Beef genomics project
BMA	Beefmaster
BON	Bonsmara
BOR	Boran
BLUP	Best linear unbiased prediction
|D'|	Measure of LD
bp	Base pairs
BV	Breeding value
CHL	Charolais
CV	Cross Validation
DAFF	Department of Agriculture Fisheries and Forestry
DGV	Direct genomic value
DNA	Deoxyribonucleic Acid
EBV	Estimated breeding value
FAO	Food and Agriculture Organization
F_{IS}	Individual Inbreeding Coefficient
F_{PED}	Pedigree-based Inbreeding Coefficient
F_{ROH}	Inbreeding Coefficient based on ROH
GCTA	Genome Wide Complex Trait Analysis
GGP	Geneseek Genomic Profiler
GRM	Genomic Relationship Matric
GS	Genomic selection
GWAS	Genome-wide association studies
H_E	Expected heterozygosity
HFD	Hereford

H_O	Observed heterozygosity
HOL	Holstein
HWE	Hardy-Weinberg Equilibrium
IBD	Identical by descent
IBS	Identical by state
kg	Kilogram
LD	Linkage disequilibrium
Logix	Livestock Operational and Genetic Information Exchange
MAF	Minor allele frequencies
MAS	Marker assisted selection
N_e	Effective population size
NGI	Nguni
PCA	Prinicple Component Analysis
PCR	Polymerase chain reaction
QC	Quality Control
QTL	Quantitative trait loci
ROH	Runs of homozygosity
r^2	Measure of LD
SA	South Africa
SNP	Single nucleotide polymorphism
TUL	Tuli

Chapter 1: Introduction and literature review

1.1 Introduction

The South African beef cattle industry comprises of more than thirty recognized breeds which can be classed as indigenous, composite, or exotic breeds (Abin *et al.,* 2016). These breeds possess different traits and breed compositions which allow them to be suited to a range of climatic regions and production environments (Makina *et al.,* 2016). The indigenous and local composite breeds represent approximately 40% of the beef cattle populations in South Africa (Nyamsuhamba *et al.,* 2017). The red meat industry contributed 17,4% to the overall worth of agriculture production in 2016/2017 (DAFF 2017), which makes it an important industry to provide much needed animal protein to a growing population, projected to reach almost seventy million people by 2050 (Worldometers, 2020).

The beef industry in South Africa is unique due to its dualistic nature (DAFF, 2017), with two major sectors in the beef industry, the commercial (developed) and non-commercial (developing) sectors (van Marle-Köster and Visser, 2018). The commercial sector is well developed in South Africa and is responsible for a large proportion of meat production and contributes to creating employment on various levels. The commercial sector generally has a high level of managerial input regarding animal recording, nutrition, animal health and disease control. The developing sector includes the smallholder and subsistence farmers. Unlike the developed sector, these sectors often have an absence of records, and animals are regularly kept and used for religious or social purposes and are a sign of wealth (Mapiye *et al.,* 2019). Although these two sectors of the beef industry are vastly different, both sectors are vital for the success of the beef industry.

The various beef cattle breeds which constitute the beef industry of South Africa have originated from various parts of the globe. The ancient ancestor of cattle is known to be the extinct Eurasian aurochs (*Bos primigenius*) which was domesticated over ten millennia ago in the upper region of the fertile crescent (Verdugo *et al.,* 2019). Evidence suggests that SA indigenous cattle originated from north Africa, now known as the Sahara Desert, over 7000 years ago (Verdugo *et al.,* 2019). The increased temperatures and climatic changes of Northern Africa forced the people in the area, along with their livestock to move south. The well-known Nguni cattle, a Sanga type, is believed to have been farmed by the Nguni speaking people in the 18th century along the eastern coastal regions of South Africa (van Marle-Köster *et al.,* 2021). The other well-known Sanga type is the Drakensberger breed. Although the exact history of the Drakensberger has not been well documented, it is believed that at the time of the Great Trek, several families took their black oxen (known as Vaderlander cattle) to travel north until settling in the Drakensberg range (Bisschoff *et al.,* 2013). The famous Uys family was among these families and started farming and developing the black cattle breed by

selecting the most productive and adapted animals in the herd, ultimately developing the Drakensberger breed known today (Drakensberger Breeder's Society of SA, 2016).

These Indigenous beef cattle breeds are known for their unique adaptive qualities to harsh environments (Mwai *et al.,* 2015) and their capability to produce and reproduce in low input systems (Nyambushama *et al.,* 2017). Traits of importance include drought tolerance, where the individual has natural attributes of good feed utilization where food may be scarce as well as good walking ability (Shabtay *et al.,* 2015). Considering climate change, heat tolerance (Scholtz *et al.,* 2013) is also a critical trait. There is a need for understanding the underlying genetic architecture which determine these adaptive traits to exploit the available genetic variation (Engelsma *et al.,* 2014; Rothschild and Plastow, 2014).

Local composite breeds were developed for their aptitude to adapt to a variation of climatic environments while maintaining a satisfactory level of performance (Bunning *et al.,* 2019). The Bonsmara breed was developed by crossbreeding Milk Short Horn, Hereford, and Afrikaner cattle in 1963 (Bonsma, 1980). The intention to develop the breed was to establish a locally adapted beef breed (Makina *et al.,* 2014). The Beefmaster breed is an international composite breed that was developed using Brahman, Shorthorn and Hereford cattle (Beefmaster SA, 2021). Although the Beefmaster was originally developed in the USA in the 1940's, some South African farmers saw the potential for the breed and started importing semen in the late 1980's (Beefmaster SA, 2021). The breed became an accepted breed in 1987 in South Africa (Beefmaster SA, 2021).

Numerous exotic *Bos taurus* beef cattle breeds (e.g., Hereford and Charolais) were brought to South Africa by European settlers, followed by routine importation into the country over centuries (van Marle-Köster *et al.,* 2021). The first Hereford cattle were imported to SA between 1892 and 1903 and the Hereford Society of South Africa was established in 1917. Charolais cattle were first established in France in 1773, however it was not until the end of World War II that the Charolais was found in other parts of the world. The breed was first introduced in South Africa in 1955 (Charolais Society of South Africa, 2021).

These breeds which were previously considered to be exotic breeds have adapted and perform well in selected regions and production systems in South Africa (van Marle-Köster *et al.,* 2015). The demand for leaner beef has increased in the last decade (DAFF, 2017), proving the exotic breeds with their superior carcass traits to be vital contributors to South African beef production. Several challenges are faced by the beef cattle industry such as climate change, reducing methane emissions, decreasing land and water resources, which all call for an increase in the efficiency and sustainability of the available genetic resources (Garrick, 2011). An insight into the architecture and variation of the genetics in South African beef cattle breeds, may be a valuable tool to warrant sustainable beef production.

The development of genomic technologies, for instance single nucleotide polymorphisms (SNP) and SNP arrays have provided the opportunity and means to explore and comprehend the genetic composition of individuals along with individual traits and genes (Akanno *et al.,* 2018; Visser *et al.,* 2020). The various bovine SNP BeadChips are an important tool used in assessing genetic diversity, by evaluating various diversity parameters (Plastow, 2016; Akanno *et al*., 2018). SNP have been used in a limited number of diversity studies on South African beef cattle (Makina *et al.,* 2014; Lashmar *et al.,* 2019). This could have been due to the lack of genotyping and an African cattle component present in SNP arrays. There is a need to investigate the diversity of SA cattle using SNP data to provide a greater comprehension into the genetic components of SA cattle. In the past six years the number of SA beef cattle which have been genotyped with SNP arrays having an indicine component has increased. This has enabled an in-depth analysis of genetic diversity using SNP.

The beef genomics project (BGP) was founded in 2015 with the primary objective to enable genomic selection (GS) by establishing reference populations for South African beef breeds (van Marle-Köster and Visser, 2018). The ideal reference population is influenced by its size and composition. The levels of relatedness within the reference populations will influence the number of animals necessary to capture all possible genetic diversity within the breed (Lashmar *et al.,* 2019). Additionally, certain population parameters, such as MAF and LD influence the minimum number of SNP that are necessary for downstream applications such as imputation (Lashmar *et al.,* 2018). Thus, the precision of genomic selection and the rate of genetic improvement is influenced by breed-specific population parameters (Lashmar *et al.,* 2018). During a three-year period (2015-2018), over four thousand bovine genotypes were generated (SA Studbook, 2020) using 80k and 150k SNP arrays resulting in useful resources for studying genetic diversity, parentage and selection signatures of cattle breeds (Burrow *et al.,* 2017). Twelve South African beef breeds were used in the BGP to establish or enable reference populations for GS.

South African beef cattle resources include indigenous (Sanga types), local composites and exotic breeds. Due to the varying sizes of the breeds, the number of the genotypes available per breed differ and not all breeds are yet able to apply genomic selection. Information on diversity parameters such as MAF, heterozygosity, LD, ROH, genomic inbreeding and admixture could however assist in providing a greater conception of the genetic structure and relatedness of South African beef breeds. Additionally, it will inform on the structure of the respective training populations.

Aim of the study:

The overall aim of this study was to assess SNP based genomic diversity and population structure of eight South African beef cattle breeds.

The objectives to achieve the aim are as follows:

1. Estimate genetic diversity within and between populations and biological types.
2. Estimate genomic inbreeding and effective population sizes per populations and biological types.
3. Investigate the population structure and admixture of the respective reference populations.

Populations that will be investigated are the Beefmaster, Bonsmara, Boran, Charolais, Drakensberger, Hereford, Nguni and Tuli.

Biological types investigated in this study are indigenous types (Drakensberger, Nguni and Tuli), composite types (Beefmaster and Bonsmara) and exotic types (Boran, Charolais and Hereford).

1.2 Literature review

1.2.1 Introduction

South African beef cattle are rich in genetic diversity and each breed is a vital contributor to the South African beef industry. Each breed has an array of traits which enable them to produce and reproduce in an environment. Some breeds possess adaptation traits which are vital in drought prone areas. Investigating the genetic diversity of these breeds can enable breeding strategies to be made or adjusted with the preservation of some genetic resources.

Genetic technologies, such as SNP arrays, have provided the opportunity to examine the level of genetic diversity in SA beef populations (Plastow, 2016). These technologies enable the ancestry of breeds to be understood. Furthermore, the technologies enable the inbreeding of a population to be investigated and managed to ensure a high level of diversity is maintained in the breed. This literature review aims to review the beef cattle breeds of South Africa, investigation of genetic diversity and implementation of genomic selection in beef cattle so far.

1.2.2 South African beef cattle breeds

Literature suggests that the origin of cattle in Southern Africa was a consequence of two or three migration routes, as shown in figure 1 (Verdugo *et al.,* 2019). The migration routes are based on archaeological evidence, such as bone morphology and rock paintings (Marshall, 2000). Genetic evidence suggests two domestication events from the Near East Africa and India with subsequent crossings with two Wild Auroch subspecies (Gebrehiwot *et al.,* 2020). Mitochondrial DNA evidence suggests that there were two genetically distinct groups of cattle before domestication, *Bos taurus* and *Bos indicus* (Verdugo *et al.,* 2019). *Bos taurus* cattle likely spread across Northern Africa and made their way down to the lush areas of the Sahara (Gebrehiowt *et al.,* 2020). Presumably one of the initial cattle groups to appear in Africa in the Sahara area were humpless *Bos taurus* animals around 4500 and 4000 BC and were used for farming, as supported by archaeological evidence (Blench and McDonald, 2006). The farmers of the time migrated south about four thousand years ago to escape the increasing temperatures, decline in useable land and solar radiation conditions of the Sahara plains with their livestock (Orton *et al.,* 2013). The earliest indication of *Bos indicus* in Sub-

Saharan Africa was found in East Africa where genetic evidence dated back to between 2500 and 2000BC (Grigson, 1999; Gebrehiwot *et al.,* 2020).

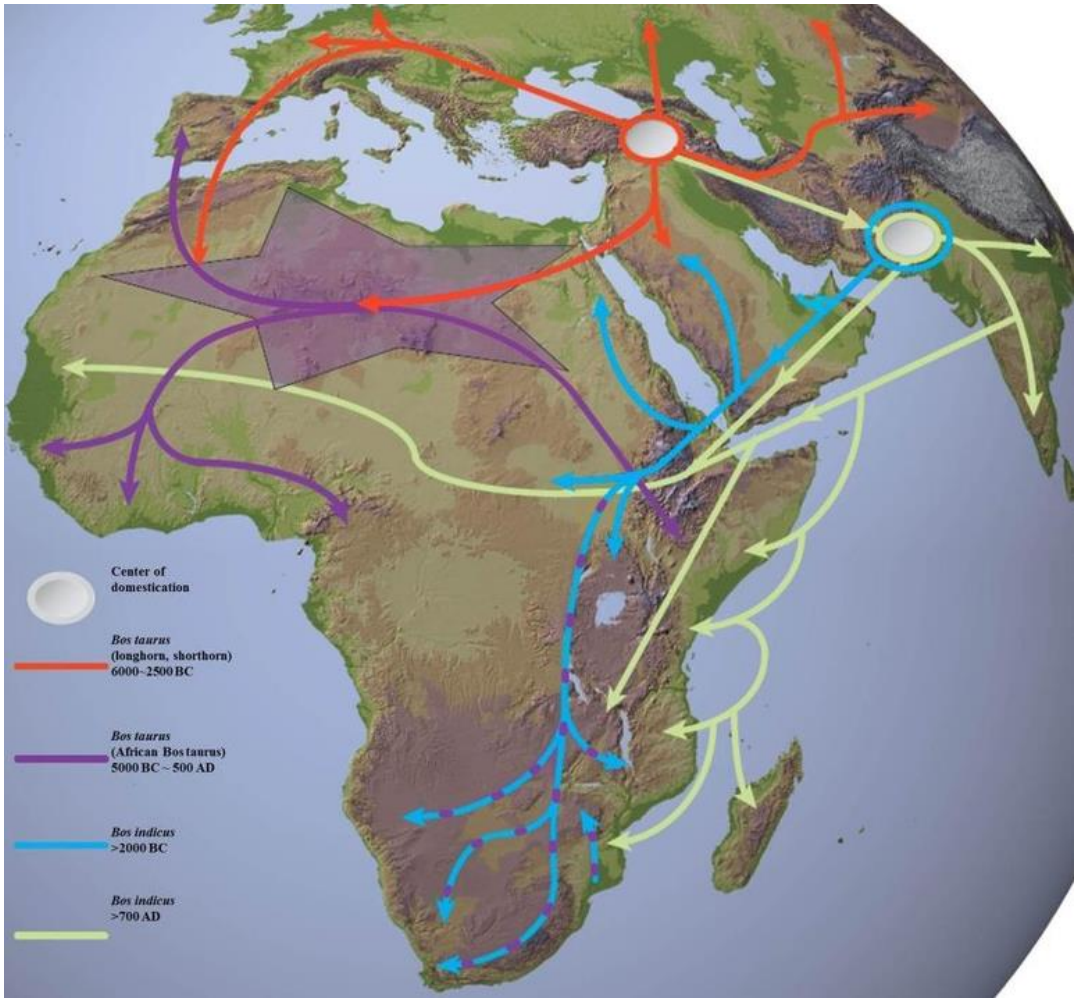

Figure 1.1: Migration routes and the origin of African domestic cattle (Mwai *et al.,* 2015).

The cattle breeds of Africa cattle can be categorized into two broad types; Taurine (*Bos Taurus*) and Indicine cattle (*Bos indicus*) which can be phenotypically distinguished by placement of the thoracic hump (Hanotte *et al.,* 2002). Each breed type that migrated to South Africa is known to have unique characteristics which are advantageous in different climatic regions or production systems. Table 1.1 provides a broad overview of the different traits that the different breeds possess in South Africa for cattle breeds which will be included in this study.

Table 1.1: Broad overview of South African beef cattle classifications and their traits.

Biological types	Breed example	Traits of importance
Indigenous (Sanga)	Drakensberger, Nguni and Tuli	Drought and heat tolerance. Lower susceptibility to ticks with morphological coat characteristics such as colour and thickness of coat and skin.
Local composites	Beefmaster and Bonsmara	Adaptation traits and carcass traits High growth efficiency for any environment.
Exotic (*Bos taurus*)	Charolais and Hereford	Larger framed animals enabling superior growth and carcass traits. Good fertility traits and mothering ability
Exotic (Zebu)	Boran	Zebu cattle have large dewlaps enabling heat dissertation and tick resistance.

Indigenous breeds

Indigenous African cattle breeds were derived from the crossbreeding of *Taurine* cattle with *Indicine* cattle (Verdugo *et al.,* 2019). It is suggested African cattle are divided into three categories of African indigenous breeds; African *Bos taurus* (Sanga types), *Bos indicus* (Zebu types), and Zenga types (Gebrehiwot *et al.,* 2020). The African indigenous populations displayed good growth and adaptation traits which was proven useful for the harsher climates of the regions in which they lived (Makina *et al.,* 2016; Mwai *et al.,* 2015).

Sanga types are a resultant of cross breeding of the humpless *Bos taurus* x humped *Bos indicus* hybrid to create a breed type with a significant body mass and production in regions where animal adaption was mandatory e.g tsetse-free areas (Zwane *et al.,* 2019; Gebrehiwot *et al.,* 2020). The San and Sudanic Bantu tribes pioneered Sanga cattle to South Africa during their migration to Southern Africa along with the European settler's arrival in the fifteenth century (Strydom, 2008).

Most SA indigenous breeds are small to medium-framed allowing the cattle to have lower nutrient requirement for maintenance compared to larger framed animals (Mapiye *et al.,* 2019). Some cattle have good walking and foraging ability to select for optimal nutrition which enables the feed intake level to be sufficient for production (Mwai *et al.,* 2015). Tick resistance is associated with the coat qualities such as colour, density and hair length, making it challenging for ticks to attach to the skin (Marufu *et al.,* 2011). Indigenous cattle are known to be heat resistant (Mkize *et al.,* 2020) causing these animals to perform well in harsh temperature environments with little to no effects of heat stress (Scholtz *et al.,* 2013).

Composite breeds

Composite breeds tend to adapt well to most climates while maintaining a satisfactory level of performance (Bunning *et al.,* 2019). Traits which are known to have low heritability (e.g., fertility), tend to have a higher level of heterosis (Bunning *et al.,* 2019) in composite breeds. In both the Beefmaster and the Bonsmara, the breeds were developed using animals which possessed adaption traits to harsh South African climates (e.g., Afrikaner and Brahman breeds) and carcass and growth traits which are of vital importance to meet the demands for beef in SA (e.g., Hereford breed).

Exotic breeds

Exotic breeds in SA can be classified as European *Bos taurus.* The exotic *Bos taurus* breeds are typically larger framed animals, allowing a greater potential for increase in carcass mass (Soji and Muchenje, 2016). These breeds have been selected and are widely utilized under South African farming conditions, due to their carcass and growth traits. The breeds which were previously considered to be exotic breeds; have adapted and perform well in selected regions and production systems (van Marle-Köster *et al.,* 2021). The demand for leaner beef has increased in the last decade (DAFF, 2017), proving the exotic breeds to be vital contributors to South African beef production. Larger framed animals typically have larger hind quarters and back muscles, enabling larger amounts of lean meat to be produced (Soji and Muchenje, 2016). These breeds also have good fertility traits and mothering ability. These traits are important for reproducing in the harsher climatic regions of South Africa.

The *Bos indicus* Zebu breed included in this study (Boran) can also be classified as an exotic breed. The Borans ancestors from Asia arrived in the horn of Africa around 1500 years ago (Decker *et al.,* 2014). The cattle were situated around semi-arid areas of Ethiopia and Kenya and were used for beef and dairy production (DAGRIS, 2010). The breed was imported to South Africa in the late 1990s and was bred for beef production. The breed is known for high fertility traits and adaptability traits such as drought resistance (Gaughan *et al.,*1999).

South Africa is one of the several nations in Africa where a countrywide program for animal recording has been implemented and is used (Van Marle-Köster *et al.,* 2015). Animal recording plays an essential role to managing the diversity among the breeds and assists in genomic predictions. SA Studbook's Logix database is used by some breeders' associations in South Africa for animal recording and genetic analysis. According to the 2016 annual report of SA Studbook, 1859 herds are registered with SA Studbook and 69% participate in Logix. There are approximately just over 284 600 beef cattle that participated in Logix beef in 2016 (SA Studbook, 2016). Tables 1.2 and 1.3 below illustrate the different cattle breeds with the number of animals participating in Logix with the average for each trait measured.

Table 1.2: Beef cattle breeds, number of animals participating in Logix Beef with averages of respective weights measured by Logix Beef.

Breed	Number of Animals	BW (kg)		WW (kg)		12MW (kg)		18MW (kg)	
		Bull	Cow	Bull	Cow	Bull	Cow	Bull	Cow
Afrigus	570	30	22.0	245.2	237.2	312.9	271.4	-	-
Afrikaner	6937	30.1	28.4	202.2	191.3	235.4	240.7	322.3	299.7
Afrisim	1020	34.4	32.0	205.6	186.9	-	264.6	-	246.2
Angus S.A.	20716	35.8	34.0	233.5	224.1	333.7	307.8	506.7	406.3
Ankole	429	24	-	182.0	178.0	311.0	-	-	-
*Beefmaster	47517	35.5	33.8	241.5	224.5	302.4	277.0	420.6	350.7
Beef shorthorn	402	40.6	38.6	232.6	217.4	387.2	244.5	467.3	327.5
*Bonsmara	118758	36.3	34.3	226.5	210.3	278.7	257.9	375.0	325.3
*Boran	16294	30.4	28.8	202.8	185.8	261.9	216.0	349.4	273.6
Borguni	305	30.1	29.9	187.1	177.0	-	-	286.0	289.3
Braunvieh	2121	38.1	35.7	242.2	226.1	333.2	242.6	442.0	336.1
Charbray	83	-	-	255.3	243.6	331.5	297.7	489.0	379.0
*Charolais	4913	40.7	39.4	236.8	223.3	351.7	306.6	468.1	361.1
Chianina	100	-	-	-	-	-	-	-	-
Dexter	765	25.3	24.1	155.1	144.7	199.5	207.9	-	202.2
*Drakensberger	12862	36.0	34.1	215.3	200.1	248.9	234.9	359.6	309.1
Gelbvieh	541	38.1	36.0	242.4	234.1	282.0	237.6	417.6	302.7
*Hereford	5996	38.3	36.4	216.2	201.7	304.5	280.1	453.6	360.3
Hugenoot S.A.	2383	40.8	38.9	222.1	210.0	234.0	250.4	339.1	310.1
*Nguni	15901	27.3	25.8	162.5	148.9	218.9	182.2	271.1	239.8
Pinzgauer	1498	37	30.0	239.4	210.3	-	-	-	-
Pinzyl	1597	-	-	195.8	181.9	-	-	292.0	300.5
Red Poll	560	34.1	32.3	182.5	168.7	280.0	200.3	282.8	357.8
Romangnola	1004	43.5	37.6	260.5	242.4	417.0	-	-	-
Senepol	4488	37.8	36.1	219.4	213.9	286.3	268.2	383.0	332.6
South Devon	1251	38.0	36.8	215.5	204.2	346.0	260.6	-	385.0
Sussex	6190	39.6	37.2	235.7	216.1	329.4	283.6	482.3	367.3
*Tuli	9486	32.7	30.8	189.7	176.6	240.2	206.8	338.7	270.6

BW: Birth weight, WW: weaning weight, 12MW: 12-month weight, 18MW: 18-month weight,

A * indicates breeds included in this study

Table 1.3: Beef cattle breeds, number of animals participating in Logix Beef with averages of respective production and fertility traits measured by Logix Beef.

Breed	AFC	ICP	ADG (g)	FCR (kg/kg)
Afrigus	31	408.9	-	-
Afrikaner	35.5	459.8	1245	7.36
Afrisim	32.1	403.5	1552	7.16
Angus S.A.	31.3	408.6	1779	6.20
Ankole	33.7	380.2	-	-
*Beefmaster	30.6	404.6	1737	5.59
Beef shorthorn	36.2	380	1497	8.03
*Bonsmara	31.4	413.7	1721	5.97
*Boran	33.6	436.5	1308	5.80
Borguni	36.5	475.6	-	-
Braunvieh	33.6	450.9	1808	5.66
Charbray	37	469.5	-	-
*Charolais	33.7	422.5	1874	6.09
Chianina	45.8	475.7	1518	7.65
Dexter	26.9	440.9	1136	7.32
*Drakensberger	34.1	424.1	1612	5.96
Gelbvieh	33.2	458.1	1759	5.51
*Hereford	31.9	391.3	1746	6.02
Hugenoot S.A.	35.5	479.8	1538	6.35
*Nguni	32.3	415.4	657	-
Pinzgauer	36.2	467.1	1813	6.21
Pinzyl	34.7	448.7	1300	7.89
Red Poll	34.9	494.9	-	-
Romangnola	36	497.6	1851	5.62
Senepol	31.8	433.1	1863	6.21
South Devon	35	393.3	-	-
Sussex	32.4	408.3	1867	5.67
*Tuli	34.9	423.0	1074	-

AFC: Age at first calving (months), ICP: Inter calving period, ADG: average daily gain, FCR: feed conversion ratio

Growth performance at numerous stages of the growth curve directly influences profitability in the beef industry (Koch *et al.,* 2003). Weight measurements in table 1.2 indicate

the *Bos taurus* breeds, such as the Charolais and Hereford had the largest eighteen-month weight. Individuals belonging to the *Bos taurus* family are typically large framed, hence the large weight measurements. The larger frame of these individuals enables a larger amount of muscle to be added on to the frame which consequently adds to the individuals having superior carcass traits. The average daily gain (ADG) measured in grams, further explains the individuals having superior carcass traits by the *Bos taurus* breeds having the highest ADG measurements in table 1.3. The larger framed animals perform particularly well in feedlots where they can put on weight and muscle mass without any potential food or water shortage.

The indigenous Sanga types have an advanced tolerance for heat and drought compared to the *Bos taurus* breeds (Shabtay, 2015). In certain regions of South Africa, heat stress becomes one of the main challenges for animals in the summer months particularly in the savannah regions of the country. Heat stress decreases growth and production performance such as average daily gain, reproductive performance, and meat quality. The animal's ability to perform everyday physiological processes that are necessary for survival is influenced by the amount of energy needed to maintain these processes (Hoffmann, 2010). The small to medium frame of the Sanga types occurred from the genotype of individuals becoming adapted to available feed resources and temperatures which assists in the maintenance of important processes (Hasen, 2004). The weight measurements in table 1.2 indicate the Sanga types are smaller framed compared to the *Bos taurus* breeds, which enables the physiological and cellular characteristics to enable Sanga types to be more adapted to tropical environments (Kim et al., 2017). Smaller framed animals have lower energy requirements for body weight and weight gain enabling them to perform well in environments where food or water may be limited such as the savannah (Scholtz *et al.,* 2013).

The weight measurements for the composite breeds such as the Beefmaster and Bonsmara have a higher average compared with Sanga types, but smaller measurements versus the *Bos taurus* types which is to be expected due to the composite nature of the breeds. This enables the breeds to be farmed extensively in the savannah regions but still to produce the desired meat quality. There are many composite breeds in South Africa where animal recording is done by other service providers such as the LRF, however they will not be investigated in this study.

All South African breeds play a crucial role in the beef industry, from their larger frames for increased muscle gain, to adaptation traits enabling production in stressing environments (Scholtz et al., 2013). Therefore, the investigation of the diversity of these breeds will provide insight to the genetic resources of these breeds for potential conservation and future breeding strategies to be implemented.

1.2.3 Genetic diversity

Genetic diversity can be defined as the set of genetic differences within populations and individuals within a population that are present in the population's genetic makeup (Eusebi *et al.,* 2020). In recent years, genetic diversity studies have become an important tool to evaluate genetic resources and understand population structure (Jemma *et al.,* 2018; Upadahay *et al.,* 2018) allowing breeding decisions and objectives to be set (Lenstra *et al.,* 2012). Breed ancestry and diversity also provide a background of the origin of the breed, and this can allow for certain genes to be identified (Signer-Hasler *et al.,* 2017). Genetic diversity plays an essential role in natural selection where populations and individuals require to adapt to a variety of environmental stressors (Jemma *et al.,* 2018). Understanding how artificial and natural selection influences the genetic diversity of a population allows for successful breed management to maintain the level of diversity (Engelsma *et al.,* 2014; Mapiye *et al.,* 2019).

Early artificial selection of beef cattle in the late 1700s could have caused a decrease in diversity within populations as a result of uncontrolled inbreeding with individuals with superior traits for an increase in production or fertility (Engelsma *et al.,* 2014; Rothschild and Plastow, 2014). Artificial selection results in selection of a gene or many genes on a particular part of the genome to improve a trait of interest, or a variety of traits (Rothschild and Plastow, 2014). The specific traits selected for, often lead to an increase in the favourable alleles, with allele fixation and reduction in differences at that section of the genome as a consequence (Zhao *et al.,* 2015). Using DNA-marker technology is a method of managing selection in a population by analyzing the variation present in the population.

1.2.3.1 Development of DNA-marker technology.

Over the past four decades there has been substantial enhancements and technological advances in the sphere of molecular genetics (Ducrocq *et al.,* 2018). Technologies such as the development of Sanger sequencing in the late 1970s which was based on chain-termination methods (Dekkers & Hospital, 2002), led to the development of the polymerase chain reaction (PCR), which was created in 1983. PCR permitted for accurate and reliable magnification of small segments of DNA or particular regions on the genome (Fore *et al.,* 2006). This contributed to discoveries in 1989 on the human genome led to the further developments of automatic sequencing instruments and associated software which pioneered discovery of DNA markers, such as microsatellites and SNP, which are commonly used in livestock (Fore *et al.,* 2006; Mrode *et al.,* 2019).

The first genome to be mapped was the human genome which was completed in 2001. This groundbreaking achievement unleashed the possibilities of genomes of other species to be mapped (Eggen, 2012). The mapped bovine genome was sequenced using a female

Hereford cow and was completed in 2009 (Fan *et al.,* 2010). This allowed scientists to gain understanding of the architecture of the genome (Matukumalli *et al.,* 2009). The mapped genome provided the tool for an in-depth understanding of the progression of species, breed structure and offered new assumptions of population genetics and diversity (Fan *et al.,* 2010; Ducrocq *et al.,* 2018). The reference genome allowed for DNA marker discovery (microsatellites and SNP) to be used for further in-depth analysis of the genome (Plastow, 2016; Ducrocq *et al.,* 2018).

Microsatellite markers are numerous copies of short tandem repeats that are consistently dispersed throughout an animals' genome (Sabir *et al.,* 2014). Microsatellite markers can be useful in parentage verification, valuation of genetic distance within and between breeds, mapping of genes, marker assisted selection and population genetics (Cole *et al.,* 2013; Malteca *et al.,* 2020). Some diversity studies on cattle in South African have been conducted using microsatellites, which includes Pienaar *et al.,* (2018) where microsatellites were used to estimate the levels of heterozygosity among the Afrikaner breed in South Africa. Six Nguni ecotypes in South Africa were investigated by Sanarana *et al.,* (2015), with the aim to understand genetic diversity for management and conservation purposes. Madilindi *et al,* (2019) also investigated the diversity and relationships among South African Nguni ecotypes. Van der Westhuizen *et al.,* (2020) assessed breed genetic variation to estimate associations among the different classifications of cattle in South Africa. These studies provided insight into the diversity and relationships of South African beef cattle breeds. All studies mentioned, concluded that there is a moderate to high level of genetic variation within and between breeds which puts the various populations in a favorable position for genetic improvement (Pienaar *et al.,* 2018; Madilindi *et al.,* 2019; van der Westhuizen *et al.,* 2020). Although the mentioned studies provide positive conclusions, microsatellite markers have various limitations (Fernăndez and Bennewitz, 2017). Such limitations include high costs associated with the development of the marker panels, or heterozygotes can be misclassified as homozygotes which results in inaccurate conclusions of diversity in a population (Cole *et al.,* 2013). Some microsatellite markers are known to cause underestimation of genetic divergence with the markers only cover a small portion of the genome (Yang *et al.,* 2013; Plastow, 2016).

Single nucleotide polymorphisms (SNP) can be characterized as a difference at the same point on a genome between individuals or between individual chromosome pairs (Goddard, 2009). Over 2.3 million SNP were discovered subsequent to the completion of the bovine genome (Williams *et al.,* 2009). SNP markers are typically bi-allelic and are relatively easily interpretable, making SNP an ideal tool for genomic studies (Fan *et al.,* 2010; Malteca *et al.,* 2020). SNP arrays have provided the opportunity and means to explore and perceive the genetic composition of individuals and individual traits and genes (Akanno *et al.,* 2018; Visser *et al.,* 2020). SNP markers were primarily developed for association mapping,

admixture plotting with studies aimed at distinguishing phenotype and genotype relations (Goddard & Hayes, 2009). SNP markers are known to be located abundantly and are relatively consistently spaced throughout the genome (Hayes *et al.,* 2009). This is advantageous in identifying and studying quantitative/polygenic traits (Edea *et al.,* 2013). SNP panels were initially developed using European *Bos taurus* breeds (Lwin *et al.,* 2018). This became a challenge for accurate studies being conducted on *Bos indicus* breeds as the novel SNP of the *indicus* breeds were often not included in the SNP array (Edea *et al.,* 2018). The development of the 80k SNP Indicus array by GeneSeek in 2014 included some of the novel SNP from *Bos indicus* breeds genome, leading to accurate estimates of diversity parameters for the breeds (Akanno *et al.,* 2018; Edea *et al.,* 2018). Table 1.4 indicates the different available commercial SNPbead arrays for cattle:

Table 1.4: A non-comprehensive summary of some available commercial SNPbead chips (Adapted from Geneseek® and Illumina® 2021)

Company	Bead chip	Number of SNP
Geneseek®	GeneSeek Dairy Ultra LD v2 GGP-LD	7 049
	- version 1 (GGP9K)	8 610
	- version 2 (GGP20K)	19 721
	- version 3	26 151
	GGP-indicus 50k	54 791
	GGP-HD	76 879
	GGP-150K	139 480
	GGP- F250	221 115
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
	- version 3	53 218
	Bovine HD 770k	777 962

1.2.3.2 Diversity studies for South African beef cattle.

The various available SNP arrays present the opportunity of studying genetic architecture in a variety of populations by identifying genes and understanding relevant physiological pathways for adaptation and production (Edea *et al.,* 2018; Lwin *et al.,* 2018). Over the past decade, SNP arrays have become more affordable and utilized, giving scientists an opportunity to estimate diversity parameters and population structure (Makina *et al.,* 2014, Plastow, 2016, Akanno *et al.,* 2018). Limited studies using SNP panels have been conducted

on South African cattle populations to estimate the levels of diversity present (Makina *et al.,* 2014; Lashmar *et al.,* 2018). A summary of these studies from 2014 to present are shown in table 1.5.

Table 1.5: A non-comprehensive summary of SNP-based genetic diversity studies on South African beef cattle.

Aim of study	Breeds	Diversity parameters	Reference
Investigate the population composition within and between six South African cattle breeds	AFR, ANG, BON, HOL, DRB, NGI	H_e, H_o, allelic richness F_{IS}, F_{ST}, Allelic sharing and genetic distance Genetic structure using PCA	Makina *et al.,* 2014
To provide a clearer analysis of patterns of admixture and ancestry in South African cattle breeds	AFR, BON, DRB, NGI	Genetic relatedness between animals Population structure	Makina *et al.,* 2016
Investigate within the SA Drakensberger breed, quantifying inter-chromosomal variation.	DRB	Genetic relatedness between animals Inter-chromosomal variation (MAF and LD) ROH and F_{ROH}	Lashmar *et al.,* 2018
Identify novel SNP in three SA indigenous breeds.	AFR, DRB, NGI	Pedigree analysis. DNA sequencing Variant discovery SNP annotation Assessed SNP density. Identification of selective sweeps.	Zwane *et al.,* 2019
Identify breed informative SNP among three SA beef cattle breeds using genotypic data from the Bovine50SNP and GGP-80k assays.	AFR, DRB, NGI	Allele frequency Genetic structure analysis with PCA. Breed-specific markers. Pairwise F_{ST} values FLK statistic	Zwane *et al.,* 2016

AFR: Afrikaner, ANG: Angus, BON: Bonsmara, DRB: Drakensberger, HOL: Holstein, NGI: Nguni

The studies indicate the moderate levels of genetic diversity among and between different populations. The studies provide details on the admixture of the different populations and indicates that South African beef cattle populations can still respond well to selection. The

studies in table 1.5 further comment on the need for future research to be done due to the potential the populations have to respond to selection and downstream genomic applications. The ability to respond to selection is important in the beef industry due to the ongoing climate change and fluctuations in consumer preferences. Additional genetic diversity studies have been conducted throughout Africa and worldwide (Edea *et al.,* 2018; Fabbri *et al.,* 2019). The studies highlight the importance of indigenous populations acting as reservoirs of biodiversity for adaptation traits which may be valuable in the future. Literature further indicates the importance of diversity to develop breeding strategies for the future (Cesarani *et al.,* 2018; Edea *et al.,* 2018). Managing breeding programs and breeding objectives are key factors in creating a sustainable use of genetic resources.

Genetic resources should be sustainably managed to ensure components of livestock diversity are exploited at a rate that does not cause a decline of genetic resources in the long-term and has the potential to satisfy the needs and desires of forthcoming generations (Engelsma *et al.,* 2014; Mapiye *et al.,* 2019). Genetic diversity is therefore important for conservation purposes of indigenous breeds with lower population numbers (Mastrangelo *et al.,* 2018). These indigenous breeds may hold unique genetic information which may become of importance in the future (Jemma *et al.,* 2018).

1.2.4 Genetic diversity parameters

There are several parameters which can be utilized to assess the genetic diversity of a population using the various SNP arrays. These parameters include observed and expected heterozygosity, runs of homozygosity (ROH), inbreeding coefficient, linkage disequilibrium (LD) and effective population size (Al-Mamun *et al.,* 2015). Estimation of genomic population parameters (such as H_e) is important to investigate the diversity within a breed and the association between target traits (Akanno *et al.,* 2018).

The extent of genetic heterozygosity observed could illustrate the genetic variation between and among animals in populations and individuals of a certain cattle breed (Bahbahani *et al.,* 2017). In genomic studies, expected (H_E) and observed (H_O) heterozygosities are estimated (Eusebi *et al.,* 2020). H_O is the frequency of observed heterozygotes in a population and is negatively associated with inbreeding depression (Lenstra *et al.,* 2012; Bahbahani *et al.,* 2017). H_E is the likelihood that two alleles at a specific locus from any two individuals selected at random in a population are distinct (Nei, 1973). H_E is calculated using the formula:

$$H_E = 1 - \sum p^2$$

Where H_E is the expected heterozygosity and p is the allele frequency (Melka and Schenkel, 2012). A populations' ability to respond to either natural or artificial selection within

a short time frame is measured by estimating H_E (Fernández and Bennewitz, 2017). It is often presumed that populations that have a higher H_E have an increased genetic variation and can therefore be better adapted to changes in environment and respond well to selection (Fernández and Bennewitz, 2017; Eusebi *et al.,* 2020). A gain in genetic variability is seen when the H_O is larger than H_E. A loss in variability occurs when the H_E is larger than the H_O ((Melka and Schenkel, 2012).

Linkage disequilibrium (LD) between two markers indicates the extent of non-random association between them (Al-Mamun *et al.,* 2015). Genomic LD is a product of various genetic occurrences such as selection, mutation, and genetic drift, however, it can also be as a result of non-genetic causes such as migration (Eusebi *et al.,* 2020). Traditionally, LD has been measured using the estimates of one of two parameters: |D'| and r^2 (Corbin *et al.,* 2010). For small population samples and for populations who may have rare alleles, |D'| may result in an overestimation of LD in the population (Bohmanova *et al.,* 2010). r^2 is therefore often the preferred measurement for LD estimation in livestock studies (Van Liere and Rosenberg, 2008). This LD measure (r^2) is defined as the square of the correlation coefficient between two indicator variables – each variable represents the presence or absence of a particular allele at each position on the chromosome (Van Liere and Rosenberg, 2008). It has been shown that r^2 values are dependent on the population size and portion of recombination (Al-Mamun *et al.,* 2015). r^2 has a frequency dependent range, this means that the maximum value of r^2 declines with the extent of the MAF, thus r^2 will have a range of 0 to 1 (Eusebi *et al.,* 2020).

LD-based association studies use the r^2 statistic for the identification of informative markers. LD is first determined over the whole genome to indicate the number of markers required for specific chromosomal analysis (Boichard *et al.,* 2016). Once this is done, certain fragments of interest can be analyzed (Al-Mamun *et al.,* 2015). For example, fewer markers are required when LD is high over longer chromosomal portions. LD decay and the pattern in which it occurs, provides insight into the evolutionary history of a population (Boichard *et al.,* 2016). This can assist in estimation of the effective population size (Edea *et al.,* 2018).

Minor allele frequencies (MAF) are often estimated in diversity studies and provides information to distinguish among the more frequent and rare variants in a population. Often SNP that have low-MAF values are excluded from analysis due to potential genotyping errors (Malomane *et al.,* 2018). Furthermore, low-MAF SNP have been shown to negatively impact downstream genomic applications such as imputation (Schrooten *et al.,* 2014). A greater proportion of polymorphic SNP present in the population can often be applied from a higher MAF estimate (Qwabe *et al.,* 2013). Rare variants could be important for future implications for breeding objectives and may become traits of importance (Engelsma *et al.,* 2014).

Population structure and admixture provide an indication of the genetic diversity between populations and can be represented using a principal component analysis plot (PCA). PCA plots are characterized by the clustering of populations. These indicate partition of genomes of individuals into a specific number of clusters. The number of clusters that form, indicate the number of distinct populations. Furthermore, PCA plots provide an indication of the time of divergence for breeds (Edea *et al.,* 2013). Populations that derived more recently tend to form tight clusters, suggesting a common ancestor (Gautier *et al.,* 2007). Populations that have admixture are populations that have ancestry from a variety of previous generations. Admixture can be seen on an admixture plot where an individual is represented by a line displaying different colours for each population that is being studied (Alexander *et al.,* 2009). K-values are used in admixture plots to determine the maximum likelihood estimation of ancestry to determine the true number of distinct genetic populations.

1.2.5 Genomic inbreeding and effective population size

Inbreeding occurs when two individuals which share at least one common ancestor are bred with each other (Fleming *et al.,* 2018). Inbreeding can diminish the level of heterozygosity which consequently decreases the genetic diversity in a population (Goddard, 2012; Bunning *et al.,* 2019). Inbreeding can further cause adjustments in the distribution of genetic variation, resulting in allelic fixation. An increase in inbreeding in a population can occur for several reasons such as intense directional selection over several generations, the use of artificial insemination (AI) with a select few superior sires and the use of EBVs in conjunction with artificial selection which often leads to parents being related individuals (Robertson, 2007; Howard *et al.,* 2017). Inbreeding should be managed to maintain the desired level of diversity (Burrow *et al.,* 2017). An increase in inbreeding could result in a decline in animal performance, resulting in inbreeding depression (Du Toit *et al.,* 2012; Eusebi *et al.,* 2020). Inbreeding depression influences fitness traits for instance, reproduction and adaptation traits, this has an economic impact due to resulted decreased in performance by the animals (Felming *et al.,* 2018; Howard *et al.,* 2017).

In the early 1910s, the initial attempts to measure inbreeding was done using pedigree information (Curik *et al.,* 2014). The parameter F_{PED} was used as an estimate of the pedigree inbreeding coefficient (Ferenčaković *et al.,* 2011). Malecot and Blaringhem (1948) explained that if no selection or mutations are occurring, it is assumed that all loci are separating in the same hereditary configuration and thus likely to have a similar inbreeding coefficient (F_{PED}). F_{PED} is the estimated fraction of the genome that is identical by decent (IBD) and the random probability distribution nature of inheritance is not taken into account (Curik *et al.,* 2014). Furthermore, F_{PED} is an approximate estimate of the individual autozygosity. F_{PED} is also estimated using the reference generation as the founder generation. However, should the

founder generation be poorly characterized, future generations may also be poorly characterized as a result, therefore inbreeding information could be considered inaccurate. Pedigree information requires an extensive and accurate animal recording program to be in place (Abin *et al.,* 2016). South Africa has an animal recording scheme for stud breeders, however, some on farm errors and lack of pedigree depth of records can still decrease the accuracy of pedigree information and influence the inbreeding.

The development of SNP arrays has guided an escalating interest in estimating the inbreeding coefficients on a molecular level rather than a statistic obtained from pedigree data (Ferencakovic *et al.,* 2013; Kim *et al.,* 2018). The inbreeding coefficient is the probability of any two randomly selected alleles at the same locus from two gametes that are IBD from a shared ancestor (Peripolli *et al.,* 2017). The inbreeding coefficient (F_{IS}) is a popular parameter to estimate inbreeding of a population (Lenstra *et al.,* 2012). F_{IS} is specified as the proportion of the overall inbreeding within a population due to inbreeding within subpopulations (Ewens, 1969). F_{IS} estimated based on heterozygosity estimates, both observed and expected heterozygosity, using the formula:

$$F_{IS} = \frac{H_E - H_O}{H_E}$$

Runs of homozygosity (ROH) are utilized to quantify individual autozygosity due to the high correlation (r = 0.7) between ROH and individual autozygosity (Zavarez *et al.,* 2015). ROH occurs when an individual inherits chromosomal fragments that are IBD from both parents (Ferencakovic *et al.,* 2011; Peripolli *et al.,* 2017). This results in homozygous segments in the progeny's genome which gives rise to ROH (Peripolli *et al.,* 2017; Upadhay *et al.,* 2018). The first livestock analyses on ROH were conducted on Simmental which were all genotyped using the Illumina Bovine SNP50K array (Sölkner *et al.,* 2010; Ferencakovic *et al.,* 2011).

When statistical analyses are conducted, the density of the SNP array employed to generate ROH data can cause a variation in ROH identification (Mastrangelo *et al.,* 2016). Other influences of the size of the ROH segments include genetic drift, bottle necks, mutation rate, recombination, LD and popular sires (Periplolli *et al.,* 2016; Cesarani *et al.,* 2018). In recent years, studies pertaining to ROH have shifted focus to the identification and characterisation on its relationship with population structure and inbreeding.

Estimations from genotypic information indicate definite relatedness among individuals of a population through individuals being IBD (Zavarez *et al.,* 2015). The statistical parameter of F_{ROH} can be used to estimate inbreeding coefficients as the parameter has a strong correlation with homozygosity (Fleming *et al.,* 2018). F_{ROH} is a more accurate prediction (than

F_{PED}) of the current autozygotic proportion of the genome and distinguishes autozygosity because of common ancestry (Keller *et al.,* 2011; Periplolli *et al.,* 2016). The autozygosity distribution of the genome can be studied to locate specific areas that may have high levels of autozygosity (Keller *et al.,* 2011). This study of the genome further allows the F_{ROH} for each chromosome to be estimated (Zavarez *et al.,* 2015). F_{ROH} estimates can reveal the incidence of inbreeding according to the length of runs of homozygosity (ROH) (Periplolli *et al.,* 2016). Long segments of ROH indicate inbreeding due to a recent ancestor where short segments represent inbreeding from distant generations (Upadhyay *et al.,* 2018). F_{ROH} estimates have practical implications for conservation purposes (Herrero-Medrano *et al.,* 2013). Individuals with elevated levels of F_{ROH} can be omitted or used sparingly in mating programs for populations with small effective population sizes (Periplolli *et al.,* 2016). The F_{ROH} can be calculated as follows (McQuillan *et al.,* 2008):

$$F_{ROH} = \frac{\sum F_{ROH}}{\sum L_{AUTO}}$$

Where: $\sum F_{ROH}$ = the length of ROH in one individual

$\sum L_{AUTO}$ = the length of the genome covered by SNP

The effective population size (N_e) is a significant population parameter that illustrates the genetic drift of a population (Barbato *et al.,* 2015). N_e is often influenced by the fluctuation of the census population, the number of different individuals used for breeding and the variation in reproductive success (Wang *et al.,* 2015). The estimation of N_e can be demographic-based, pedigree-based or marker-based. The important and somewhat limiting aspect with a pedigree-based approach is the fact that the pedigrees must be complete (Meuwissen *et al.,* 2016). The preferred, and more accurate method of N_e estimation is the marker-based approach (Barbato *et al.,* 2015). The marker-based approach uses genomic data and can often be estimated using LD (Eusebi *et al.,* 2020). A population that has a lower effective population size is more at risk of inbreeding (Lwin *et al.,* 2018). Comparison of the magnitude of LD and effective population size is considered to be informative regarding the total diversity level in a species and can assist in the understanding of the different selection pressures placed on a population (Al-Mamun *et al.,* 2015). N_e and inbreeding coefficients are correlated, meaning that when the N_e of a population is low, the risk of inbreeding increases (Engelsma *et al.,* 2014). Inbreeding can cause an increase in homozygosity and could potentially decrease the level of diversity (Al-Mamun *et al.,* 2015).

1.2.6 Genomic selection of South African beef breeds

Meuwissen *et al,* (2001) put forward an approach that suggested the idea that the breeding value (BV) could be estimated from markers which could span over the whole genome. The approach proposed that the genetic effects are estimated for each marker and then totaled to predict the overall BV of an individual (Boichard *et al.,* 2016). This approach is called “Genomic Selection” (GS). GS makes use of dense SNP markers covering the entire genome to estimate the Direct Genomic Value (DGV) and QTLs are assumed to be in LD with no less than one SNP on a SNP chip (Meuwissen *et al.,* 2001). Hence markers can be used to select for favorable alleles without identifying the actual gene or allele (Ducrocq *et al.,* 2018). GS assumes that all genetic variation is explained by the markers and therefore the accuracy of GS is dependent on marker density (Goddard, 2012; Eusebi *et al.,* 2020). This means that the markers should be dense enough so that all QTLs, or alleles of interest, are in high LD with a couple or at least one of the markers (Pryce *et al.,* 2010). The consequence if this does not occur is that some of the genetic variance may not be detected by the markers (Goddard, 2012; Boichard *et al.,* 2016). GS should increase genetic gain while maintaining genetic variation (Mrode *et al.,* 2019).

Young animals can be selected, with decisions being made more accurately without their phenotypic records, and consequently, young bulls can be selected for mating the moment they are physiologically capable to reproduce (Pryce *et al.,* 2010, Meuwissen *et al.,* 2016, Mrode *et al.,* 2019). This decreases the generation interval for a population, which is another advantage of GS due to genetic progress being made at a faster rate (Goddard, 2012). However, a decrease in generation interval could potentially increase inbreeding estimates per generation due to strong selection of a limited number of young sires with the desired breeding values (Fleming *et al.,* 2018). There are methods, such as optimal contribution selection, to balance the genetic gain and potential rate of inbreeding which allows for better conservation or increase of genetic diversity when centered around genomic data (van der Westhuizen *et al.,* 2017).

Successful implementation of GS for a breed is dependent on the accurate genotyping of a reference population for each breed to gage a level of diversity that is comparable in future (Burrow *et al.,* 2017; Meuwissen *et al.,* 2020). The reference population size is dependent on individuals who have complete phenotypes and genotypes. The prediction equation is estimated based on these complete genotypes and phenotypes (Goddard, 2012). Accurate pedigree and phenotypic records should be provided with the corresponding genotypic information for each animal (Boichard *et al.,* 2016; Lashmar *et al.,* 2019). The establishment of a reference population often requires a large population size to sample individuals that will represent the national herd (Lashmar *et al.,* 2019). Although a minimum number of one thousand individuals are used as a guideline for a reference population this number may differ due to breed dynamics

(Meuwissen *et al.,* 2001; Habier *et al.,* 2007). It is important to note the reference population is not solely dependent on the number of individuals, but also the composition of the population (Mrode *et al.,* 2019). Thus, breed-specific population parameters will influence the accuracy of genomic selection and the rate of genetic improvement (Lashmar *et al.,* 2018). It is advised to select genetically influential individuals, such as grandparents, cows and sires with superior BV accuracies for reference population establishment (Daetwyler *et al.,* 2008; Ducroq *et al.,* 2018). This information from the reference population is applied to derive prediction equations to estimate the phenotype from marker genotypes.

Genotype imputation is a statistical methodology that could be beneficial to low-income countries where scientific funds are limited by enabling GS (Lashmar *et al.,* 2019). This methodology encompasses the prediction and simulation of absent SNP genotypes from observed or non-missing genotypes using model-based methods (Marchini *et al.,* 2007). Imputation is specific for each population and the accuracy with which genotypes can be imputed is dependent on the continuous presence of LD between individuals in the reference and test populations (Lashmar *et al.,* 2019). Previous research has found that accuracy of imputation usually improves when the reference population is larger (Hayes *et al.,* 2013and Ogawa *et al.,* 2016). However, once a population is selected the effect of genomic parameters on population size and structure should be investigated (Mrode *et al.,* 2019). Imputation could be successful in indigenous South African breeds by creating a low-density SNP chip for the indigenous breeds from which the information could be imputed to a higher density chip (Ogawa *et al.,* 2016). Hence diversity of these breeds could be studied more in-depth for future applications (Lashmar *et al.,* 2019).

There have been many advancements in the application of genomics and has led to an increase in genetic diversity due to genotyping and management of inbreeding (Habier *et al.,* 2007). In 2015 the Beef Genomics Project (BGP) began with the primary aim to enable genomic selection by establishing reference populations for South African beef breeds (van Marle-Köster and Visser, 2018). From published results for the SA Bonsmara breed, traits which typically are low heritability, or difficult to measure traits, showed improvement accuracies of up to 30% (van der Westhuizen *et al.,* 2017). After the BGP ended, the genotyping of key individuals became the responsibly of individual farmers and breeders. Literature has reported a reduction in generation interval has led to an increase in genetic gain. A reason for this is due to animals being able to breed at a younger age due to an increase in accuracy of EBVs (Decker, 2021). Literature indicates that the genetic gain as a result of GS can be doubled per generation with the greatest genetic gain found in traits which are expressed later in an animals' life (Pryce *et al.,* 2010;

Fleming *et al.,* 2018). GS can increase the rate of genetic gain and the diversity levels in populations through accurate selection.

1.2.7 Conclusion

South Africa is home to many breeds of beef cattle which contribute to the total genetic diversity of the beef cattle population. Evaluating genetic diversity is important to understand breed structure and evaluate the genetic resources present in the populations. Genomic technologies, such as SNP, have provided a pathway for a more in-depth analysis of the diversity by evaluating diversity at a genomic level using various methodologies. Genetic diversity estimations along with potential genomic selection may perhaps be useful to maintain the level of genetic diversity found in South African beef cattle populations.

Chapter 2 - Materials and Methods:

2.1 Introduction:

The genotypic data analysed in this study have been generated within the Beef Genomics Program (BGP). The eight respective breed societies provided consent for the use of the genotypes and were provided via the SA Studbook Association. Ethical approval was granted by the University of Pretoria Ethics Committee (NAS194/2020) for use of external data.

2.2 Materials:

In this study different breeds representing local composites (Beefmaster and Bonsmara), indigenous Sanga breeds (Drakensberger, Nguni and Tuli), exotic *Bos taurus* breeds (Charolais and Hereford) and *Bos indicus* Zebu (Boran) were included. All animals in this study were genotyped as part of the BGP. Genotyping was done at the ARC Biotechnology Platform using the GGP 150k SNP array. The number of available genotypes for the different breeds varied from 1634 for Bonsmara to as low as 217 for the Tuli breed.
A minimum of 217 and a maximum of 300 genotypes were selected per breed to ensure a balanced data set for unbiased results. Excel documents were provided by SA Studbook Association containing the animal identification, the parents' identification, inbreeding coefficients as well as the postal code for the different farmers. The postal code was used to identify the different provinces the animals were from. Province was applied to ensure a representative sample of all breeds. Only South African animals were included. Table 2.1 provides a outline of the eight breeds and their available genotypes for the study.

Table 2.1 A summary of the eight breeds and their available genotypes for the study.

Breed	Total number of genotypes available	Number of genotypes selected
Beefmaster	1107	300
Bonsmara	1634	300
Boran	458	232
Charolais	279	279
Drakensberger	1126	300
Hereford	272	268
Nguni	381	294
Tuli	217	217

For the Beefmaster, Bonsmara and Drakensberger breeds for which more than 1000 genotypes were available, 300 animals were selected to represent the different regions of SA where these breeds were farmed with. For these numerically large breeds, no sibs were included in the analysis. For breeds with 300 or less available genotypes, all genotypes were included. For the Boran and Nguni breeds, maximum of three animals per sire with the lowest inbreeding values were selected for analyses.

2.3 Methods:

2.3.1 Quality Control:

Quality control was performed using PLINK (Purcell *et al.,* 2007). The *"- -mind"* command was used to remove individuals with more than 10% (0.1) missing genotypes for each of the eight breeds. Step two used the *"- -geno"* command to remove SNP with an average SNP call rate lower than 95%. Further analysis was done to remove SNP with a MAF of lower than 5% (MAF < 0.05) and that deviated from Hardy-Weinberg equilibrium (HWE < 0.001). Table 2.2 provides a summary of the number of SNP and animals that were included in the study after quality control was performed.

Table 2.2: Summary of the number of SNP and animals available per breed after quality control.

Breed	SNP before quality control	SNP after quality control	Animals remaining after QC
Beefmaster	158322	108732	300
Bonsmara	164956	121320	300
Boran	140113	121422	224
Drakensberger	156023	126466	300
Charolais	156023	129617	223
Hereford	143607	129853	268
Nguni	140113	122572	292
Tuli	143607	125702	203

2.3.2 Population genetic diversity parameters

PLINK (Purcell *et al.,* 2007) was used to calculate the following parameters per breed, the observed heterozygosity, expected heterozygosity, MAF and LD using r^2 to investigate the within population genetic diversity. For the minimum r^2 no restrictions were set; *--ld-window-r2 0* and for inter-SNP distance; *--ld-window-kb 99.99*. Table 2.3 indicates the PLINK commands which were used for analysis as described by Purcell *et al.,* (2007).

Table 2.3: Summary of parameters, PLINK commands and calculation of MAF, observed and expected heterozygosity and LD genetic diversity parameters.

Parameter	PLINK command	Parameter calculation
Observed heterozygosity (H_O) Expected heterozygosity (H_E)	- -het	$HET\ (O\ or\ E) = \frac{N(NM) - HOM\ (O\ or\ E)}{N(NM)}$ H_O and H_E is the observed and expected heterozygosity. N(NM) is the quantity of non-missing genotypes.
LD (r^2)	- -r^2 --ld-window-r2 0 --ld-window-kb 99.99 –make-founders --nonfounders	Mean r^2 values.
MAF	- - freq	Mean MAF values.

2.3.3 Genomic inbreeding and effective population size (N_e)

The command for ROH was - - *homozyg* using PLINK (Purcell *et al.,* 2007) for each population. PLINK makes use of a sliding window of 50 SNP, in one SNP interval across the genome to estimate homozygosity. No more than two probable heterozygous genotypes were permitted using the *--homozyg-window-het 2*, command and no more than five missing genotypes were permitted per window through the use of the *--homozyg-window-missing 5*, command. The minimum SNP density was set to at least one SNP per 50kb. The maximum interval length for two consecutive SNP was set at no more than 1000kb. ROH lengths were broken up into five different length categories, 0 to 3.99 MB, 4 to 7.99 MB, 8 to 11.99 MB, 12 to15.99 MB and >16 MB, under the assumption that there would be many ROH segments for the smaller length categories. The percentage for each length category per breed was analyzed.

The inbreeding coefficient (F_{IS}) and runs of homozygosity (F_{ROH}) were calculated using PLINK (Purcelll *et al.,* 2007). The PLINK command for F_{IS} was *--het* and the value was calculated by using the average of all the F values. The values for F_{ROH} were calculated as described by McQuillan *et al.,* (2008) by dividing the total ROH length by the total length of autosomal chromosomes covered by SNP as described by the formula:

$$F_{ROH} = \frac{\Sigma L_{ROH}}{\Sigma L_{auto}}$$

Where: ΣL_{ROH} = the length of ROH in one individual

ΣL_{auto} = the length of the genome covered by SNP

N_e analysis

N_e was estimated for each population using SNeP version 1.1 as described by Barbato *et al.,* (2015). SNeP estimates the N_e from genome-wide LD (Corbin *et al.* ,2012). SNEP makes use of the following formula for N_e estimation:

$$N_{T(t)} = \frac{1}{(4f\ (c_t))} \frac{1}{(E[r^2_{adj}|c_t])} - \alpha.$$

Where $N_{T(t)}$ is the effective population size estimated t generations ago. c_t defines the recombination rate for a particular physical distance between markers and r^2_{adj} is the LD value adjusted for sample size (Corbin *et al.,* 2012; Barbato *et al.,* 2015). QC files which were originally generated from PLINK were converted into map and ped files using the *-- convert* command in PLINK. The new map and ped files for each breed were used to estimate the effective population size.

2.3.4 Population genetic diversity parameters between biological types

Raw data files were used to select the same 203 to 300 animals per breed as in section 2.3.1 and merged into the three different breed groups without any quality control being performed as shown in table 2.4. The separate breed files were merged using the command *--bmerge* in PLINK (Purcell *et al.,* 2007).

Table 2.4: Summary of the three biological types, with their respective breeds and number of individuals in each biological type.

Biological Type	Breeds	Number of animals
Indigenous (Sanga)	Drakensberger	795
	Nguni	
	Tuli	
Composite	Beefmaster	600
	Bonsmara	
Exotic (*Bos taurus* and Zebu)	Charolais	715
	Hereford	
	Boran	

Once the breeds were merged into groups, quality control was performed using the following procedures: removal of autosomes (--*auto*) and filtering call rates using the commands --*mind 0.05* and --*geno 0.001* were done before MAF and LD and inbreeding analysis.

2.3.4.1 Inter-chromosomal variation

The low-MAF per autosome was estimated using PLINK (Purcell *et al.,* 2007) software and incorporated all SNP with call rates above 95%. The files produced from PLINK were used to generate graphs to display the percentage of low-MAF SNP per autosome and the average MAF per autosome.

LD per chromosome was calculated using the r^2 measure using PLINK software with the command --*r2*. No restrictions were set for the minimum r^2; --*ld-window-r2 0* and for inter-SNP distance; --*ld-window-kb 99.99*. The average r^2 per chromosome was estimated as well as the SNP pairs with an LD of r^2 greater than 0.2 (r^2<0.2), to assess LD for breed groups for downstream genomic applications.

2.3.4.2 Genomic inbreeding for each biological type

Genomic inbreeding using the estimates F_{IS} and F_{ROH} were calculated for each breed group using PLINK (Purcell *et al.,* 2007). The PLINK command for F_{IS} was - -het and the value was calculated by using the average of all the F values. The values for F_{ROH} were calculated as described by McQuillan *et al.,* (2008) by dividing the total ROH length by the total length of autosomal chromosomes covered by SNP as described by the formula:

$$F_{ROH} = \frac{\Sigma L_{ROH}}{\Sigma L_{auto}}$$

Where: ΣL_{ROH} = the length of ROH in one individual

ΣL_{auto} = the length of the genome covered by SNP

2.3.5 Between-breed population diversity parameters

Datasets produced after quality control in section 2.3.1 per breed were merged using the command *--bmerge* in PLINK (Purcell *et al.,* 2007), to perform between-breed analysis and to assess population structure.

2.3.5.1 Principal component analysis

Principal component analysis (PCA) was completed for the combined dataset of the eight cattle breeds. GCTA version 1.24 (Genome-wide Complex Trait Analysis) was used to estimate SNP based genetic relatedness between individuals (Yang *et al.,* 2011). A genetic relationship matrix was calculated using the commands; *--autosome --autosome-num 29 --make-grm* with the GCTA program. PCA was performed using the genetic relationship matrix using the commands; *--grm --pca 2* in the GCTA program. The *eigenvec* folder produced was used to construct the PCA plot in excel.

2.3.5.2 Population structure

The ADMIXTURE 1.23 software (Alexander *et al.,* 2009) was used to determine the population structure of the animals through the maximum likelihood estimation of ancestry to determine the true number of genetic populations (K-values). ADMIXTURE uses cross validation (CV) procedure to estimate the best K-value (Alexander *et al.,* 2009). The preferred K-value exhibits the lowest CV error compared to other K-values. K-values from 2 to 15 were tested to identify the optimal K-value using the *--cv .bed x |tee logx.out* command. Where x indicates the various K-values tested. From this command, a .Q (fam file) and a .P (phenotype file) file were generated. An admixture plot to visually display the admixture was generated using the program Genesis and the .Q file which was previously produced.

A phylogenetic tree was created using the F_{ST} data produced by ADMIXTURE in RStudio using the ape package and cladogram option for the visual display of the phylogenetic tree.

Chapter 3: Results

3.1 Summary of within breed genetic diversity statistics.

A summary of the average MAF values as well as expected and observed heterozygosity values after quality control and LD estimates are presented in table 3.1 using the measure r^2.

Table 3.1: Summary statistics including minor allele frequency (MAF), expected (H_e) and observed (H_o) heterozygosity and LD (r^2) for eight beef cattle breed populations.

Breed	Average MAF	H_e	H_o	Average r^2
BMA	0.317	0.405	0.404	0.165
BON	0.264	0.350	0.350	0.174
BOR	0.210	0.297	0.295	0.159
CHL	0.269	0.362	0.362	0.190
DRB	0.263	0.349	0.343	0.170
HFD	0.265	0.360	0.359	0.249
NGI	0.238	0.320	0.320	0.154
TUL	0.247	0.332	0.335	0.176

BMA: Beefmaster, BON: Bonsmara, BOR: Boran, CHL: Charolais, DRB: Drakensberger, HFD: Hereford, NGI: Nguni, TUL: Tuli

MAF were calculated for all eight populations after quality control. This was done to examine the distribution of the complete set of SNP within the various MAF categories for each population. From this, an average MAF value for each population was obtained as shown in table 3.1. The results indicate that the average MAF values ranged from has high as 0.317 to as low as 0.21 between the eight populations. The Beefmaster population had the highest average MAF value (0.317) and the Boran had the lowest average MAF (0.21) while the other populations had fairly similar MAF values. Figure 3.1 illustrates the percentage of SNP with MAF values ranging from 0.06 to 0.5.

A limited loss of heterozygosity ($H_e > H_o$) was observed for Beefmaster, Drakensberger and Hereford populations. A slight gain in diversity ($H_o > H_e$) can be seen in the Boran and Tuli populations. The Boran population had the lowest heterozygosity value (0.297), while the highest value was observed in the Beefmaster population (0.405). Results for LD between populations ranged between 0.154 and 0.190 for all populations except the Hereford (r^2=0.249) population.

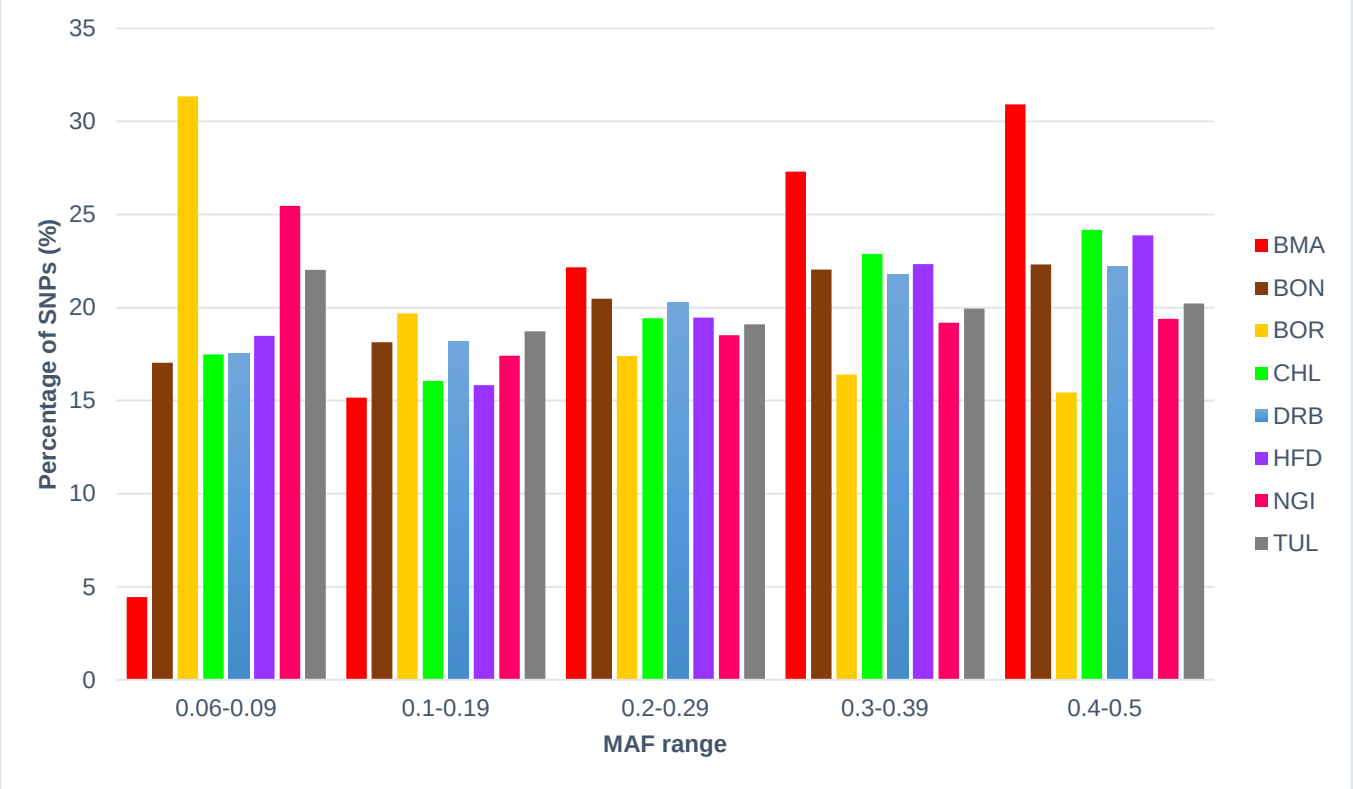

Figure 3.1 Bar graph indicating the percentage of SNP for MAF values in categories from 0.06 to 0.5 for eight beef cattle populations.

Figure 3.1 indicated that most of the populations had a higher percentage of SNP in the MAF categories 0.2 to 0.5. The Beefmaster population showed a low percentage of SNP in the range 0.06-0.09 and an increase in percentage of SNP as the MAF range value increased. This was the opposite of the Boran population, who had a decrease in percentage of SNP over the increasing MAF categories.

3.2 Population genetic diversity parameters between biological types

Table 3.2 summarizes the MAF, expected and observed heterozygosity estimates for each biological type after quality control took place. The table indicates that the indigenous biological type had the lowest average MAF (0.261) while the exotic biological type had the highest average MAF (0.300). There was little variation between breed groups for heterozygosity estimates. The heterozygosity estimates showed a limited loss of heterozygosity ($H_e > H_o$) for all breed groups.

Table 3.2. Summary of the average MAF, H_e (expected heterozygosity) H_o (observed heterozygosity for 3 biological types.

Biological type	MAF	H_e	H_o
Indigenous	0.261	0.347	0.330
Composite	0.289	0.384	0.375
Exotic	0.300	0.388	0.331

Table 3.3 provides a summary of the specific autosome per breed group on which the lowest and highest MAF values were observed respectively.

Table 3.3. Summary of the specific autosome indicating the lowest and highest MAF estimates for each biological type.

Biological type	Lowest MAF average		Highest MAF average	
	Autosome	MAF	Autosome	MAF
Indigenous	BTA14	0.252	BTA18	0.276
Composite	BTA14	0.270	BTA23	0.300
Exotic	BTA14	0.277	BTA23	0.312

Further analysis was performed to estimate the variation of MAF per autosome when the SNP with low MAF (<5%) were included in analysis. The estimation of the percentage of low-MAF SNP per autosome was also conducted and can be seen in figures 3.2, 3.3 and 3.4 for each biological type respectively.

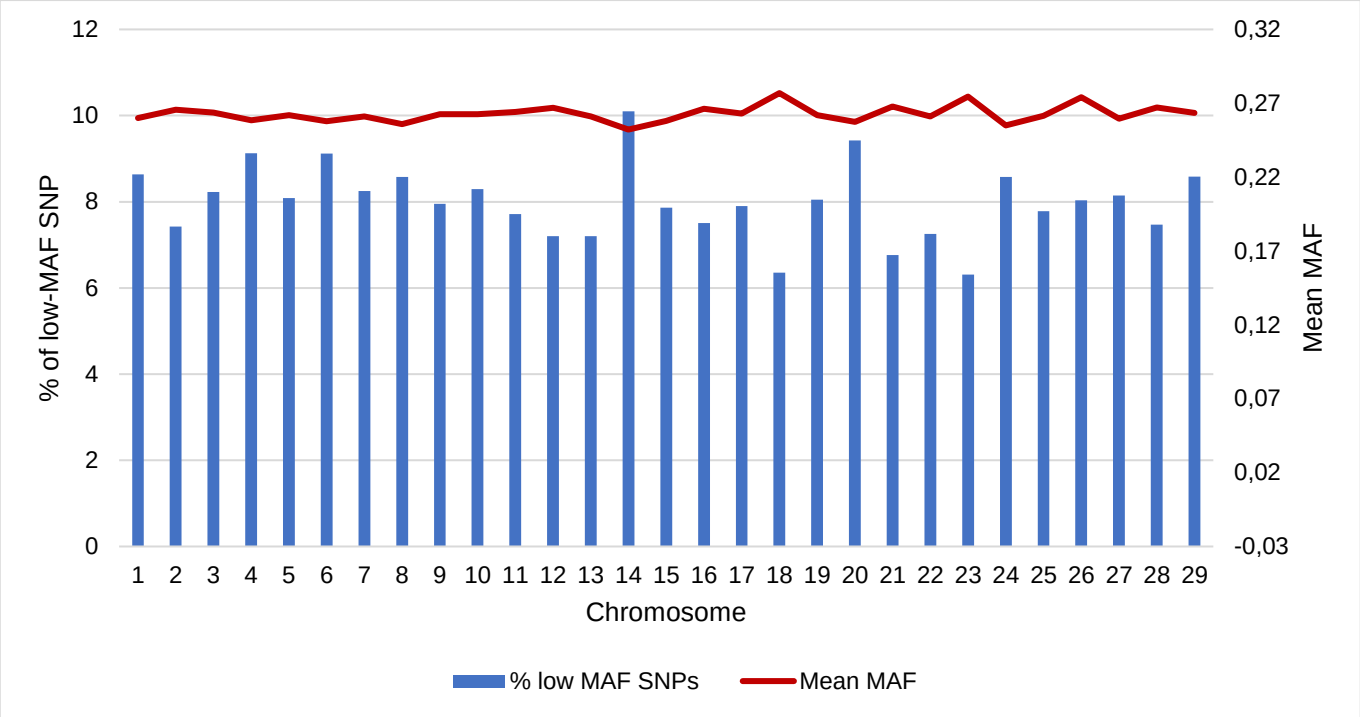

Figure 3.2 Variation in percentage of low-MAF SNP and average low-MAF for each autosome for the indigenous biological type.

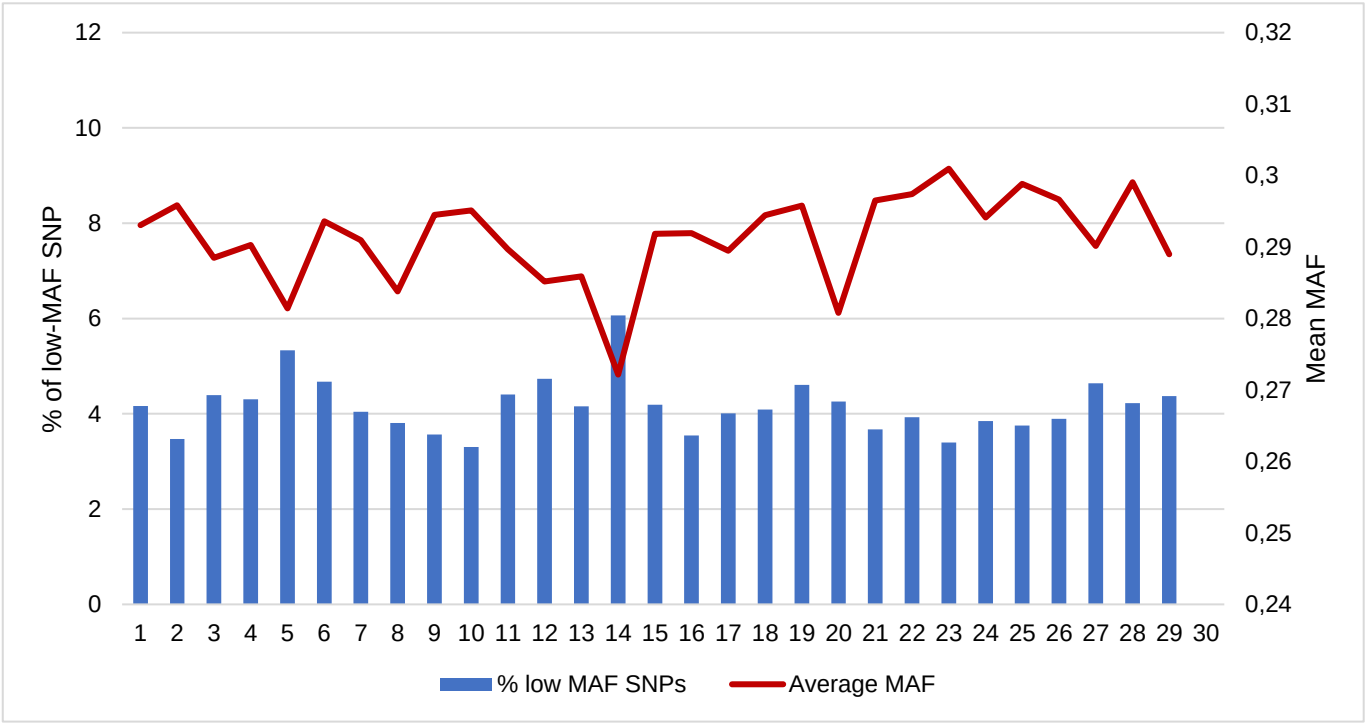

Figure 3.3 Variation in percentage of low-MAF SNP and average of low-MAF for each autosome in the composite biological type.

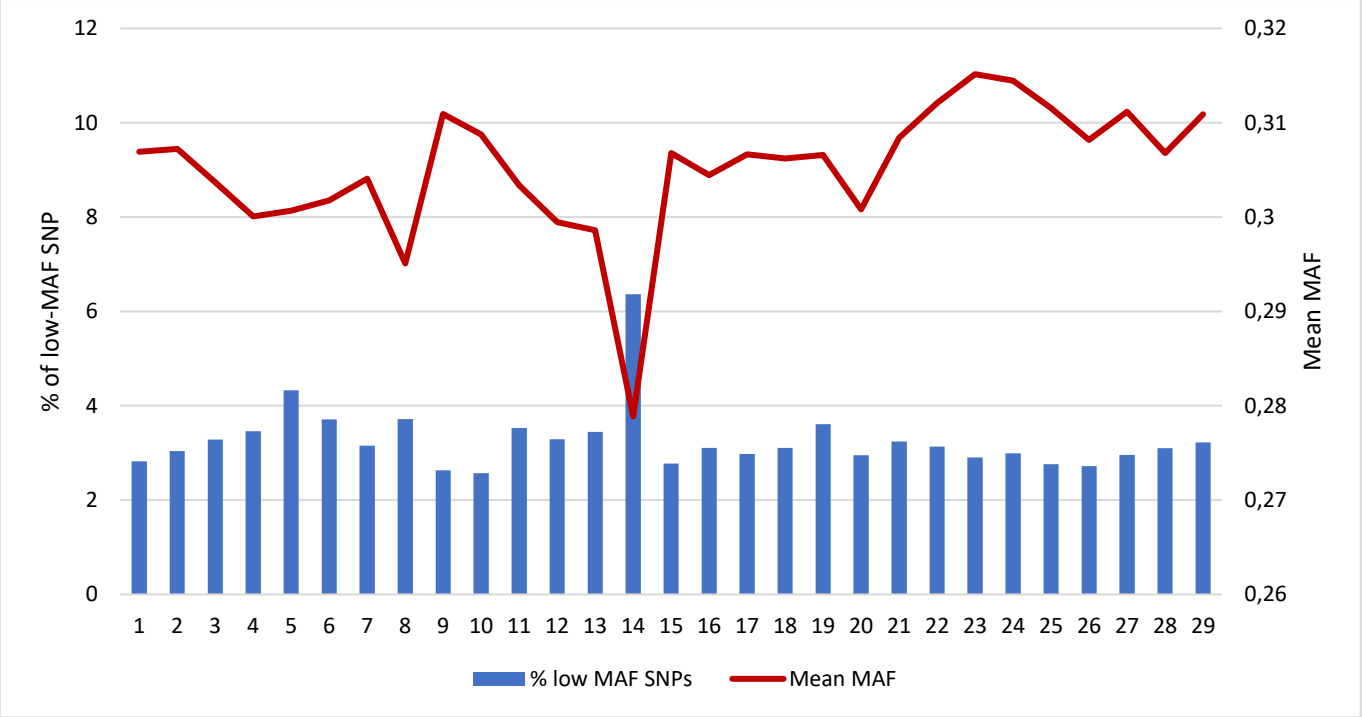

Figure 3.4 Variation in percentage of low-MAF SNP and average low-MAF for each autosome for the exotic biological type.

All figures indicate that the highest percentage of SNP exhibiting low MAF (<5%) was observed on BTA14 for all three biological types. This is consistent with table 3.1 when the SNP with low MAF were removed during quality control and the average MAF per autosome was estimated. The figures also indicated that the exotic biological type had the lowest average low-MAF per autosome while the indigenous biological type had the highest.

Further within group population diversity parameters was calculated for linkage disequilibrium using the parameter r^2. Table 3.4 summarizes the average r^2 estimate for each autosome and the autosome with the lowest and highest r^2 estimates for each biological type after quality control was completed.

Table 3.4 Summary of the average r^2 estimate and specific autosome indicating the lowest and highest MAF estimates for each biological type.

Biological type	Average r^2	Lowest r^2 estimate		Highest r^2 estimate	
		Autosome	r^2	Autosome	r^2
Indigenous	0.153	BTA28	0.087	BTA14	0.152
Composite	0.169	BTA28	0.089	BTA14	0.155
Exotic	0.181	BTA28	0.104	BTA14	0.173

Table 3.4 indicates that there was limited variation between all biological types for average r^2 values (0.153-0.181). It can also be seen in the table that all three biological types followed

the same trend with the highest and lowest estimates. BTA18 which had the lowest LD estimate, has a size of 66.35 Mb with 66346785 base pairs (NCBI, Genome assemble and Annotation report, 2021) and had an average of 2254.3 SNP on the autosome across all biological types. BTA14 had the highest LD estimate with a length of 85.01 Mb with 85007120 base pairs and an average of 5068 SNP on the autosome across all biological types. Figures 3.5, 3.6 and 3.7 indicate the proportion of SNP pairs which had an r^2 estimate of larger than 0.2 (r^2>0.2) for each autosome and the average r^2 estimate for each autosome per biological type.

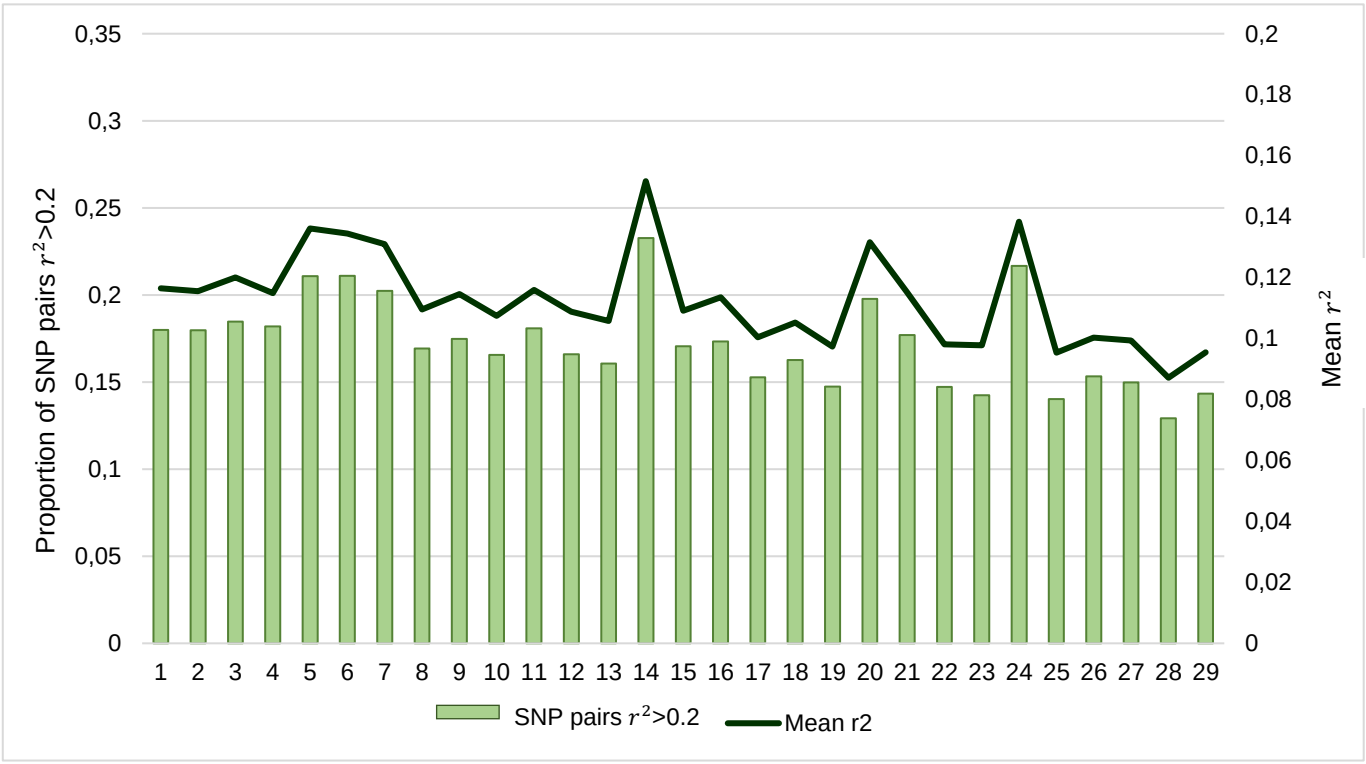

Figure 3.5 Variation in proportion of SNP pairs where r^2 is greater than 0.2 and the average r^2 estimate per autosome for the indigenous biological type.

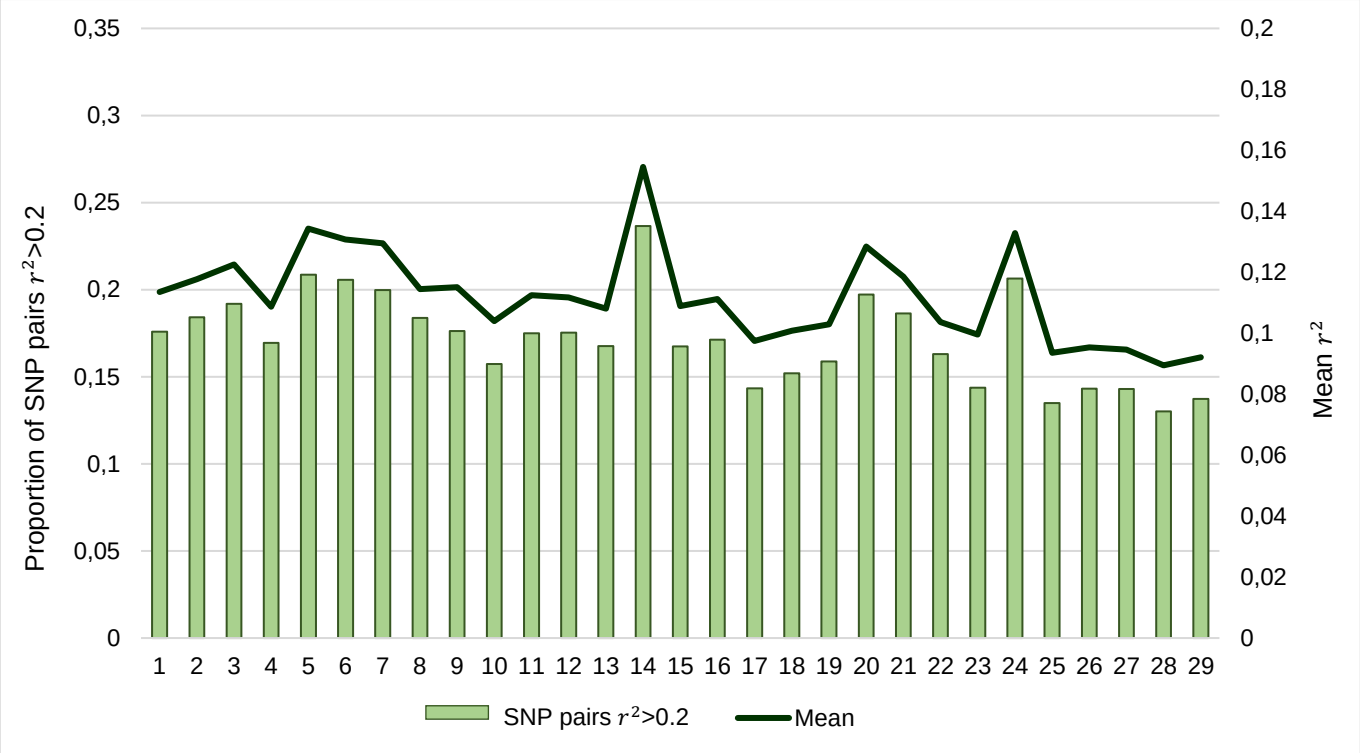

Figure 3.6 Variation in proportion of SNP pairs where r^2 is greater than 0.2 and the average r^2 estimate per autosome for the composite biological type.

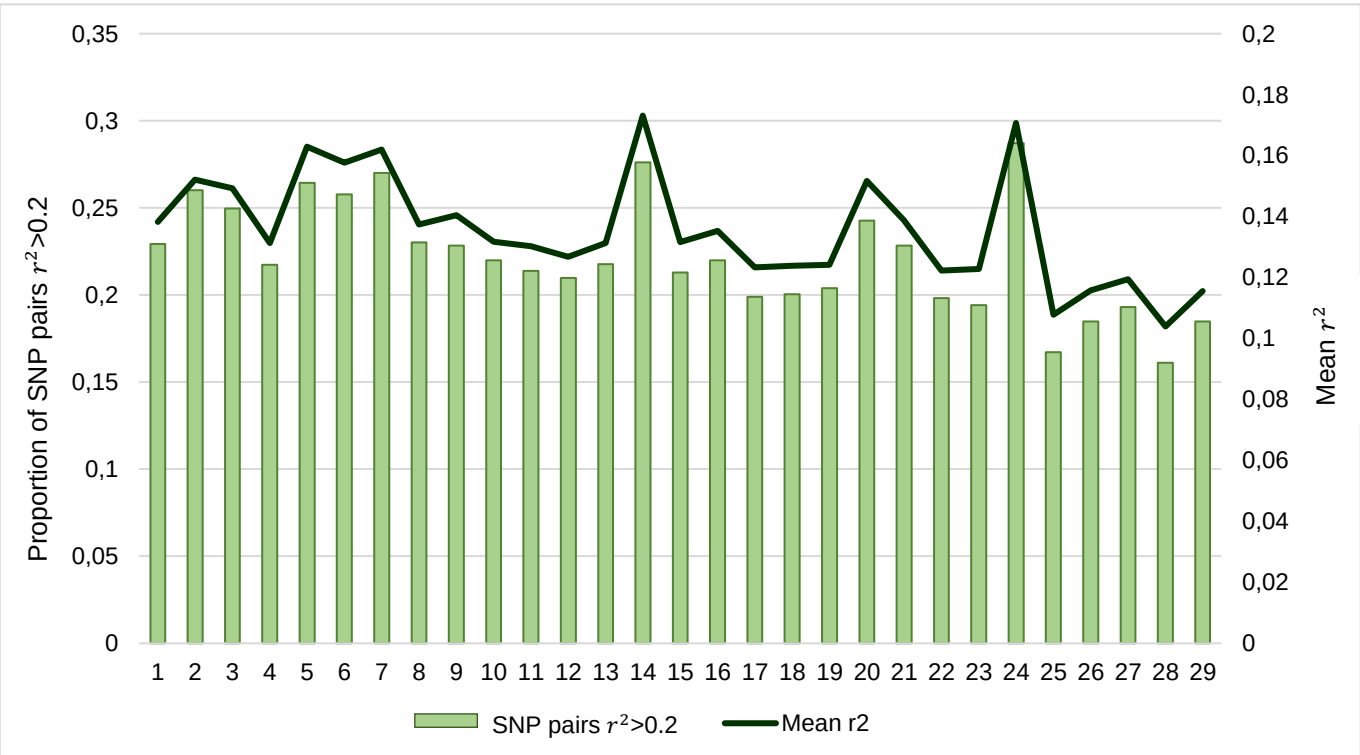

Figure 3.7 Variation in proportion of SNP pairs where r^2 is greater than 0.2 and the average r^2 estimate per autosome for the exotic biological type.

There is a general trend for all breed groups that when the mean r^2 value of an autosome increases or decreases, the proportion of SNP pairs with r^2 > 0.2 also increases. BTA14 had the lowest MAF value for all breeds and the highest mean LD for all breeds. By comparing the low MAF graphs and LD graphs the proportion of SNP pairs with r^2 >0.2 increased when there was a higher level of low-MAF SNP present for each autosome.

3.3 Genomic inbreeding and effective population size

3.3.1 Genomic inbreeding

Table 3.5 shows a summary of the average inbreeding coefficients, F_{IS} and F_{ROH} for the eight respective populations including the inbreeding coefficients for the greatest and least inbred individuals in each population.

Table 3.5 Average inbreeding coefficients (F_{IS}, F_{ROH}) for eight beef populations as well as the most and least inbred individual for each breed.

Breed	**Average F_{IS}**	**F_{IS} Least inbred individual**	**F_{IS} Most inbred individual**	**AverageF_{ROH}**
BMA	0.001	-0.024	0.114	0.005
BON	0.000	-0.129	0.135	0.002
BOR	-0.007	-0.176	0.148	0.001
CHL	-0.002	-0.091	0.229	0.002
DRB	0.016	-0.082	0.218	0.003
HFD	0.002	-0.131	0.192	0.002
NGI	-0.001	-0.197	0.160	0.002
TUL	-0.009	-0.142	0.203	0.003

Table 3.5 indicates there were very small differences between the F_{IS} inbreeding coefficient estimates between all the breeds ranging from -0.001 (NGI) to 0.016 (DRB). The low variation presented in the F_{IS} results were supported by the F_{ROH} values, where F_{ROH} values ranged from 0.001 (BOR) to 0.005 (BMA). The lowest inbred individual belonged to the Nguni breed (-0.197). The most inbred individual belongs to the Charolais breed (0.229) with the Drakensberger and Tuli populations having individuals with inbreeding coefficients above 0.2 as well. The negative F_{IS} values indicate an excess of observed heterozygotes present in the population. Negative F_{IS} values are possible to obtain if there is a sight with poor mapping or part of the population is not in HWE.

ROH was calculated to distinguish the degree of recent versus ancient inbreeding. The average ROH within each length classification was calculated for each population and displayed in percentage format, as shown in figure 3.8.

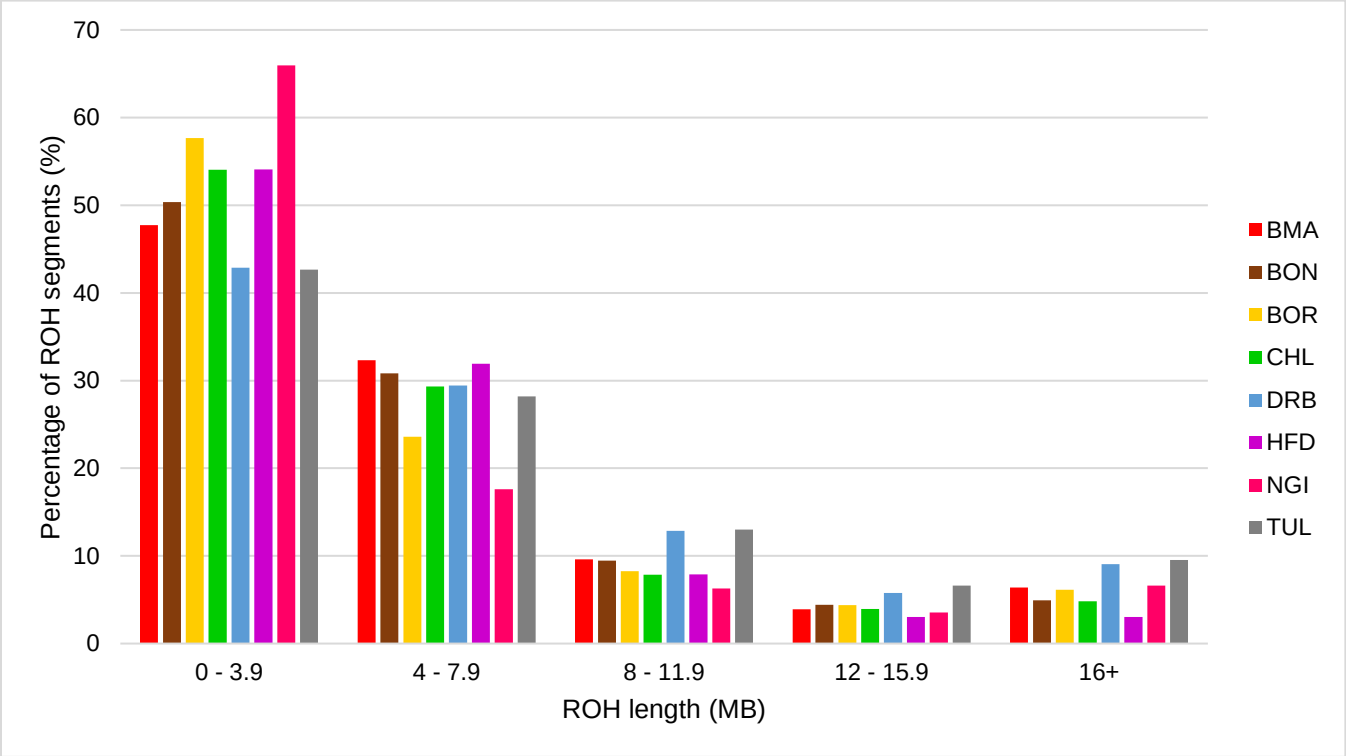

Figure 3.8 Percentage of ROH segments for different length segments for eight beef cattle populations.

The figure clearly shows the majority of ROH segments were short segments (0 to 3.9MB) for all eight populations. The Nguni displayed the largest number of short segments while the Drakensberger and Tuli had the lowest number of large segments. There was a sharp decline in percentage of ROH segments for the segment's lengths 4-15.99MB for all the populations with a slight increase in percentage for the 16+ MB length category.

3.3.2 Genomic inbreeding between biological types

Genomic inbreeding using the inbreeding coefficients, F_{IS} and F_{ROH}, were estimated for the three biological types with the results are presented in table 3.6.

Table 3.6 Genomic inbreeding estimates (F_{IS} *and* F_{ROH}) for three biological types

Biological type	F_{IS}	F_{ROH}
Indigenous	0.002	0.002
Composite	0.001	0.002
Exotic	-0.003	0.001

The exotic biological type had the lowest inbreeding estimate (-0.003) while the indigenous biological type had the highest inbreeding estimate (0.002) which is displayed in table 3.6. The difference between F_{ROH} estimates is very small, which does agree with the F_{IS} results presented in the table.

3.3.3 Effective population size

Figure 3.9 illustrates the effective population size (N_e) for the eight respective beef cattle populations. N_e was plotted for all populations from 995 to roughly 13 generations ago.

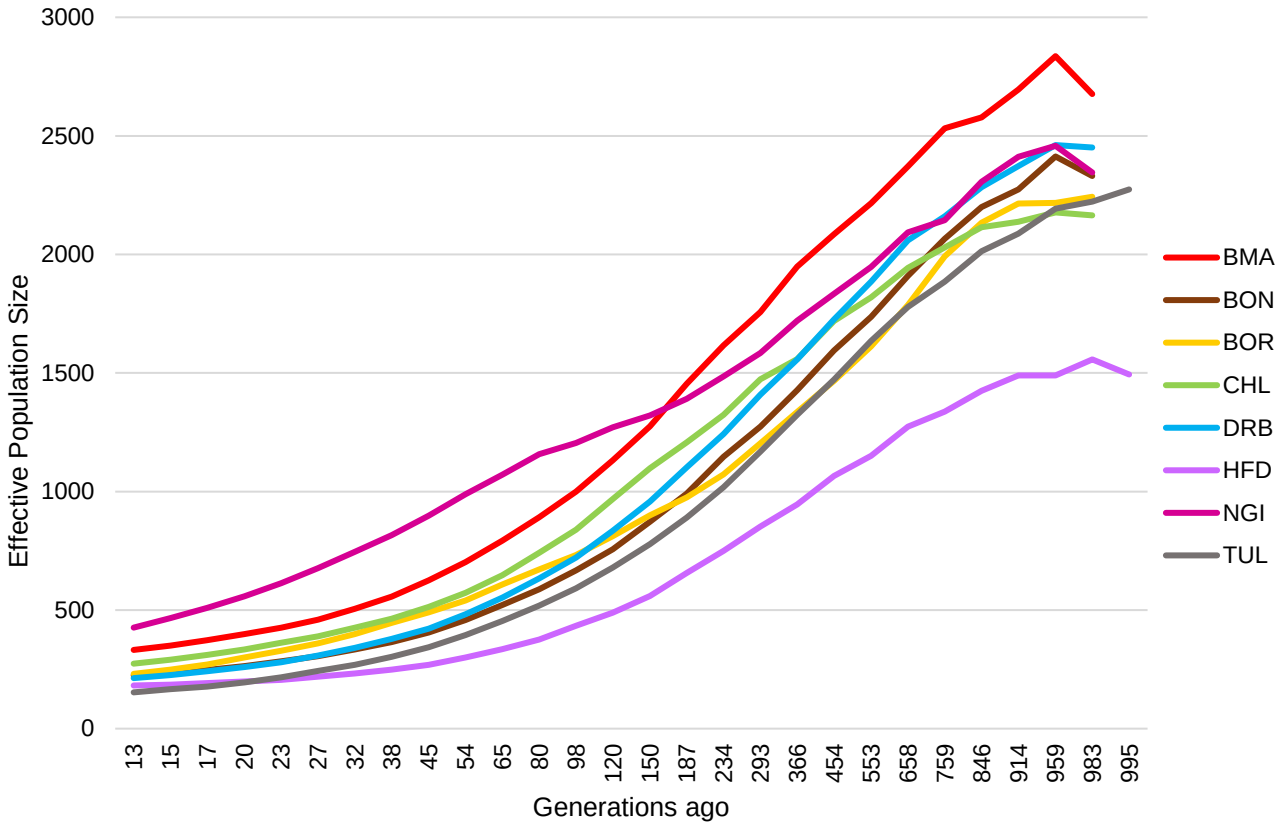

Figure 3.9 Trends in historic effective population size (Ne) for eight beef cattle populations.

A general downward trend in N_e could be noticed for all eight populations, indicating an overall decrease in genetic diversity. The Beefmaster population displayed the largest decline N_e from 2676 animals 983 generations ago to 332 animals 13 generations ago. The Tuli population had the smallest N_e, with 153 animals, for all eight populations followed by the Hereford population with an N_e of 184. The Nguni population had the largest effective population with 426 animals.

3.4. Between population genetic diversity

3.4.1 Principal component analysis

The genetic relatedness between the eight different cattle breeds was assessed using a principal component analysis (PCA). The first (PCA1) and second (PCA2) were plotted against each other in figure 3.10.

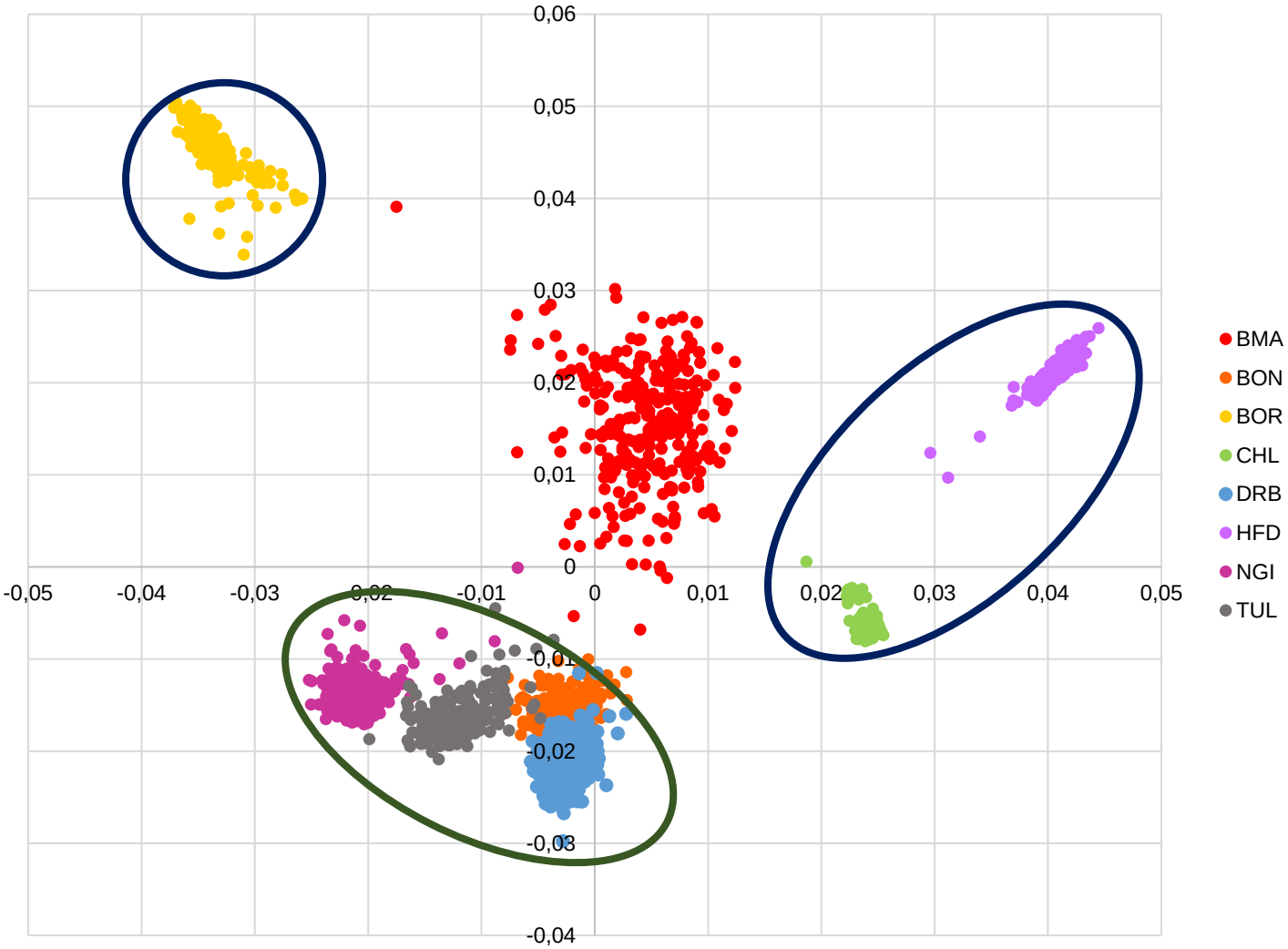

Figure 3.10 The genetic relatedness between eight beef cattle populations as seen when plotting the first (PCA1) and second (PCA2) principal components.

Figure 3.10 firstly illustrates that the individuals from each population cluster together within their respective populations. The Beefmaster population produced a loose cluster with some outliers being presented. In all the other populations, a few outliers can be seen. The three indigenous populations (DRB, NGI and TUL) clustered close together as a group as shown by the dark green oval on the graph. The exotic *Bos taurus* breeds (CHL and HFD), and the Zebu

(BOR) did not cluster as a distinct group, but rather each breed formed their own cluster spread out from each other as shown by the dark blue circles on the graph. The Bonsmara breed, a composite breed, (BON) is clustered very closely to the DRB and TUL breeds.

3.4.2 Population structure analysis

Cross-validation (CV) scores for K-values 2 to 16 were plotted on a line graph to obtain an inflection point. The inflection point was used to choose the most suitable K-value for population structure. A K-value of eight was used to construct a population structure plot. A K-value of eight was chosen as it was the first inflection point in the CV plot (0.544).

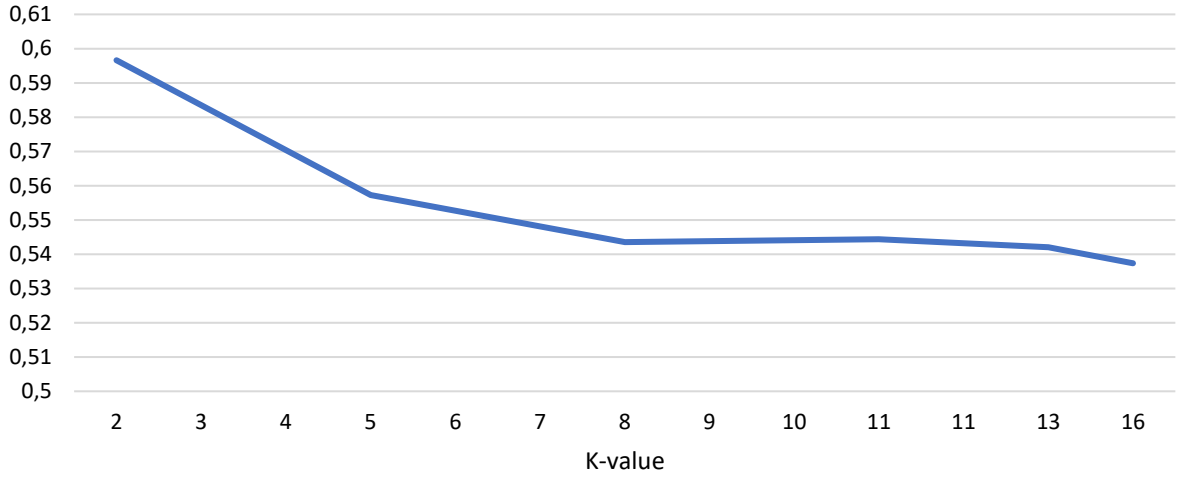

Figure 3.11 Cross validation error graph displaying the most appropriate K-value with the first inflection point present.

Figure 3.12 illustrated that the eight beef cattle breeds had their own distinct ancestral populations, even though some admixture could be seen. This agrees with the results from the PCA plot in figure 3.10. The Beefmaster population showed the highest admixture with all other breeds. TUL, NGI and DRB, the indigenous breeds, exhibit some admixture with each other. The BOR, HFD and CHL show the least admixture.

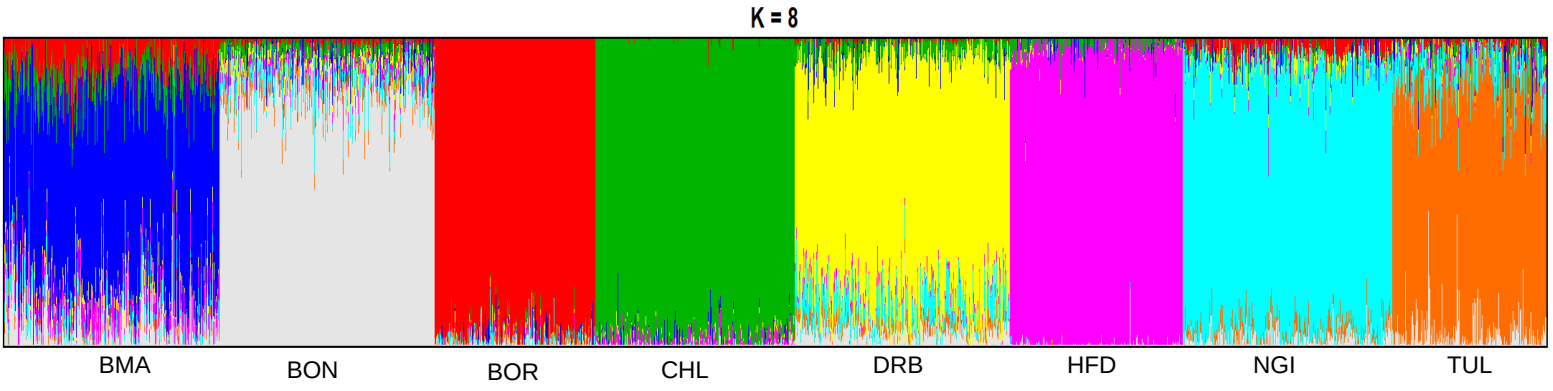

Figure 3.12 Population structure plot (K=8) of the different beef populations.

The phylogenetic tree in figure 3.13 illustrated the ancestral relationship between the eight beef cattle populations.

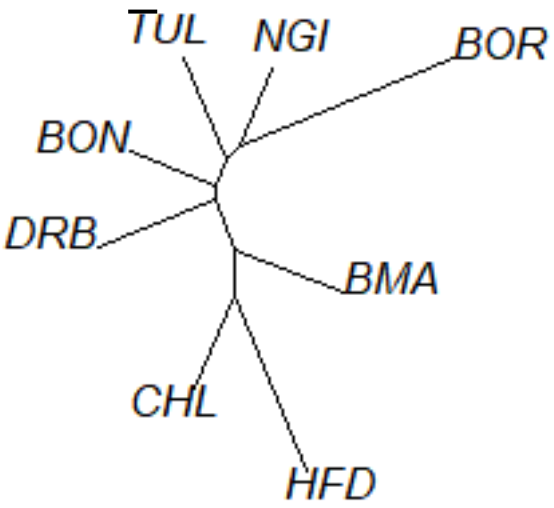

Figure 3.13 A phylogenetic tree representing the ancestral relationship between the eight beef cattle populations.

The phylogenetic tree illustrates that the exotic breeds (CHL and HFD) originated from a common ancestor. The Bonsmara, Tuli, Nguni and Boran all share common ancestry.

Chapter 4: Discussion

4.1 Introduction

Genetic diversity studies using genomic SNP information of livestock are growing and important for the monitoring and maintenance of diversity in South African breeds. Evaluation of diversity contributes to understanding the breed structure and to ensure there is no loss in diversity of populations. This is important for future farming and sustainable breeding practices. Cattle farming is a fundamental source of income and animal protein in South Africa (DAFF, 2017). Therefore, it is vital to assess the genetic diversity of the cattle breeds to ensure current farm practices are not hindering the present or future diversity.

It has been previously established that in South Africa, various cattle breeds have an international origin and have made their way to South Africa via migration (Verdugo et al., 2019), in addition to the well-adapted indigenous breeds (Zwane et al., 2019). With such a variety of indigenous, local composites and exotic cattle breeds, many genotypes are available for selection and genetic improvement (Abin et al., 2016). It has therefore been of vital importance that genetic parameters such as MAF, heterozygosity, LD, ROH, genomic inbreeding and effective population size along with breed structure are investigated to improve or maintain the genetic diversity of South African cattle breeds. The aim of the current study was to assess SNP based genomic diversity and population structure of eight SA beef cattle breeds.

4.2 Within population genetic diversity between breeds

The eight populations included in this study all had a moderate to high level of genetic diversity based upon MAF, as well as observed and expected heterozygosity estimates. This also indicates that the breeds are not at risk of a loss in genetic diversity with the current farming practices that are in place in SA.

MAF is often used to distinguish between common and rare variants in a population (Engelsma *et al.,* 2014). Lower MAF estimates for *Bos indicus* populations or African indigenous populations have often been accredited to ascertainment bias due to the earlier developed low-density arrays that mainly incorporated *Bos taurus* breeds (McKay *et al.,* 2008). This could be seen in a study by Qwabe *et al*. (2013) where MAF estimates for the NGI population was 0.17 using the 50K array compared to this study where the MAF estimate for NGI was 0.238 using the GGP 150K array. In the current study the African indigenous populations had similar MAF estimates to the *Bos taurus* populations, although the *Bos taurus* estimates were slightly higher (DRB = 0.263 vs. HFD = 0.265). Therefore, any variation in MAF values are probably not be due to ascertainment bias but rather to the variation of genetic

diversity between the populations which could have been caused by selection practices (O'Brien *et al.,* 2014). Out of the three indigenous types, the Drakensberger population had the highest MAF (DRB = 0.263, NGI = 0.238, TUL = 0.247) estimate which could be due to the Drakensberger sharing ancestry with taurine breeds (Makina *et al.,* 2014). The highest average MAF was for the Beefmaster breed in this study (0.317) consistent with history and development of the breed in SA. The Beefmaster was established in the late 1980's in South Africa deeming the breed to be fairly "young" in the country (Beefmaster SA, 2021). The Boran population (BOR) had the lowest MAF average (0.21) which could be an indication of a higher proportion of alleles which could be fixed in a population and a small effective population size. The MAF averages for the remaining populations (BON, CHL, DRB, HFD, NGI and TUL) were consistent with other studies on SA populations, indicating variation may not be due to ascertainment bias (Zwane *et al.,* 2016).

The Beefmaster population also had high heterozygosity levels of 0.405. The lower diversity levels for the BOR population seen in the low MAF estimate were supported by the lowest H_e value of 0.297. This is estimate is lower compared to previous studies performed by Edea *et al.,* (2018) and Msalya *et al.,* (2017) whose estimates for H_e for the BOR populations were both 0.4. This may be because of smaller sample sizes used in the previous studies as well as the different SNP arrays used for the previous studies (GGP 80K indicine). However, Msalya *et al.* (2017) also indicated that the BOR population had the lowest H_e estimate of all the populations included. This is consistent with other studies indicating lower heterozygosity estimates for *Bos indicus* populations (Lin *et al.,* 2010, Edea *et al.,* 2013, Makina *et al.,* 2014). This corresponds to the history of the BOR breed in SA. BOR embryos were imported into South Africa from Kenya where there was a narrow gene pool and possible inbreeding causing a decrease in genetic diversity (Ajmone-Marsan *et al.,* 2010). Heterozygosity is an important parameter and one of the most extensively used in the analysis of genetic diversity. A shortage of heterozygous genotypes can indicate allele fixation which could have resulted from genetic drift, bottle necks, mutation, migration or selection (Chan *et al.,* 2010; Purfield *et al.,* 2019) When specific traits are selected for, the frequency of the favourable allele increases until the frequency of the allele becomes close to 1 or 1, resulting in allele fixation in the population (Makina *et al.,* 2016). Small populations are particularly at risk of allele fixation, or a loss in heterozygosity when strong selection pressures are put in place (Purfield *et al.,* 2019). This further hinders the populations' ability to adapt to any change in climate or environment (O'Brien *et al.,* 2014). In the current study the H_o for the TUL and BOR populations was slightly higher than the H_e, this could be due to small population sizes. It is clear that indigenous types should be carefully managed in order to conserve their diversity in SA.

LD aids as a predictor of the density of the SNP array required to produce accurate GEBVs. LD furthermore serve as a tool for assessing the possible achievable imputation accuracy of specific genomic regions or breeds (Bohmanova *et al.,* 2010). The parameter that was used to estimate LD was r^2, which is dependent on allele frequencies (O'Brien *et al.,* 2014). It can also be noted that varying sample sizes, LD measures, marker densities and population demographics makes it challenging to compare LD levels between different studies. Results from this study were comparable to LD results reported by Lashmar *et al*. (2018) and Zwane *et al.* (2016) for DRB, HFD and NGI populations. In the current study, the Hereford population was a clear outlier and had the highest LD estimate of all populations. This high LD could be a consequence of the first SNP arrays being developed from a Hereford cow causing most SNP to be in high LD with each other (Boichard *et al.,* 2012).

4.3 Population genetic diversity parameters between biological types

The eight beef cattle populations analyzed in this study were grouped into indigenous, composite and exotic populations to study the diversity of the various biological types. Heterozygosity estimates were very similar between the biological types, and all displayed moderate levels of diversity. These results coincide with those reported by Gebrehiwot *et al.,* (2020) who specified relatively high and similar heterozygosity estimates for European *Bos taurus* (exotic), African *Bos taurus* (indigenous, Tuli) and *Bos indicus* (Boran) indicating high levels of genetic diversity within the biological types. However, Lin *et al.,* (2010) reported higher genetic variability for *Bos taurus* populations compared to *Bos indicus* populations. This inconsistency could be due to ascertainment bias from the SNP array used by Lin *et al.,* (2010) as the earlier developed SNP arrays were based on *Bos taurus* genotypes. There has been considerable improvement and development of higher density SNP arrays in the last decade which also include indicine genotype content (Edea *et al.,* 2013).

In the current study it was found that the indigenous biological types had an increased number of low-MAF SNP compared to the remaining two types. SNP with low-MAF (MAF < 0.05) values were included in analyses for the three biological types. Low-MAF SNP are usually excluded from genomic analysis in quality control, as they are associated with genotyping errors and influence LD estimations (Calus *et al.,* 2014). However, SNP with a low-MAF value may be an indication of fixated SNP that can be important for future use and should be considered in GWAS or imputation (Tabangin *et al.,* 2009; Malomane *et al.,* 2018). SNP arrays tend to lack inclusion of rare variants and have a tendency to be biased towards variants found in the population which were used for array development (Geibel *et al.,* 2021). It could be for this reason that the indigenous biological type had more low-MAF SNP than the other biological types. However, in this study, it is rather likely that the low-MAF SNP are SNP which

could be rare alleles specific to the breed type and of importance for the population (Engelsma *et al.,* 2014), for instance SNP associated with adaptability traits. Indigenous populations are known for their adaptive traits such as disease or heat resistance (Nyamushamba *et al.,* 2017). If loci associated with these genes were not included on the SNP arrays when complied, a lower MAF value for the indigenous biological type would be expected (Geibel *et al.,* 2021).

When the idea of GS was put forward, Meuwissen *et al.* (2001) suggested that GS could be applied to a breed with an average r^2 value of 0.2 or greater. It was also stated that the prediction accuracy of GS may increase by up to 85% when the average LD between SNP pairs is greater than 0.2 (Fan *et al.,* 2010). In this study, assessing the proportion of SNP pairs with an r^2>0.2 provides an indication if the reference populations for SA breeds and biological types are sufficient to enable GS and other genomic applications. The exotic breed group had an average r^2 closest to 0.2 with the highest proportion of SNP pairs with an LD of r^2>0.2 for each chromosome. This implies that the exotic population can apply GS as most QTLs will be in LD with at least one SNP. The indigenous population had the lowest average r^2 (0.153) and the lowest proportion of SNP pairs with an LD of r^2>0.2. However, it could still be suggested that genomic applications, for instance, imputation or GS can be investigated due to the small effective population sizes of these populationss (Brito *et al.,* 2011; Makina *et al.,* 2016). The availability of high-density SNP arrays has favoured the analysis of LD patterns across the genome (Eusebi *et al.,* 2020).

Imputation is a methodology that deduces missing genotypic information based on shared alleles or sections of the genome within a population using a representative sample of a population or breed (Berry *et al.,* 2014). The imputation of absent genotypic data relies on the fact that lengthier haplotypes are shared over short expanses between individuals which share a common ancestor (Antolin *et al.,* 2017). Imputation is also population specific, with a large part of imputation accuracy being dependent on the continuity of LD between animals in the reference population. In order for imputation to be accurate in populations, the LD estimates need to be moderate to high. Furthermore, the effective population size provides an indication of LD and the recombination distances between SNP (Barbato *et al.,* 2015). The smaller N_e tends to display higher within population LD, indicating that the populations share larger haplotypes. This relationship between N_e and LD results in SNP being more accurately imputed. The indigenous biological type had a moderate LD estimate (0.153) as well as two indigenous populations having lower N_e (TUL = 153, DRB = 213) and therefore the populations could be considered for imputation. The benefit of imputation for the indigenous biological types would be enabling cost effective low-density genotyping, thus creating a large reference population for GS at a lower cost (Lashmar *et al.,* 2019).

The 35K GGP chip was especially designed for *Bos indicus* cattle (Ferraz *et al.,* 2020). This array would allow for more SA indigenous animals to be genotyped at a low density for investigation of genetic diversity. The genomic information could be imputed to higher density chips such as the 150K GGP chip (Berry *et al.,* 2014). This imputed information could potentially allow for more SA breeds to undergo GS. Although imputation provides promising results for GS (Berry *et al.,* 2014; Antolin *et al.,* 2017), previous studies have mentioned the possible inaccuracy of imputation when low-MAF SNP are imputed. This is a consequence of low frequency alleles not often been represented in the reference haplotypes (Schrooten *et al.,* 2014; Lashmar *et al.,* 2019). Thus caution could be taken when indigenous populations undergo imputation as these populations tend to have higher frequencies of low-MAF SNP as seen in this study.

4.4 Genomic inbreeding and effective population size

The inbreeding estimates produced in this study for each population and biological type did not deviate from zero significantly, proving that although there has been significant selection there is no immediate risk for decline in genetic diversity (Mastrangelo *et al.,* 2016). Genomic inbreeding specifies the probability that alleles chosen at random on the genome are identical by decent (IBD). High levels of homozygosity can be due to selection and breeding for favorable alleles in a population as well as natural phenomena such as bottlenecks and genetic drift. In the current study, the inbreeding coefficient was firstly assessed using F_{IS} which estimates the genomic inbreeding value. The average F_{IS} estimates for four populations (BOR, CHL, NGI, TUL) were negative while the other four populations (BMA, BON, DRB, HFD) had very low positive inbreeding estimates. These low inbreeding results could be an indication of successful on farm management against inbreeding along with the usage of local and international bulls, e.g., the SA Hereford breed makes use of semen from bulls from the USA to ensure there is no loss in genetic diversity (SA Hereford Cattle Society, 2021). The inbreeding estimates from this study are comparable to that of Makina *et al.,* (2014) who indicated that F_{IS} estimates for BON, DRB and NGI were very low (-0.017 to 0.005). Similar results were reported by Zwane *et al.,* (2016) for DRB, HFD and NGI breeds using genomic data. Makina *et al.,* (2014) suggested inbreeding should be assessed every five years to ensure that appropriate steps for breeding management and selection strategies are taken to prevent a loss in diversity. Therefore, the results from this study contribute towards the continuous assessment of diversity in SA cattle populations.

There was an abundance of short ROH segments compared to long ROH segments for all eight populations which indicates that ancient inbreeding occurred for all populations. This is consistent with results estimated by Lashmar *et al.,* (2018) for the DRB breed, where 35.7% of ROH segments were indicative of ancient inbreeding. This ancient inbreeding could

be due to breed establishment, where specific traits, alleles and individuals were favoured above others for a breed to become distinct, thus creating a founder effect (Purfield *et al.,* 2012). ROH results generated in this study are indicative of a founder effect or breed establishment due to the abundance of short ROH segments (0.0 to 3.9 Mb). The lack of recent inbreeding could be due to the desire to maintain a sustainable level of genetic diversity within populations (Bosse *et al.,* 2012; Peripolli *et al.,* 2018).

ROH length is an important characteristic for inbreeding estimation and may be applied to deduce population history (McQuillan *et al.,* 2008). The density of the SNP array to detect ROH plays a significant role in autozygosity estimates (Singer-Hasler *et al.,* 2017). Peripolli *et al.,* (2018) reported that the SNP array used to produce data for ROH analysis, as well as the frequency of genotyping errors influences the ROH identified in cattle. It has therefore been suggested to ensure prudent and critical interpretation when other ROH estimates are compared as the above factors contribute to ROH-based estimate accuracy.

The second inbreeding coefficient estimated was F_{ROH}. F_{ROH} more accurately depicts the current autozygotic proportion of the genome because of common ancestry. F_{ROH} estimates in this study were very low negative, indicating low inbreeding (Purfield *et al.,* 2012). Both inbreeding coefficient estimates indicated limited inbreeding within all eight populations. This agrees with the limited number of short ROH segments described previously. These results further indicate that although SA cattle breeds have been placed under significant selection limited inbreeding occurred maintaining a sufficient level of genetic diversity in the population.

The N_e for eight SA beef cattle populations in this study decreased substantially from 995 to 13 generations ago, which could indicate a loss of genetic diversity. The results in this study are comparable to the results estimated by Abin *et al.,* (2016) who also indicated a decrease in N_e over generations, using pedigree data for BOR, DRB, NGI and TUL populations. N_e estimates by Gebrehiowt *et al.,* (2020) indicated that *Bos indicus* and African indigenous populations showed a higher N_e over 1000 generations ago compared to *Bos taurus* populations. The current study supports this where the HFD population is at an N_e of 1557 compared to the DRB population of 2451 over nine hundred generations ago. This could be due to *Bos taurus* breeds being domesticated earlier than *Bos indicus* and African indigenous breeds (Gebrehiwot *et al.,* 2020). N_e estimates indicate the amount of genetic drift in a population and ensures sufficient diversity within the breed in order to enhance the breed in the future (Barbato *et al.,* 2015).

The substantial reduction in the effective population sizes over time could be a consequence of intense selection for specific traits of importance (Gebrehiowt *et al.,* 2020). This would have caused in an increase in genetic gain but consequently an increase in allele

fixation, therefore decreasing the effective population size. This is supported by the high number of short ROH segments observed in this study indicating ancient inbreeding which could have resulted allele fixation due to breed establishment (Eusebi *et al.,* 2020). An N_e of 50 individuals is considered as the minimum limit to maintain genetic diversity (FAO, 2010; Abin *et al.,* 2016). With the lowest N_e estimated in this study being 153 individuals for the TUL breed it can be deduced that SA beef cattle populations are not at risk of a loss of genetic diversity. The use of young sires and genomic technologies allow for accurate estimates of EBVs and therefore decreases the generation interval while increasing the genetic gain (Dekker *et al.,* 2021).

4.5 Between population diversity

Overall, the results of the PCA plot, admixture plot and phylogenetic tree revealed distinctiveness among the eight SA beef populations studied. This agrees with the different time divergence of the breeds. This also indicates that genetic diversity could be associated to the regions of origin, suggesting that breeds which segregated in more recent years have a closer relationship compared breeds which were derived a longer time ago (Makina *et al.,* 2016). *Bos taurus* populations are believed to have been developed a longer time ago, and from different areas of origin compared to *Bos indicus* populations. The Hereford breed (Bos taurus) originated in England where in 1742 the breed was properly established and was the first English group of cattle to be recognized as a true breed. The Charolais is a *Bos taurus* French breed which was established in 1773 and only after the second world war could the breed be found in other parts of the world. The distinct clusters of the Hereford and Charolais breeds could therefore be a consequence of the breeds having distinctly different regions of origin and were derived long ago (Edea *et al.,* 2013).

The *Bos indicus* Zebu population, Boran had been well established and adapted in, Kenya for many decades (Ajmone-Marsan *et al.,* 2010). However, it was not until the mid-1990s that the breed was imported to SA from Kenya, with the SA Boran society being formed in 2003. This could be a reason as to why the Boran clustered on its own. The SA indigenous populations (Sanga types) in this study are known to share similar migration routes, even though some breeds migrated along the eastern side of Southern Africa while others on the western side (Verdugo *et al.,* 2019). Furthermore, the Drakensberger, Nguni and Tuli breeds were all properly established in South Africa in the 1940's, and have the same *Bos taurus africanus* ancestry, providing explanation for the closer relationship between breeds present in this study (Drakensberger's Breeders' Society of SA, Nguni Breeders' Society of SA and Tuli Breeders' Society of SA). The results generated from this study show that the indigenous populations share common ancestry which could be due to the similar time when the breeds were established in South Africa.

The current study indicated population differentiation, with the eight populations each forming their own clusters. The SA indigenous Sanga populations (Drakensberger, Nguni and Tuli) clustered closer together, while the exotic populations (Zebu: Boran, *Bos taurus*: Charolais and Hereford) formed their own distinct clusters separate from the indigenous populations. This was supported by the admixture results which indicated eight distinct populations with some admixture. These results were similar to analyses performed by Makina *et al.,* (2015) and Zwane *et al.,* (2016). Both the current study and Makina *et al.,* (2015) further indicated an overlap between the Drakensberger and Bonsmara populations which could be due to the Bonsmara having Sanga ancestry. The Bonsmara breed was first developed using Afrikaner cattle which is one of the oldest Sanga populations in SA. The relatively close clustering of the SA indigenous populations could be an indication that the populations had ancestors of similar origin (Scholtz *et al.,* 2011). The lose clustering of the composite populations was to be expected since both Bonsmara and Beefmaster have *Bos taurus* and either Zebu, or Sanga ancestry. Partial population overlap of populations with similar origin or ancestors is consistent with results shown by Decker *et al.,* (2014) and can indicate a low to moderate level of homogeneity among the populations.

Results obtained in this study can serve as insight for formulation and updating breeding objectives for the eight populations. Based on genomic analyses, genetic diversity is moderate to high within and between the populations. This study can serve as a reference for the potential of the eight South African cattle populations to undergo further selection and participate in downstream genomic applications. This in turn will improve the overall cattle population in South Africa and assist farmers to maintain a genetically strong and diverse national herd of cattle.

Chapter 5: Conclusion

In this study genomic information from a total of 2110 beef cattle, representing eight South African beef breeds across the country was analysed. The cattle were originally genotyped with the GGP 150K Bovine array within the Beef Genomics Program (BGP). The BGP had the primary aim of establishing reference populations for GS in the country with the program running for a three-year period. The aim of this study was to assess SNP based genomic diversity and population structure of the eight beef cattle breeds.

Gaining insight into the genetic diversity of beef cattle breeds could assist in breeding strategies, understanding the origin of the breed and evaluate how breeds may respond to future genomic applications. The diversity parameters evaluated in this study also provided an indication if any breeds or breed type was at risk of a loss in genetic diversity.

Heterozygosity estimates in this study indicated a limited loss in genetic diversity and were comparable to other studies. The LD estimates (r^2>0.2) between the biological types provided an indication that GS may be applied for the exotic and composite biological types. The LD estimate of the indigenous biological type indicated the potential use of genotype imputation in order for further application of GS. The potential application of GS in the SA beef cattle breeds is of value for accurate predictions and breeding decisions to be made.

The low genomic inbreeding estimates for the F_{IS} parameter is an indication of unique gene pools for each population as well as good on farm breeding practices. It has been recommended to update inbreeding estimates every five years to ensure there is no loss in diversity. The F_{ROH} values for estimated inbreeding displayed similar results. The numerically large amount of short ROH segments is indicative of ancient inbreeding which could be due to a founder effect for the eight breeds. The lack of long ROH segments further highlights the low recent inbreeding levels, with all breeds having a desired level of diversity. Although inbreeding estimates were low, they highlight the importance of breeding practises to ensure the levels do not increase.

The PCA and admixture plots further highlighted the unique gene pools of the eight breeds when eight different clusters were formed. The eight clusters indicate a level of genetic diversity for each breed with the potential for each unique breed to contribute to the South African beef industry.

South African beef cattle populations exhibited moderate to high levels of genetic diversity. Results from this study further indicate the SA beef populations investigated are not at risk of a loss of diversity with the current farm practises and selection.

References

Abin, S., Theron, H.E. & van Marle-Köster, E. 2016. Population structure and genetic trends for indigenous African beef cattle breeds in South Africa. S. Afr. J. Anim. Sci. 46(2),152-156.

Ajmone-Marsan, P., Garcia, J.F. & Lenstra, J.A. 2010. On the origin of cattle: how aurochs became cattle and colonized the world. Evol. Anthropol.: News Issues Rev. 19, 148-157.

Akanno, E.C., Chen, L., Abo-Ismail, M.K., Crowley, J.J., Wang, Z., Li, C., Basarab, J.A., MacNeil, M.D. & Plastow, G.S. 2018. Genome-wide association scan for heterotic quantitative trait loci in multi-breed and crossbred beef cattle. Genet. Sel. Evol. 50(1), 48.

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Gen. Res. 19, 1655-1664.

Al-Mamun, H.A., Clark, S.A., Kwan, P. and Gondro, C. 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47(1), 1-14.

Antolín, R., Nettelblad, C., Gorjanc, G., Money, D. & Hickey, J.M. 2017. A hybrid method for the imputation of genomic data in livestock populations. Genet. Sel. Evol. 49, 30.

Bahbahani, H., Tijjani, A., Mukasa, C., Wragg, D., Almathen, F., Nash, O., Akpa, G.N., Mbole-Kariuki, M., Malla, S., Woolhouse, M. & Sonstegard, T. 2017. Signatures of selection for environmental adaptation and zebu x taurine hybrid fitness in East African Shorthorn Zebu. Front. Genet. 8, 68.

Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M.W., 2015. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front. Genet. 6.

Beefmaster Breeders' Society of SA & SA Stud Book, 2017. Joint media release: Genomic breeding values for South African Beefmaster cattle. Available online at: http://www.sastudbook.co.za/images/photos/News-Beefmaster-Genomic-EBVs.pdf

Beefmaster SA. 2021. Beefmaster SA History. http://www.beefmastersa.co.za/p12/about-the-breed/beefmaster-sa-history.html Accessed 22 August 2021.

Berry, D.P., McClure, M.C. & Mullen, M.P., 2014. Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. J. Anim. Breed. Genet. 131, 165-172.

Bisschoff, C. and Lotriet, R., 2013, July. The Drakensberger as competitive breed of cattle in the South African beef industry. International Farm Management Congress 19, 39-49.

Blasco, A. & Toro, M.A. 2014. A short critical history of the application of genomics to animal breeding. Livest. Sci. 166, 4-9.

Blench, R. & MacDonald, K. 2006. The origins and development of African livestock: archaeology, genetics, linguistics and ethnography. Routledge, London, U.K.

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

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

Boichard, D., Ducrocq, V., Croiseau, P. and Fritz, S., 2016. Genomic selection in domestic animals: principles, applications and perspectives. C.R. Biol. 339(7-8), 274-277.

Bonsma, J.C., 1980. Crossbreeding, breed creation and the genesis of the Bonsmara. Livestock production. A Global Approach. Ed. J.C. Bonsma. Tafelberg, Cape Town, South Africa. pp 126-136.

Bosse, M., Megens, H.J., Madsen, O., Paudel, Y., Frantz, L.A., Schook, L.B., Crooijmans, R.P. and Groenen, M.A. 2012. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. PLos Genet. 8(11), p.e1003100.

Brito, F.V., Neto, J.B., Sargolzaei, M., Cobuci, J.A. and Schenkel, F.S., 2011. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. BMC Genet. 12(1),1-10.

Bunning, H., Wall, E., Chagunda, M.G., Banos, G. & Simm, G., 2019. Heterosis in cattle crossbreeding schemes in tropical regions: meta-analysis of effects of breed combination, trait type, and climate on level of heterosis. J. Anim. Sci, 97(1), 29-34.

Burrow, H.M., Wolcott, M.L., Maiwashe, A., Makgahlela, M.L., Hayes, B.J., Rees, J.G. & Bradfield, M.J. 2017. Can grazing livestock in developing countries benefit from use of genomic selection. Proc. Assoc. Advmt. Anim. Breed. Genet. 22, 353-360.

Calus, M.P.L., Bouwman, A.C., Hickey, J.M., Veerkamp, R.F. and Mulder, H.A. 2014. Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. Animal. 8(11), 1743-1753.

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

Chan, E.K., Nagaraj, S.H. & Reverter, A. 2010. The evolution of tropical adaptation: comparing taurine and zebu cattle. Anim. Genet. 41, 467-477.

Charolais Society of South Africa. 2021. Origin of the Charolais breed. https://www.charolais.co.za/Breed-Origin.htm Accessed 22 August 2021

Cole, J. B., Lewis, R. M., Maltecca, C., Newman, S., Olson, K. M. & Tait Jr., R. G. 2013. Identifying relationships among markers, genes, and phenotypes. J. Anim. Sci. 91, 521-522.

Corbin L.J., Blott S.C., Swinburne J.E., Vaudin M., Bishop S.C. & Woolliams J.A. 2010. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. Anim. Genet. 41, 8–15.

Curik, I., Ferenčaković, M. & Sölkner, J. 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 166, 26-34.

Daetwyler, H.D., Schenkel, F.S., Sargolzaei, M. & Robinson, J.A.B. 2008. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J. Dairy Sci, 91(8), 3225-3236.

DAFF, 2017. Abstract of Agricultural statistics 2016. Department of Agriculture, Forestry and Fisheries.

DAGRIS (Domestic Animal Genetic Resources Information System), 2010. Domestic Animal Genetic Resources Information System: http://dagris.ilri.cgiar.org.

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

Decker, J., 2021. Decreasing generation interval to increase genetic progress. Doi: https://mospace.umsystem.edu/xmlui/handle/10355/85172

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

Dekkers, J.C.M. & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3, 22-32.

Drakensberger Cattle Breeders' Society of SA. 2011. Drakensberger Handbook, first ed. Volksrust, Mpumalanga, South Africa. Accessed 22 August 2021. https://drakensbergers.co.za/wp-content/uploads/2021/02/DRAKENSBERGER-HANDBOOK-3rd-edition-April-2017-small.pdf

Du Toit, J., Van Wyk, J.B. & Maiwashe, A. 2012. Assessment of inbreeding depression for functional herd life in the South African Jersey breed based on level and rate of inbreeding. S. Afr. J. of Anim. Sci. 42(1), 55-62.

Ducrocq, V., Laloe, D., Swaminathan, M., Rognon, X., Tixier-Boichard, M. & Zerjal, T. 2018. Genomics for ruminants in developing countries: from principles to practice. Front. Genet. 9, 251.

Edea, Z., Bhuiyan, M. S. A., Dessie, T., Rothschild, M. F., Dadi, H. and Kim, K. S. 2015. Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. Animal, Cambridge University Press. 9(2), 218–226.

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

Edea, Z., Dadi, H., Kim, S.W., Dessie, T., Lee, T., Kim, H., Kim, J.J. & Kim, K.S. 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. Front. Genet. 4, 35.

Eggen, A. 2012. The development and application of genomic selection as a new breeding paradigm. Anim. Front. 2, 10-15.

Engelsma, K.A., Veerkamp, R.F., Calus, M.P.L. & Windig, J.J. 2014. Consequences for diversity when animals are prioritized for conservation of the whole genome or of one specific allele. J. Anim. Breed. Genet. 131(1), 61-70.

Eusebi, P.G., Martinez, A. & Cortes, O. 2020. Genomic tools for effective conservation of livestock breed diversity. Diversity. 12(1), 8.

Ewens, W.J., 1970. Evolution and the Genetics of Populations. Vol. 2, The Theory of Gene Frequencies. Sewall Wright. University of Chicago Press, Chicago, 1969. viii+ 512 pp., illus.

Fabbri, M.C., Gonçalves de Rezende, M.P., Dadousis, C., Biffani, S., Negrini, R., Souza Carneiro, P.L. and Bozzi, R. 2019. Population structure and genetic diversity of Italian beef breeds as a tool for planning conservation and selection strategies. Animal. 9(11), 880.

Fan, B., Du, Z.Q., Gorbach, D.M. & Rothschild, M.F. 2010. Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian Austral. J. Anim. Sci. 23(7) 833-847.

FAO, 2010. http://www.fao.org/unfao/procurement/statistics-from-2010-2015/statistics-2010/en Accessed 7 July 2020.

Ferraz, J.B.S., Wu, X.L., Li, H., Xu, J., Ferretti, R., Simpson, B., Walker, J., Silva, L.R., Garcia, J.F., Tait, R.G. and Bauck, S. 2020. Development and evaluation of a low-density single-nucleotide polymorphism chip specific to Bos indicus cattle. Anim. Prod. Sci. 60(15), 1769-1776.

Ferenčaković, M., Hamzić, E., Gredler, B., Curik, I. & Sölkner, J. 2011. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. Agriculturae Conspectus Scientificus. 76(4), 325-329.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I. & Sölkner, J. 2013a. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. J Anim. Breed. Genet. 130, 286–293.

Ferenčaković, M., Sölkner, J., & Curik, I. 2013b. Estimating autozygosity from high-throughput information: Effect of SNP density and genotyping errors. Genet. Sel. Evol. 45, 42.

Fernández, J. & Bennewitz, J. ed., 2017. Chapter 2 “Defining genetic diversity based on genomic tools”. In: Genomic management of animal genetic diversity. Publisher, Wageningen Academic. pp. 55-62.

Fleming, A., Abdalla, E.A., Maltecca, C. & Baes, C.F. 2018. Invited review: Reproductive and genomic technologies to optimize breeding strategies for genetic progress in dairy cattle. Arch. Anim. Breed. 61(1), 43-57.

Fore, J., Wiechers, I.R. & Cook-Deegan, R. 2006. The effects of business practices, licensing, and intellectual property on development and dissemination of the polymerase chain reaction: case study. J. Biomed. Discovery Collab.1, 7-24.

Garrick, D.J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet. Sel. Evol., 43(1), 17.

Gaughan, J. B., T. L. Mader, S. M. Holt, M. J. Josey, & K. J. Rowan. 1999. Heat tolerance of Boran and Tuli crossbred steers. J. Anim. Sci. 77, 2398-2405.

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

Gebrehiwot, N.Z., Strucken, E.M., Aliloo, H., Marshall, K. & Gibson, J.P., 2020. The patterns of admixture, divergence, and ancestry of African cattle populations determined from genome-wide SNP data. BMC genomics. 21(1), 1-16.

Geibel, J., Reimer, C., Pook, T., Weigend, S., Weigend, A. and Simianer, H. 2021. How imputation can mitigate SNP ascertainment Bias. BMC genomics. 22(1), 1-13.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long-term response. Genetics. 136, 245–257

Goddard, M.E. & Hayes, B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10, 381-391.

Goddard, M.E., 2012. Uses of genomics in livestock agriculture. Anim. Prod. Sci. 52(3), 73-77.

Grigson, C. 1991. An African origin for African cattle? – Some archaeological evidence. Afr. Archaeol. Rev. 9, 119–144.

Habier, D, R.L. Fernando, & J.C.M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. 177(4),2389–2397.

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. Sci. 296, 336-339.

Hansen, P. J. 2004. Physiological and cellular adaptations of Zebu cattle to thermal stress. Anim. Reprod. Sci. 82, 349-360.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92, 2, 433-443.

Hayes, B.J., Lewin, H.A. & Goddard, M.E. 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. Trends Genet. 29, 206-214.

Herrero-Medrano J.M., Megens H.-J., Groenen M.A.M., Ramis G., Bosse M., P_erez-Enciso M. & Crooijmans R.P.M.A. 2013. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. BMC Genetics. 14, 106.

Hoffmann I. 2010. Climate change and the characterization, breeding and conservation of animal genetic resources. Anim Genet. 41. 32-46.

Howard, J.T., Pryce, J.E., Baes, C. & Maltecca, C. 2017. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100, 8, 6009-6024.

Jemaa, S.B., Rahal, O., Gaouar, S.B.S., Mastrangelo, S., Boussaha, M. and Ciani, E. 2018. Genomic characterization of Algerian Guelmoise cattle and their genetic relationship with other North African populations inferred from SNP genotyping arrays. Livest. Sci. 217, 19-25.

Keller, M.C., Visscher, P.M. and Goddard, M.E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189(1), 237-249.

Kim, J., Hanotte, O., Mwai, O.A., Dessie, T., Bashir, S., Diallo, B., Agaba, M., Kim, K., Kwak, W., Sung, S. and Seo, M. 2017. The genome landscape of indigenous African cattle. Genome Biol. 18(1), 34.

Koch, R. M., Cundiff, L. V., Gregory, K. E. & Van Vleck, L. D. 2004. Genetic response to selection for weaning weight or yearling weight or yearling weight and muscle score in Hereford cattle: Efficiency of gain, growth, and carcass characteristics. J. Anim. Sci. 82, 668-682.

Lashmar, S.F., Muchadeyi, F.C. and Visser, C. 2019. Genotype imputation as a cost-saving genomic strategy for South African Sanga cattle: A review. S. Afr. J. Anim. Sci. 49(2), 262-280.

Lashmar, S.F., Visser, C., van Marle-Köster, E. and Muchadeyi, F.C. 2018. Genomic diversity and autozygosity within the SA Drakensberger beef cattle breed. Livest. Sci., 212, 111-119.

Lenstra, J.A., Groeneveld, L.F., Eding, H., Kantanen, J., Williams, J.L., Taberlet, P., Nicolazzi, E.L., Sölkner, J., Simianer, H., Ciani, E. & Garcia, J.F. 2012. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. Anim. Genet. 43(5), 483-502.

Lin, B.Z., Sasazaki, S. and Mannen, H. 2010. Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers. Anim. Sci. J. 81(3), 281-289.

Lwin, M., Mon, S.L.Y., Yamanaka, H., Nagano, Y., Mannen, H., Faruque, M.O., Kawabe, K., Okamoto, S. and Shimogiri, T. 2018. Genetic diversities and population structures of four popular Myanmar local cattle breeds. Anim. Sci. J. 89(12), 1648-1655.

Madilindi, M.A., Banga, C.B., Bhebhe, E., Sanarana, Y.P., Nxumalo, K.S., Taela, M.G., Magagula, B.S. and Mapholi, N.O. 2019. Genetic diversity and relationships among three Southern African Nguni cattle populations. Trop. Anim. Health Prod. 1-10.

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

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. and Maiwashe, A. 2016. Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide. Genet. Sel. Evol. 48(1), 88.

Malécot, G. & Blaringhem, L.F. 1948. The mathematics of heredity. Masson, Paris. pp 1-64.

Malomane, D.K., Reimer, C., Weigend, S., Weigend, A., Sharifi, A.R. and Simianer, H. 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. BMC genomics. 19(1), 1-16.

Maltecca, C., Tiezzi, F., Cole, J.B. and Baes, C. 2020. Symposium review: Exploiting homozygosity in the era of genomics—Selection, inbreeding, and mating programs. J. Dairy Sci. 103(6), 5302-5313.

Mapiye, C., Chikwanha, O.C., Chimonyo, M. and Dzama, K., 2019. Strategies for Sustainable Use of Indigenous Cattle Genetic Resources in Southern Africa. Diversity. 11(11), 214.

Marandure, T., Mapiye, C., Makombe, G., Nengovhela, B., Strydom, P., Muchenje, V. and Dzama, K. 2016. Determinants and opportunities for commercial marketing of beef cattle raised on communally owned natural pastures in South Africa. African Journal of Range & Forage Science. 33(3), 199-206.

Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906-913.

Marshall, F., 2000. The origins and spread of domestic animals in East Africa. The origins and development of African livestock: Archaeology, genetics, linguistics and ethnography. 191-221.

Marufu, M.C., Chimonyo, M., Mapiye, C. & Dzama, K., 2011. Tick loads in cattle raised on sweet and sour rangelands in the low-input farming areas of South Africa. Trop. Anim. Health Prod. 43, 307-313.

Mastrangelo, S., Ciani, E., Marsan, P.A., Bagnato, A., Battaglini, L., Bozzi, R., Carta, A., Catillo, G., Cassandro, M., Casu, S. and Ciampolini, R., 2018. Conservation status and historical relatedness of Italian cattle breeds. Genet. Sel. Evol. 50(1), 1-16.

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M.T. & 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.

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. & Van Tassell, C.P. 2009. Development and characterization of a high-density SNP genotyping assay for cattle. PloS One. 4(4), 5350.

McKay, S.D., Schnabel, R.D., Murdoch, B, M., Matukumalli, L.K., Aerts, J. & Coppieters, W. 2008. An assessment of populations structure in eight breeds of cattle using a whole genome SNP panel. BMC Genet. 9, 37. doi: 101186/1471-2156-9-27.

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. & MacLeod, A.K. 2008. Runs of homozygosity in European populations. AM. J. Hum. Genet. 83(3), 359-372.

Melka, M.G. & Schenkel, F.S. 2012. Analysis of genetic diversity in Brown Swiss, Jersey and Holstein populations using genome-wide single nucleotide polymorphism markers. BMC Res. 5(1), 161.

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

Meuwissen, T.H., Sonesson, A.K., Gebregiwergis, G. and Woolliams, J.A. 2020. Management of genetic diversity in the era of genomics. Front. Genet. 11, 880.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genet. 157, 1819-1829.

Mkize, L.S. and Zishiri, O.T. 2020. Population genetic structure and maternal lineage of South African crossbred Nguni cattle using the cytochrome b gene in mtDNA. Trop. Anim. Health Prod.1-11.

Mrode, R., Ojango, J.M.K., Okeyo, A.M., Mwacharo, J.M. 2019. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. Front. Genet. 694, 1–11.

Msalya, G., Kim, E.S., Laisser, E.L., Kipanyula, M.J., Karimuribo, E.D., Kusiluka, L.J., Chenyambuga, S.W. and 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), p.e0171088.

Mwai, O., Hanotte, O., Kwon, Y.J. and Cho, S. 2015. African indigenous cattle: unique genetic resources in a rapidly changing world. Asian-Australasian J. Anim. Scis. 28(7), 911.

Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M.S. & Bernabucci, U., 2010. Effects of climate changes on animal production and sustainability of livestock systems. Livest. Sci. 130, 57-69.

NCBI, Genome Assembly and Annotation Report. 2021. Bos taurus (cattle) Reference genome sequence. Accessed 28 October 2021. https://www.ncbi.nlm.nih.gov/genome/82?genome_assembly_id=262553

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings Nat. Academy Sci. USA 70, 3321-3323.

Nyamushamba, G.B., Mapiye, C., Tada, O., Halimani, T.E. and Muchenje, V. 2017. Conservation of indigenous cattle genetic resources in Southern Africa's smallholder areas: turning threats into opportunities—A review. Asian-Australasian J. Anim. Scis. 30(5), 603.

O'Brien, A.M.P., Mészáros, G., Utsunomiya, Y.T., Sonstegard, T.S., Garcia, J.F., Van Tassell, C.P., Carvalheiro, R., da Silva, M.V. and Sölkner, J., 2014. Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high-density SNP chip data and different minor allele frequency distributions. Livest. Sci. 166, 121-132.

Ogawa, S., Matsuda, H., Taniguchi, Y., Watanabe, T., Takasuga, A., Sugimoto, Y. & Iwaisaki, H. 2016. Accuracy of imputation of single nucleotide polymorphism marker genotypes from low-density panels in Japanese Black cattle. Anim. Sci. J. 87, 3-12.

Orton, J., Mitchell, P., Klein, R., Steele, T. and Horsburgh, K.A. 2013. An early date for cattle from Namaqualand, South Africa: implications for the origins of herding in southern Africa. Antiquity. 87(335), 108-120.

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R. and Baldi, F. 2017. Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 48(3), 255-271.

Peripolli, E., Stafuzza, N.B., Munari, D.P., Lima, A.L.F., Irgang, R., Machado, M.A., do Carmo Panetto, J.C., Ventura, R.V., Baldi, F. and 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.

Pienaar, L., Grobler, J.P., Scholtz, M.M., Swart, H., Ehlers, K., Marx, M., MacNeil, M.D. and Neser, F.W. 2018. Genetic diversity of Afrikaner cattle in southern Africa. Trop. Anim. Health Prod. 50(2), 399-404.

Plastow, G.S. 2016. Genomics to benefit livestock production: improving animal health. Revista Brasileira de Zootecnia. 45(6), 349-354.

Pryce, J.E., Goddard, M.E., Raadsma, H.W. & Hayes, B.J. 2010. Deterministic models of breeding scheme design that incorporate genomic selection. J. Dairy Sci. 93(11), 5455-5466.

Purcelll, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. & Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. AM J. Hum. Genet. 81(3) 559-575.

Purfield D.C., Evans R.D., Carthy T.R. and Berry D.P. 2019. Genomic Regions Associated with Gestation Length Detected Using Whole-Genome Sequence Data Differ Between Dairy and Beef Cattle. Front. Genet. 10, 1068.

Purfield, D.C., Berry, D.P., McParland, S. and Bradley, D.G. 2012. Runs of homozygosity and population history in cattle. BMC genetics. 13(1), 1-11.

Qwabe, S.O., Maiwashe, A. & Muchadeyi, F.C. 2013. Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. S. Afr. J. Anim. Sci. 43(1) 64-67.

Robertson, A., 2007. Inbreeding in artificial selection programmes. Gen. Res. 89(5-6), 275-280.

Rothschild, M.F. and Plastow, G.S. 2014. Applications of genomics to improve livestock in the developing world. Livest. Sci. 166, 76-83.

SA Hereford Cattle Society. 2021. Hereford Breed Standards. https://www.hereford.co.za/Breed-Standards.htm Accessed 15 September 2021.

SA Studbook, 2016. SA Studbook annual report 2016. SA Studbook. https://studbook.co.za/ci86/SA-Stud-Book-Annual-Report-2016.html

Sabir, J., Mutwakel, M., El-Hanafy, A., Al-Hejin, A., Abdel Sadek, M., 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. J. Food, Agric. Environment 12(2), 541-553.

Scholtz, M. M. 2010. Beef Breeding in South Africa, 2nd Edn. Pretoria: ARC Pretoria.

Scholtz, M., Bosman, D. J., Erasmus, G. J., and Maiwashe, A. 2010. “Selection as the base of improvement in beef cattle,” in Beef Breeding in South Africa, 2nd Edn., ed M. Scholtz (Pretoria: Agricultural Research Council), 2–10

Scholtz, M.M., McManus, G., Leeuw, K-C., Louvandini, H., Seixas, L., Demelo, C.B., Theunissen, A. & Neser, F.W.C. 2013. The effect of global warming on beef production in developing countries of the southern hemisphere. Nat. Sci. 5, 106-119.

Schrooten, C., Dassonneville, R., Ducrocq, V., Brøndum, R.F., Lund, M.S., Chen, J., Liu, Z., González-Recio, O., Pena, J. & Druet, T. 2014. Error rate for imputation from the Illumina BovineSNP50 chip to the Illumina BovineHD chip. Genet. Sel. Evol. 46, 10.

Scovronick, N., Sera, F., Acquaotta, F., Garzena, D., Fratianni, S., Wright, C.Y. and Gasparrini, A. 2018. The association between ambient temperature and mortality in South Africa: A time-series analysis. Environ. Res. 161, 229-235.

Shabtay A. 2015 Adaptive traits of indigenous cattle breeds: The Mediterranean Baladi as a case study. Meat Sci. 109,27-39.

Signer-Hasler, H., Burren, A., Neuditschko, M., Frischknecht, M., Garrick, D., Stricker, C., Gredler, B., Bapst, B. & Flury, C. 2017. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet. Sel. Evol. 49(1), 83.

Soji, Z. and Muchenje, V. 2016. Effect of genotype and age on some carcass and meat quality traits of beef carcasses subjected to the South African classification system. Meat Sci. 117, 205-211.

Sölkner, J., Ferenčaković, M., Gredler, B. & Curik, I. 2010. Genomic metrics of individual autozygosity, applied to a cattle population. In: 61st Annual Meeting of the European Association of Animal Production.

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

Tabangin, M.E., Woo, J.G. and Martin, L.J. 2009. The effect of minor allele frequency on the likelihood of obtaining false positives. BMC proceedings. 3(7), 1-4

Tuli Cattle Breeders Society of South Africa. 2021. The Tuli cattle story. http://www.tulicattle.co.za/p2/history/the-tuli-cattle-story.html . Accessed 22 August 2021

Upadhyay, M., Bortoluzzi, C., Barbato, M., Ajmone-Marsan, P., Colli, L., Ginja, C., Sonstegard, T.S., Bosse, M., Lenstra, J.A., Groenen, M.A. and Crooijmans, R.P. 2019. Deciphering the patterns of genetic admixture and diversity in southern European cattle using genome-wide SNP. Evol. Appl. 12(5), 951-963.

Van der Westhuizen, L., MacNeil, M.D., Scholtz, M.M., Neser, F.W., Makgahlela, M.L. and van Wyk, J.B. 2020. Genetic variability and relationships in nine South African cattle breeds using microsatellite markers. Trop. Anim. Health Prod., 52(1), 177-184.

Van der Westhuizen, R.R.; Van der Westhuizen, J.; Van Marle-Köster, E. 2017. Estimation of genomically enhanced estimated breeding values for SA beef cattle: In 50th SASAS Congress. Port Elizab. 48, 18–22.

Van Liere, J.M. and Rosenberg, N.A. 2008. Mathematical properties of the r2 measure of linkage disequilibrium. Theor. Popul Biol. 74(1), 130-137.

Van Liere, J.M. and Rosenberg, N.A. 2008. Mathematical properties of the r2 measure of linkage disequilibrium. Theor. Popul Biol. 74(1), 130-137.

Van Marle-Köster, E., Visser, C., Makgahlela, M. & Cloete, S.W.P. 2015. Genomic technologies for food security. A review of challenges and opportunities in southern Africa. Food Res. Int. 76, 971-979.

Van Marle-Köster, E., Visser, C., Sealy, J. and Frantz, L., 2021. Capitalizing on the Potential of South African Indigenous Beef Cattle Breeds: A Review. Sustainability.13(8), 4388.

Van Marle-Köster, E.; Visser, C. 2018. Genetic Improvement in South African Livestock: Can Genomics Bridge the Gap Between the Developed and Developing Sectors? Front. Genet. 9, 1–12

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Delser, P.M., Hare, A.J., Burger, J., Collins, M.J., Kehati, R. and Hesse, P. 2019. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science. 365(6449), 173-176.

Visser, C., Van Marle-Köster, E., Myburgh, H.C., De Freitas, A. 2020. Phenomics for sustainable production in the South African dairy and beef cattle industry. Anim. Front. 10(2), 12-18.

Wang, D.M.; Dzama, K.; Hefer, C.A.; Muchadeyi, F.C. 2015 Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genom. 16, 984.

Williams, J.L., Dunner, S., Valentini, A., Mazza, R., Amarger, V., Checa, M.L., Crisa, A., Razzaq, N., Delourme, D., Grandjean, F. and Marchitelli, C. 2009. Discovery, characterization and validation of single nucleotide polymorphisms within 206 bovine genes that may be considered as candidate genes for beef production and quality. Anim. Genet., 40(4), 486-491.

Worldometers. 2020. Growth rate. Accessed: 5 July 2020. https://www.worldometers.info/world-population/#growthrate

Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. 2011. GCTA: a tool for genome-wide complex trait analysis. Amer. J. Hum. Genet. 88, 76-82.

Yang, W., Kang, X., Yang, Q., Lin, Y. & Fang, M. 2013. Review on the development of genotyping methods for assessing farm animal diversity. J. Anim. Sci. Biotech. 4(1), 2.

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. and da Silva, M.V. 2015. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. Front. Genet. 6, 5.

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. Genet. Sel. Evol. 47(1), 1-12.

Zwane, A.A., Maiwashe, A., Makgahlela, M.L., Choudhury, A., Taylor, J.F. & Van Marle-Köster, E., 2016. Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds. S. Afr. J. Anim. Sci. 46, 302-312.

Zwane, A.A., Schnabel, R.D., Hoff, J., Choudhury, A., Makgahlela, M.L., Maiwashe, A., Marle-Koster, V. and Taylor, J.F. 2019. Genome-wide SNP discovery in indigenous cattle breeds of South Africa. Front. Genet. 10, 273.