A Genome-wide association study of fertility and maternal traits in South African Bonsmara cattle

by

Jason Jack Reding

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae Agriculturae

MSc Agric Animal Science: Animal Breeding and Genetics

in the Faculty of Natural and Agricultural Sciences

Department of Animal Sciences

University of Pretoria

Pretoria

May 2020

Supervisor: Prof E. van Marle-Köster

Co-supervisor: Dr R.R. van der Westhuizen

Extra-ordinary Professor Dept. Animal science: Prof D.P Berry

Declaration

I, Jason Jack Reding, hereby declare that this thesis, submitted for the degree MSc(Agric) Animal Science: Animal Breeding and Genetics at the University of Pretoria, is my own work and has not been previously submitted by me for a degree at any other University or institution of higher learning.

Jason Reding

Pretoria

May 2020

“When I run after what I think I want, my days are a furnace of stress and anxiety; if I sit in my own place of patience, what I need flows to me, and without pain. From this I understand that what I want also wants me, is looking for me and attracting me. There is a great secret here for anyone who can grasp it.”

-Rumi

Acknowledgements

The assistance of SA Stud Book and the SA Bonsmara Breeders Society in the contribution of the data towards this research is hereby acknowledged.

This masters project is dedicated to my grand-father, Malcolm Christopher Neels Reding (23/09/1930 - 27/03/2019), the patriarch of our family. The words of encouragement towards striving to do something fulfilling with my life will never be forgotten. Christopher, Jonathan, Veryan, Jason, Ruby, Eddy, Janan and Janade. He lives in us.

To my mom Veryan, thank you for allowing me to make my own decisions from a young age, albeit with a guiding hand. I love you. Gran, thank you for being there when I felt most frustrated. Here is to the countless great conversations that have been instrumental in developing my sense of wonder of the world around us and pushing me to be the best person I can be. To Byron, my spirit brother, for being the best friend any young man would hope to share their journey through University with. May we continue to learn from each other's respective disciplines and inspire ourselves for greatness. Stuart, my longest friend, thank you for grounding me in times of hardship and the result of our shenanigans always leaving a smile on my face. Maike, thank you for opening my heart to an avenue of life that I never thought I would find myself walking down and gaining a deeper appreciation of what life means to others. Thank you Flippie, for providing me with a place of escape from the stresses of my day to day life, and the future adventures that await us in the Cape.

To my supervisor, Prof E. van Marle- Köster. From our first interaction in the 4th year of my studies, thank you for seeing the potential that I could not yet see in myself. Thank you for allowing me to flourish as a post-graduate student in my own space and time and putting in a good word with my current employers. Bobbie, thank you for the time so far at SA Stud Book and for helping me with this project when I was most annoyed. May our list of collaborations continue to grow. A thanks of support to Dr Donagh Berry and Tara for helping me the Methods, especially with deciphering the WOMBAT manual. Rulien, the most Afrikaans person I have ever been friends with, I cannot express in words the gratitude I have for you. For allowing me to help in the lab with your project and making me feel welcome within our little department.

Anyone I have not written a small mention of; Warren, Keegan, Keldt, Martin, Kayla, Anel, Peter, Tasmin, Taylor, Nokuthula, Scott, Astle, Helena, Bernice, Donne, Suretha, Japie and Oom Schalk, thank you for everything you have done and will do in my life.

Finally, to myself. To defeat the mind you need two weapons, Education and Discipline. Education in that you always strive to learn more, and Discipline in that you will not allow yourself to falter, especially in the moments that require you the most. Be proud.

Abstract

Reproductive wastage is of major economic importance and limits overall herd efficiency. Fertility traits are lowly heritable but are an essential component in selection programs for the genetic improvement of all livestock species. In order to gain a better understanding of the underlying genetic basis of these traits a GWAS was conducted to investigate fertility and maternal traits within the SA Bonsmara breed. Traits of interest were age at first calving (AFC), inter-calving period (ICP), scrotal circumference (SC) and weaning weight maternal (WW_{MAT}). Estimated breeding values, pedigrees and genotypes for 3 291 SA Bonsmara animals were available for the study. Three different commercial arrays underwent quality control (QC), principal component analysis (PCA), with the amalgamation of the three arrays via imputation to a density of 128 793 SNPs and finally an association analysis by single SNP regression. Gene annotation was done for significant SNPs ($\leq 1x10^{-8}$), with four associated with ICP (BTA 4, BTA11, BTA17, BTA19), twenty with AFC (BTA1, BTA2, BTA4, BTA5, BTA7, BTA8, BTA9, BTA11, BTA12, BTA16, BTA20, BTA24, BTA27, BTA28, BTA29), twenty-two with SC (BTA1, BTA2, BTA3, BTA4, BTA5, BTA6, BTA8, BTA10, BTA11, BTA15, BTA16, BTA20, BTA22, BTA24, BTA26, BTA28) and forty-four with WW_{MAT} on all chromosomes except BTA10, BTA13, BTA16, BTA17, BTA23, BTA24, BTA25 and BTA26. The three SNPs significantly associated with AFC on BTA3 were all in gene regions. Of the six SNPs associated with WW_{MAT} on BTA15, five were identified in four genes, with two SNPs annotating to the same gene (*LRRC4C*). Three genes for SC, eight genes for AFC and nineteen genes for WW_{MAT} were identified. Genes associated with SC (*PPP3CA*), AFC (*AKT3*, *GRM8*, *KIF1B*, *OVOS2*) and WW_{MAT} (*BMP1*, *LRRC4C*, *MACROD1*, *RBM47*, *THSD7B*) were reported for cattle in other studies. Genes associated with AFC (*AKT3*; BTA16) and SC (*PPP3CA*, *TRPM6*; BTA6, BTA8) where shown to share serine/threonine biological pathway processes. Chromosomes and SNPs that yielded novel associations with previously uncharacterised genes and previously not reported in literature may possibly be unique to the SA Bonsmara beef breed and will require further investigation.

Table of Contents

List of Abbreviations

A2M	Alpha-2-macroglobulin
ACR	Allele concordance rate
AFC	Age at first calving
ARC	Agricultural Research Council
BD	Bechet's diseases
BCS	Body condition score
BGP	Beef Genomics Program
BLUP	Best linear unbiased prediction
BMP	Bone morphogenic protein
BTA	*Bos taurus* genome
BW	Birth weight direct
BW_{MAT}	Birth weight maternal
CE	Calving ease
CS	Calving Success
DAFF	Department of Agriculture, Forestry and Fisheries
DAYTC	Days to calving
dEBV	Deregressed estimated breeding value
DNA	Deoxyribonucleic acid
DSCAML1	DS cell adhesion molecule like 1
EBV	Estimated breeding value
ECM	Extracellular matrix
ER	Endoplasmic reticulum
ERC	Effective record contribution
f	Frequency
FA	Favourable allele
FER	Tyrosine-protein kinase
FGFR1	Fibroblast growth factor receptor
FILIP1	Filamin A interacting protein 1

FRS2	Fibroblast growth factor receptor substrate 2
GCR	Genotype concordance rate
GEBV	Genomically enhanced estimated breeding value
GGP	GeneSeek Genomic Profiler
GL	Gestation length
GL_{MAT}	Gestation length maternal
GnRHR	Gonadotrophin releasing hormone receptor
GRM	Genomic relationship matrix
GRM8	Metabotropic glutamate receptor 8
GS	Genomic selection
GWAS	Genome wide association study
GyE	Genotype x environmental effects
h^2	Heritability
HD	High(er) density
HMM	Hidden Markov models
HWE	Hardy-Weinberg equilibrium
H_z	Heterozygosity
IA	Imputation accuracy
IBD	Identity by descent
ICBF	Irish Cattle Breeding Federation
ICP	Inter-calving period
ID	Identification
IDB	International Dairy and Beef
kg	Kilogram
KPH	Kidney, pelvis and heart
LAWR	Lifetime annual weaning rate
LD	Linkage disequilibrium
LE	Linkage equilibrium
LRRC4C	Leucine rich repeat containing 4C
MACE	Multiple across country evaluation
MACROD1	ADP-ribose glycohydrolase MACROD1

MAF	Minor allele frequency
MAS	Marker assisted selection
Mg^{2+}	Magnesium
MGLL	Monoglyceride lipase
mm	millimetres
MME	Mixed model equations
n	Sample size
NGS	Next-generation sequencing
NTM	Neurotrimin
NUP210	Nucleoporin gp210
OSW	Overlapping sliding windows
PCA	Principal component analysis
PCG	Pre-conjugated gradient
PPP3CA	Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform
PR	Pregnancy rate
PREG	Pregnancy
QC	Quality control
QTL	Quantitative trait loci
RBM47	RNA binding motif protein 47
RDH16	Microsomal NAD+-dependent retinol dehydrogenase 4
RNA	Ribonucleic acid
RTN4RL1	Reticulon 4 receptor like 1
SA	South Africa
SC	Scrotal circumference
SNP	Single nucleotide polymorphism
ssGBLUP	single-step genomic best linear unbiased prediction
TCA	Thalamocortical axon
TGF	Transforming growth factor
THSD7B	Thrombospondin type 1 domain containing 7B
TRPM6	Transient receptor potential cation channel subfamily M member 6
UBAC2	Ubiquitin like modifier activating enzyme 2

UBE2J1	Ubiquitin conjugating enzyme E2 J1
WDPCP	WD repeat-containing planar cell polarity effector
WGS	Whole genome sequence
WW	Weaning weight
WW_{MAT}	Weaning weight maternal
WWOX	WW domain containing oxidoreductase
ZGRF1	Zinc finger GRF-type containing 1

List of Tables

List of Figures

Chapter 1 Introduction and Literature Review

1.1 Introduction

Reproductive wastage is of major economic importance and limits subsequent efficiency from conception to weaning due to its influence on herd productivity. The estimated calving percentage of South Africa's beef commercial sector is approximately 62% (Grobler et al., 2014). Furthermore, total lifetime productivity is a primary factor that needs to be improved within beef cattle populations (Burns *et al.*, 2010). As the beef industry primarily relies on reproductive efficiency, a reduction in unproductive periods during a female's reproductive life would positively impact production costs and income.

Improvement in fertility traits has typically been slow due to the low heritability, binomial nature of a short-controlled breeding season and/or the late expression of fertility traits, at least for traits such as inter-calving period (ICP) (Meyer *et al.*, 1990; Cammack *et al*., 2009; Hawken *et al.*, 2012). An understanding of the genetic basis of fertility traits is required to implement selection programs that may increase reproductive efficiency (Miar *et al*., 2015). The true underlying genetic architecture of most traits remains unknown; genomic tools have now opened opportunities for different approaches to improve our understanding (Meuwissen *et al*., 2001).

South Africa formed a consortium consisting of breed societies, their respective industry service providers (SA Stud Book and Breedplan), research institutions (Agricultural Research Council) in collaboration with the University of Pretoria, as well as other universities, known as the SA Beef Genomics Program (BGP). This resulted in the collection of many biological samples from registered beef cattle that were participating in performance recording. Although the cost of genotyping has drastically decreased over the last few decades, it remains an expensive undertaking for SA breeders due to volatile exchange rates which can make the true cost and economic viability unpredictable.

Lower-density panels have become more feasible, due to their low cost, and these genotypes may be imputed up to higher density (HD) panels. With the constant addition and consolidation of a HD reference population, this enables the same task at a reduced cost. Identification of the most important SNPs may aid in the decision-making of creating a minimal SNP density panel for industry that enables sufficient panel overlap and facilitates current imputation software.

The South African (SA) Bonsmara, classified as a Sanga type, is a unique composite breed of 3/8 exotic (Milk Shorthorn, Hereford) and 5/8 Afrikaner (Bonsma, 1980). The SA Bonsmara was

established through a well-documented crossbreeding program which had the aim of founding a local composite well adapted to the challenges of a diverse SA climate.

According to Stud Book SA, there were approximately 108 000 Bonsmara cattle participating in Logix Beef in 2019. Bonsmara make up 41,7% of the total number of recorded animals currently participating in Logix Beef (SA Stud Book, 2019).

The SA Bonsmara was the first beef breed to implement genomically enhanced estimated breeding values (GEBVs; van der Westhuizen *et al.*, 2017) in South Africa. These are now predicted using single-step genomic best linear unbiased prediction (ssGBLUP; Legarra *et al.*, 2014) and are used in selection programs to aid seedstock breeders. Currently, as of 1 January 2020, there are approximately 4000 SA Bonsmara genotypes across three different genotyping platforms managed by SA Stud Book.

Large regions of localised DNA that influence traits are referred to as quantitative trait loci (QTL). The effect of individual genes on most complex traits, is likely to be small and therefore a large number of markers must be used to identify QTL (Hayes & Goddard, 2010). Single Nucleotide Polymorphisms (SNPs) are bi-allelic single base pair mutations, where most are neutral to trait variation while others can be in linkage disequilibrium (LD) with causative mutations (Russel, 2010). SNPs on genotyping panels are evenly spread throughout the genome and are therefore more likely to be present in regulatory gene regions. Due to this distribution, SNPs can have significant associations with traits of low heritability or traits that are hard to measure as they may only be expressed later in life or are sex-limited traits (Hayes & Goddard, 2010).

A genome-wide association study (GWAS) is used to coarsely identify the location of causal genes with the identification of SNP markers located near these QTL (Dekkers & Hospital, 2002). A GWAS evaluates molecular data alongside estimated breeding values (EBV) or phenotypic data of available genotyped animals. This may lead to the detection of significant associations between these SNP markers and observed trait variation (Stranger *et al*, 2011). The use of this has shed additional light on the mechanisms of complex traits (Sharma *et al*., 2015), as well as quantifying diversity among populations (Tellam *et al.*, 2009).

The majority of GWAS involving the use of SNPs in cattle species have been focussed on traits such as milk production (Saowaphak *et al.*, 2017), growth (Snelling *et al.*, 2009; Martínez *et al.*, 2017) and carcass characteristics (Hay & Roberts, 2018). The main difference between these studies seems to be the definition of trait of interest. This is most probably due to differences between breeding goals, recording measures and species or breed differences.

Reviews by Hawken *et al.*, (2012) and Fortes *et al.*, (2013) highlight the need for a fertility genome reference and concluded that the X chromosome is associated with male and female reproductive traits. Regatieri *et al.*, (2017) reported that SNPs on seven chromosomes were significantly associated with sexual precocity in Nellore heifers. Scrotal circumference was assessed alongside the expression of hormones released that affect its subsequent performance (Fortes *et al*., 2012). The numerous findings of similar genomic regions for different fertility traits indicates the complexity of reproduction and that further research is needed for a greater understanding.

Aim of the study

The SA Bonsmara breed produces weaners with favourable growth traits for feedlot conditions, which is evident due to their dominant position within the SA beef industry (SA Stud Book, 2019). The reputation of the SA Bonsmara is world renowned, with foreign herds occurring in Argentina, Brazil, Namibia, Zambia and the USA (SA Stud Book, 2016). The number of pedigree recordings, accurate phenotypic information, correct implementation of EBVs across numerous herds and the number of genotyped animals make the SA Bonsmara breed a perfect candidate for GWAS.

In this study the focus will be on fertility and maternal traits, as a reduction in non-pregnant periods in the female's reproductive life would positively impact production. The traits to be studied include age at first calving (AFC), inter-calving period (ICP), weaning weight maternal (WW_{MAT}) which is also known as "milk" to beef breeders and scrotal circumference (SC). ICP is known as the number of days between successive calvings. Berry *et al.*, (2014b) states that ICP involves the ability of the animal to be able to resume oestrous as soon as possible post-calving, express oestrous sufficiently for detection, to conceive, establish pregnancy and maintain pregnancy in the required gestation length. AFC is recorded as it is genetically correlated with calving interval and SC. SC is a measure of bull fertility and is negatively correlated with AFC, indicating that sires with larger SC produce heifers that calve at a younger age.

WW_{MAT} indicates a cow's ability to provide milk to the growing calf and reflects her ability to gain or lose weight during this physiologically demanding period. Animals that gain weight in this period are more sought after as this indicates that the cow had sufficient body reserves to supply adequate milk, as well as to gain body mass. This in turn results in the cow coming into oestrous cycling sooner, affecting the ICP as cows that need to replenish body reserves will return to

oestrous at a later period (Berry *et al.*, 2014b). This increases the likelihood of non-pregnant years, reducing overall herd reproductive performance.

The overall aim of this study was to conduct the first Genome-wide association study on the SA Bonsmara breed to investigate the underlying genetics of fertility and maternal traits.

The following objectives were set in order to reach the aim of this study:

1. Perform a GWAS to evaluate genetic markers significantly associated with fertility and maternal traits of interest.
2. Annotation of biological processes and molecular functions of genes potentially associated with significant genetic markers.

1.2 Literature Review

1.2.1 Introduction

Observable traits of economic importance are typically known to be affected by multiple genes, which contribute to the polygenic expression of these traits. The application of genomics, with regards to the accessibility and cost-factor of molecular analysis, has become a viable tool in the investigation of these polygenic traits. The true underlying genetic architecture of most traits remains unknown (Dekkers & Hospital, 2002; Stranger *et al.*, 2011) and genomic tools have now opened opportunities for various approaches to improve our understanding (Meuwissen *et al.*, 2001), which may help explain any physiological or pathogenic conditions that remain unanswered.

An understanding of the genetic basis of fertility traits is required to implement appropriate selection programs that may increase reproductive efficiency (Miar *et al.*, 2015). The application of genomic selection (GS) in livestock breeding allows for the reduction of the generation interval by up to 2 years, which could likely result in a 60 to 120% increase in the rate of genetic gain (Hayes *et al.*, 2013).

In this chapter, a brief overview of the SA beef industry will be provided and a discussion of fertility and maternal traits of interest. This will be followed by a review of relevant literature regarding the current and potential applications of GWAS and the use and manipulation of available phenotypic data with regards to appropriate mixed model equations (MME).

1.2.2 Overview of the South African beef industry

Sub-Saharan Africa has been recognized as an important resource in terms of genetic diversity within the Bovidae species. There are over 180 cattle breeds present, of which 150 are

deemed indigenous, in Southern Africa. The introduction of major foreign *Bos taurus* beef and dual-purpose breeds in the first half of the 20th century to South Africa was aimed at trying to establish herds with known economically favourable carcass traits (van Marle, 1974).

The exotic breeds imported at the time proved to be poorly adapted to the more subtropical regions of SA (van Marle, 1974) and the need was recognised to develop a composite breed where the superior traits of the imported breeds could be combined with the local adapted types (Bonsma, 1980). This resulted in the creation of various SA composite breeds.

The livestock industry contributes 34.1% to the total domestic agricultural production, contributing over R127 million to the Gross Domestic Product of SA between 2016-2017 (van Marle-Köster & Visser, 2018) and provides 36% of the population's protein needs (Organisasie & RPO, 2017). With an estimated per capita beef consumption of 19.2 kg/year (ARC, 2016), this amounts to 3 678 000 cattle being slaughtered annually (DAFF, 2017). More than 75% of beef cattle slaughtered in the formal sector are finished in feedlots (Organisasie & RPO, 2017), although primary beef cattle farming (the cow-calf production cycle) is mostly extensive in SA. The commercial sector mainly focuses on maximising the number of weaned calves for a given number of cows under the prevailing environmental and management conditions (Rust & Groeneveld, 2001), that will be subsequently sold to the feedlot. The sire-dam complementarity derived from the seedstock sector allows growth and carcass traits to dominate the breeding objective and selection criteria within this sector. Thus, the combination of highly fertile and adaptable females with fast growing males produces a calf that is highly suitable to conditions in the SA feedlot industry.

The common practice of crossbreeding indigenous and exotic breeds has resulted in 66,4% of herds in the emerging sector to be comprised of crossbred/non-descript cattle (Scholtz *et al.*, 2008). Table 1.1 summarises the cattle types and breeds used in beef production in South Africa. Composite breeds, Bonsmara and Beefmaster, have the highest head counts (SA Stud Book, 2019), indicating the industry's perception that these are superior adapted medium maturing type animals. The local Sanga types include the Nguni, Drakensberger, Tuli and Afrikaner breeds, which are known to be highly adaptable and resistant to harsh climate and high disease prevalence. The Drakensberger and Tuli breeds are early-maturing, medium-framed animals and are characterized to show good performance on a range of different grazing conditions. The Nguni is historically a multipurpose breed (Musemwa *et al.*, 2008), used for its milk, meat and the hide. These Sanga cattle are present in both small holder (developing) and commercial sectors in SA.

Table 1.1 Summary of the number of active stud registered animals (SA Stud Book, 2019) and respective averages (SA Stud Book, 2016) for traits of interest in this study.

Breed	Females	Males	Total	Maturing Type	AFC (Months)	ICP (Days)	WW_{MAT} (kg)	SC (mm)
British Origin								
Angus SA	13 414	4 883	18 297	Early	31.3	408.6	541	344.4
Hereford	3 752	1 498	5 250	Early	31.9	391.3	568	328.8
Sussex	4 299	1 705	6 004	Early	32.4	408.3	585	318
European Origin								
Braunvieh	581	171	752	Late	33.6	450.9	491	344.8
Charolais	3 240	1 286	4 979	Late	33.7	422.4	596	347
Pinzgauer	489	192	681	Late	36.2	467.1	503	341.9
Sanga Types								
Drakensberger	9 496	3 892	13 388	Intermediate	34.1	424.1	507	327.1
Nguni	16 037	5 269	21 306	Early	32.3	415.4	366	-
Tuli	5 572	1 987	7 559	Early	34.9	423	429	-
Composite Types								
Beefmaster	32 457	13 944	46 401	Intermediate	30.6	404.6	509	352.2
Bonsmara	77 917	30 281	108 198	Intermediate	31.4	413.7	506	339.6
Total*	201 108	69 305	270 413		32.1	417.3	500.1	337

*Includes animals not listed in this table

Development of the SA Bonsmara

An initial climatological study carried out at the Messina Research Station was tested on progeny of varying *Bos taurus* x *Bos indicus* compositions. The genesis of the Bonsmara breed commenced at the Mara research Station in 1940 (Bonsma, 1980). The Indicine proportion was decided to be allocated to the Afrikaner breed, a Sanga type that originates from Central Africa and shares an ancestral population with the Nguni breed, another Sanga type (Makina *et al.*, 2014). During the development of the Bonsmara, considerable emphasis was placed on the selection for adaptive ability (Bonsma, 1980).

The two Taurine breeds selected were Dairy Shorthorn and Hereford. The Shorthorn was chosen for its good milk production, earlier maturing age and excels in sweet veld pasture utilization, while the Hereford was selected due to its good temperament as well as good utilization of natural sour veld pastures. These parental breeds were further selected according to the following characteristics; smooth coats and thick hides, outstanding beefiness in two of the

parental breeds, cows with good milk-ability and bulls from highly fertile bulls that showed good temperament (Bonsma, 1980). The first-generation crosses were Afrikaner cows mated to Herford and Shorthorn bulls. These resultant F1 bulls were again mated to Afrikaner cows to obtain large numbers of 3/4 Afrikaner and 1/4 exotic. Mating's 7 & 8 produced the original 5/8 Afrikaner and 3/8 exotic breed. From mating's 9, 10 and 11 onwards we speak of these progenies as Bonsmara, a composite breed in its own right (Bonsma, 1980).

Currently one of the most prominent beef cattle breeds in SA, with over 108 000 registered cattle, the Bonsmara has had economically important traits positively selected as well as visual evaluation for selection in functional efficiency. This concept of "functional efficiency" is based on the hypothesis that strict selection of phenotypic traits may have an influence on an animal's environmental adaptability and consequently improving the animal's production and reproduction potential (Bonsma, 1980; Webb *et al.*, 2017). To enable a greater understanding for the premise of adaptive traits in an animal's own environment, knowledge of breed composition of specific cattle breed populations may be useful in the prediction of the degree of heterosis and in turn enabling long-term sustainability of genetic resources through proper management (Gorbach *et al.*, 2010). It is notable that when Sanga cattle were originally brought to Southern Africa by the indigenous Khoi-San, the Nguni breed settled on the eastern side of SA and the Afrikaner breed on the western side respectively (Scholtz *et al.*, 2011). However, the development of the SA Bonsmara breed occurred in the eastern region of SA, which was mainly populated by Nguni cattle at that time. This was illustrated by Makina *et al.*, (2014) when comparing population structure of indigenous Sanga types, the Bonsmara shared a higher proportion of genetic links with the Nguni breed (3%) compared to the Afrikaner breed (0,5%), respectively.

1.2.3 Selection for fertility and maternal traits

The focus of this section will be an overview on maternal and fertility traits that are measured in beef cattle. Beef cattle traits are categorised into multiple groups, which can be mainly grouped into fertility, production, quality and disease traits (Kinghorn *et al.*, 2014). Fertility traits differ, in terms of definition and method of measurement, on the breed type of cattle and/or regulations set out by specific breed societies.

Fertility is a measure of reproductive success. Male fertility is the ability of the bull to produce semen that would result in a successful pregnancy (Foote, 2003; Nino-Soto & King, 2004). Female fertility can be defined as a cow's ability to conceive successfully, calve down and provide sufficient milk to wean a suckling calf (Davis, 1993; Nino-Soto & King, 2004; Berry & Evans, 2014). Fertility traits can be classified into three broad categories, namely; interval traits, binary traits and count traits (Berry *et al.*, 2014b). Binary traits are phenotypes that are controlled by quantitative genetic interactions with environmental effects determining the threshold for how the trait is expressed qualitatively. Non-return rate is the proportion of mated cows that do not return to oestrous within a certain timeframe and are determined to be pregnant. Thus, non-return rate is measured as pregnant (1) or not pregnant (0) and is therefore considered a binary trait. Count traits, such as number of services or inseminations to pregnancy, are used to determine the level of management needed per cow, with less labour-intensive cows being favoured and selected for. An example would be animals that conceive after only one service, which are more favourable than cows needing more than one service to achieve conception. Interval traits tend to have greater heritability estimates compared to count and binary traits (Berry *et al.*, 2014b).

Some fertility traits are easy to record and involve low cost managerial input from farmers (Rust & Groeneveld, 2001), while most fertility traits, like maternal weaning weight or new traits using ultrasound (Corbet *et al.*, 2018), are quite difficult to consistently measure and weigh correctly. Accurate recording is essential for fertility traits as current genetic variation within herds aids in on-farm management (Berry *et al.*, 2014b) and is used to decide the optimal breeding objective to maximise genetic gain on these lowly heritable traits.

The role of measurement and animal recording in breeding programs depends on the degree of intensiveness of the livestock enterprise (James & Roberts, 1979; Holst, 1999). Highly intensive husbandry systems allow for greater opportunities for recording of animals to occur compared to more extensive systems. Traits that influence productivity influence the income and cost of a beef herd. The beef production industry mainly relies on reproductive efficiency, thus a reduction in unproductive periods in the female's reproductive life would positively impact

production costs. This has typically been slow due to the low heritability among other previously mentioned shortfalls when it comes to improving fertility traits (Cammack *et al.*, 2009; Granleese *et al.*, 2015). Therefore, genomic association analyses of fertility traits may result in the identification of new or existing SNPs that iterate there may be an underlying gene or QTL action influencing their expression. Well defined traits are vital for accurate quantification of the heritability, possible correlations and any future genetic associations.

Beef breeding covers two essential areas, these being genetic improvement of the weaning cow herd and the use of appropriate sires with the overall goal of achieving fertility targets to improve herd productivity (www.teagasc.ie). Genetic gains achieved within a breeding herd are cumulative and remain in the herd indefinitely. Every beef farmer needs to have a breeding objective to ensure that the next generation of seedstock is genetically and phenotypically superior to that of the previous generation. An ideal breeding objective includes selecting traits that will in future, be of the most influence towards the breeding goal (Garrick, 2011). This is essential in order for a farmer to maximise future return on investment (Ponzoni, 1986).

The Logix national evaluation scheme has created a standardised method of trait recording for all participating herds across Southern Africa. This enables researchers to model specific population and genetic effects across multiple breeds in order to quantify breed and genetic diversity within the cattle genome (Makina *et al.*, 2014, 2016; Zwane *et al.*, 2016; Gororo *et al.*, 2018; Lashmar *et al.*, 2018a; Pienaar *et al.*, 2018).

Predictor traits are traits that are highly genetically correlated with traits that are expressed later in life or are hard to measure. These are used to predict the future performance of an animal and enables early estimation of an animal's phenotypic potential. Due to the low heritability of fertility traits and extended time periods it takes to measure them, and indicates that the use of predictor traits would be a benefit to selection accuracy (Berry & Evans, 2014). Actual recording of reproductive phenotypes is necessary to resolve any genetic or phenotypic antagonisms with other traits and enable breeders to achieve a higher accuracy of genetic selection.

Related traits are recorded and evaluated in conjunction with the weaning of the calf and encompasses the overall productivity of the cowherd (du Plessis *et al.*, 2006). Eler *et al.*, (2008) states that cows that wean consistently light calves must be culled, indicating that direct weaning weight has a high repeatability. These traits allow us to identify females that are exceptionally good at cycling after calving as well as those that wean heavier calves compared to the breed average.

An unfavourable genetic correlation exists between birth weight (BW) and calving ease (CE), with higher BW of calves associated with higher incidences of dystocia (Cammack *et al.*, 2009). Calving success (CS) is defined as the birth of a live calf without birthing difficulties (van der Westhuizen *et al.*, 2001a). The recording of performance traits has been essential in the calculation of EBVs. Genetic conditions and disorders that affect the overall health of livestock animals have an unfavourable effect on animal welfare and decreases the potential profitability within the industry. The investigation via GWAS into genomic regions associated with these diseased conditions are underway, but mainly on European breeds (Visscher *et al.*, 2017; Freebern *et al.*, 2020), with the goal of eradicating affected individuals from the reproductive population.

In Tables 1.2a and b, genetic and phenotypic correlations for female fertility traits from several studies that included different breeds have been summarized.

Table 1.2a Summary of genetic and phenotypic correlations between maternal and fertility traits

Traits	Genetic	Phenotypic	Breed	Reference
AFC/Calving Date	0.09		Composite beef	(van der Westhuizen et al., 2001b)
	0.88	0.05	Pooled data	(Corbet et al., 2006)
AFC/ICP	-0.03		Composite beef	(van der Westhuizen et al., 2001b)
	0.44	-0.28	Pooled data	(Corbet et al., 2006)
	-0.05		Brahman	(Cavani et al., 2015)
	0.22		Crossbred beef	(Berry & Evans, 2014)
AFC/WW_{MAT}	0.35		Crossbred beef	(Berry & Evans, 2014)
AFC/Stayability	-0.38		Brahman	(Cavani *et al*., 2015)
CS/Calving Date	-0.95		Afrikaner	(Beffa, 2005)
CS/Stayability		0.14	Composite beef*	(van der Westhuizen *et al.*, 2001a)
		0.04	Composite beef*	(van der Westhuizen *et al.*, 2001a)
		0.2	Composite beef*	(van der Westhuizen *et al.*, 2001a)
DAYTC/Pregnancy	-0.99	-0.42	Tropical Composites	(Burrow, 2001)
Direct				
Dystocia/AFC	0.28		Crossbred beef	(Berry & Evans, 2014)
Dystocia/ICP	0.18		Crossbred beef	(Berry & Evans, 2014)
Dystocia/Survival	-0.06		Crossbred beef	(Berry & Evans, 2014)

*measured at 36, 60 and 84 months of age respectively
DAYTC = Days to calving

Table 1.2b Summary of genetic and phenotypic correlations between maternal and fertility traits

Traits	Genetic	Phenotypic	Breed	Reference
Maternal				
Dystocia/AFC	-0.22		Crossbred beef	(Berry & Evans, 2014)
Dystocia/ICP	0.32		Crossbred beef	(Berry & Evans, 2014)
Dystocia/Survival	-0.65		Crossbred beef	(Berry & Evans, 2014)
ICP/Calving Date	0.75		Composite beef	(van der Westhuizen et al., 2001b)
	0.01	-0.02	Pooled Data	(Corbet et al., 2006)
ICP/MWW	-0.21		Crossbred beef	(Berry & Evans, 2014)
ICP/Survival	-0.63		Crossbred beef	(Berry & Evans, 2014)
SC/Age at Puberty	-0.41		Brahman#	(Johnston *et al.*, 2014b)
	-0.15		Assorted	(Martinez-Velazquez *et al.*, 2003)
SC/DAYTC	-0.25		Hereford	(Meyer *et al.*, 1991)
	-0.28		Angus	(Meyer *et al.*, 1991)
	-0.41		Zebu crossbreds	(Meyer *et al.*, 1991)
	0.32		Tropical Composites	(Burrow, 2001)
SC/AFC	0.15		Assorted	(Martinez-Velazquez *et al.*, 2003)
SC/Pregnancy Rate	0.27		Brahman+	(Johnston *et al.*, 2014b)
	-0.12		Tropical Composites	(Burrow, 2001)
	-0.17		Tropical Composite+	(Johnston *et al.*, 2014b)
SC/Weaning Rate	0.4		Brahman+	(Johnston *et al.*, 2014b)
	0.35		Tropical Composite+	(Johnston *et al.*, 2014b)

*measured at 36, 60 and 84 months of age respectively
+measured at 18 months of age
#measured at 12 months of age
DAYTC = Days to calving

Residual variation is caused by both known and unknown environmental effects, with yet unexplained additive and non-additive genetic effects. The interaction effects between environmental and genetic effects (GyE) varies, due to genetic differences between breed and epigenetics. Epigenetics refers to all alterations in DNA function while no changes have been made in the DNA sequence and cause variation in trait expression. This is evident due to the threshold nature of pregnancy, with the commercial production need for high pregnancy rates exacerbating this interaction (Cammack *et al.*, 2009). Environmental effects include temperature, exposure to radiation, level of nutrition, ecto-and-endoparasites, disease type and prevalence. These contribute and influence the expression and variability of all traits.

Female fertility traits for various different breeds are summarised in Tables 1.3a and b, with the respective heritability estimates.

Table 1.3a Summary of heritability estimates for female fertility traits, according to breed

Trait	Breed	Heritability	Reference
Age at First Calving	Afrikaner	0.27	(Meyer *et al.*, 1990)
	Bonsmara cross	0.13	(Corbet *et al*., 2006)
	Brahman	0.1	(Cavani *et al*., 2015)
	Composite beef	0.4	(van der Westhuizen *et al*., 2001b)
	Assorted	0.16	(Martinez-Velazquez *et al.*, 2003)
	Crossbred beef	0.31	(Berry & Evans, 2014)
	Drakensberger	0.3	(Meyer *et al*., 1990)
	Nellore	0.09-0.28	(Costa et al., 2015)
	Nellore	0.11	(Kluska *et al.,* 2018)
	Nellore	0.08	(Melo et al., 2018)
Calving Date	Bonsmara cross	0.02	(Corbet *et al*., 2006)
	Composite beef	0.04	(van der Westhuizen *et al.*, 2001b)
Calving Interval	Bonsmara cross	0.04	(Corbet *et al*., 2006)
	Brahman	0.02	(Cavani *et al*., 2015)
	Composite beef	0.01	(van der Westhuizen *et al.*, 2001b)
	Crossbred beef	0.02	(Berry & Evans, 2014)
Calving Success	Afrikaner	0.08	(Beffa, 2005)
	Composite beef	0.03	(van der Westhuizen *et al.*, 2001a)
	Hereford	0.08	(Meyer *et al.*, 1990)
	SA Angus	0.02	(Meyer *et al.*, 1990)
	Zebu crosses	0.08	(Meyer *et al.*, 1990)
Calving Rate	Hereford	0.07	(Meyer *et al.*, 1990)
	SA Angus	0.02	(Meyer *et al.*, 1990)
	Zebu crosses	0.17	(Meyer *et al.*, 1990)
Days to Calving	Hereford	0.05	(Meyer *et al*., 1990)
	SA Angus	0.08	(Meyer *et al*., 1990)
	Tropical Composite	0.07	(Burrow, 2001)
Dystocia	Crossbred beef	0.25	(Berry & Evans, 2014)

Table 1.3b Summary of heritability estimates for female fertility traits, according to breed

Trait	Breed	Heritability	Reference
Lifetime Annual Calving Rate	Brahman	0.16	(Johnston *et al.*, 2014a)
	Tropical Composite	0.04	(Johnston *et al.*, 2014a)
Lifetime Annual Weaning Rate	Brahman	0.11	(Johnston *et al.*, 2014a)
	Tropical Composite	0.07	(Johnston *et al.*, 2014a)
Longevity	Composite beef	0.08	(van der Westhuizen *et al.*, 2001a)
Weaning Weight Maternal	Afrikaner	0.21	(Beffa, 2005)
	Belmont Red	0.14	(Corbet *et al*., 2006)
	Bonsmara	0.19	(Corbet *et al*., 2006)
	Nguni	0.16	(Norris *et al.*, 2004)
	SA Brangus	0.11	(Neser *et al.*, 2012)
Pregnancy	Tropical Composite	0.04	(Burrow, 2001)
Stayability	Brahman	0.1	(Cavani *et al*., 2015)
	Composite beef	0.03 - 0.11	(van der Westhuizen *et al.*, 2001a)
	Crossbred beef	0.02	(Berry & Evans, 2014)

Male fertility traits include SC, libido and others. SC is a trait that is easy to measure and in Table 1.4, heritability estimates for SC that were previously reported in literature were summarized.

Table 1.4 Summary of heritability estimates for Scrotal Circumference (SC)

Breed	Heritability	Reference
Bonsmara	0.44 - 0.46	(SA Studbook, 2016)
Tropical Composite*	0.41	(Burrow, 2001)
Brahman*	0.75	(Corbet *et al.*, 2013)
Tropical Composite*	0.42	(Corbet *et al.*, 2013)
Assorted breeds	0.41	(Martinez-Velazquez *et al.*, 2003)
Brahman+	0.75	(Melo *et al.*, 2018)
Nellore+	0.41	(Melo *et al.*, 2018)
Hereford	0.29	(Meyer *et al.*, 1991)
Hereford	0.71	(Evans *et al.*, 1999)
SA Angus	0.47	(Meyer *et al.*, 1991)
Nellore#	0.48	(Kluska *et al.*, 2018)
Zebu crosses	0.28	(Meyer *et al.*, 1991)

*measured at 36, 60 and 84 months of age respectively
+measured at 18 months of age
#measured at 12 months of age

1.2.3.1 Fertility traits

Female traits

It is well-known that *Bos taurus* females reach puberty at younger ages than *Bos indicus* females. Puberty is experienced earlier in composite and crossbred animals compared to their purebred counterparts with the same trend seen for early-maturing versus late-maturing beef breeds (Bourdon, 2000; Sartori *et al.*, 2010). This trait has an overall effect on reproductive performance as the heifer will become productive at an earlier age. This late onset of puberty in all beef heifers is one of the main factors that increases production costs (Costa *et al.*, 2015). A reduction in this would allow for earlier insemination and AFC, but negative correlations can result in an increased ICP due to a lower WW_{MAT} and may result in reducing longevity. The continuous development of selection indices within the beef industry is ongoing as these indices simultaneously account for a multitude of different traits and consider both the biological impact and subsequent effect on production levels from an economic perspective (Parish *et al.*, 2011; Kluska *et al.*, 2018). As with most selection indexes, there is a need to fine-balance selection ing conjunction with genetic and phenotypic co_rrelations in order to instil maximum genetic gains.

Age at first calving is measured as the number of days from birth to first calving (Berry & Evans, 2014), and as this trait is easy to measure it is often included as a selection objective. Early-maturing breeds are known to experience higher levels of dystocia, as AFC is normally experienced at 85% mature body weight, with the beef breed industry aiming to calve heifers

between 23 to 25 months of age (Wathes *et al.*, 2014). Corbet *et al.*, (2006) analysed a pooled data set of beef cattle and observed a genetic correlation (r_g) between AFC and ICP of 0.44. This indicates that animals with earlier AFC experience shorter ICPs. This is corroborated by Berry & Evans, (2014) observation of an r_g of 0.22 in crossbred beef cattle. This is reported differently in other studies (van der Westhuizen *et al.*, 2001b; Cavani *et al.*, 2015), that report a slight negative r_g between AFC and ICP. Inter-Calving Period (ICP) is known as the number of days between successive calvings and has an industry target of around 365 days. Maintaining this interval is vital to maximise the utilisation of lower-cost grazed grass, especially in SA where most beef cattle exist in natural pasture based extensive systems. Days to Calving (DAYTC) is the number of days between successful insemination or service and subsequent calving. This trait has been reported to have a low h^2. Animals with shorter DAYTC will have subsequently shorter ICP's, with a high favourable genetic correlation indicating this inference.

Berry *et al.*, (2014b) states that ICP involves the ability of the animal to resume normal cyclicity post-calving, to express oestrous sufficiently for detection, to conceive, establish and maintain pregnancy in the required gestation length. Inter-calving period is estimated to have a low heritability (Table 1.3a). Berry & Evans, (2014) found a low heritability for calving interval in beef cattle. Mkhize *et al.*, (2019) indicated that the ICP of first-calf cows is generally longer than that of multiparous cows. The postpartum period is essential for the re-establishment of ovarian activity with proper nutrition allowing the preparation of re-conception to be shortened (Mukasa-Mugerwa, 1989). Seasonal breeding of herds is common practice in South Africa. This method allows for strict observation of cows that are reproducing and by ensuring a cow is in good body condition post-partum will reduce calf mortalities (Mkhize *et al.*, 2019) and ensuring an appropriate ICP. ICP of older cows is known to be shorter compared to younger cows but this may be attributed to less productive cows being culled before reaching a certain age. Therefore, this shorter ICP may be due to the cows being measured being the superior reproductive animals.

An interesting phenotypic correlation in heifers was observed (Bourdon, 2000), that those with a later AFC, have subsequently shorter ICPs in the following calvings. This indicates that the animal's overall reproductive efficiency may increase through prolonging the initial breeding of late maturing heifers. The length of gestation is highly associated with ICP, with shorter gestation lengths associated with shorter ICP (Kirkpatrick, 2014). Therefore, we can assume that selecting animals for a shorter ICP may negatively affect AFC (Berry & Evans, 2014).

Maternal weaning weight (WW_{MAT}) indicates the cow's ability to provide sufficient milk to the growing calf and her ability to gain or lose weight during this physiologically demanding period

(Meyer, 1997). In models used to determine heritability's and correlations, it is included as a genetic component of the dam and permanent maternal environmental effect on her calf. Animals that gain weight in this period are more sought after as this indicates that the cow had sufficient body reserves to supply adequate milk. This in turn results in the cow coming into oestrous cycling sooner, affecting the ICP as cows that need to replenish body reserves will return to oestrous at a later period (Berry *et al.*, 2014b). This is increasing the period of unproductivity and reduces overall herd reproductive performance. This trait affects cow-calf efficiency, which can be defined as kg calf weaned per Large Stock Unit (KgC/LSU) mated, with the Bonsmara breed showing a 10.0% increase over a period of 33 years (Mokolobate *et al.*, 2018) due to proper selection practices and management of breeding resources. Mokolobate *et al.*, (2018) stated that cow productivity will be improved if the weaning weight of the cow can be increased in relation to the weight of the cow.

Inter-calving period and WW_{MAT} are negatively favourably correlated, with a r_g of -0.21 being reported (Berry & Evans, 2014). This implies cows that experience a gain in WW_{MAT} will experience a shorter ICP and will recycle sooner, allowing for earlier insemination for subsequent pregnancy. A loss in WW_{MAT} will result in an unwanted extension of the ICP. Messine *et al.*, (2004) noted that excessive suckling was a factor that contributed to longer ICPs, which correlates to the reasoning that a loss in WW_{MAT} will cause an extended ICP.

Correlations between AFC and WW_{MAT} are limited to a few studies, but evidence suggests that animals that experience earlier AFC will have a decreased WW_{MAT} (Berry & Evans, 2014). This will negatively affect the ensuing ICP, increasing unproductivity as a result of the cow needing to gain enough weight before recycling. This may be an indicator for the r_g between AFC and ICP (Corbet *et al.*, 2006; Berry & Evans, 2014), where a younger AFC results in a decreased WW_{MAT} that unfavourably increases the ICP. Calving rate, defined as the number of calves born divided by the number of opportunities the cow has had to calf (Rust & Groeneveld, 2001) is a good indication of overall lifetime productivity. Calving Ease (CE), the component trait that indicates the level of dystocia, is mainly correlated with the size of the dam and the BW of the calf. Difficult calvings are known to increase the ICP in cows, thus high levels of dystocia will negatively affect the overall calving rate. Dystocia is selected against by selecting terminal sires with favourable EBVs for CE and BW as well as breeding of heifers at the appropriate age.

Primiparous females can experience dystocia via feto-pelvic disproportion sometimes due to the high birthweight of the calf. Foetal malpresentation is the most common source of dystocia in mature, multiparous cows (Kirkpatrick, 2014). Medium and late-maturing beef breeds

experience lower rates of dystocia and thus are more highly favoured in high-throughput breeding programs. Gutiérrez *et al.*, (2002) stated that a decreased lifetime performance is associated with a later AFC in a dual-purpose Taurus breed. As observed in Table 1.3a, AFC has the highest reported h^2 estimate, but also has the largest range, indicating high genetic and environmental variability of AFC h^2 estimates across various breeds. This trait is recorded as it is highly genetically correlated with calving interval and with age at subsequent calving's.

Body Condition Score (BCS) is a trait that describes the relative fatness of a cow. It is subjectively measured but guidelines do exist to aid recorders in determining the animals BCS. BCS ranges from 0 to 5, with 5 being overly fat and 2,5 being average (Nephawe *et al.*, 2004). The genetic correlations between BCS and reproductive performance are favourable, with overly heavier animals becoming compromised (Rasby *et al.*, 2014; Berry & Evans, 2014). Animals with poor BCS are also associated with poor reproductive performance as they may require a longer time to resume normal oestrous after calving, which will negatively affect the ICP (Berry *et al.*, 2014b; Rasby *et al.*, 2014; Hlatshwayo, 2015). To maintain the desirable 365-day ICP, a cow must recycle back into oestrous and be serviced by day 83 after calving. This, in addition to the 282-day average gestation length, will meet the profitable ICP. Average length of the postpartum interval for cows with poor BCS that calve is about 80 days, compared to 55 days for cows at the desirable BCS who express higher rebreeding rates (Rasby *et al.*, 2014). Berry & Evans, (2014) state that the absence of genetic studies attempting to associate body fatness (BCS) with reproductive performance of beef cows' points to a gap in vital knowledge. A few phenotypic studies (Selk *et al.*, 1988; Berry & Evans, 2014) have shown some evidence that demonstrates an association between reproductive performance and change in BCS during parities which can be corroborated with the multiple studies on BCS in dairy cattle (Roche *et al.*, 2009). A negative genetic correlation of -0.44 to -0.31 between ICP and carcass subcutaneous fat depth was observed (Berry & Evans, 2014), indicating that carcass traits and reproductive traits are antagonistically correlated. The genetic merit of ICP is primarily deteriorating due to aggressive selection for a larger mature body size, carcass conformation and carcass yield (Phocas, 2009; Berry *et al.*, 2014b).

Webb *et al.*, (2017) concluded that production and reproductive efficiency of Bonsmara cows was affected by the production region (bioregion) that they were present in. This indicates a geographical influence on cow size and is corroborated by farmers who believe that there is a tendency for Bonsmara cows in the western regions of SA to be larger and more reproductively efficient than their eastern counterparts (Webb *et al.*, 2017). WW_{MAT} and WW direct have a

repeatability ranging from 0.7054 to 0.7182 (Silva *et al.*, 2015), which one can deduce that the calf will be directly affected by the dams' ability to supply adequate milk-ability while maintaining bodyweight.

Pregnancy Rate (PR), a binary trait, is determined as either pregnant = 1(PREG); or not pregnant = 0. Burrow, (2001) defines pregnancy (PREG) as a threshold trait. This can allow the farmer to identify cows that don't become pregnant after multiple conception attempts. These cows should be culled as animals with poor genetic merit for fertility must be eliminated from the breeding herd. Cows that become pregnant after one insemination or service are highly sought after. Lifetime PR is the total number of pregnancies a cow has had across all her possible mating years (Berry *et al.*, 2014b). Pregnancy rates are highly genetically correlated with WW_{MAT}, inferring that the condition of the dam post-weaning affects her ability to subsequently become pregnant again. A negative genetic correlation between pregnancy and scrotal circumference (SC) in Tropical Composite beef breeds (Burrow, 2001; Johnston *et al.*, 2014b) in Table 1.2b indicates that sires with larger SC produce heifers with higher pregnancy rates.

Lifetime annual weaning rate (LAWR) is the total number of calves weaned from all matings divided by the total number of mating's experienced, for all cows in the herd (Johnston *et al.*, 2014a). Stayability is known as the probability of an animal surviving to a specific age, given every opportunity to reach that age (van der Westhuizen *et al.*, 2001a). This trait is often construed with longevity of the animal, the total lifetime of production up to death. A few heritability estimates are given for both traits in Table 1.3b.

These highly variable correlations across all traits illustrates the complexity of accurate record keeping, proper genetic and environmental characterisation as well as the evaluation of GyE interactions. The low number of records for certain breeds could lead to a biased interpretation (SA Stud Book, 2016), indicating that only better performing animals are actually a majority of the recording process. SA Bonsmara, with over 108 000 recorded animals, provide a more accurate estimation of population trait averages. The high number of recorded phenotypes provides a more accurate assessment of the variation amongst the Bonsmara breed (Makina *et al.*, 2016).

Male traits

Fertility measures for bulls are limited and mainly include testicular characteristics used primarily to determine sperm output as sperm production per unit testes volume is constant (Parkinson, 2004). The male measured traits associated with reproduction are moderately to highly correlated with each other, but these correlations are population dependant. Differences exist and can be attributed to many factors such as sampling variation and certain population specific characteristics like selection pressure (Berry *et al.*, 2013). Certain genetic correlations may lack precision due to a combination of the lowly heritability of reproductive traits and an insufficient number of individuals with accurate phenotypes (Berry *et al.*, 2014b) to constitute an acceptable population size.

Bulls that produce daughters with incomplete records and length of exposure to a group of females have unreliable genetic estimates. The smaller the group of females, the more superficially increased these measures are, due to the effects of sampling bias and small sample sizes. Reproduction is a function of the underlying genetic forces that affect the endocrine system and physiological factors that may contribute towards the overall expression of these traits. Animals that fail to produce offspring every year are not profitable. An animal will only compensate the cost of its maintenance through either selling of the offspring it produces or selling of that animal itself.

SC is the most used trait for male fertility. It is easily measured, either at 12, 18 or 24 months of age, using a standard measuring tape at the widest point of the scrotum. A SC of 28-30cm is associated with an onset of puberty in 52-97% of bulls (Parkinson, 2004). The quality and quantity of spermatozoa producing tissue directly contributes to a bulls' overall fertility. Table 1.4 indicates that h^2 estimates for SC are quite high.

SC has been negatively genetically correlated with days to calving as indicated by Table 1.2b. A highly negative genetic correlation between SC and age at puberty has been reported (Evans *et al.*, 1999), with a high variability across experiments and breeds. This decreased days to puberty could potentially increase lifetime productivity rates of dams (Burrow, 2001). As SC is directly related to the reproductive performance of bulls and the sexual precocity of their female progeny (da Silva Romero *et al.*, 2018), finding an optimum SC will increase production efficiency. The genetic correlation between SC and reproductive performance of female animals has been recorded to be weak (Martinez-Velazquez *et al.*, 2003), which contradicts the aforementioned statements and indicates the variability of results in similar studies.

SC has been positively genetically correlated with certain growth traits, which are favourable in medium to late maturing beef breeds used in current feedlot conditions. Burrow, (2001) reported a genetic correlation of 0.37 - 0.40 between SC and mature BW. Meyer *et al.*, (1991) noted that the genetic and phenotypic correlations between SC of sires and growth rates in their progeny ranged from 0.24 – 0.52 for Angus and Hereford breeds. It was discovered that Zebu type crosses had higher genetic correlation of 0.65 – 0.69. This indicates that sires with larger SC produce progeny with higher growth rates.

Beef fertility tends to vary due to genetic and non-genetic factors, with reported per service calving rates of between 50-60% (Cammack *et al.*, 2009). Grobler *et al.*, (2014) estimated the calving percentage of SAs beef commercial sector to be 62%. As discussed, there are many unfavourable genetic correlations between production and reproductive traits with production being highly selected for in previous generations, and reproduction being subsequently negatively affected. Correlation estimates differ between populations, mainly due to differences in trait definition, environment and the statistical model used (Berry & Evans, 2014).

1.2.4 Estimated breeding values (EBVs)

Bourdon, (2000) defines the breeding value as the value of an individual as a contributor of genes to the next generation. In other words, it can be described as the individual's genotypic value or genetic merit that is due to independent additive genetic effects that are transmitted from parent to offspring.

The calculation of EBVs at SA Stud Book makes use of Best Linear Unbiased Prediction (BLUP) and is a method of estimating random effects. There are multiple derivations of BLUP available, with (Robinson, 1991) describing a range of them. These being Hendersons Justification (Henderson, 1950), use of a Bayesian Derivation (Dempfle, 1977) or Goldbergers Derivation (Goldberger, 1962). The BLUP linear model, in context can be calculated as:

$$y = X\beta + Z\mu + e$$

Where

y is a vector of *n* observable random variables;

β is a vector of *p* unknown parameters having fixed values (fixed effects);

X and Z are known matrices;

and *μ* and *e* are vectors of *p* and *n*, respectively (Robinson, 1991).

BLUP was originally developed for the ranking and selection of animals within breeding programs. As mentioned, BLUP comprises the use of fixed and random effects. Fixed effects model can be described as;

$$y = X\beta + \varepsilon$$

Where

y is a vector of N observations;

X is a known matrix;

β is a vector of p fixed, unknown constants (fixed effects)

and ε is a vector of possible random effects.

We take all elements of ε to be uncorrelated with one another with the same variance σ_ε^2, in doing so the variance-covariance matrix of ε is;

$$var(\varepsilon) = \sigma_\varepsilon^2 I_N,$$

With I_N the identity matrix of order N.

The most common fixed effects in beef cattle analysis are herd, year and season. These are classed in order to create contemporary groups between animals of different ages and those that were raised in different environmental conditions. Other fixed effects include sex, age of dam, sire and parity number. Random effects are attributed to the yet unknown and unquantifiable environmental effect on the animals' genotype.

The interpretation of EBVs differs between traits and the direction of selection that is being placed on a specific trait. An EBV of 0 is the average trait expression within a population. A negative EBV for AFC and ICP is favoured as this indicates a reduction in the time of these traits and as previously discussed this would reduce the time of unproductivity. WW_{MAT} and SC EBVs that show a positive score are selected as these would result in an increased weaning weight and daughters of sires with positive EBVs for SC would most likely have a shortened AFC.

The premise for the initial calculation of deregressed breeding values was presented by Goddard (1985), to allow for the comparison of breeding values of dairy sires across multiple countries. Multiple across country evaluation (MACE) allows for the prediction of a bull's EBV from his country of origin to predict his true EBV in another country. Implementation of this method was successful in both single trait (Jairath *et al.*, 1998) and multiple trait (Schaeffer, 2001) models. Initially this method was developed for dairy sire evaluations, but Goddard (1985) proposed that it could equally be used for beef cattle and other livestock species.

The availability of genomic data has resulted in the development of efficient methods that allow for the estimation of thousands of marker effects simultaneously in order to increase the reliability of EBVs (VanRaden, 2008). These methods require high computing power to allow programs and algorithms to run appropriate analysis. The benefits of incorporating genotypic data may be equivalent to about 20 daughter progeny records with this level of information increasing as the genotypes of more relatives are added (VanRaden, 2008).

As large datasets can cause excessive computational difficulties for the evaluation and calculation of EBVs, Strandén & Lidauer, (1999) proposed the use of an iterative procedure based on preconditioned conjugate gradient (PCG) to aid in the solving of linear models. This allowed the assessment of multiple root finding methods, these being, bisection, secant and Broydens method in order to accelerate the implementation of deregression into existing model software for use in calculating EBVs (Strandén & Mäntysaari, 2010). These root finding methods are used in order to find the general mean, calculated from the mean EBVs for each trait respectively. Strandén & Mäntysaari, (2010) observed that the secant and Broyden methods had the lowest number of PCG calls, indicating these methods are more appropriate for analysis. The secant method is a one dimensional root-finding algorithm that uses a succession of roots of secant lines to better approximate a root of a function (Wang *et al.*, 2018), and can be interpreted as a method in which the derivative is replaced by an approximation and is thus a quasi-Newton method.

Genomically enhanced breeding values (GEBVs) are increasingly being used to calculate values for all animals in the pedigree using single step models (Legarra *et al.*, 2014). ssGBLUP calculations require the input of available pedigree information and phenotypic records in conjunction with genomic information. In single step models, non-genotyped individuals are imputed implicitly. GEBVs are calculated using a genomic relationship matrix (GRM) in conjunction with mixed model equations (MME) (Taskinen *et al.*, 2013). Koivula *et al.*, (2016) observed that the inclusion of genotypic and phenotypic cow data within the reference population caused a persistent increase in validation reliability and that bias was minimised. Thomasen *et al.*, (2014) established that annual genetic gain as well as the reliability of genomic predictions when more cows were included in the reference population proved to be slightly higher.

Deregression is a non-linear problem (Strandén & Mäntysaari, 2010)**.** Due to the presence of major genes, linear predictions may be less efficient on not normally distributed data allowing for the use of non-linear predictions (Henderson, 1963). As major genes may exist on some chromosomes, genetic variance may not be equal across markers or chromosomes (VanRaden, 2008). The solving of deregressed EBVs requires fitting the general mean as an unknown fixed

effect within a mixed model equation. The definition of genetic groups affects the convergence of the general mean effects, with more genetic groups resulting in a faster convergence (Strandén & Mäntysaari, 2010). BLUPF90 (Aguilar *et al.*, 2018) and MiX99 (Lidauer *et al.*, 2017a) are routinely used for applying such models on large datasets.

It should be noted that the reliability of an older animals' breeding value is mainly determined by the information of its' direct progeny (Erbe *et al.*, 2018). This can be more accurately estimated if the information of direct descendants and other possible relatives that may contribute significantly towards the genotyped individual(s), are included. The reliability approximation program ApaX99 (Lidauer *et al.*, 2017b), which expands on algorithms proposed by Harris & Johnson, (1998) allows for the estimation of effective record contributions (ERCs) for a pre-selected subset of individuals. These ERCs are used as weights during the estimation of deregressed EBVs. Multiple recent studies (McGovern *et al.*, 2019; Purfield *et al.*, 2019a; Ring *et al.*, 2019; Twomey *et al.*, 2019) have used the deregressed EBV as a more independent variable in GWA studies, with various recent studies using MiX99 (Lidauer *et al.*, 2017a) as the appropriate software program.

1.2.5 Overview on the development of the Bovine genome

The first *Bos Taurus* whole genome assembly, a female Hereford, was published just over 10 years ago (Zimin *et al.*, 2009). They were successfully able to construct an assembly with large-scale continuity that covered around 91.5% of the bovines 30 chromosomes. This opened new and exciting avenues in animal science. Currently, along with the UMD 3.1 (Zimin *et al.*, 2009), two other whole genome assemblies are available for reference, these being BTAU 4.2 (Partipilo *et al.*, 2011) and BTAU 4.6 (Elsik *et al.*, 2016)**.**

Single nucleotide polymorphisms are universal, as well as the most abundant form of genetic variation among individuals of the same species, even though they are less polymorphic than microsatellites due to their bi-allelic nature but easily compensate by being abundant, ubiquitous and highly amendable to automation. Due to being evenly spread across the genome, every 100bp or so, SNPs are more likely to be present in regulatory regions. This allows for SNPs to be more likely to be located near a putative QTL or candidate gene. Single nucleotide polymorphisms are also highly favoured by industry as they have great repeatability across labs and are cheaper than microsatellites.

Single nucleotide polymorphisms evenly spread across the genome can be used to detect and map mutations that cause variation in the expression of traits of interest by means of a

genome-wide association (GWA) analysis (Bader, 2001; Visscher *et al.*, 2012). This GWA approach can assess every SNP independently or simultaneously to determine a level of association against the trait of interest whilst minimising population stratification that may be a result of pedigree and/or breed composition discrepancies between animals. The assumption is that each SNPs variance is explained and relates to the effect size of the hidden causal polymorphism affecting the trait, the degree of association between the SNP and the polymorphism, and any experimental and/or residual error associated with measuring the trait of interest (Bader, 2001). This GWA approach requires a considerable number of SNPs to be genotyped in a large sample of individuals that enables any analysis to retain significant power in order to detect and identify true associations (Mapholi *et al.*, 2016).

Due to their high distribution, SNPs can have significant associations with traits of low heritability, such as fertility and adaptability traits that are hard to measure as they may only be expressed later in life or are sex-limited traits. Some traits linked to adaptability struggle to have an exact definition that is reproducible across herds. Use of SNPs would greatly increase the genetic gain through marker-assisted selection (MAS). Especially if the allele frequency of the favourable allele and level of heterozygosity in the population is low (Meuwissen *et al.*, 2001). SNPs can be extremely informative with regards to local indigenous breeds in Southern Africa, with studies showing their current application and future potential (Zwane *et al.*, 2016; Makina *et al.*, 2016; Lashmar *et al.*, 2018a). The use of high density (HD) platforms, that genotype ~800,000 SNPs, combined with WGS of high impact indigenous individuals may uncover an abundance of QTL, candidate genes, new polymorphisms. This may result in possible identification of causative mutations that may be included for GS (Fortes *et al.*, 2013; van Marle-Köster *et al.*, 2013).

1.2.5.1 Linkage disequilibrium

The knowledge of linkage disequilibrium (LD) between SNPs is essential to estimate the necessary number of SNPs for genomic selection (GS), accurate association studies and investigating variation between breeds (Pritchard & Przeworski, 2001; Marchini & Howie, 2010; van Marle-Köster *et al.*, 2013; Zwane *et al.*, 2016). Linkage occurs when two or more genetic markers are linked on a chromosome and not separated during recombination, resulting in LD (Nielsen & Slatkin, 2013).

Linkage disequilibrium is defined as the non-random association between alleles at two or more loci and is influenced by historical population-wide evolution (Ardlie *et al.*, 2002; Porto-Neto

et al., 2014). Therefore, regions with low recombination rates have a higher LD extent indicating that segments of DNA with a high LD are inherited as a unit.

D' and r^2 are two statistical properties most used to assess LD between diallelic markers like SNPs. D' aids in understanding population specific long-range LD that can be used to represent historical recombination patterns (Gurgul *et al.*, 2014). A disadvantage is that small sample sizes, the presence of rare alleles and/or low haplotype frequencies tend to inflate D' (McRae *et al.*, 2012; Nielsen & Slatkin, 2013). Long-range LD can't be utilised across breeds as animals from different breeds do not share a recent common ancestor (Goddard & Hayes, 2009). r^2 describes the correlation between haplotypes at two loci, and is useful in predicting the power of association mapping. Genome-wide average LD (r^2) decreases with increasing genomic distance in all breeds, which is commonly referred to as LD decay (Porto-Neto *et al.*, 2014). Increased ancestral recombination can lead to a loss in statistical power due to a reduced LD (Nielsen & Slatkin, 2013), across all genotyping platforms.

LD estimates depend on factors such as the history and structure of the population under investigation (Porto-Neto *et al.*, 2014; Lashmar *et al.*, 2018a). Other factors include marker type(s), the density and distribution of the markers, sample size of the population (Matukumalli *et al.*, 2009; Judge *et al.*, 2016), accurate method(s) used for haplotype inference and construction (Sargolzaei *et al.*, 2014) and correct implementation of SNP filtering (Gurgul *et al.*, 2014). Causal markers have proven to be difficult to recognise, and it has been an even more imposing task in confirming their functional causality (Dekkers, 2010; Sharma *et al.*, 2015).

SNPs found to be linked with causal variants between highly-related individuals, but are in linkage equilibrium (LE) with the majority of the population are not effective for use in GS (Goddard & Hayes, 2009; Dekkers, 2010). van Marle-Köster *et al.*, (2013) reported that due to the limited use of AI by SA beef breeders, limited linkage across the entire population is expected. In contrast, SNPs confirmed to be in LD with known causal variants grant an alternative approach for the deployment of MAS in livestock populations. These may be discovered in GWAS by analysing genomic regions that have been significantly associated with economically important traits (Zalesky *et al.*, 1984; Sharma *et al.*, 2015; Lashmar *et al.*, 2018a). The ability to discover these regions has been facilitated by the advances in genomic technologies (Meuwissen *et al.*, 2001), including whole genome sequencing of most livestock species (McKay *et al.*, 2007; Goddard & Hayes, 2009; Zimin *et al.*, 2009).

Multiplex SNP genotyping has allowed for high precision sequencing of hundreds of thousands of loci, at an effective cost that allows for sufficiently large-scale GWAS. The density

of SNPs required for an efficient GWA scan depends on the average length of chromosomal "blocks" possessing high levels of LD. What varies between species and amongst species-specific populations is the intermediate length of genomic segments to which the r^2 between terminal loci achieves a precalculated value (Matukumalli *et al.*, 2009).

A haplotype is defined as any combination of alleles or SNP pairs that occur at a specific locus (Nielsen & Slatkin, 2013). Haplotype blocks are a combination of alleles found in a genomic region, where 5% or less of comparisons between informative SNP pairs showing strong evidence of historical recombination within a population (Gabriel *et al.*, 2002). Haplotype block-based methods grant a greater level of information content compared with single SNP methods in GWAS (Kong *et al.*, 2008; Qanbari *et al.*, 2010; Sargolzaei *et al.*, 2014). When LD is weak between marker loci (r^2 <0.2), haplotype-block methods are then used in conjunction with scanning and sequencing of the whole genome. Identity by descent (IBD) at a given chromosomal location can be predicted with numerous markers spanned across that position (Meuwissen *et al.*, 2001).

Observations of species-specific LD/haplotype blocks has resulted in the observation that cattle have on average, longer length blocks compared to that for humans. This may be due to population bottlenecks that happened during domestication and/or establishment of modern cattle breeds (McKay *et al.*, 2007). An average block length of 100,000bp in *Bos taurus* appears to achieve an r^2 of 0.25, with this LD block size being three times greater than those in human populations (Matukumalli *et al.*, 2009).

A consequence of this high level of LD in *Bos taurus* indicates that it is essential to only genotype a subgroup of polymorphic loci to capture most of the variation common in the genome (Berry *et al.*, 2014a), ranging between 30 000 and 50 000 SNP) markers (McKay *et al.*, 2007) in pure *Bos taurus* or *Bos indicus* breeds. Matukumalli *et al.*, (2009) indicated in cattle, the goal is to develop an array that would exceed the minimal number of markers needed to span the bovine genome, achieves an average marker density of 100kb and attains an average minor allele frequency (MAF) of at least 0.15. Appropriate utilisation of LD information in differing populations requires significant population-wide disequilibrium between SNPs and a medium or large-effect QTL. In order to quantify enough marker allele diversity to predict any potentially significant QTLs, individuals need to be genotyped from across the entire population (Goddard & Hayes, 2009). SNPs that are in high LD suggest that they may all be associated with the same QTL. Other reasons for upward bias due to high levels of LD, may be the presence of more than one QTL in that genomic region (Allison *et al.*, 2002). Earlier studies prove that LD is definitely population specific and can be heterogenous between populations depending on their respective histories

(Thévenon *et al.*, 2007; McKay *et al.*, 2007; Edea *et al.*, 2014). Successful applications that rely on LD across populations depends on allele phase relationship preservation between SNPs and QTLs across these populations (Goddard & Hayes, 2009).

Using SNP data of three different breeds, de Roos *et al.*, (2008) demonstrated that when two populations separate and became more genetically divergent from one another, the allele phase relationships between them are less inclined to be preserved. McKay *et al.*, (2007), using 2 670 SNPs, assessed the extent of LD in genotyped populations across eight breeds of *Bos taurus* and *Bos indicus* cattle. This revealed moderate LD (r^2 = 0.2) extended to 40-60kb in cattle and indicated that around 50 000 SNP markers could secure the majority of the LD information required for GWAS in the European *Bos taurus* breeds. Thévenon *et al.*, (2007) evaluated the extent of LD in a *Bos indicus* x *Bos taurus* breed, n = 364, of Western African origin using 42 microsatellite markers on BTA1, BTA4 and BTA7. These results illustrated that LD extended for shorter distances in African cattle compared to what had already been observed in European cattle breeds. Edea *et al.*, (2014) assessed the extent of LD in indigenous Sanga and Zebu breeds of Ethiopia using a 50 000 SNP array (Matukumalli *et al.*, 2009) and discovered that the degree of LD was lower in Sanga and Zebu breeds in comparison to that of European *Bos taurus*, and indicated that any potential GWAS of these hybrid breed would require genotyping platforms with 75 000 to 300 000 SNPs. The reduced degree of LD observed in these breeds was attributed to the SNP ascertainment bias that was resultant from the original detection of SNPs in European *Bos taurus* breeds used in the design of this assay (Zimin *et al.*, 2009).

The extent of genome-wide LD in all SA cattle breeds has yet not been determined. The SA Sanga breeds (Afrikaner, Bonsmara, Drakensberger and Nguni) genetic diversity was previously described by Makina *et al.*, (2014) and were proved to be genetically distinct from Angus and Holstein breeds. Lashmar *et al.*, (2018a) reported low LD ranging from r^2= 0.11 (BTA28) to r^2= 0.17 (BTA14) for SNPs separated by ≤ 1Mb and r^2= 0.20 extended only up to < 30 kb in SA Drakensberger breed. This means that well over 50 000 SNPs are needed to capture any significant LD relationships in local Sanga breeds. Sanga cattle are hypothesised to be a hybrid of *Bos taurus* and *Bos indicus* origin. This most likely transpired as *Bos taurus* Egyptian longhorn cattle migrated south from Egypt and the Sudan, and *Bos indicus* Lateral Horned Zebu cattle which originated from Arabia and India (Scholtz *et al.*, 2011). The four Sanga breeds analysed (Makina *et al.*, 2014) were shown to share some level of co-ancestry, but with clear distinguishable genetic relatedness. From the evidence given above, it has been deduced that the degree of LD and the persistence of allelic phase relationships be estimated specifically for populations in which

genomic studies and applications are to be implemented. This information is essential for the identification of the optimal array to apply regarding cost, potentially increasing the number of genotyped individuals and development of low-density SNP panels that achieve a satisfactory level of inter-marker LD (Zalesky *et al.*, 1984; Wang *et al.*, 2012; Berry *et al.*, 2014a).

1.2.5.2 Prerequisites for genome-wide association studies (GWAS)

The primary consideration of a GWAS should be its statistical power, the probability of detecting a variant assumed to be causal, which can only be assessed via simulation (Zalesky *et al.*, 1984; Hayes *et al.*, 2003; Spencer *et al.*, 2009). To be included within genotyping platforms, SNPs must have conversion rates of higher than 50%. Other incorporation factors include having a call rate greater than 99%, must be accurately mapped to either sex chromosomes or autosomes and maintain Hardy-Weinberg equilibrium (HWE) with a pre-determined p-value for each species. In calculating power, data is simulated under the assumption that an allele is causal and then observed to see if any SNPs on the SNP array, within a large region flanking the causal allele, attain a high level of significance (Spencer *et al.*, 2009; Zhou & Stephens, 2012). Increasing SNP density will increase the power of detecting QTLs and may increase the precision of genome mapping to a certain degree. Although, if LD is high over a specific chromosomal segment, an increased SNP density may still not allow for the accurate description of a QTLs position within this segment (Goddard & Hayes, 2009; Matukumalli *et al.*, 2009). SNPs must be spaced no greater than 10kb apart to consistently show LD phase across breeds (Goddard & Hayes, 2009).

SNP information is available from a wide variety of open source or public domains. These include HapMap and Interbreed to name a few. If the true causative SNP is not on the array, there may typically be a few flanking SNPs on the array that are associated with it due to their proximity. One or more of these SNPs will generate a signal of association and thus allow for the detection of new or existing causative loci (Hirschhorn & Daly, 2005; Wang *et al.*, 2005; Spencer *et al.*, 2009; Stranger *et al.*, 2011). As the whole genome sequences (WGS) of most domesticated animals is known, an extensive number of previously undiscovered SNPs were identified, as a result of sequencing and/or resequencing. Table 1.5 provides a summary of all the available SNP arrays currently available. These array sets differ in the way in which the SNPs are selected, and the total number of SNPs genotyped (Spencer *et al.*, 2009; Matukumalli *et al.*, 2009; Nicolazzi *et al.*, 2015).

Table 1.5 Summary of the currently available genotyping arrays

Company	Name of Genotyping Array	Number of SNPs
Illumina	Bovine3k BeadChip	2 900
	BovineLD BeadChip	6 909
	BovineLD .v2 BeadChip	7 931
	BovineSNP50.v2 BeadChip	54 609
	Bovine HD BeadChip	777 962
	GeneSeek Dairy Ultra LD .v2	7 049
	GGP LD .v3	26 151
	GGP HD .v2	139 480
Affymetrix	Axiom Bovine BeadChip	648 875*
Independent	ICBF IDB v2	17 807
	ICBF IDB v3	53 262

*SNP probes

Polymorphic SNPs in many populations were primarily derived from WGS reads representing five taurine and one indicine breed to compare against the reference genome assembly from a slightly inbred Hereford cow (Zimin *et al.*, 2009; Bovine HapMap Consortium *et al.*, 2009). The majority use of WGS of taurine breeds indicates that ascertainment bias exists within most genotyping platforms as taurine breeds will have higher levels of informativeness and inclusion in comparison to the indicine breeds (Spencer *et al.*, 2009; Matukumalli *et al.*, 2009). Zwane *et al.*, (2016) indicated there is ascertainment bias in the BovineSNP50 towards the European breeds as the overall MAF in the indigenous SA breeds she studied was lower. The lower representation of the *Bos indicus* populations in the design of these platforms indicates a smaller proportion on indicine loci compared to the taurine loci (Chan *et al.*, 2009; Espigolan *et al.*, 2013; Edea *et al.*, 2014). This poses a dilemma for the use of these SNP chips in local SA indigenous breeds. Most breeds are an admixture of both taurine and indicine descent, with the local Sanga breeds showing wide variation in genetic and breed composition.

The availability of cost-effective HD SNP chips that can assay more than half a million loci offers a marker density that generates a high statistical power. Recent studies (Bolormaa *et al.*, 2015; Do Nascimento *et al.*, 2018; Kluska *et al.*, 2018; Lopes *et al.*, 2018; Melo *et al.*, 2018) have detailed the behaviour of LD using the set of 777 000 SNPs on the BovineHD platform (Illumina

Inc, San Diego). A major advancement with regard to this HD array is that it allows for a more accurate estimation of LD over shorter chromosomal distances as it carries an increased number of marker pairs that are separated by 10 kb or less (Porto-Neto *et al.*, 2014).

In general, Affymetrix arrays have a higher level of redundancy compared to Illumina arrays, in that they contain a greater number of SNP sets that are more highly correlated with each other. This results in a lower coverage and power for the same number of SNPs, but this redundancy can be advantageous. The loss of specific SNPs to quality control (QC) thresholds may not be too costly as significant signals of association are likely to include more SNPs, which make them easier to differentiate from false-positive associations or any genotyping artefacts (Spencer *et al.*, 2009).

Qwabe *et al.*, (2013) reported that SA cattle breeds have a higher number of SNPs with low MAF (MAF<0.05) and thus have fewer polymorphic loci compared to European Angus and Holstein breeds. This lower number of polymorphic loci among SA cattle breeds can be associated to the ascertainment bias correlated with the creation of the BovineSNP50 BeadChip (Qwabe *et al.*, 2013), as the SNPs used in the creation of this assay were detected in European *Bos taurus* breeds, resulting in the observation of low MAF in *Bos indicus* breeds. SNPs with low allelic frequencies contribute to the underestimation of r^2 assessments of the LD between SNP markers.

1.2.5.3 Imputation

The rapid development of cost-effective genotyping platforms over the last few years, coupled with the ongoing discoveries of new SNPs, means that a wide range of HD and low-density SNP arrays are now readily available (Table 1.5). Thus, a methodology to combine separate genotyped populations to create a larger training population for appropriate GWAS for trait analysis and use of GS was needed. The release of newer lower density arrays, complementary to pre-existing HD arrays (Judge *et al.*, 2016), has resulted in most breeds having individuals genotyped on a range of different genotyping platforms. With the early establishment of a reference population on the GGP 150K array through the BGP, SA Stud Book have been marketing the use of two lower density panels, namely the GGP 80K and ICBF IDB v3 arrays, as a cost-effective way of including more animals in estimation of GEBVs (van der Westhuizen *et al.*, 2017).

The benefits linked to increasing SNP density and/or from using available imputation methods are greatest for low MAF SNPs. A clear consequence is that an increased sample size is going to have a greater effect on power than an increased SNP density. Three MAF thresholds

(0.05, 0.1, and 0.2) were outlined to investigate the effect of allele frequencies on LD extent (Anderson *et al.*, 2008), and LD was observed to increase as the MAF increased across all breeds, notably for SNP pairs separated by distances shorter than 100kb in length. This was previously reported by Khatkar *et al.*, (2008) in Australian Holstein-Friesian cattle. It was observed that the mean r^2 was higher (0.70) for SNPs with a MAF ≥ 0.2 compared to the r^2 0f 0.59 for SNPs with MAF ≥ 0.05 separated by lengths of 1–10 kb. This was accredited to the fact that as the MAF threshold increases, there is a corresponding increase in SNP pairs with comparable allele frequencies which results in an increase in r^2 (Khatkar *et al.*, 2008).

Imputation is the prediction of missing genotypes which are estimated from a reference population (Scheet & Stephens, 2006) and is considered a zero-cost method of inferring SNPs between genotyping platforms (Anderson *et al.*, 2008). This results in increasing the statistical power of any potential association studies. Halperin *et al.*, (2005) states that the use of economically and computationally efficient low-density SNP marker panels is essential for quantifying breed specific trait variation. Thus, imputation is best used for increasing the size of the population under investigation.

Imputation can be a powerful means of combining data sets genotyped with different arrays, provided sufficient overlap exists between the panels. This overlap consists of common SNPs between the SNP chips. How genotyping costs are reduced, is by way of breed societies tending to genotype young individuals with a cost-effective low-density panel. Low-density panels cover the genome uniformly, and are used in imputation, coupled with HD panel reference population information to infer the genotype of untyped loci (Weigel *et al.*, 2010; Zhang & Druet, 2010; Berry & Kearney, 2011; Dassonneville *et al.*, 2012; Wang *et al.*, 2012; Berry *et al.*, 2014a; Sargolzaei *et al.*, 2014; Chud *et al.*, 2015; Judge *et al.*, 2016; Lashmar *et al.*, 2018b). This reference population is normally derived from high-impact animals, mainly sires, that contribute significantly to the genetic variation within breeds. Imputing of SNP haplotype sequences uses the current patterns of genetic variation within this reference population as well as any relevant pedigree information that is available (Kong *et al.*, 2008; Sargolzaei *et al.*, 2014). Berry *et al.*, (2014a) pointed out that superior imputation accuracies are achieved within breeds in comparison with between breeds.

An effective imputation strategy involves capitalising on existing pedigree relationships between individuals and historical population-wide LD by seeking for shared haplotypes from longest to the shortest. This is the focal point of the method(s) used in FImpute v2.2 software (Sargolzaei *et al.*, 2014). FImpute, described as a deterministic program, employs both family-based and population-based methods. It utilises two separate methods, which are phasing and

the use of overlapping sliding windows (OSW). The current version (v2.2) of FImpute is able to correlate and complement homologous autosomes on multi-core Linux systems (Sargolzaei *et al.*, 2014).

Phasing, a type of haplotype-block method, is the process of correctly aligning SNP pairs located on homologous chromosomes. Initially proposed by Kong *et al.*, (2008), local phasing exploits the strong correlations between SNP alleles within LD blocks. The main limitation here being that SNPs that are far apart cannot be reliably phased due to weak LD correlations. The OSW approach, proposed by Sargolzaei *et al.*, (2014), in a way presents a solution to the aforementioned limitation. The OSW approach utilises accurate pedigree records and by moving long windows over an autosome. As information from more distant relatives is factored in, the size of the OSW shrinks with each autosomal sweep. The first window is a maximum of 1000 SNPs, with each subsequent window being reduced by a factor between 0.1-0.2, with the minimum window size being 2 SNPs. Each window has a haplotype library reference, that is then used for phasing and eventual imputation within each window.

The OSW method allocates haplotypes based on the number of crossover events that occurred since a most recent common ancestor. Closely related individuals will have fewer crossover events and thus will share longer haplotype blocks than individuals who are separated by multiple generations. The higher the number of recombination events between two individuals results in a decrease in the length of the haplotype block. As these windows consistently overlap, the accuracy of correctly identifying shorter-shared haplotypes between distantly related individuals is increased. Genealogy plays a significant role in phasing and imputation (Kong *et al.*, 2008) but as the panel density of the reference population increases, the importance of the family information decreases. Increasing panel density also increases the likelihood of correctly identifying shared segments, especially for shared haplotypes (Sargolzaei *et al.*, 2014).

Imputation accuracy (IA) is affected by many factors. These include the different SNP panels used, MAF, the imputation software utilised, population structure and level of relatedness between the reference and target populations (Howie *et al.*, 2009; Chud *et al.*, 2015; Pausch *et al.*, 2017). Genomic regions with low IA may occur due to greater observed heterozygosity than expected, incorrect SNP genomic position, poor genotyping calling, or a recombination hotspot may occur nearby (Berry *et al.*, 2014a).

Accuracy of correctly imputing genotypes from an individual is increased if one or both parents are included in the reference population. Any missing information is inferred when haplotypes of the parents are matched against progeny haplotypes. These detected matches

allow for crossover events to be identified and the chromosomal location they occurred. Non-genotyped parents with sufficient progeny data may have their genotypes phased and imputed (Li *et al.*, 2009). This was verified (Berry & Kearney, 2011), where a minimum of five offspring are needed for accurate imputation. These *"in silico"* genotypes can then be used as if the SNPs involved were directly genotyped. Intermittent missing genotypes may also be imputed which will improve the call rate of poorly genotyped SNPs (Moser *et al.*, 2009; Marchini & Howie, 2010).

In order to use imputed sequences in breed specific association studies, verification of FImpute is needed to ensure the highest accuracy possible (Lashmar *et al.*, 2019). This involves the splitting of the reference population into two groups, a reference and target population. The target population has >80% of their genotype removed and is then imputed, with several statistics calculated. These are done by comparing the target animals imputed genotype with its original genotype (Berry & Kearney, 2011; Chud *et al.*, 2015; Lashmar *et al.*, 2018b). Concordance rate is the proportion of correctly imputed genotypes (GCR) or alleles (ACR) and the correlations between actual and imputed genotypes.

Table 1.6 summarises the main imputation software currently available. BEAGLE and IMPUTE2 utilize hidden Markov models (HMM), while assuming individuals are unrelated. These calculate the posterior distribution of SNP haplotypes, and infer missing genotypes from that distribution. They have shown to produce consistent predictions but are mainly used in human population studies (Marchini & Howie, 2010; Browning & Browning, 2016) with a few done in livestock (Berry & Kearney, 2011; Dassonneville *et al.*, 2012; Berry *et al.*, 2014a). These programs are known to be computationally and time demanding. A recent review (Whalen *et al.*, 2017), indicated that there is a possibility of improving HMM methods.

Table 1.6 Summary of the available imputation software and the methodology used

Imputation Software	Methodology	Reference
BEAGLE v3.3.2	Hidden Markov Models	(Browning & Browning, 2009, 2016)
IMPUTE2	Hidden Markov Models	(Marchini *et al.*, 2007; Howie *et al.*, 2009)
FImpute	Open Sliding Windows and Long Range-Phasing	(Sargolzaei et al., 2014)

FImpute is not as computationally demanding as BEAGLE (Chud *et al.*, 2015) and this, paired with the utilisation of Linux multi-core systems, is the reason as to why FImpute is now

widely used. This is evident as FImpute has been successfully applied across multiple species. These include sheep (Bolormaa *et al.*, 2017), horses (Pereira *et al.*, 2017), pigs (van Son *et al.*, 2017), beef and dairy cattle (Do *et al.*, 2018; Iung *et al.*, 2018; Lopes *et al.*, 2018) and even salmon (Yoshida *et al.*, 2017).

Rare alleles present a challenge regarding accurate imputation. These variants may substantially contribute to what is being called “missing heritability” (Zuk *et al.*, 2012). As MAF decreases, genotyping errors create sensitivity within association tests. Therefore, high accuracy of imputation for rare alleles is vital. As most rare variants tend to be recent mutations, they tend to be associated with longer haplotypes. The use of pedigree information should increase the imputation accuracy of these rare variants. Inaccurate pedigree records will result in the lowering of imputation accuracy. Sargolzaei *et al.*, (2014) did mention that as the reference population becomes bigger, the accuracy of imputed genotypes for SNP with low MAF will increase.

The importance of the bovine X-chromosome in GWAS and GS has been largely ignored, as males are heterogametic at the SNPs. The X-chromosome accounts for about 6% of the total physical genome (Zimin *et al.*, 2009) and according to *Ensembl* carries 4.2% of protein encoding genes, thus its omission from GWAS studies is questionable.

A study by Qanbari *et al.*, (2010) reported that by minimizing SNP density by only including SNPs that were polymorphic (MAF > 0.1) for all breeds, the LD decay for Bovidae X-chromosome became more homogeneous across all breeds which didn’t differ much from the results the autosomes obtained. Due to the bottlenecks experienced by cattle populations since their domestication, during more recent breed formation and in part a result of the frequent and intensive use of AI, it is feasible to expect extensive LD on BTAX (Porto-Neto *et al.*, 2014). Indicine cattle continued to have a lower LD when the distances between markers were large, in comparison to the other breeds under investigation, which implies that they originated from a larger ancestral population (Porto-Neto *et al.*, 2014). Thus the inclusion of the X-chromosome in GWAS can be beneficial and may produce novel results. Mao *et al.*, (2016) proposed a method of imputing X-chromosome in ungenotyped sires by treating the pseudo autosomal region as autosomal using FImpute v2.2 (Sargolzaei *et al.*, 2014). The IA was further increased when including more females in the reference population.

The use of FImpute has been shown to be successful in imputing from low-density to HD panels in composite beef cattle (Chud *et al.*, 2015; Lopes *et al.*, 2018). The benefits of dense marker panels for GWAS can be compromised when the IA is too low (Khatkar *et al.*, 2012; Pausch *et al.*, 2017). More recently, a study conducted at the University of Pretoria, provided verification

of the accuracy of FImpute on a local Sanga breed, the SA Drakensberger (Lashmar *et al.*, 2018b) using the GGP HD (150k) bovine beadchip. Genotype and allele call rates of greater than 95% were observed (Lashmar *et al.*, 2018b) This indicates that FImpute may be suitable for use in local Sanga breeds to increase the training population for use in GWAS. Milanesi *et al.*, (2015) argued that IA is robust to updates to the cattle reference genome and the use of the UMD.3.1. (Zimin *et al.*, 2009).

With the advancement of Next-Generation Sequencing (NGS) (Behjati & Tarpey, 2013), the ability to capture WGS has become a possibility. This has occurred in Brown Swiss, Holstein, Jersey breeds (Naderi *et al.*, 2018). Recent studies using whole-genome sequences as a reference population for imputing a range of smaller SNP panels (Brøndum *et al.*, 2014; Frischknecht *et al.*, 2017; McGovern *et al.*, 2019; Purfield *et al.*, 2019a; Ring *et al.*, 2019; Twomey *et al.*, 2019) have occurred. The program used was Minimac2 (Fuchsberger *et al.*, 2015), which used FImpute generated HD SNP chip data from individuals imputed up from 50k SNP chip data.

There has been impressive innovation in creating computational statistical models that are effective in capturing breed specific patterns of LD (Scheet & Stephens, 2006; Marchini & Howie, 2010; Sargolzaei *et al.*, 2014). Dassonneville *et al.*, (2012) stressed that the development of low-density panels should be advanced to maximise imputation accuracy, with Berry *et al.*, (2014a) investigating and reporting that they have been. This has aided in the apprehension of the confounding nature of correlations between SNPs that exist within increasingly larger genotyped populations. Large studies today are still underpowered, unable to detect most SNP effects. A combination of data across genotyping platforms will be essential to facilitate these meta-analytic approaches. Accurate imputation is key to ensuring that the benefits of a larger population with more common SNPs exceeds the imputation loss as gains from HD are small (VanRaden *et al.*, 2013). Lashmar *et al.*, (2018b, 2019) stated that imputation of Sanga breeds is possible, with strict adherence to ensuring quality input data, proper implementation of QC and pedigree completeness.

1.2.5.4 Application of genome wide association studies (GWAS)

A GWAS involves the evaluation of molecular data alongside phenotypic data to detect any significant associations between genetic markers spread across the genome, and trait variation (Stranger *et al.*, 2011). The use of this has shed additional light on the mechanisms of complex traits (Sharma *et al*., 2015), as well as quantifying diversity among local indigenous breeds populations (Makina *et al.*, 2014, 2016; Sanarana *et al.*, 2016; Lashmar *et al.*, 2018a). Several

studies have been conducted on European and tropically adapted beef cattle breeds on traits of economic importance (Snelling *et al.*, 2009; Bolormaa *et al.*, 2011; Hawken *et al.*, 2012; Berry *et al.*, 2013; McDaneld *et al.*, 2014; Costa *et al.*, 2015; Regatieri *et al.*, 2017; Melo *et al.*, 2018).

In literature, two perspectives are described with regards to GWAS. The first is to consider a GWAS as a self-contained experiment, with the statistical inference of a formal hypothesis test assuming a null hypothesis of no association. From this aspect, the end goal of a GWAS is to determine if SNPs are, or are not, associated with the phenotype or trait of interest (Meuwissen *et al.*, 2001; Spencer *et al.*, 2009).

Secondly, is to regard a GWAS as an experiment that highlights SNPs of interest, and then to include as many as possible in further replication studies. This increases the probability that at least one SNP will have a level of significance imposed by a predetermined threshold. According to Spencer *et al.*, (2009) and Zhou & Stephens, (2012), it is recommended to keep the number of SNPs consistent in order to reduce the amount of false positive associations that could occur. The level of significance doesn't affect the relative performance of the arrays, or the corresponding effect of the sample. For a SNP array, overall false positives rates will differ as they are dependent on the population in which the GWAS is being conducted, due to alternating patterns of LD between the SNPs on the chip (Matukumalli *et al.*, 2009; Lashmar *et al.*, 2018a).

GWAS requires efficient and accurate tag SNP selection, proper interpretation of results generated by multiple comparisons and strategies to manipulate pairwise SNP-by-SNP interactions (Thomas, 2006; Anderson *et al.*, 2010; Voorman *et al.*, 2011). Quality control of data is one of the most important steps that minimizes error in a GWAS, especially those of false positive associations (Anderson *et al.*, 2010; Sharma *et al.*, 2015). Two steps for minimizing false positives include that the population must be genetically homogenous, have no population stratification, and all individuals in the sample must be independently drawn from the population. An alternate scenario is that related individuals may share both causal and non-causal alleles and the corresponding LD between these sites can leave artefacts (Nielsen & Slatkin, 2013).

VanRaden *et al.*, (2009) reported large gains in reliability occur from having large training populations and large numbers of SNPs because most traits are influenced by many genes with small effects. Most GWAS use mass-produced genotyping platforms to capture a given proportion of genetic variation genome wide.

Most GWAS on the Bovidae species have focused on dairy breeds or European beef breeds. Recent identification of breed-informative SNPs within three local, indigenous breeds has occurred (Zwane *et al.*, 2016), but there is a need to continue these studies into all breeds to

enable us to further understand the genetic architecture of these breeds. The Drakensberger breed has recently participated in a few genetic studies (Lashmar *et al.*, 2018a; b). Major efforts were made to verify the Celtic allele as the causal allele for horned or polledness in two Sanga breeds (Grobler *et al.*, 2018). This poses as a promise to reduce the need for dehorning and portrays the idea that genomic information will become invaluable for issues such as welfare and longevity.

Mapping and confirmation of QTL in independent populations of cattle, increases the confidence of reported results (Fortes *et al.*, 2013). Other studies have identified SNPs associated with traits that influence reproductive efficiency in *Bos taurus* cattle, which are mostly a variety of dairy breeds. Obvious candidates for fertility traits are genes that encode for a variety of reproductive hormones and genes with functional processes related with fertility. Traits such as days to first calving and calving interval, were associated with polymorphisms of Gonadotrophin Releasing Hormone Receptor (GnRHR) in dairy cattle (Derecka *et al.*, 2010). McDaneld *et al.*, (2014) commented that high-ranking SNP effects on dairy reproductive traits including daughter pregnancy rate, heifer conception rate, and cow conception rate were confirmed in 87 SNPs. These were distributed across 26 autosomes and the X chromosome.

Casas *et al.*, (2004) stated that on BTA29, a region was associated with their estimated age at puberty, testicular volume and weight. When a bull would achieve a SC of 28cm, their age at puberty was the predicted bulls age in days. An IGF1 polymorphism was associated with age at a SC of 28cm in Angus bulls (Lirón *et al.*, 2012). Between these studies reported for various dairy breeds, similar regions of the genome harbour genetic variation that influence fertility and maternal traits and this warrants further assessment. That QTLs or any associated genes have been recognized in virtually all chromosomes, including a chromosomal anomaly that Y chromosomal segments present in female cattle are associated with decreased female fertility, confirms the complexity of reproductive traits (Hawken *et al.*, 2012; Fortes *et al.*, 2013).

In order to identify polymorphisms that promote variation in complex traits (Bolormaa *et al.*, 2014; Melo *et al.*, 2018), QTLs detected in a GWAS meta-analyses must be validated against evidence from independent studies of related traits. Studies investigating the underlying influences of genomic regions that affect WW_{MAT} and ICP are scarce, while those on AFC and SC are more common.

A GWAS performed on Nellore cattle (*Bos indicus*) identified 5 SNPs on five separate chromosomes were associated with sexual precocity in heifers (Nascimento *et al.*, 2016). Using the *Ensembl* database, 3 SNPs on BTA5, 10 and 22 were identified to be associated with genes

that encode for U6 spliceosomal RNA that forms part of the main spliceosomal complex within the nucleus. The evidence of different chromosomes affecting the same trait, as well as similar gene regions between the autosomes, verifies the idea that multiple QTLs each with a small effect contribute additively towards the quantitative expression of reproductive performance.

There is a lack of GWA studies on the genetic characterisation of maternal traits in beef cattle. Studies on dairy cattle have been done, but as fertility traits are labour-intensive to measure, the lack of studies points to a lack of appropriate and accurate recordings for genomic statistical analyses.

Frischknecht *et al.*, (2017) analysed maternal gestation length (GL_{MAT}) and maternal birth weight (BW_{MAT}). No significant SNPs for BW_{MAT} were found, but a large QTL located on BTA13 was associated with GL_{MAT}. GL is a large proportion of the ICP and an optimisation of this may result in a positive decrease of the ICP. This variant is located at 65.5 Mb, located in the *CPNE1* gene. This gene encodes for a calcium-dependant phospholipid binding protein.

Costa *et al.*, (2015) reported 19 SNPs to be linked with AFC, which explained 6,42% of the trait's phenotypic variance in Nellore heifers. A GWAS on weaning weight for both direct and maternal effects in Colombian Brahman associated 5 chromosomes with 15 QTLs to the expression of WW direct and WW_{MAT} (Martínez *et al.*, 2017). It was also indicated that the reason these studies may not have discovered any significant SNPs on BTA 14 (*DGAT* or *PLAG*) is due to the genetic differences that exist between *Bos taurus* and *Bos indicus* types.

Hawken *et al.*, (2012) determined that in Tropical Composite and Brahman female cattle's age of puberty was linked to 20 SNPs on 12 different chromosomes. Two SNPs are located on BTA 14, with one SNP being at 61,9Mb which is 0.1 Mb from the QTL identified for SC in Brahman cattle (Fortes *et al.*, 2012b).

The use of this trait, SC, as an indicator trait for AFC in daughter heifers is used extensively across the beef breeding world. Extensive genome analysis on SC by McClure *et al.*, (2010) using 390 microsatellite markers identified 39 SNPs on 22 different chromosomes. The 19^{th} chromosome had 4 regions between 11.8 Mb and 59.4Mb. This indicates that there may be a large QTL here that contributes to the additive variance in the expression of SC. With the decrease in cost of SNP genotyping platforms, the use of SNPs to identify causal gene regions has now replaced the use of microsatellites.

Fortes *et al.*, (2012b) identified a region on the X-chromosome at 62-96 MB associated with SC at 12 months in Brahman cattle. Age at puberty, which is highly correlated to SC, was found to be associated with QTLs on both the 14^{th} and X chromosome by Fortes *et al.*, (2012a)

respectively. The X-chromosome QTL is located at 86Mb, which is in the same region as the QTL for SC, indicating these two traits may be influenced by the same QTL.

Melo *et al.*, (2018) identified 108 significant SNPs on 19 BTA using a multi-trait meta-analysis method. The two relevant traits analysed were AFC and SC in Brahman and Nellore cattle. The overall finding was these SNPs were located around QTLs or gene regions associated with physiological mechanisms that control the expression of sexual precocity in composite cattle.

da Silva Romero *et al.*, (2018) identified 4 regions associated with SC in Canchim bulls. 39 QTLs were observed to be in these regions using *CattleQTLdb* database, while 37 were observed when referring to the *Ensembl* database. Two candidate genes were located on BTA21 and one on BTA12. As previously stated, these genes affect the physiological membranes that control expression of reproductive traits.

1.2.6 Conclusion

Fertility traits are essential in selection programs for the genetic improvement of all livestock species and respective breed types. The discovery of SNP markers and the rapid development of high throughput sequencing technology resultant in analyses on WGS, has led to a multitude of various GWA studies to gain a better understanding of the true underlying genetic mechanisms that affect traits of interest. Most studies have focussed on growth and production traits, with only more recently the investigation of fertility traits has been given major attention. A similar examination of genomic regions affecting reproductive traits, as explored in this literature review and to my knowledge, has not yet been done in our indigenous SA Bonsmara cattle.

Chapter 2 Materials and Methods

2.1 Introduction

This study involved the analyses of a dataset provided by the SA Stud Book and Animal Improvement Association in conjunction with Bonsmara SA. Animal recording is compulsory for the breeders of SA Bonsmara and a number of fertility and growth traits are recorded. The data available for the study consisted of a pedigree file, estimated breeding values (EBVs) of four different traits and the genotypes of the animals with estimated breeding values. The four fertility traits included were, age at first calving, inter-calving period, weaning weight maternal and scrotal circumference with 4 171 animals having EBVs for all four traits. Genotypes have been generated within the Bovine Genomics Project (BGP), aimed at building a reference population for the implementation of genomic selection. Individual breeders have also submitted biological samples for genotyping. Genotypes on 3 291 animals were available, genotyped using one of three genotype arrays, namely, the GGP 150K, GGP 80K and the ICBF IDB platforms.

Consent was provided by the Bonsmara breeders society, for use of the phenotypic and genotypic data and ethical approval was granted from the Research/Ethics Committee (EC-180000127), Faculty of Natural and Agricultural Sciences at the University of Pretoria for the use of external data.

2.2 Materials

The number of animals with pedigree data, estimated breeding values and genotypes is summarized in Table 2.1, with pedigree records collected over a 70-year period (06/01/1949 to 18/08/2019).

Table 2.1 Summary of animals with pedigrees, estimated breeding values (EBVs) and genotypes for the SA Bonsmara breed available for this study

Number of Animals	With Pedigrees	Genotyped	EBVs
Male	917 322	1 522	2 098
Female	1 102 926	1 769	2 073
Total	2 020 248	3 291	4 171

In Table 2.2 the genotypic data originating from three separate arrays are shown. The number of SNPs, animals and mean call rate per genotyping array is included as well as the respective laboratories where genotyping was performed.

Table 2.2 Summary of genotypic data provided for the analyses

Genotyping panel	SNPs	Animals	Mean Call Rate	Laboratory
GeneSeek GGP 150K	140 113	1 949	0.977258	ARC BTP (South Africa)
GeneSeek GGP 80K	76 883	597	0.990379	GeneSeek (USA)
ICBF IDB V3	55 445	745	0.979629	Weatherbys (Ireland)

2.3 Methods

2.3.1 Initial data editing

Data was provided in two separate files containing the pedigrees, trait EBVs and associated accuracy values. RStudio (RStudio Team. RStudio., 2019) was used to separate the traits individually and appropriately format them for further downstream analysis. The pedigree file was edited into block code format required for use in MiX99 (Lidauer *et al.*, 2017b) and the given accuracies were converted into reliabilities in RStudio (RStudio Team. RStudio, 2019).

Genotypic data was received in PLINK-format input files (MAP and PED files) for the respective genotyping arrays. Duplicate animals across or within SNP arrays were identified and removed using RStudio (RStudio Team. RStudio, 2019) and animal IDs and sexes were updated in PLINK (v1.9) (Purcell *et al.*, 2007). SNPConvert (Nicolazzi *et al.*, 2016) was used to update the genotype files with a revised file downloaded from SNPChiMP (Nicolazzi *et al.*, 2015) using the UMD 3.1.1 build (GCF_000003055.6; Zimin *et al.*, 2009), to accommodate any changes in SNP names and/or base pair positions. The ICBF and GGP 80K genotyping platforms underwent the same protocol followed as above for the HD chip.

All data analysis was undertaken using the Linux Ubuntu software platform (Free Software Foundation, 2018) and a multitude of other software applications or platforms.

2.3.2 Deregression of estimated breeding values

Estimated breeding values for the fertility traits (AFC, ICPs 1, 2 and 3, WW_{MAT} and SC) were deregressed to remove any potential bias that may have occurred during the calculation of EBVs. The use of MiX99 (Strandén & Lidauer, 1999) requires multiple input files and is a command line software package requiring the use of the terminal in Unix systems. The use of APaX99 (Lidauer *et al.*, 2017a) requires a complete block code pedigree, a reliability file and an executable file. As

previously mentioned, the pedigree and reliability file were constructed in RStudio (RStudio Team. RStudio, 2019).

The effective record contributions (ERCs) were generated using the reversed reliability approximation method in APaX99 (Lidauer *et al.*, 2017b). Relevant ERCs were extracted using a Perl (Wall *et al.*, 2000) script to be used as a weighting factor in the deregression of the EBVs. The deregression step was done using the Secant method (Strandén & Mäntysaari, 2010) which requires the creation of a MIX.99.DIR file using the combination of the MiX99 pre-processor (Lidauer *et al.*, 2017b) and a .CLIM file (Lidauer *et al.*, 2017c). This involves the use of a block code pedigree including phantom parent groups, which, in the present study consisted of 60 separate pedigree clusters that grouped unknown parents from sire-sire, sire-dam and dam-dam lines and those from specific decades. SA Stud Book provided these genetic groups and phantom parent groups were appropriately allocated.

The data file contained the block code ID of an animal, a column of ones for each animal with an EBV and its respective ERC. Only animals with an ERCs of at least 0.5 were retained and the deregression was repeated. Table 2.3 indicates the data filtering that occurred with the initial number of animals with EBVs being 4 171 and the total number of animals with EBVs and an ERC >= 0.5 that underwent deregression.

Table 2.3 Summary statistics of the traits and animals with EBVs included in the deregression process

Trait	Heritability	Animals with ERC ≥ 0.5	Median ERC	ERC Range
AFC	0.22	3 230	1.3971	0.5033 - 58.056
ICP 1	0.13	3 259	2.478	0.5054 - 34.1130
ICP 2	0.09	3 221	3.6619	0.5026 - 45.6430
ICP 3	0.03	3 255	10.674	0.5146 - 258.97
MWW	0.15	3 385	4.3913	0.5016 - 114.85
SC	0.37	2 939	1.3752	0.5003 - 32.663

Lastly, the solve.deregression line of code using the MiX99 solver generates a Solani file containing the newly deregressed proofs. Animals with both deregressed EBVs (dEBVs) and post quality control genotypes were extracted from the Solani file using an appropriate Perl (Wall *et al.*, 2000) script. Plots generated in RStudio (RStudio Team. RStudio, 2019) indicate ERCs against reliabilities and EBVs against dEBVs.

2.3.3 Quality Control

A preliminary analysis and quality control (QC) of the 150K population was performed prior to the imputation of the lower density platforms using PLINK v1.9 (Purcell *et al.*, 2007). The given datasets were firstly transformed into binary files for easier manipulation and pruning of the genotype files. The following threshold parameters were applied. 1) Only autosomal SNPs were retained. 2) Any SNP with an unknown position was removed. 3) Only animals with a call rate greater than or equal to 0.9 were retained. 4) Only SNPs with a call rate of greater than or equal to 0.9 were retained. The filtered datasets were then recoded into PLINK format (MAP and PED) files for further downstream analysis. Genotyping rates as well as the number of remaining animals and SNPs were recorded after each QC step.

Due to the process of imputation, SNPs that occur on the lower density platforms must be present on the reference panel (Lashmar *et al.*, 2019). These are referred to as the common SNPs (Sargolzaei *et al.*, 2014). Thus, the high-density platform must first undergo pre-imputation QC before establishing the final list of SNPs that are common across all three genotyping arrays. Table 2.4 summarizes the number of SNPs on the genotyping platform, the loss of SNPs at each step, the number of animals in the dataset and the resultant change in mean GCR as QC proceeds.

Table 2.4 Summary of the quality control undertaken on the GGP 150 HD data frame using PLINK (Purcell *et al.*, 2007)

150 HD Dataset	Number of SNPs	Loss of SNPs	Number of Animals	Mean GCR
Original	140113	-	1949	0.977258
Common SNPs*	138898	1215	-	0.982249
Autosomal SNPs**	132294	6604	-	0.982448
Post-PCA	132294		1944	0.982445
Unknown/Duplicate BP	132236	58		0.98264
90% Individual Call Rate	132236	-	1932	0.983222
90% SNP Call Rate	128793†	3443	-	0.988047
Total Change		11320	17	0.010789

*SNPs that are on genotyping platform v1 and updated v2
**Removal of the X-Y sex chromosomes
†Final number of SNPs used for imputation using FImpute (v3) (Sargolzaei *et al.*, 2014)

Analysis and QC of the populations on the two lower density platforms occurred after the QC of the population genotyped with the Geneseek GGP 150K platform. The Geneseek GGP 150K array will be referred to as the reference panel from here on as it is the highest density that imputation will occur up to in this study.

The common SNPs among the three genotyping platforms were identified. Initial SNP densities, number of animals per platform and genotyping rates were recorded for a later comparison with the final dataset post-QC. QC of the respective platforms was completed using PLINK (v1.9) (Purcell *et al.*, 2007).

Eight and thirteen duplicate animals originating from the Geneseek GGP 80K panel and the ICBF IDB v.3 array respectively, were removed using the *--remove [file.txt]* command in PLINK (v1.9) (Purcell *et al.*, 2007).

Table 2.5 below indicates the number of SNPs retained and final number of animals genotyped on the Geneseek GGP 80K HD platform.

Table 2.5 Summary of the quality control undertaken on the GGP 80 HD data frame using PLINK (Purcell *et al.*, 2007)

80 HD Dataset	Number of SNPs	Loss of SNPs	Number of Animals	Mean GCR
Original	76883	-	597	0.990379
Duplicate Animals	-	-	589	0.988035
Common SNPs*	69430	7453	-	0.993053
90% SNP Call Rate	68865	565	-	0.994818
Total Change		8018	8	0.004439

*SNPs that are present on the reference panel

Table 2.6 indicates the final number of animals genotyped and number of SNPs remaining on the ICBF IDB v3 array.

Table 2.6 Summary of the quality control undertaken on the ICBF IDB v.3 data frame using PLINK (Purcell *et al.*, 2007)

ICBF IDB v3 Dataset	Number of SNPs	Loss of SNPs	Number of Animals	Mean GCR
Original	52445	-	745	0.979629
Duplicate Animals	-	-	732	0.979681
Common SNPs*	36629	15816	-	0.98713
90% SNP Call Rate	35706	923	-	0.991721
Total Change		16739	13	0.012092

*SNPs that are present on the reference panel

Genotyping rates as well as the number of remaining animals and SNPs were recorded after each QC step.

2.3.4 Principal component analysis (PCA)

The main goal of principal component analysis (PCA) is to minimize the confounding of genotypes through population stratification or cryptic relatedness and exclude random outliers. The software GCTA (Yang *et al.*, 2011) was used to calculate eigenvectors and eigenvalues. This involved the use of binary files generated in PLINK (Purcell *et al.*, 2007), these being .bin, .fam and .bam files respectively. These serve to generate a genomic relationship matrix (GRM), which is then used to calculate the principal components. RStudio (RStudio Team. RStudio, 2019) was used to produce both 2D and 3D plots using the (rgl) package in order to visualize the degree of genetic relatedness between the individuals genotyped for the three respective platforms.

The identification and removal of five animals on the 150K GGP HD SNP chip of Namibian origin occurred as they lay -0.3 eigenvectors away from the average of the population. A new PCA plot was created after a rerun of the aforementioned QC with the initial unfiltered dataset was re-run as the resultant allele frequencies would change. A plot was also generated for the total combined population pre- and post-imputation.

2.3.5 Imputation

Table 2.7 below indicates the number of common SNPs between the different SNP arrays, respectively.

Table 2.7 Summary of the number of common SNPs across different arrays

Genotyping Platform	Number of common SNPs with Geneseek GGP 150K
Geneseek GGP 150K	128 793
Geneseek GGP 80K	69 417
ICBF IDB v3	36 606
All Three Platforms	23 646

The use of FImpute (v3) (Sargolzaei *et al.*, 2014) requires four different input files (Sargolzaei *et al.*, 2014), which include a .ped file with the pedigree for the Bonsmara breed, a .snp_info_file that designates SNP IDs, their respective bp position on the chromosome as well as whether the SNP is on the LD platform. SNPs absent on the LD platforms are designated a 0 indicating it must be imputed, otherwise it is given the number it has on the SNP chip. The .geno file assigns an animal ID with its respective genotype in the 0, 1, 2 or 5 formats. The .ctr_file contains the executable information necessary for FImpute (v3) to run. All necessary file creation or manipulation was completed using Unix (Free Software Foundation, 2018), Perl (Wall *et al.*, 2000) or RStudio (RStudio Team. RStudio, 2019) respectively.

The imputation of 1 321 animals, being 677 males and 644 females, from the lower density SNP platforms was done using FImpute v.3 with a reference population comprising of 1 932 animals, being 818 males and 1 114 females. Table 2.8 below indicates the squared correlation between allele frequencies among chips for all SNPs.

Table 2.8 Squared correlation between allele frequencies on the three various platforms.

Array	150 GGP	80 GGP	ICBF IDB v3
150 GGP	1		
80 GGP	0.999	1	
ICBF IDB v3	0.997	0.997	1

Overall time for the completion of imputation was fourteen minutes and forty seconds.

2.3.6 Genome-wide association studies (GWAS)

Deregressed EBVs are the dependent variables used as phenotypes in this study. The source of available information allows for the weighting of these phenotypes by the following equation described in Garrick *et al.*, (2009);

$$w_i = \frac{1 - h^2}{[c + \frac{1 - r_i^2}{r_i^2}]h^2}$$

Where

w is the weighting factor of the ith animal with a deregressed EBV;

h^2 is the heritability estimate for the respective traits;

r^2 is the reliability of the ith animal for a specific trait

and c is the proportion of genetic variance not accounted by the SNPs with a value of 0.9 being used for all weighting factors between all the traits under analysis.

These weightings were calculated in RStudio (RStudio Team. RStudio, 2019) and will be used in conjunction with the dEBVs in further downstream analysis.

Inter-calving period consists of three sequential traits, which are known to be highly genetically correlated and therefore ICPs 1, 2 and 3 were weighted equally and merged into one trait. Only animals with EBVs across all three ICP traits remained for analysis.

A GRM was constructed among all animals for each trait respectively using the VanRaden method 1 (VanRaden, 2008). A GRM for each trait analysed was created as the number of genotyped animals per trait differs due to the aforementioned steps taken to arrive at this point.

Table 2.9 indicates the number of animals in the four GRMs for each trait created using HGInv program (Strandén, 2014) within the MiX99 software package.

Table 2.9 Summary of number of animals in genomic relationship matrix (GRM) creation

Trait	Number of Animals in GRM
AFC	2 620
ICP	2 245
WW_{MAT}	2 730
SC	2 318

WOMBAT (Meyer, 2007), in conjunction with linear mixed models, was used for single SNP regression association analyses in order to calculate the SNP effects of the subset of animals being investigated within each trait. The model fitted for each SNP analysis was:

$$\textit{Deregressed EBV} = \mu + \textit{SNP} + a + e,$$

Where

the *deregressed* EBV is the weighted dependant variable;

μ is the fixed effect of the population mean;

SNP is the fixed effect of allele dosage for each SNP (coded as 0, 1 or 2);

a is the random effect of the animal, where $a \sim (0, \boldsymbol{G}\sigma_a^2)$, with σ_a^2 representing the additive genetic variance of the animal;

G is the genomic relationship matrix among animals, part of *a*;

e represents the residual, where $e \sim N(0, \mathbf{I}\,\sigma_e^2)$,

with σ_e^2 representing the residual variance

and **I** the identity matrix.

The dependant variable, the deregressed EBV, was weighted using the weightings calculated previously. The GWAS feature in WOMBAT was invoked using the **--snap** run time option (Meyer & Tier, 2012). t-test distribution statistics for all SNPs were obtained and SNPs with a $P \leq 5 \times 10^{-8}$ were considered to be genome-wide significant.

The various t-distributions were modelled in various Manhattan plots using a range of R packages. These include; *readr, plotly* and *qqman*. Output files for WOMBAT were assessed and appropriate results were displayed. Putative genes could be identified from any peaks that are indicative of suggestive ($P \leq 1 \times 10^{-5}$) and/or significant ($P \leq 1 \times 10^{-8}$) SNPs.

Figure 2.1 is a flow diagram that illustrates the number of individual steps taken with regards to the phenotypic and genotypic data that was analysed in this study.

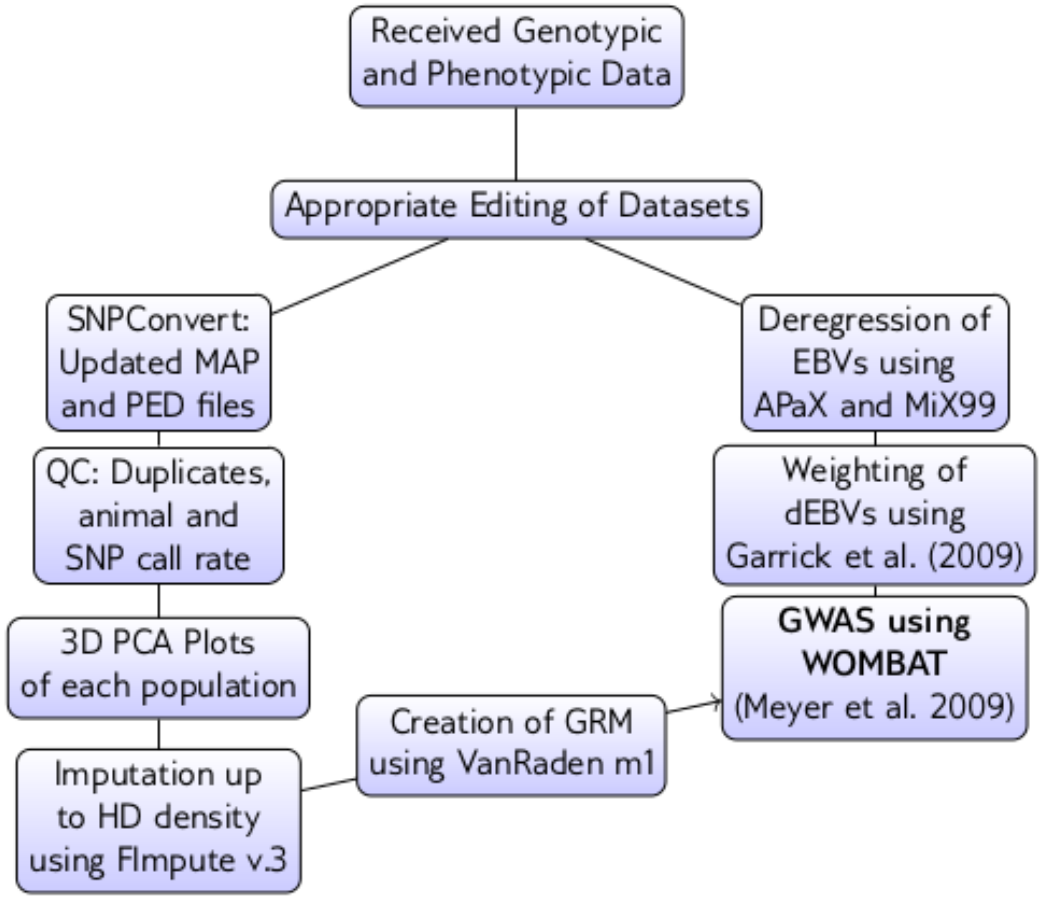


Figure 2.1 Flow diagram that indicates the approach conducted for this GWAS study

2.3.7 Identification of putative genes

SNPs with a $P \leq 1 \times 10^{-5}$ were considered suggestive while a $P \leq 1 \times 10^{-8}$ was treated as significant. SNPs that were deemed to be suggestive and/or significant were identified in RStudio (RStudio Team. RStudio., 2019) and any genes located within 5000bp boundary up and downstream from the associated SNP position. When originally updating with SNPCHiMP (Nicolazzi *et al.*, 2015), the UMD_3.1.1 (GCF_000003055.6) was used to align the SNPs and is subsequently corroborated on NCBI ('National Center for Biotechnology Information', 2020) to identify any significant SNPs from this study. Panther (Mi *et al.*, 2017) was used to list biological and metabolic functions and/or processes involved with genes identified near any significant SNPs.

Chapter 3 Results

Results are presented for all steps involved in data preparation for performing genome-wide association (GWA) analysis on the four different traits. The first step involved deregression of estimated breeding values (EBVs) followed by a principal component analysis, calculation of weightings for deregressed EBVs (dEBVs) and finally the association analysis.

3.1 Deregression of estimated breeding values (EBVs)

The calculation of effective record contributions (ERC) involved the reliabilities of 2 020 248 animals in a pedigree file dating back to 1949. Figure 3.1 illustrate the relationship between the reliability an animal has, and the ERC calculated in the reversed reliability approximation step using APaX99 (Lidauer *et al.*, 2017a) for AFC. Additional plots for the other traits of interest are in the Addendum A as Supplementary Figures A1-A3.

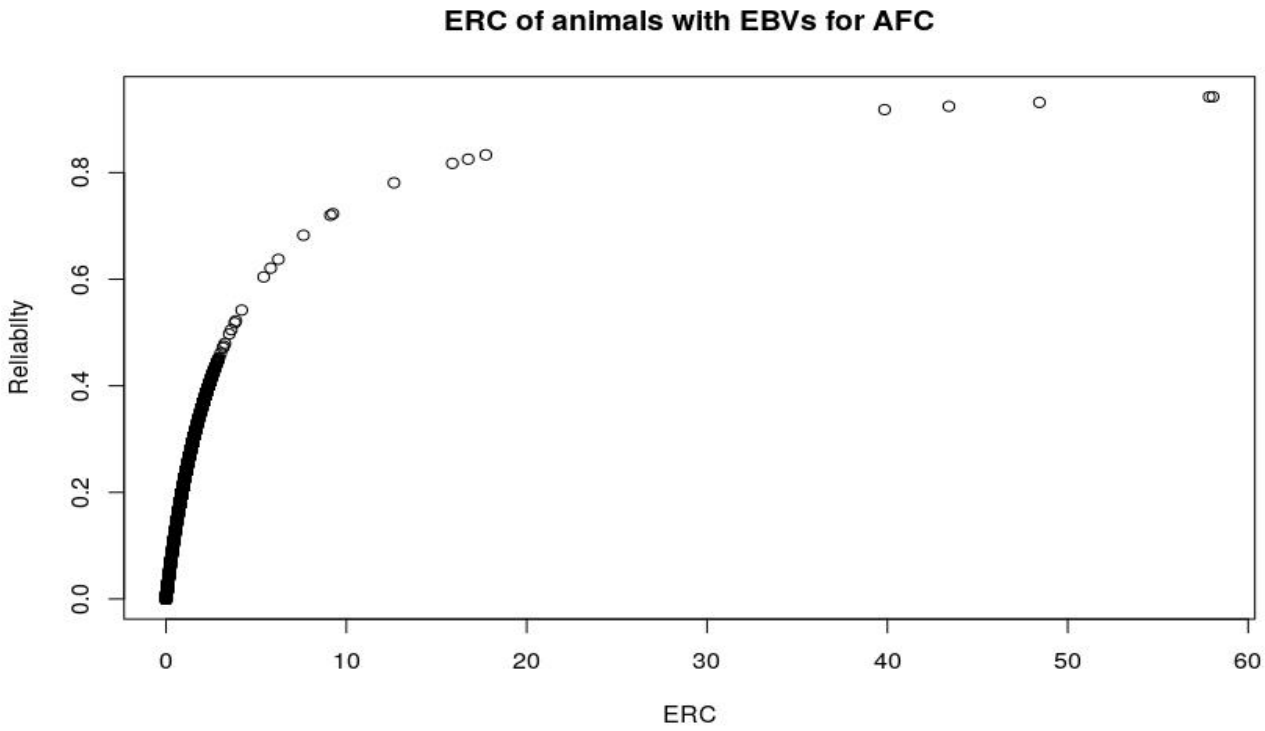


Figure 3.1 Plot of effective record contributions (ERC) against reliabilities for age at first calving

A threshold of only including animals with ERCs ≥ 0.5 was applied to the datasets. Figure 3.2 reveals the relationship between the EBVs and the dEBVs for AFC. All figures for the other traits of interest with ERC ≥ 0.5 are in Addendum A as Supplementary Figures A4-A8.

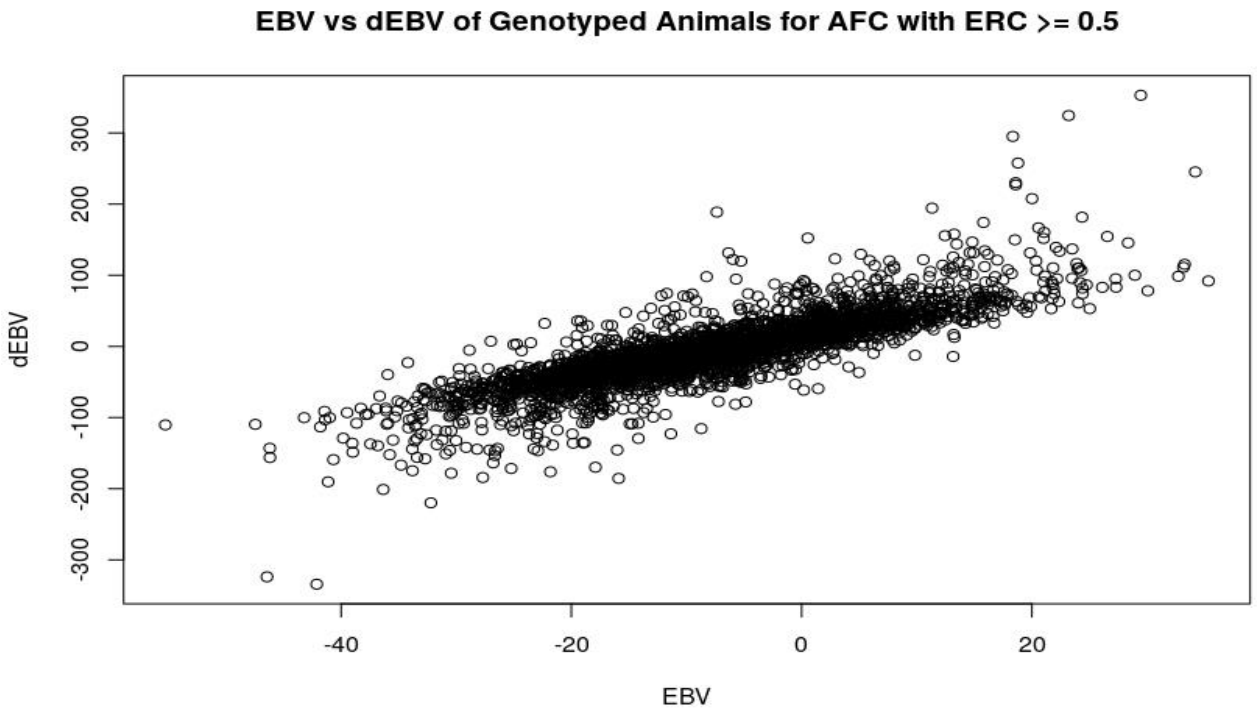


Figure 3.2 Plot of estimated breeding values against deregressed estimated breeding values for age at first calving

3.2 Principal component analysis (PCA)

Results for principal component analysis (PCA) are split into four categories. The first three categories are related to each genotyping platform individually, with the fourth category being all the animals in this study, pre- and post-imputation. The first three categories have Supplementary Figures as well as a PCA of all the imputed genotypes in the Addenda and are summarized below.

A high proportion of the reference population was shown to be tightly clustered together (Supplementary Figure B1), irrespective of country of origin. The left half of the plot indicated five male Namibian animals, -0.4 eigenvectors on PCA 2 away from the general cluster, which proved to be outliers. After the removal of these five outliers, the process of creating a GRM was repeated. Eigenvectors were generated and a new PCA plot was created (Figure 3.3). This indicated a few SA male animals as outliers, with the majority of the population clustered sufficiently close together. This was the final group of animals used as the reference population on the HD array for the imputation following technical QC. All SA female animals were clustered with the SA male animals respectively. The PC 1 and PC 2 plot for the above scenario is in Addendum B as Supplementary Figure B2.

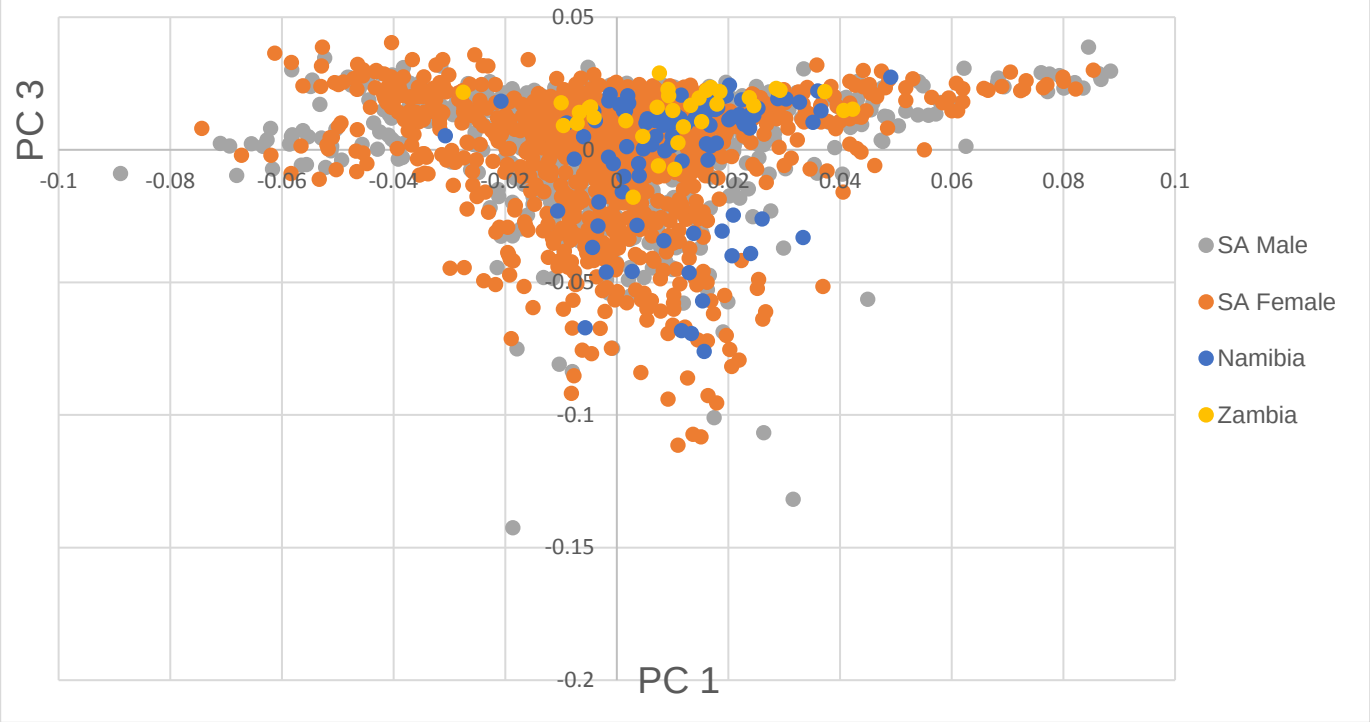


Figure 3.3 Genetic relationships between 1932 Bonsmara animals and 128 793 SNPs genotyped on the GGP 150K HD array for the first and third principal components (PC 1 and PC 3)

Figures 3.4 and 3.5 are the PCA plots for animals that are genotyped on the GGP 80K HD SNPChip. This PCA was done post technical QC. The clustering of individuals is desirable within this population with few if any outliers present.

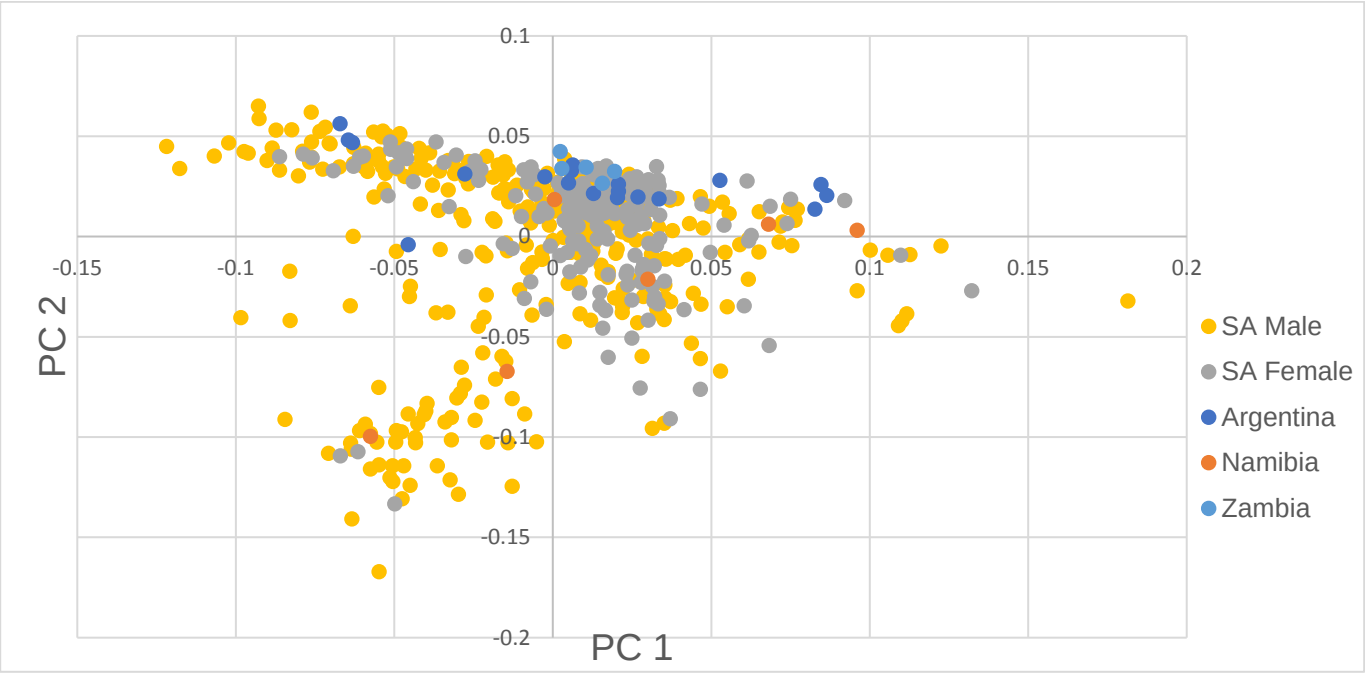


Figure 3.4 Genetic relationships between 589 Bonsmara animals and 69 417 SNPs genotyped on the GGP 80K array for the first and second principal components (PC 1 and PC 2)

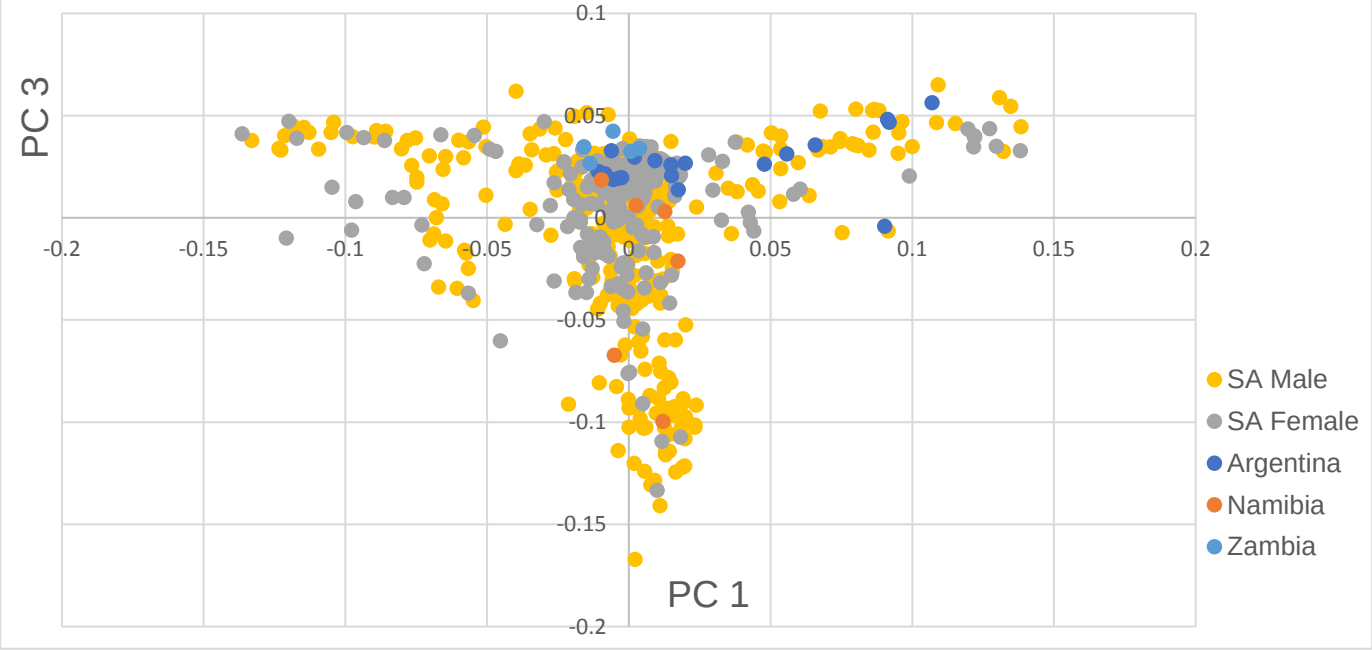


Figure 3.5 Genetic relationships between 589 Bonsmara animals and 69 417 SNPs genotyped on the GGP 80K array for the first and third principal components (PC 1 and PC 3)

Figures 3.6 and 3.7 are the principal component plots for animals that are genotyped on the IDB ICBF v3 SNPChip. These plots indicated one large cluster and two smaller ones. It was assumed due to the low number of SNPs that not all the genetic variation present in the Bonsmara population was captured and the GRM presented slightly biased results.

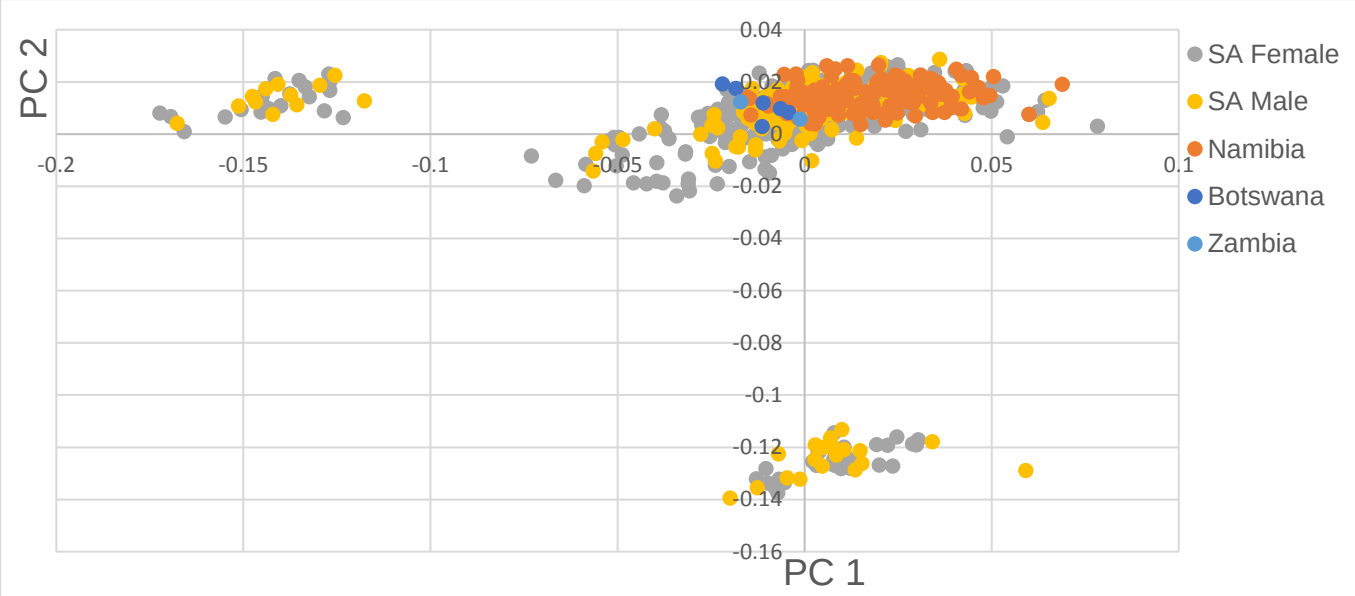


Figure 3.6 Genetic relationships between 732 Bonsmara animals and 36 605 SNPs for the first and second principal components (PC 1 and PC 2)

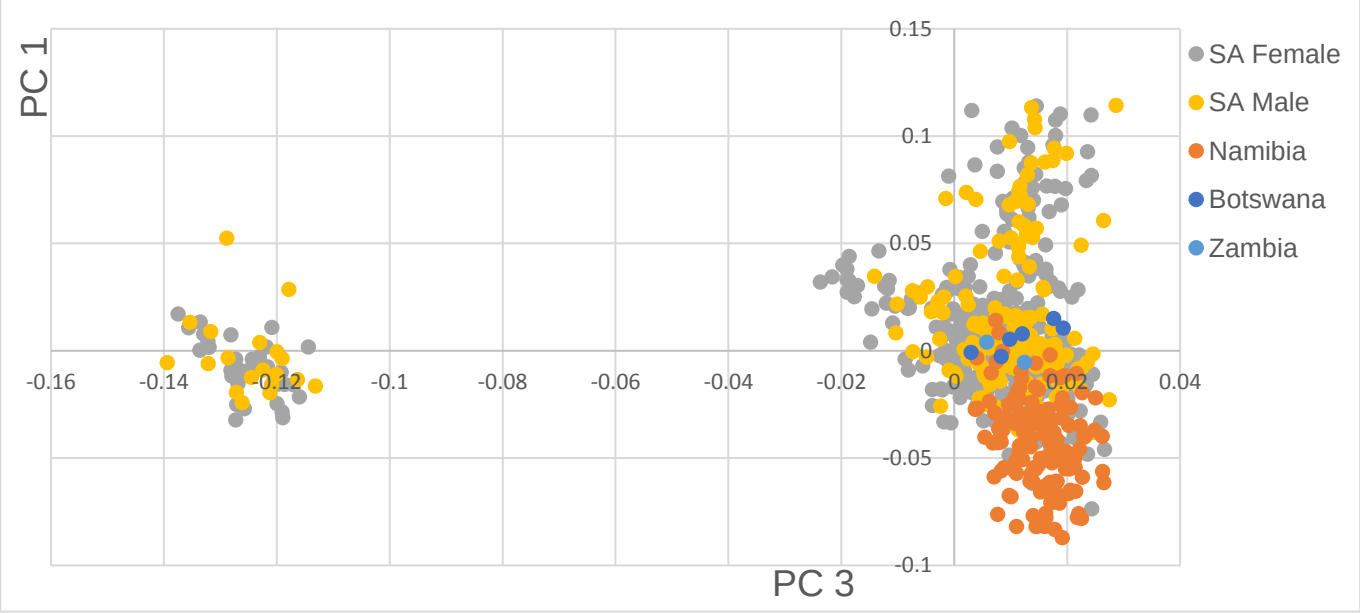


Figure 3.7 Genetic relationships between 732 Bonsmara animals and 36 605 SNPs for the first and third principal components (PC 1 and PC 3)

Thus, to ensure minimal population stratification, a PCA on all the animals on the common SNPs between the three different arrays was run, involving 23 646 markers. Figure 3.8 indicates the first and second PC for 3 253 animals originating from five countries and 216 different herds. All animals are divided into their respective country of birth and the South African population is split on sex in order to differentiate the large population of SA animals.

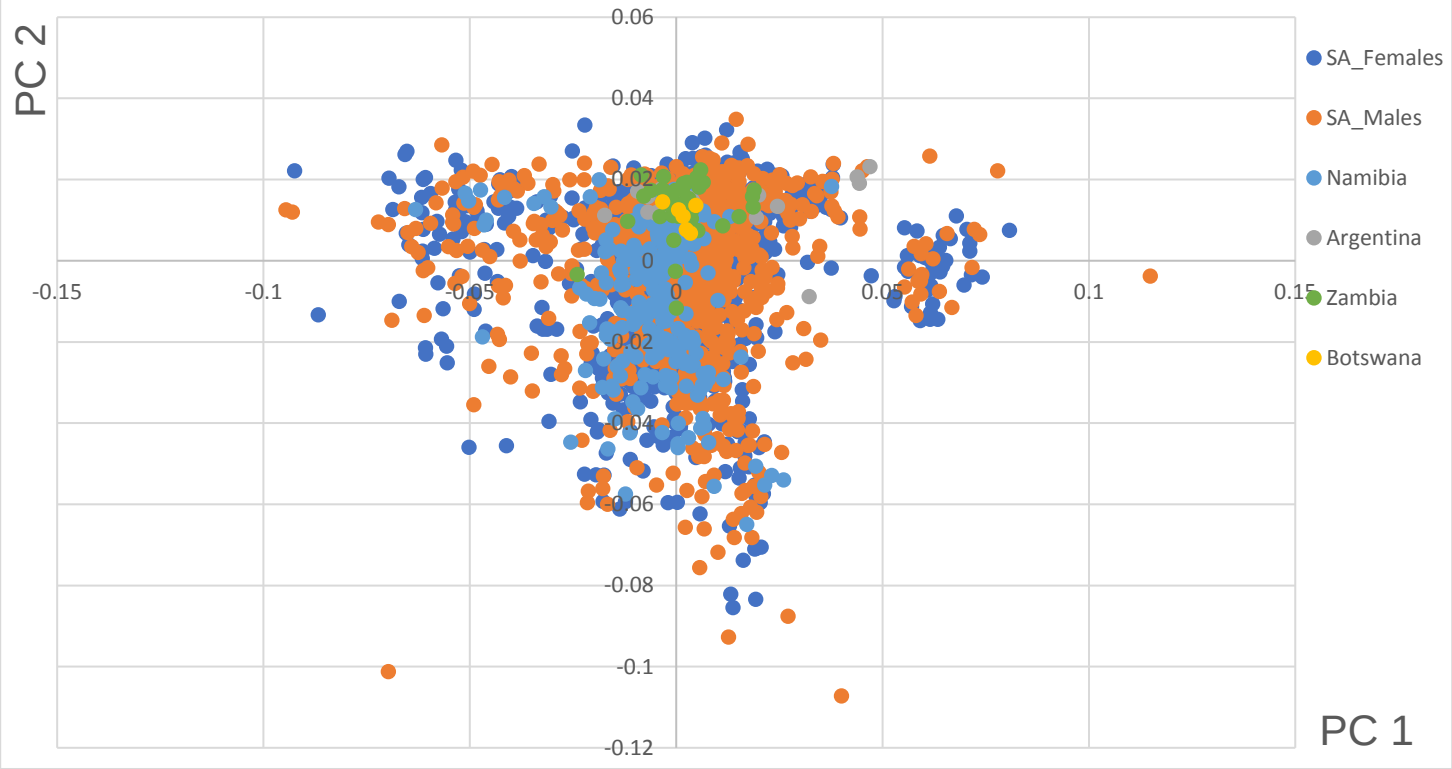


Figure 3.8 Genetic relationships between 3 253 Bonsmara animals originating from different countries on 23 646 SNPs for the first and second principal components (PC 1 and PC 2)

A PCA analysis of the population was run post-imputation where the genetic relationships between 3 253 Bonsmara animals, each with a genomic profile of 128 793 SNPs and is in Addendum B (Supplementary Figure B3).

3.3 Genome wide association studies (GWAS)

Calculation of Weightings

The weightings that were applied to the dependant variable, the dEBV, are summarized below in Table 3.1.

Table 3.1 Summary statistics of the weightings calculated by the method described in Garrick *et al. (*2009) for the respective traits of interest

Trait	Heritability	Median Weight	Weight Range
AFC	0.22	1.7724	0.4655 - 3.8525
ICP 1	0.13	1.8586	0.4733 - 6.1051
ICP 2	0.09	2.7618	0.4811 - 9.0155
ICP 3	0.03	10.468	1.299 - 33.582
WW_{MAT}	0.15	4.085	1.280 - 6.157
SC	0.37	1.103	0.4547 - 1.8502

Within-breed genome-wide association

Several Manhattan plots were generated (Figures 3.9; 3.10; 3.12 and 3.13) showing the association between each SNP and dEBV for the four different fertility traits. Tables 3.3, 3.4, 3.5a, 3.5b, 3.6a and 3.6b show the number of significant SNPs identified, their respective p-value, the favourable allele as well its frequency. The UMD 3.1 (Zimin *et al.*, 2009) build was used to identify relevant reference SNP cluster ID (rs ID) of significant SNPs. Panther (Mi *et al.*, 2017) indicates whether these SNPs were in gene regions and gives the annotation of the SNP in relation to nearby genes.

Figure 3.9, below, is the Manhattan plot for inter-calving period with a significance level of p-value < 1×10^{-8} being the red line and the blue line indicates those SNPs which are suggestive with a p-value < 1×10^{-5}. A total of 2 245 animals were used in the association analysis of ICP.

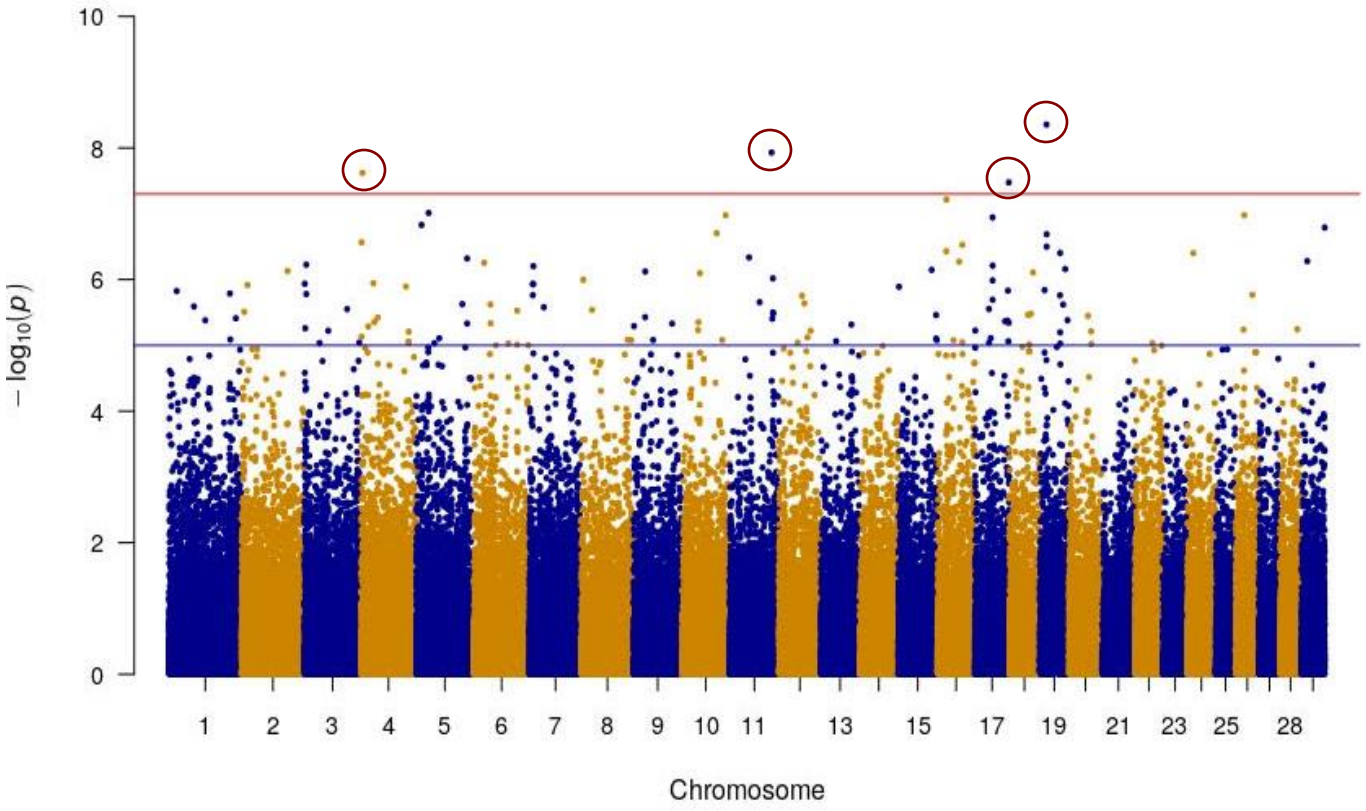


Figure 3.9 Manhattan plot for inter calving period (ICP, 2 245 animals) indicating SNPs with significant p-values above the red line (≤ $1x10^{-8}$)

Four SNPs were identified as being significantly associated with inter calving period. These are located on BTA 4, 11, 17 and 19 with Table 3.2 indicating that none of them are annotated near or within any known genes.

Table 3.2 Chromosome (BTA), position, favourable allele (FA), frequency (*f*) of the favourable allele, the rs ID and annotation of SNP for the 4 single nucleotide polymorphisms associated with inter calving period (ICP)

BTA	Position	SNP	P-value	FA	*f*(FA)	rs ID	Annotation
4	3676858	BovineHD0400000926	$2.39x10^{-8}$	A	0.86	rs137550857	intergenic
11	91069652	BovineHD1100026403	$1.17x10^{-8}$	B	0.77	rs109882606	intergenic
17	74103298	BovineHD1700021677	$3.34x10^{-8}$	B	0.6	rs109444356	intergenic
19	15107270	ARS-BFGL-NGS-40104	$4.40x10^{-8}$	A	0.75	rs109300492	intergenic

The Manhattan plot as Figure 3.10 indicates twenty SNPs associated with age at first calving at the genome wide threshold p-value of 1 x 10^{-8}.

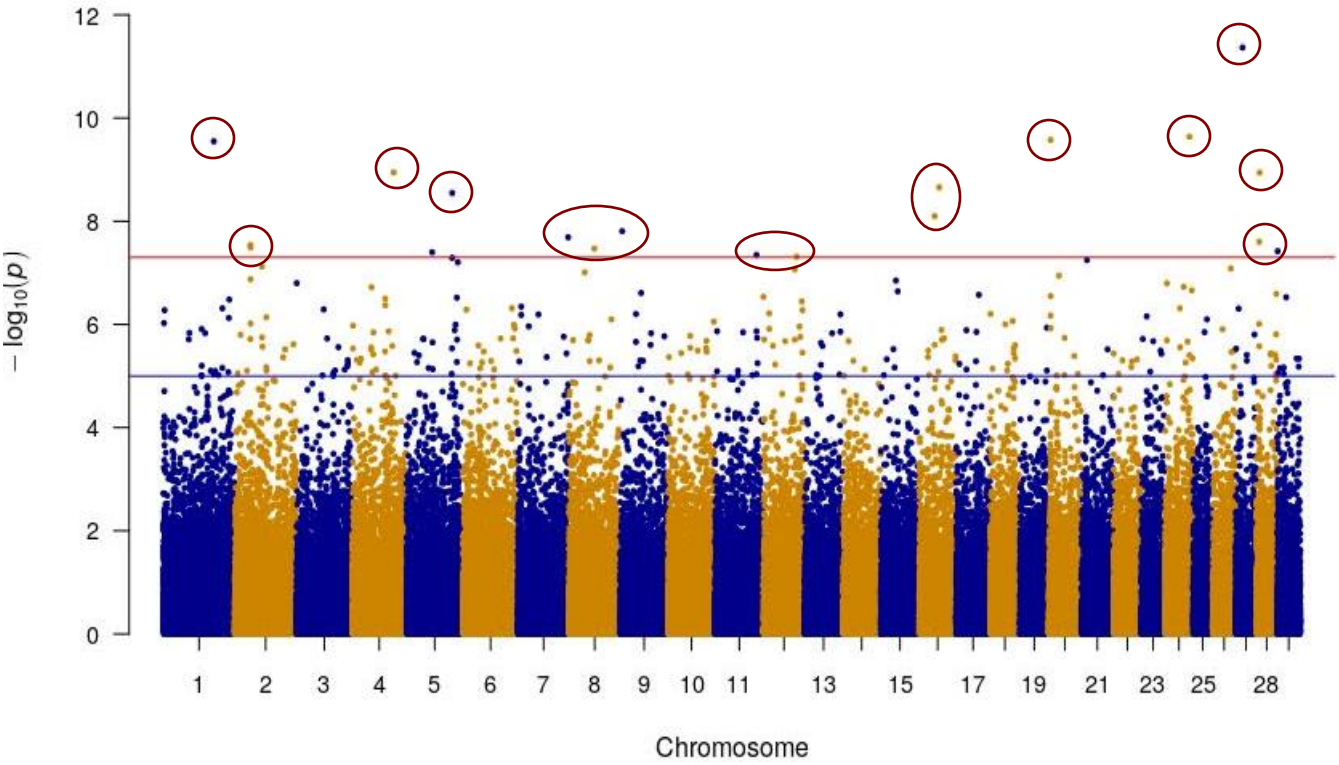


Figure 3.10 Manhattan plot for age at first calving (AFC, 2 620 animals) indicating SNPs with significant p-values above the red line (≤ 1x10^{-8})

A total of eight different genes were found across five different autosomes (BTA 4, 5, 7, 12 and 16) with the frequency of the favourable alleles being between 0.13 and 0.40 respectively. Three of these SNPs were identified on BTA 5 (Figure 3.11) and each being in the region of its own gene (*RDH16*, *OVOS2* and *A2M*) with each SNP either being upstream of the gene, in an intron or within the 5'-untranslated region (UTR) of the gene.

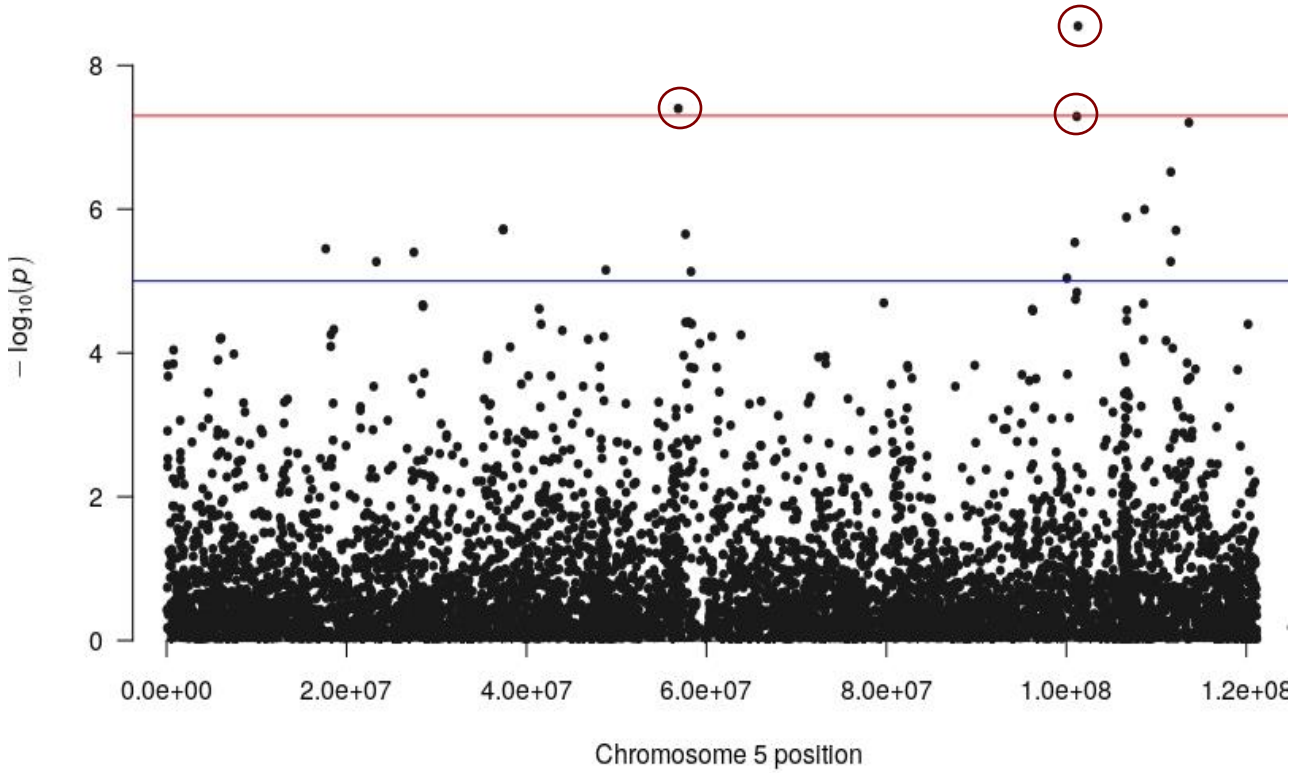


Figure 3.11 Manhattan plot of BTA 5 for age at first calving (AFC, 2 620 animals) indicating SNPs with significant p-values above the red line (≤ $1x10^{-8}$)

BTA 16 had two SNPs, rs134860307 (p-value = $7.94x10^{-9}$, *f*(FA) = 0.71) and rs41579702 (p-value = $2.21x10^{-9}$, *f*(FA) = 0.6) that were in the intron region of the genes, *AKT3* and *KIF1B*. Three autosomes (BTA 4, 7 and 12) had one SNP each occurring in genes (*GRM8*, *FER* and *UBAC2*). The four most significant SNPs, with p-value > $2.3x10^{-10}$ were all found to be intergenic (Table 3.3).

Table 3.3 Chromosome (BTA), position, favourable allele (FA), frequency (*f*) of the favourable allele, the rs ID, annotation of SNP and possible genes of the 20 single nucleotide polymorphisms associated with age at first calving (AFC)

BTA	Position	SNP	P-value	FA	*f*(FA)	rs ID	Annotation	Gene
1	111917015	BovineHD0100031651	$2.82x10^{-10}$	A	0.73	rs137537403	intergenic	-
2	34955867	BovineHD0200010311	$2.88x10^{-8}$	A	0.99	rs133156474	intergenic	-
2	35004695	ARS-BFGL-NGS-100791	$3.21x10^{-8}$	A	0.53	rs109816434	intergenic	-
4	**92201419**	**BTB-00201722**	**$1.13x10^{-9}$**	**B**	**0.84**	**rs43415362**	**intron**	**GRM8**
5	**56862044**	**BovineHD0500016118**	**$3.98x10^{-8}$**	**B**	**0.71**	**rs110633639**	**upstream gene**	**RDH16**
5	**101162000**	**BovineHD0500028941**	**$5.13x10^{-8}$**	**B**	**0.8**	**rs135781979**	**intron**	**OVOS2**
5	**101298140**	**BovineHD0500028996**	**$2.84x10^{-9}$**	**B**	**0.87**	**rs135044484**	**5-UTR**	**A2M**
7	**110620779**	**BovineHD0700032252**	**$2.05x10^{-8}$**	**B**	**0.86**	**rs43027016**	**intron**	**FER**
8	56151592	ARS-BFGL-NGS-118192	$3.39x10^{-8}$	B	0.53	rs110050395	intergenic	-
9	4141659	BovineHD0900000806	$1.56x10^{-8}$	B	0.54	rs136309097	intergenic	-
11	91868770	BovineHD1100026672	$4.51x10^{-8}$	A	0.97	rs110939063	intergenic	-
12	**73981915**	**ARS-BFGL-NGS-34730**	**$4.85x10^{-8}$**	**A**	**0.67**	**rs109026379**	**intron**	**UBAC2**
16	**34310513**	**BovineHD1600009813**	**$7.94x10^{-9}$**	**B**	**0.71**	**rs134860307**	**intron**	**AKT3**
16	**44223723**	**Hapmap39326-BTA-38925**	**$2.21x10^{-9}$**	**B**	**0.6**	**rs41579702**	**intron**	**KIF1B**
20	4373876	BovineHD2000001407	$2.65x10^{-10}$	B	0.86	rs110185739	intergenic	-
24	54578972	BovineHD2400015563	$2.31x10^{-10}$	B	0.53	rs109536476	intergenic	-
27	14901236	Hapmap57076-ss46526296	$4.29x10^{-12}$	B	0.83	rs41255435	intergenic	-
28	6606376	BovineHD2800001980	$2.52x10^{-8}$	B	0.81	rs43684018	intergenic	-
28	7516559	BovineHD2800002256	$1.14x10^{-9}$	A	0.54	rs42132791	intergenic	-
29	1310025	BovineHD2900000229	$3.81x10^{-8}$	A	0.66	rs42153312	intergenic	-

*SNPs highlighted bold were found in gene regions

An association analyses of SC resulted in twenty-two SNPs having significant p-values across 16 autosomes (BTA 1, 2, 3, 4, 5, 6, 8, 10, 11, 15, 16, 20, 22, 24, 26 and 28) as shown in Figure 3.12.

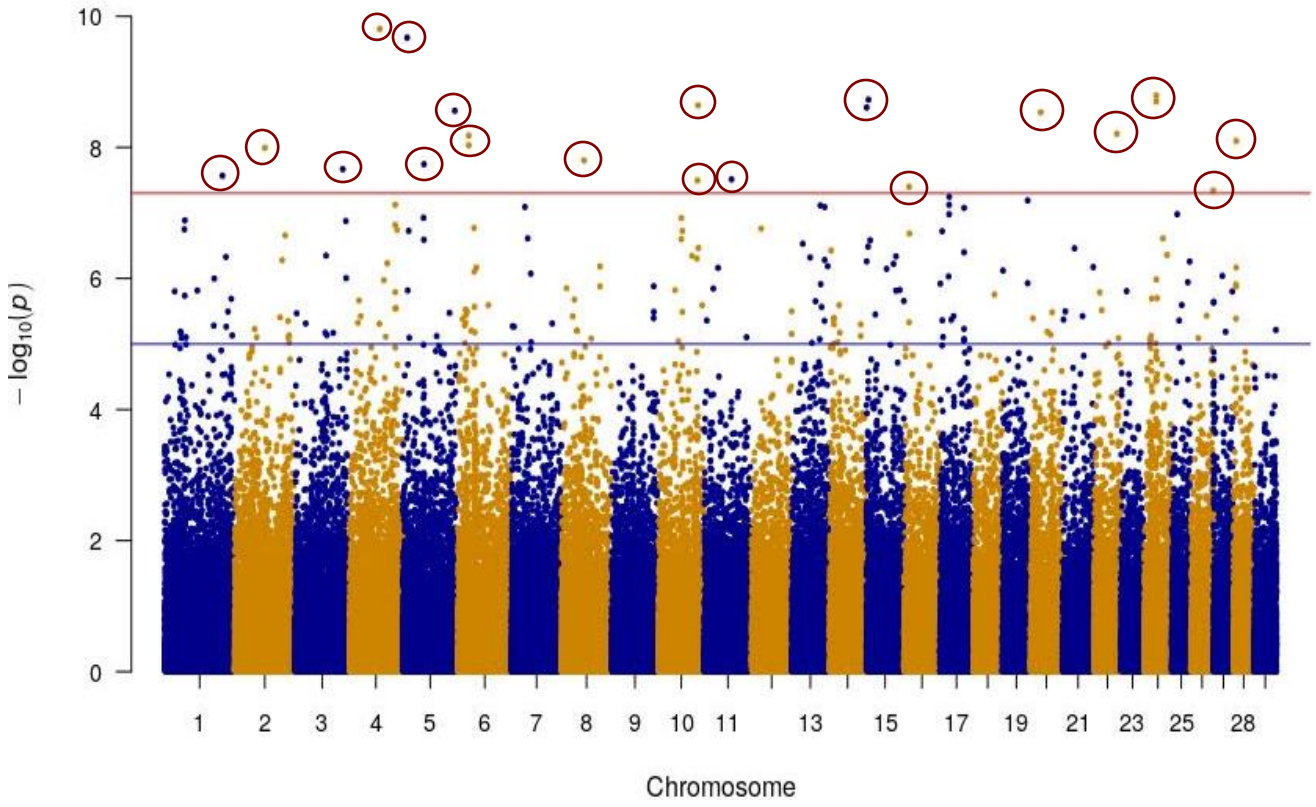


Figure 3.12 Manhattan plot for scrotal circumference (SC, 2 318 animals) indicating SNPs with significant p-values above the red line (≤ $1x10^{-8}$)

For SC, three of the SNPs were within intron regions of BTA 6 (p-value = 6.58×10^{-9}, *f*(FA) = 0.55), BTA 8 (p-value = 1.58×10^{-8}, *f*(FA) = 0.53) and BTA 11 (p-value = 3.07×10^{-8}, *f*(FA) = 0.75) respectively. The most significant SNP, BovineHD0100022236 (p-value = 1.18×10^{-15}, *f*(FA) = 0.65) was on BTA 1 and was found to be intergenic. Table 3.4 below, summarizes SNPs associated with SC.

Table 3.4 Chromosome (BTA), position, favourable allele (FA), frequency (*f*) of the favourable allele, the rs ID, annotation of SNP and possible genes of the 22 single nucleotide polymorphisms associated with scrotal circumference (SC)

BTA	Position	SNP	P-value	FA	*f*(FA)	rs ID	Annotation	Gene
1	77175969	BovineHD0100022236	1.18x10-15	B	0.65	rs137056215	intergenic	-
2	68483026	BovineHD0200019795	1.02x10-8	A	0.59	rs43763538	intergenic	-
3	108353435	BovineHD0300031128	2.15x10-8	A	0.5	rs135183004	intergenic	-
4	69143072	BovineHD0400019016	1.55x10-10	A	0.89	rs135196244	intergenic	-
5	10363213	BovineHD0500002967	2.11x10-10	B	0.65	rs43434439	intergenic	-
5	48521028	BovineHD0500014005	1.8x10-8	B	0.55	rs135351569	intergenic	-
5	118456584	BovineHD0500034431	2.76x10-9	B	0.56	rs135354351	intergenic	-
6	**24885679**	**BovineHD0600006847**	**6.58x10-9**	**B**	**0.55**	**rs42547321**	**intron**	**PPP3CA**
6	25127396	6_25127396	9.3x10-9	B	0.99	-	-	-
8	**51227380**	**Hapmap51395-BTA-113897**	**1.58x10-8**	**A**	**0.53**	**rs41573073**	**intron**	**TRPM6**
10	88909135	BovineHD1000025336	3.20x10-8	B	0.61	rs42342728	intergenic	-
10	90164914	BovineHD1000025763	2.29x10-9	A	0.57	rs135646907	intergenic	-
11	**61772774**	**BovineHD1100017519**	**3.07x10-8**	**A**	**0.75**	**rs134599584**	**intron**	**WDPCP**
15	787417	BovineHD1500000208	2.46x10-9	A	0.7	rs110874670	intergenic	-
15	4998555	BovineHD1500001208	1.87x10-9	A	0.52	rs137457807	intergenic	-
16	11816910	BovineHD1600003156	3.98x10-8	A	0.55	rs137316280	intergenic	-
20	23201811	BovineHD2000006992	2.92x10-9	B	0.51	rs110125425	intergenic	-
22	52138159	BovineHD2200014828	6.20x10-9	B	0.71	rs110310571	intergenic	-
24	27636658	BovineHD2400007539	1.62x10-9	B	0.78	rs43074384	intergenic	-
24	27637509	BovineHD2400007540	1.98x10-9	B	0.7	rs133851661	intergenic	-
26	51324600	BovineHD2600014822	4.58x10-8	A	0.55	rs42415608	intergenic	-
28	6314273	BovineHD2800001887	7.95x10-9	A	0.61	rs132650071	intergenic	-

*SNPs highlighted bold were found in gene regions

Weaning weight maternal (WW_{MAT}) was the trait that had the most significant SNPs as well as having the most SNPs in gene regions between all the traits of interest in this study. A total of forty-four SNPs were identified with a p-value < 1 x 10^{-8}, spread across twenty-one autosomes (BTA 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 14, 15, 18, 19, 20, 21, 22, 27, 28 and 29). None of the fourteen SNPs observed on BTA 1, 3, 4, 7, 11, 12, 20, 21, 27 and 28 were based near or within a gene region. Of the thirty remaining SNPs on eleven autosomes, twenty were found to be in nineteen

different gene regions. The Manhattan plot for WW_{MAT} is below as Figure 3.13 where a total of 128 793 SNPs across 29 autosomes were analysed in this study.

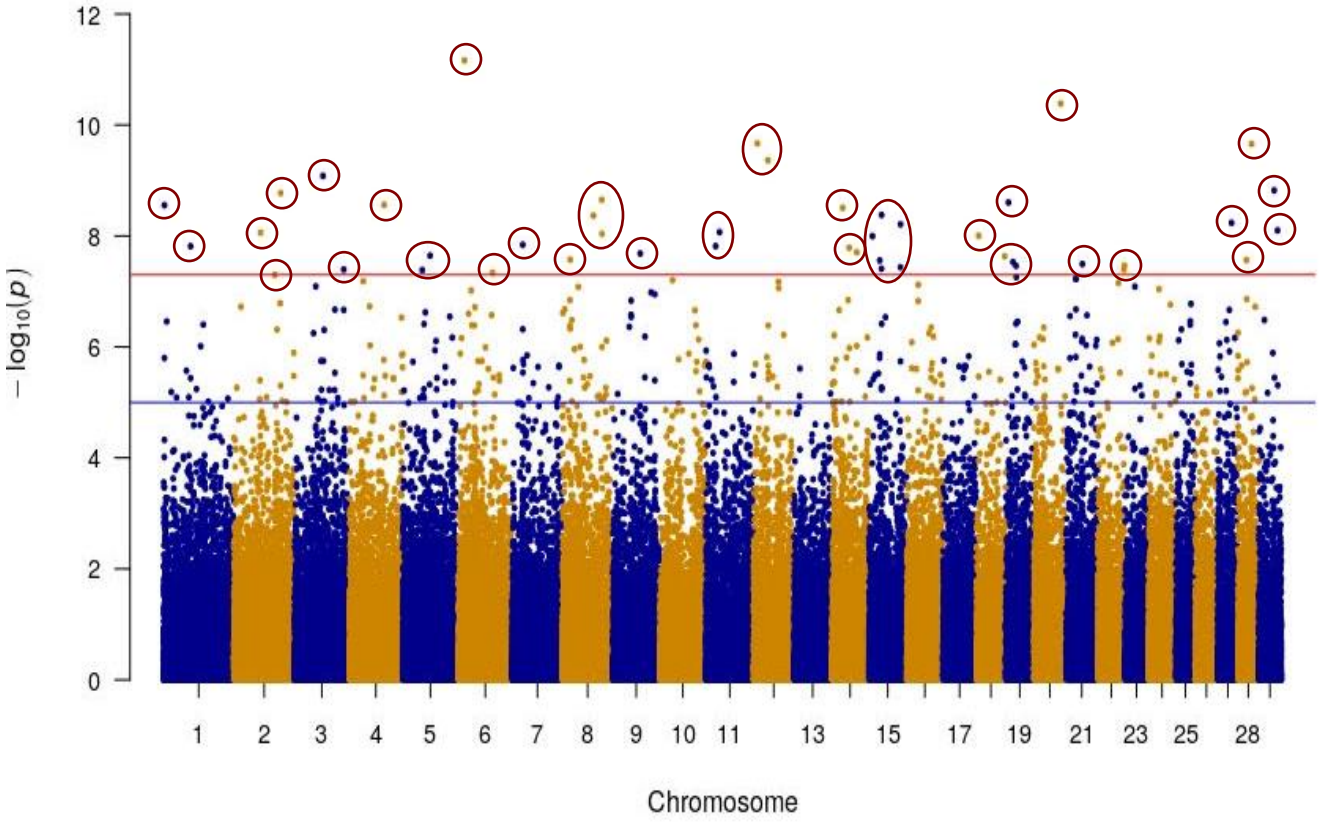


Figure 3.13 Manhattan plot for weaning weight maternal (WW_{MAT}, 2 730 animals) indicating SNPs with significant p-values above the red line (≤ $1x10^{-8}$)

The most significant SNP (p-value = $2.11x10^{-18}$, *f*(FA) = 0.77) was located on BTA 9 with the corresponding ID, rs110877106, being in an intron of the *FILIP1* gene. BTA 15 (Figure 3.14) had the most significant SNPs (six) across all the autosomes, with three SNPs (rs41614029, rs135994818, rs134654371) all within the intron of a different gene (*LOC101902845*, *CEP164*, *DSCAML1*). The two SNPs, rs134510261 (p-value = $3.67x10^{-8}$, *f*(FA) = 0.69) and rs41576433 (p-value = $6.17x10^{-9}$, *f*(FA) = 0.95) are in introns within the same gene, *LRRC4C*.

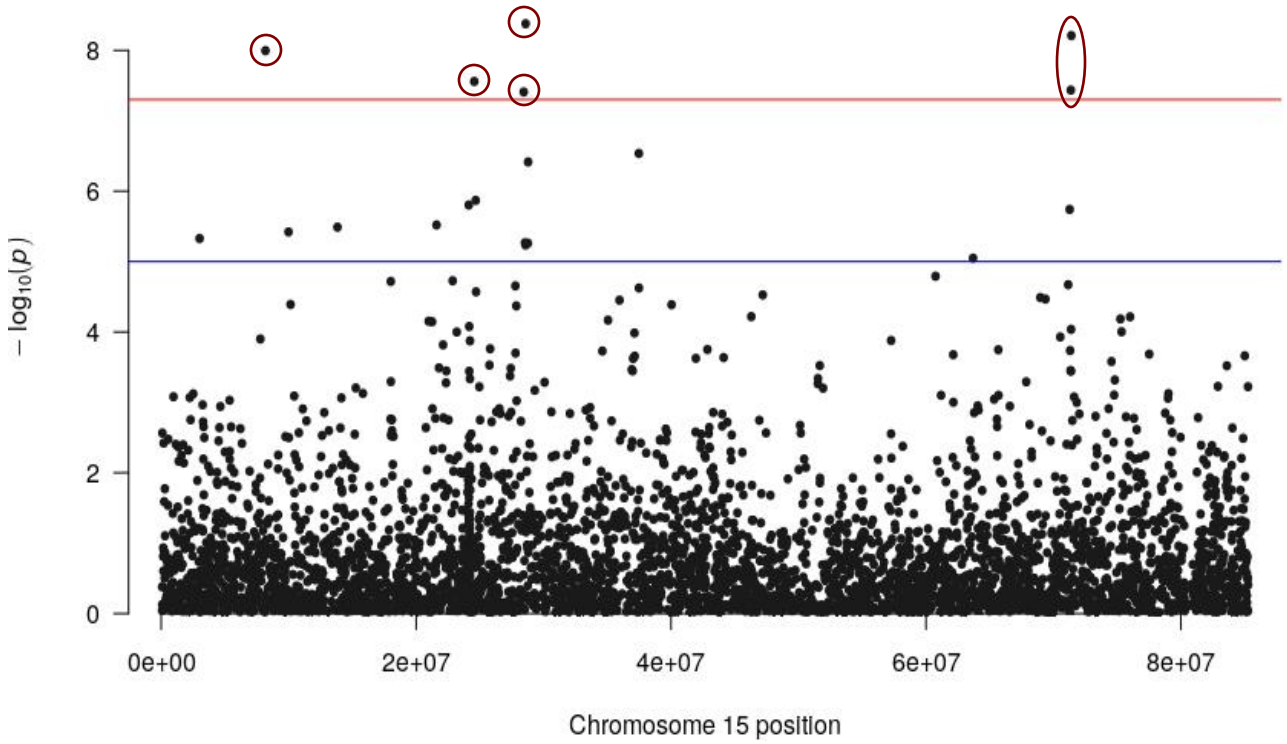


Figure 3.14 Manhattan plot of BTA 15 for weaning weight maternal (WW_{MAT}, 2 730 animals) indicating SNPs with significant p-values above the red line (≤ $1x10^{-8}$)

rs110633222 (p-value = $1.97x10^{-8}$, *f*(FA) = 0.67) on BTA 14 was based in the protein coding region of the *PKHD1L1* gene and is associated with the synonymous mutation from the codon GTG to GTA, which both encode for the amino acid (AA) valine. All three SNPs on BTA 19 (p-value < 2.49×10^{-9}) were in gene regions (*LOC528282*, *RBM47*, *RTN4RL1*) with a range of 0.29 - 0.46 for the minor allele frequency (MAF) within this population. BTA 9, 22 and 29 have two SNPs each located within six different gene regions (9: *FILIP1*, *UBE2J1*; 22: *NUP210*, *MGLL*; 29: *NLM*, *MACROD1*). The remaining five SNPs on BTA 2, 5, 6, 8, and 18 were all in intron regions of their respective genes. Tables 3.5a and b summarizes the relevant information for all the significant SNPs associated with WW_{MAT}.

Table 3.5a Chromosome (BTA), position, favourable allele (FA), frequency (*f*) of the favourable allele, the rs ID, annotation of SNP and possible genes of the 44 single nucleotide polymorphisms associated with weaning weight maternal (WW_{MAT})

BTA	Position	SNP	p-value	FA	*f*(FA)	rs ID	Annotation	Gene
1	2023687	BovineHD0100000619	2.8x10-9	A	0.85	rs109876482	intergenic	-
1	61009968	BovineHD0100017256	1.53x10-8	B	0.68	rs132713619	intergenic	-
2	**60434373**	**BovineHD0200017259**	**8.76x10-9**	**B**	**0.77**	**rs43057522**	**intron**	**THSD7B**
2	105666691	ARS-BFGL-NGS-100821	1.7x10-9	A	0.7	rs110274393	intergenic	-
3	64097592	BTB-00131847	8.30x10-10	A	0.55	rs43344079	intergenic	-
4	78976354	ARS-BFGL-NGS-11194	2.75x10-9	B	0.96	rs110432448	intergenic	-
5	**44228389**	**BovineHD0500012658**	**4.13x10-8**	**B**	**0.76**	**rs43439350**	**intron**	**FRS2**
5	61848831	BovineHD0500017333	2.27x10-8	B	0.8	rs41624983	intergenic	-
6	**14179168**	**ARS-BFGL-NGS-113393**	**6.87x10-12**	**A**	**0.85**	**rs43448385**	**intron**	**ZGRF1**
6	77611218	BovineHD0600021534	4.62x10-8	B	0.57	rs109874421	intergenic	-
7	23194080	BovineHD0700006354	1.44x10-8	B	0.93	rs132854838	intergenic	-
8	17685712	BovineHD0800005543	2.68x10-8	A	0.82	rs133608204	intergenic	-
8	**70005148**	**BovineHD0800021069**	**4.31x10-9**	**A**	**0.65**	**rs136017680**	**intron**	**BMP1**
8	88738557	BTB-00365129	2.23x10-9	A	0.69	rs43566673	intergenic	-
8	89061216	BTB-01579150	9.18x10-9	A	0.77	rs42693255	intergenic	-
9	**15312685**	**Hapmap27634-BTA-158717**	**2.11x10-18**	**B**	**0.77**	**rs110877106**	**intron**	**FILIP1**
9	**61759178**	**BovineHD0900016976**	**2.08x10-8**	**A**	**0.79**	**rs110019124**	**intron**	**UBE2J1**
11	21887840	Hapmap38359-BTA-87537	1.53x10-8	B	0.87	rs41660929	intergenic	-
11	30568083	BovineHD1100009120	8.55x10-9	B	0.93	rs134116765	intergenic	-
12	8617041	BovineHD1200002389	2.15x10-10	B	0.87	rs109564431	intergenic	-
12	32395753	BTA-19643-no-rs	4.35x10-10	A	0.97	rs41577829	intergenic	-
14	25812326	BovineHD1400007459	3.12x10-9	B	0.93	rs43770972	intergenic	-
14	40316657	BovineHD1400011561	1.64x10-8	B	0.85	rs134784743	intergenic	-
14	**57178588**	**ARS-BFGL-NGS-67086**	**1.97x10-8**	**B**	**0.67**	**rs110633222**	**synonymous**	**PKHD1L1**
15	**8180827**	**Hapmap46909-BTA-103259**	**1.01x10-8**	**A**	**0.69**	**rs41614029**	**intron**	**LOC101902845**
15	24563671	ARS-BFGL-NGS-106449	2.77x10-8	A	0.57	rs110637595	intergenic	-
15	**28442203**	**BovineHD1500007623**	**3.9x10-8**	**A**	**0.82**	**rs135994818**	**intron**	**CEP164**
15	**28590540**	**BovineHD1500007663**	**4.19x10-9**	**A**	**0.91**	**rs134654371**	**intron**	**DSCAML1**
15	**71363838**	**BovineHD1500020594**	**3.67x10-8**	**B**	**0.69**	**rs134510261**	**intron**	**LRRC4C**
15	**71406547**	**Hapmap47274-BTA-37477**	**6.17x10-9**	**B**	**0.95**	**rs41576433**	**intron**	**LRRC4C**
18	**5441355**	**BovineHD1800001736**	**9.95x10-9**	**B**	**0.82**	**rs110414980**	**intron**	**WWOX**
18	63896972	BovineHD1800018524	2.32x10-8	A	0.8	rs137006362	intergenic	-

*SNPs highlighted bold were found in gene regions

Table 3.5b Chromosome (BTA), position, favourable allele (FA), frequency (*f*) of the favourable allele, the rs ID, annotation of SNP and possible genes of the 44 single nucleotide polymorphisms associated with weaning weight maternal (WW_{MAT})

BTA	Position	SNP	p-value	FA	*f*(FA)	rs ID	Annotation	Gene
19	**6632246**	**BovineHD1900001710**	**2.50x10-9**	**B**	**0.71**	**rs132903182**	**intron**	**LOC528282**
19	**16779459**	**ARS-BFGL-NGS-6298**	**2.96x10-8**	**A**	**0.54**	**rs109567043**	**intron**	**RBM47**
19	**23605621**	**ARS-BFGL-NGS-106886**	**3.46x10-8**	**B**	**0.68**	**rs41904179**	**intron**	**RTN4RL1**
20	60255461	BovineHD2000016910	4.11x10-11	B	0.78	rs134749828	intergenic	-
21	37181339	BovineHD2100010919	3.21x10-8	A	0.68	rs134033628	intergenic	-
22	**59201628**	**BovineHD2200017188**	**4.03x10-8**	**B**	**0.57**	**rs110831540**	**intron**	**NUP210**
22	**60421134**	**BovineHD2200017578**	**3.43x10-8**	**B**	**0.86**	**rs109624265**	**intron**	**MGLL**
27	30798670	BovineHD2700008558	5.83x10-9	B	0.77	rs110130281	intergenic	-
28	18888651	BovineHD2800005152	2.73x10-8	A	0.5	rs109305830	intergenic	-
28	30741301	BovineHD2800008123	2.20x10-10	B	0.72	rs136982395	intergenic	-
29	**35232674**	**ARS-BFGL-NGS-85977**	**1.51x10-9**	**B**	**0.65**	**rs109693172**	**intron**	**NTM**
29	**42903278**	**BovineHD2900015547**	**7.95x10-9**	**B**	**0.95**	**rs134545156**	**intron**	**MACROD1**

*SNPs highlighted bold were found in gene regions

3.4 Gene Ontology

The protein identification (ID), molecular functions and biological processes for the respective genes were described using Panther (Mi *et al.*, 2017). Tables 3.6a, 3.6b, 3.7 and 3.8a to c summarize this information for the three traits that were found to be associated with genes, these being age at first calving, scrotal circumference and maternal weaning weight.

A total of eight genes were associated with AFC and Tables 3.6a and b show the functions and processes of these genes.

Table 3.6a Summary of relevant function of genes identified for age at first calving

Gene	Protein ID	Molecular Function	Biological Process
GRM8	A0A3Q1JN2	• Glutamate receptor activity; • Amino acid binding; • G protein-coupled receptor activity; • Adenylate cyclase activity;	• Glutamate receptor signalling pathway; • Synaptic transmission, glutamatergic; • G protein-coupled receptor signalling pathway; • Regulation of synaptic transmission, glutamergic;
RDH16	Q0V8D0	• Oxidoreductase activity;	
OVOS2	F1MB90	• Endopeptidase inhibitor activity;	
A2M	Q7SIH1	• Endopeptidase inhibitor activity; • Protease binding; • Serine-type endopeptidase inhibitor activity;	• Protease binding;
AKT3	F1MYJ4	• ATP binding; • Protein serine/threonine kinase activity;	• Brain morphogenesis; • Homeostasis of number cells within a tissue; • Intracellular signal transduction; • Peptidyl-serine phosphorylation; • Positive regulation of angiogenesis; • Positive regulation of artery morphogenesis; • Positive regulation of cell size; • Positive regulation of TOR signaling;

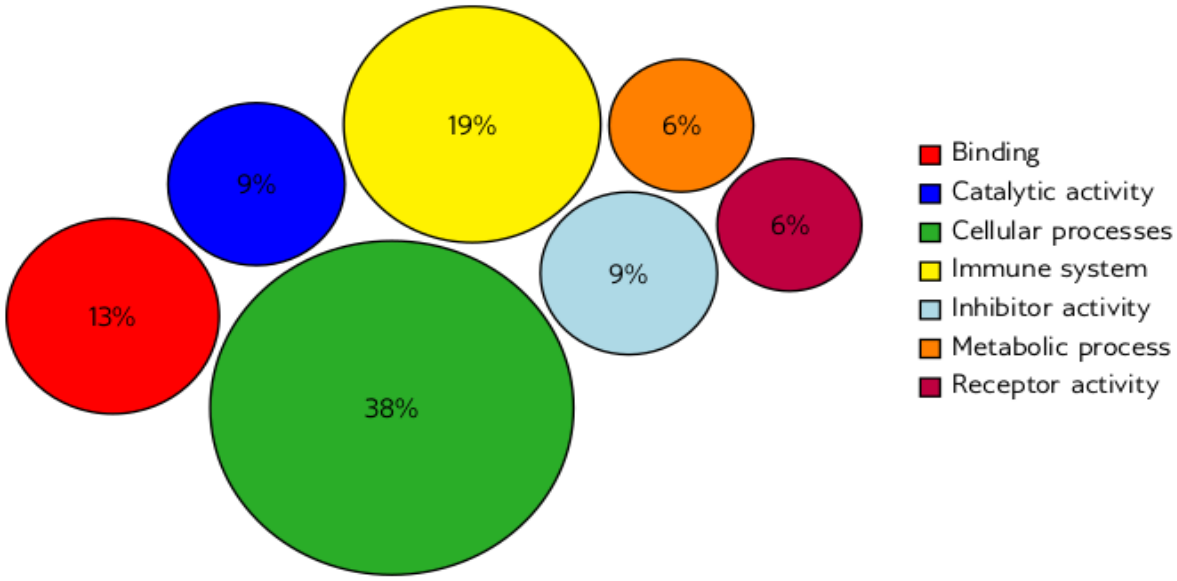
Table 3.6b Summary of relevant function of genes identified for age at first calving

Gene	Protein ID	Molecular Function	Biological Process
UBAC2	A0A3Q1LNA4		• Negative regulation of canonical Wnt signaling pathway • Negative regulation of retrograde protein transport, ER to cytosol • Protein localization to endoplasmic reticulum
FER	E1BNE0	• Regulation of exocytosis • Leukocyte mediated immunity • Lysosome localisation • Regulation of leukocyte activation • Positive regulation of immune system process • Chemotaxis • Regulation of cell population proliferation • Regulation of immune response • Intracellular transport • Cell differentiation • Cell population proliferation • Vesicle fusion to plasma membrane • Cell adhesion • Establishment of organelle localization • Regulated exocytosis • Peptidyl-tyrosine phosphorylation • Transmembrane receptor protein tyrosine kinase signalling pathway • Regulation of cellular localization • Leukocyte activation involved in immune response	• ATP binding • Epidermal growth factor receptor binding • Lipid binding • Non-membrane spanning protein tyrosine kinase activity • Protein phosphate 1 binding • Signaling receptor binding
KIF1B	A0A3Q1MPX6	• ATP binding • Microtubule binding • Microtubule motor activity	• Microtubule-based movement • Vesicle-mediated transport • Cytoskeleton-dependent intracellular transport

In Figure 3.15(a), the molecular functions listed in Tables 3.6a and b were grouped into binding (13%), catalytic activity (9%), immune system process (19%), inhibitor activity (9%), metabolic process (6%), receptor activity (6%).

Figure 3.15(b) illustrates the distribution of biological processes. This amounts to a total of 25 functions which are grouped five different ways, these being binding (24%), catalytic activity (8%), embryonic development (4%), receptor pathways (20%) and cellular processes (44%).

(a)



(b)

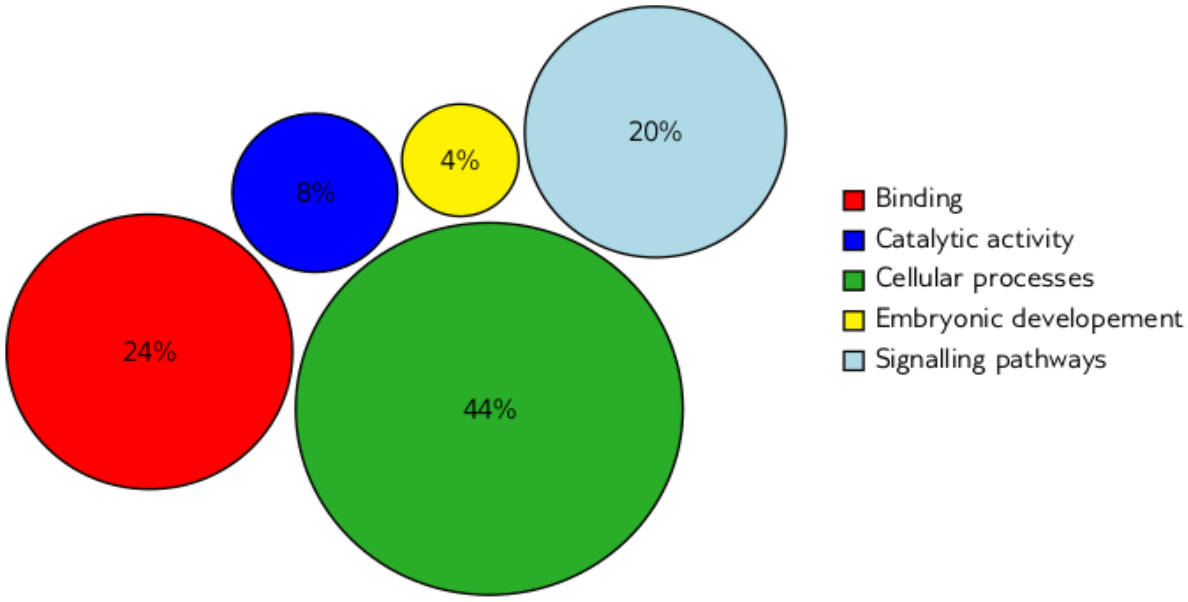


Figure 3.15 Distribution of the (a) molecular functions and (b) biological processes for age at first calving

The cellular grouping was again the largest and mainly consisted of positive and negative regulation of cellular component organization and metabolism.

A total of three genes (*PPP3CA*, *TRPM6* and *WDPCP*) were identified for significant SNPs associated with scrotal circumference. Table 3.7 summarizes the molecular functions and biological processes of the respective genes identified for SC.

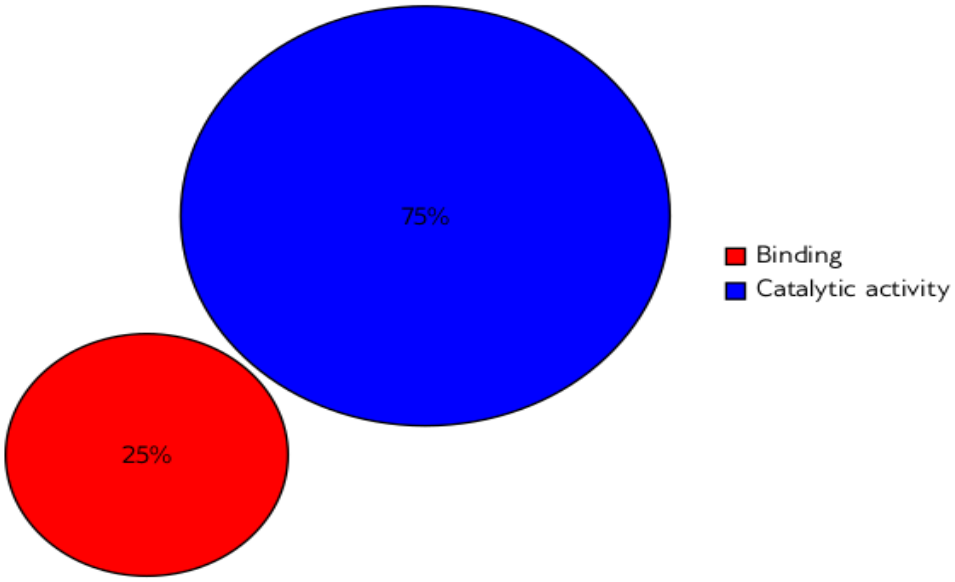
Table 3.7 Summary of relevant function of genes identified for scrotal circumference

Gene	Protein ID	Molecular Function	Biological Process
PPP3CA	P48452	• Protein serine/threonine phosphatase activity • Calmodulin binding	• Inositol phosphate-mediated signaling • Calcium-mediated signaling
TRPM6	F1MWK1	• Cation channel activity • Protein serine/threonine kinase activity	
WDPCP	E1B8R8		• Auditory receptor cell morphogenesis • Cardiovascular system development • Cilium assembly • Digestive system development • Embryonic digit morphogenesis • Establishment of protein localization • Glomerular visceral epithelial cell migration • Kidney development • Regulation of establishment of cell polarity • Regulation of fibroblast migration • Regulation of focal adhesion assembly • Regulation of ruffle assembly • Respiratory system development • Roof of mouth development • Smoothened signaling pathway

The gene WDPCP, WD repeat containing planar cell polarity effector, was only found to have a total of fifteen biological processes but no molecular functions. TRPM6 was only linked with two molecular functions. PPP3CA was related to both molecular and biological roles.

Figure 3.16(a and b) depicts the frequencies of the biological and molecular roles linked to SC and listed in Table 3.7. A total of four molecular processes were identified with one linked to binding (25%) while the remaining three were involved with catalytic activity (75%). Seventeen biological processes were allocated into five groupings for SC. These are cellular processes (35%), embryonic development (41%), localization (6%) and receptor signaling pathways (18%).

(a)



(b)

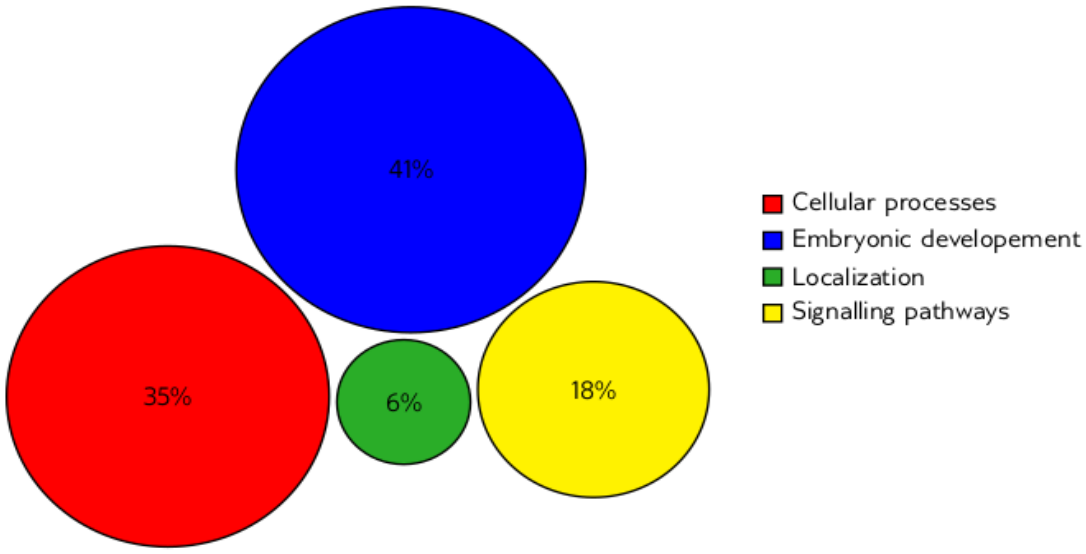


Figure 3.16 Percentage frequencies of the (a) molecular functions and (b) biological processes for scrotal circumference

Nineteen genes were identified for significant SNPs associated with WW_{MAT}. Tables 3.8a, b and c below, outline all the processes and functions identified for the nineteen genes associated with WW_{MAT}.

Table 3.8a Summary of relevant function of genes identified for weaning weight maternal

Gene	Protein ID	Molecular Function	Biological Process
THSD7B	F1MF39		• Actin cytoskeleton reorganization;
CEP164			
FRS2	E1B7I9	• Fibroblast growth factor receptor binding; • Neurotrophin TRKA receptor binding;	• Activation of MAPK activity; • Anterior/posterior axis specification, embryo; • Forebrain development; • Gastrulation with mouth forming second; • Lens fiber development; • Negative regulation of cardiac muscle cell differentiation; • Neuroblast proliferation; • Optic placode formation involved in camera-type eye formation; • Organ induction; • Prostate epithelial cord arborization involved in prostrate glandular acinus morphogenesis; • Regulation of apoptic process; • Regulation of epithelial cell proliferation; • Regulation of ERK1 and ERK2 cascade; • Ventrical septum development;
ZGRF1	E1BPG1	• Zinc ion binding;	• Double-strand break repair;
LOC101902845			
FILIP1			

Table 3.8b Summary of relevant function of genes identified for weaning weight maternal

Gene	Protein ID	Molecular Function	Biological Process
BMP1	E1BEV7	• Calcium ion bonding; • Identical protein binding; • Metalloendopeptidase activity; • Zinc ion binding;	• Positive regulation of cartilage development;
UBE2J1	F1MBQ2	• Ubiquitin-protein transferase activity; • Ubiquitin-protein ligase binding; • Ubiquitin conjugating enzyme activity;	• Negative regulation of retrograde protein transport, ER to cytosol; • Protein N-linked glycosylation vias asparagine; • Regulation of tumor necrosis factor biosynthetic process; • Spermatid development; • Ubiquitin-dependent ERAD pathway;
PKHD1L1	A0A3Q1MGJ7	• Transmembrane signaling receptor activity;	• Axon guidance; • Cell adhesion; • Regulation of cell migration; • Regulation of cell migration; • Positive regulation of axonogenesis; • Cell migration; • Regulation of cell shape; • Regulation of GTPase activity; • Cell surface receptor signaling pathway; • Negative regulation of cell adhesion;
LOC528282			
NUP210	F1MPW7		
DSCAML1	F1N4K6	• Cell-cell adhesion mediator activity;	• Axon guidance; • Central nervous system development; • Dendrite self-avoidance; • Homophilic cell adhesion via plasma membrane adhesion molecules;

Table 3.8c Summary of relevant function of genes identified for weaning weight maternal

Gene	Protein ID	Molecular Function	Biological Process
LRRC4C	F1MXH5	• Cell adhesion molecule binding; • Cell-cell adhesion mediator activity;	• Modulation of chemical synaptic transmission; • Regulation of axonogenesis; • Synaptic membrane adhesion;
WWOX	A0A3Q1LVN4	• Enzyme binding; • Transcription coactivator activity;	• Cellular response to transforming growth factor beta stimulus; • Extrinsic apoptic signaling pathway; • Intrinsic apoptic signaling pathway by p53 class mediator; • Negative regulation of Wnt signaling pathway; • Osteoblast differentiation; • Positive regulation of extrinsic signaling pathway in absence of ligand; • Positive regulation transcription by RNA polymerase II; • Skeletal system morphogenesis;
RBM47	F6RYN2	• mRNA binding; • RNA binding;	• Cytidine to uridine editing; • Hematopoietic progenitor cell differentiation;
RTN4RL1	E1BEI7	• Chondroitin sulfate binding; • Heparin binding; • Roundabout binding; • Signaling receptor activity;	• Axon guidance; • Corpus callosum development; • Negative chemotaxis; • Negative regulation of axon regeneration;
MGLL	A0A3Q1LUK0	• Lipase activity;	
NTM	Q58DA5		• Cell adhesion;
MACROD1	Q2KHU5	• ADP-ribosylglutamate hydrolase activity; • Hydrolase activity, acting on glycosyl bonds;	• Cellular response to DNA damage stimulus; • Peptidyl-glutamate ADP-deribosylation; • Protein de-ADP-ribosylation;

Four genes identified (*LOC101902845*, *CEP164*, *NUP210* and *FILIP1*) have no record of a molecular or biological role within the *Bos taurus* genome on Panther (Mi *et al.*, 2017). Two genes, *NTM* and *THSD7B*, were found to only be associated with one biological process. MGLL was only linked to a single molecular function. The remaining twelve genes were all designated as having both molecular functions as well as biological processes.

Figure 3.17a illustrates the variation in the types of molecular functions, listed in Tables 3.8a, b and c, that are abundantly associated with the genes identified for WW_{MAT}. Receptor activity and metabolic processes had the smallest percentage (4%), followed by catalytic processes and lastly binding activity (55%).

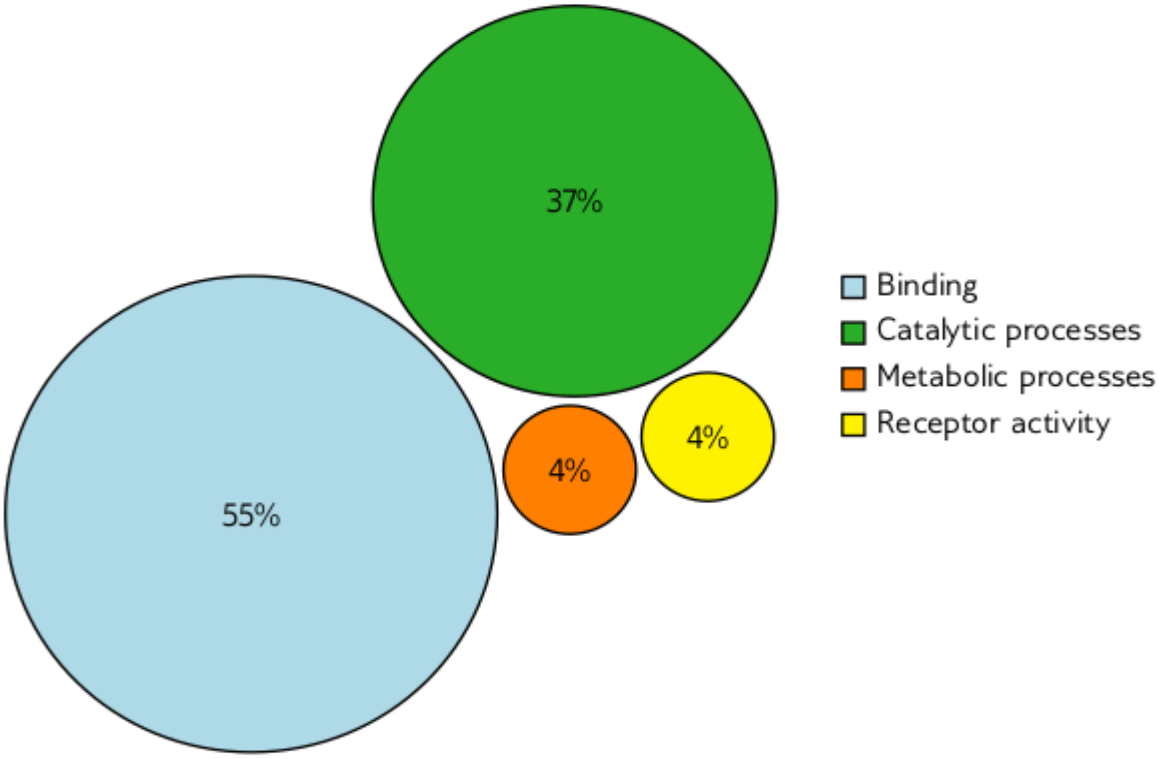


Figure 3.17a Pie chart representation of the respective molecular functions for weaning weight maternal

A total of fifty-four biological processes, Figure 3.17b, were connected to genes, listed in Tables 3.8a, b and c, significantly associated with WW_{MAT}. These consist of six small groups and two larger groups. The largest groups, cellular (41%) and embryonic development (31%)

accounted for 72% of the biological roles. The next largest was receptor signaling pathways (11%) with remaining functions having a frequency of less than 7%.

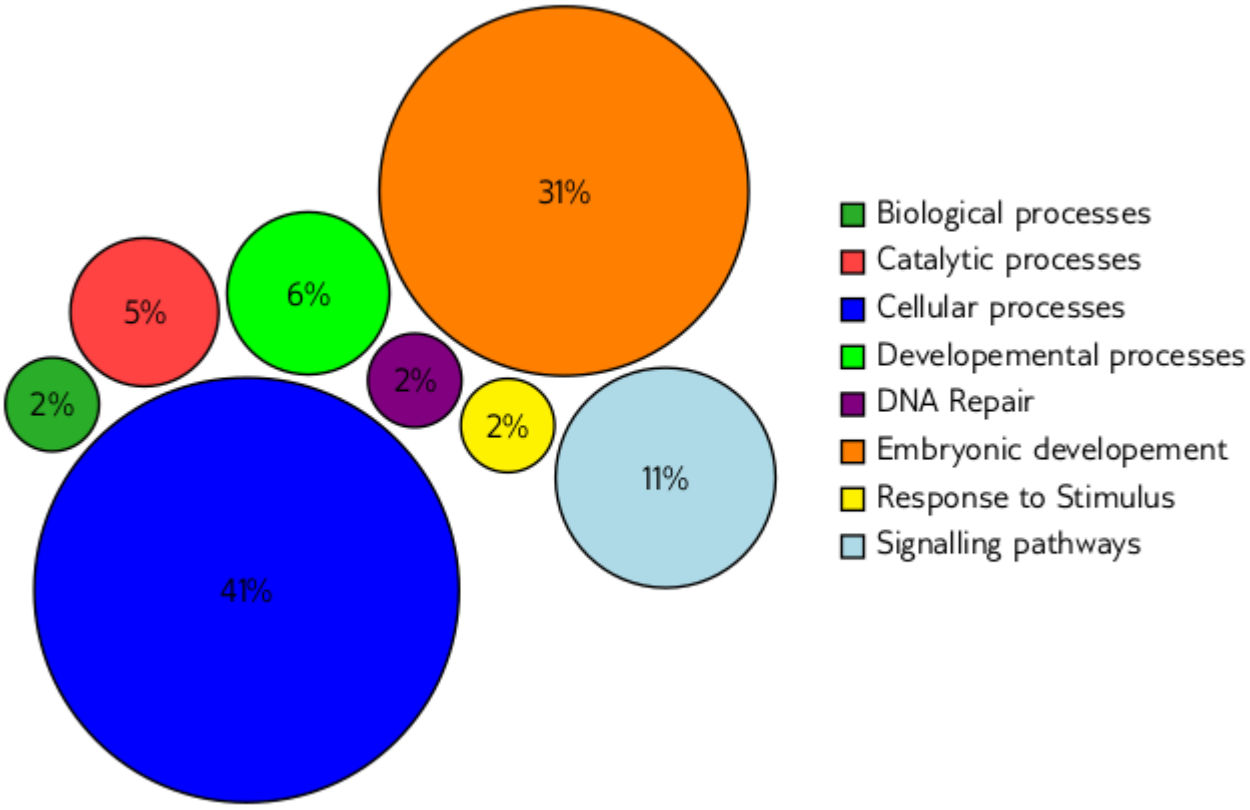


Figure 3.17b Pie chart representation of the respective biological processes for weaning weight maternal

Chapter 4 Discussion

4.1 Introduction

In this study, the core aim was to perform a genome wide association study (GWAS) on fertility and maternal traits in SA Bonsmara cattle. A GWAS requires evaluating phenotypic data in conjunction with molecular data in order to detect any significant coupling between traits and possible underlying genetic mechanisms that may be associated with phenotypic variation (Stranger *et al.*, 2011). Observable traits of economic importance are known to be affected by multiple genes, which contribute to the polygenic expression of these traits. The application of genomics has become a viable tool in the investigation of these polygenic traits. The true underlying genetic architecture of most traits remains unknown (Dekkers & Hospital, 2002; Stranger *et al.*, 2011) with a multitude of studies over the last two decades improving our genetic understanding of these complex traits (Sharma *et al.*, 2015; Mateescu, 2020). Preliminary genomic studies among local indigenous breeds (Makina *et al.*, 2014, 2015, 2016; Sanarana *et al.*, 2016; Lashmar *et al.*, 2018a) have indicated that there are potential genetic forces that may require further investigation.

The female traits of interest, these being age at first calving (AFC), inter-calving period (ICP) and weaning weight maternal (WW_{MAT}) have low heritabilities (0.03 - 0.22), with scrotal circumference (SC, heritability = 0.37) being the sole male trait of interest. SC is measured once in a Bonsmara bulls' life, around 270 days of age, and is included in this study due to its high phenotypic correlation with AFC. Phenotypic progress of these traits in most cattle populations has been slow due to these low heritabilities (Meyer *et al.*, 1990; Cammack *et al.*, 2009; Hawken *et al.*, 2012). The cow/calf operation on a beef cattle farm incurs the largest proportion of profits and costs within the herd. The times a cow is not pregnant or has a suckling calf are known as periods of unproductivity and can occur due to difficult births, poor body condition during pregnancy and/or the feeding of the calf. These incidences may lead to a cow not being able to conceive within the next breeding season and can negatively influence herd productivity. The estimated calving percentage of South Africa's beef commercial sector is approximately 62% (Grobler *et al.*, 2014). Reproductive wastage is of major economic importance and has a significant effect on efficiency from conception to weaning. An understanding of the genetic basis of fertility traits is required to implement selection programs that may increase reproductive efficiency (Miar *et al.*, 2015).

Avoiding periods of unproductivity within a cattle's life would drastically reduce maintenance costs and increase overall profit. The implementation of genetic analysis of beef cattle is still gaining traction in SA among breeds with small numbers while breeds with larger population sizes have been keenly embracing genomic technology (Bosman *et al.*, 2017). This has generated large data sets enabling potential for association analyses, especially within the SA Bonsmara breed which has a sufficiently genotyped population that allowed for the implementation of genome-enhanced estimated breeding values (GEBVs; van der Westhuizen *et al.*, 2017) to conduct these studies. Visser *et al.*, (2020) states that modern genetic improvement in livestock species is inter-dependant on the accurate phenotypic recording of traits and the availability of genomic data of suitably measured animals. The SA Bonsmara breed society has implemented strict guidelines in that the recording of specific traits, i.e. weaning weight, is mandatory for the registration of animals within this population. As a result of this, a large proportion of this population have phenotypes for the traits of interest in this study. A recent study on genetic diversity among nine South African cattle breeds using micro-satellite markers (van der Westhuizen *et al.*, 2020) indicated that the SA Bonsmara breed had the highest level of heterozygosity (H_z = 0.741). The high level of genetic variability within this composite breed, coupled with the strict selection of animals with high reliabilities being included in the genomic population (Bosman *et al.*, 2017) and accurate phenotypes, makes the SA Bonsmara breed highly suitable for GWAS protocol.

4.2 Genome wide association analysis

In order to investigate fertility and maternal traits at a genomic level, an association analysis between the genotypes and the respectively weighted dEBVs was performed. The amount of available genomic technology and progress made in mixed model analysis software allows for a multitude of ways in which these underlying genetic mechanisms can be studied and identified (Fernando & Garrick, 2008; Meyer & Tier, 2012; Lidauer *et al.*, 2017a; Aguilar *et al.*, 2018). The number of animals analysed for each trait differs depending on the EBVs and reliabilities available as well as a few filter parameters. This study utilized single SNP regression association analysis using WOMBAT (Meyer, 2007) as discussed in Chapter 2.

Inter-calving period (ICP)

The association analysis of inter-calving period yielded only four SNPs with a significant association, on BTA 4, 11, 17 and 19, with none of these being located near or within a gene region. The analysis of this trait required the blending of ICP 1, 2 and 3 on an equal weighting. As

the trait with the lowest heritability, the lack of linked genes may be attributed to the fact that management of this trait in a prevailing environment is more important than any gene that affects the phenotypic expression of ICP. The lack of previous GWA studies on this trait may point to the limitations encountered when analyzing this trait.

A study on gestation length using whole genome sequences (WGS) of beef cattle identified a few QTLs that overlapped those previously reported in literature (Purfield *et al.*, 2019b). As gestation length is a component of inter-calving period, these findings could have been inferred but as no significant SNPs were in any QTL region for ICP in this study, thus no comparison was possible.

Weaning weight maternal (WW_{MAT})

A limited number of studies have investigated the potential genetic effects of weaning weight maternal. Saatchi *et al.*, (2014) identified significant SNPs on BTA 1, 2, 3, 4, 6, 9, 11, 12, 14, 15, 17, 18, 19, 21, 22, 24, 25 and 29 for weaning weight maternal across ten different crossbred US beef cattle. Significant SNPs on BTA 7, 8, 9, 20 and 28 were linked to the maternal effect on weaning weight in Brahman cattle (Martínez *et al.*, 2017), while Hay & Roberts, (2019) identified a region on BTA24 that contributed significantly to the genetic variance of maternal weaning weight. This study had similar findings with significant SNPs identified on BTA 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 18, 19, 20, 21, 22, 28 and 29. The remaining SNPs located on BTA 5 and 27 were not previously reported to be associated with the maternal effect on weaning weight. The multitude of autosomes highlights the polygenic nature of this trait. The moderate to high frequency of most of the SNPs with associated genes observed in this study indicates that some form of indirect selection has been occurring at a genetic level.

Weaning weight maternal had the highest number of associated genes relating to important molecular and biological roles. Four different classes of molecular functions were identified with 55% of them being binding related. These binding processes are at both a cellular and DNA level. Catalytic processes (37%) are second numerous followed by metabolic processes and receptor activity both on 4%.

Cellular processes (41%) were the most numerous amongst the biological processes, followed by embryonic development (31%) and then signaling pathways (11%). The variety in biological classes for WW_{MAT} could be indicative of the polygenic expression of this trait. Other processes such as developmental processes, DNA repair and responses to stimuli were represented but in small numbers.

The genes that will now be discussed are grouped according to either binding related functions (*BMP1*, *FRS2*, *LRRC4C*, *MACROD1, RTN4RL1*, *UBE2J1*, *WWOX* and *ZGRF1*) or cellular processes (*DSCAML1*, *MGLL*, *NTM*, *PKHD1L1*, *RBM47* and *THSD7B*). A few of these genes, such as *LRRC4C*, could be linked to either grouping but for discussion purposes they were grouped according to what was identified as relevant to this study according to literature. Genes with unknown functions and processes (*CEP164*, *FILIP1*, *LOC101902845* and *NUP210*) are briefly discussed at the end of this section.

THSD7B has been linked to actin cytoskeleton reorganization and is classed as a cell adhesion molecule, but its true biological role remains unknown (Wang *et al.*, 2011). Fernández *et al.*, (2019) recently associated this gene when analysing the underlying genetic mechanism that affect AFC in a Colombian composite beef breed. A study investigating SNPs associated with phenotypic variation in serum IGF-1 concentration (Gobikrushanth *et al.*, 2018) coupled with reproductive performance in dairy heifers linked the *THSD7B* gene. IGF-1 plays a key role in the control of postnatal growth, mammary gland development, lactation, and fertility. This is an interesting find as each of these physiological processes has a marked effect on WW_{MAT}. This component trait is comprised of a cow's ability to provide enough milk to the growing calf which is dependent on pre-pubertal growth (Sejrsen *et al.*, 1982; Sejrsen & Purup, 1997) in conjunction with correct nutrition during this developmental stage.

LRRC4C is a gene that is involved in mediating cell adhesion molecule binding and is involved in the regulative function of neural progenitor cell differentiation (Zhang *et al.*, 2013b). A study on transcriptome signatures in beef cattle associated this gene with bovine uterine receptivity during early gestation (Binelli *et al.*, 2015). This affects subsequent development of the foetus during gestation. A GWAS in crossbred Holstein cattle (Saowaphak *et al.*, 2017), associated this gene with 305-days milk yield. This indicates that development of thalamocortical axons (TCA) and subsequent axonal signalling pathways may influence the production of milk through hormonal cascades.

The E2 Ubiquitin-conjugating enzyme (*UBE2J1*) has transferase and ligase binding activity. This enzyme is localised in the endoplasmic reticulum (ER), is a known substrate of MAPK signalling pathway and is vital for transient ER stress cell recovery (Elangovan *et al.*, 2017). No GWA study on cattle identified this gene but one on pigs (Do *et al.*, 2014) associated it with residual feed intake as well as backfat and weight gain.

PKHD1L1 is mainly involved in cellular biological processes and has transmembrane signalling receptor activity. Erdman *et al.*, (2017) characterised this gene in a human population

and the results indicate people that are more likely to reach a desirable age of longevity are those whose genotype is upregulated for this gene. A study in mice (Wu *et al.*, 2019) identified *PKHD1L1* as a component of the surface coat of hair cells in the cochlea and down regulation of this gene leads to progressive hearing loss.

WWOX is involved in enzymatic binding and transcription cellular activities. It is involved in the negative regulation of Wnt and apoptic signalling pathways. *WWOX* has shown capabilities of interacting with androgen and oestrogen, and it has been suggested that WWOX may act as an alternative receptor for these sex steroid hormones (Chang *et al.*, 2005; Su *et al.*, 2012). No studies in cattle breeds have associated this gene with fertility traits. Recent literature has demonstrated that *WWOX* is a candidate tumour suppressor gene in human breast tumours (Abdeen *et al.*, 2011). A preliminary study in mice (Ludes-Meyers *et al.*, 2007), where the *WWOX* gene was knocked-out, indicated reduced fertility in affected males through atrophic development of seminiferous tubules. Female knockout mice showed reduced lifespans and increased B-cell lymphomas in mammary tissue which could subsequently affect milk production.

FRS2 has been reported to be linked to receptor binding as well as activity on a variety of organs during embryonic development. It is referred to as the fibroblast growth factor receptor substrate 3 but has not been identified in beef cattle according to available literature. Biochemical studies (Hoch & Soriano, 2006; Zhang *et al.*, 2013a) have discussed how fibroblast growth factor receptor (*FGFR1*) specific substrates, *FRS2* and *FRS3* are the underlying mediators of *FGFR1* signal transduction to the PI3K and MAPK pathways. The FGF receptor family belongs to a large group of protein tyrosine kinases that play an essential role in controlling cell growth, differentiation, survival and many other functions (Ong *et al.*, 2000).

ZGRF1, observed for WW_{MAT} in this study, has a zinc ion binding function and is involved in the process of double-strand DNA repair. No other GWA study on cattle has identified this gene with any traits of interest. Choudhury *et al.*, (2019) states that *ZGRF1* belongs to the family of zinc-finger GRF-type and UPF1-like RNA helicases.

Bone morphogenic protein (*BMP*) 1 is linked to ion channelling as well as metalloendopeptidase activity. It is highly involved in regulating morphogenesis by transforming enzymatic precursors and mature functional extracellular matrix (ECM) proteins and several growth factors including GFD8 and transforming growth factor (TGF-β) family members (Grgurevic *et al.*, 2011). Asharani *et al.*, (2012) tested the functions of attenuated *BMP1* function and identified that it led to compromised osteogenesis and resulted in bone fragility in humans and zebrafish. An investigation in Chinese Holsteins (Li *et al.*, 2016) identified BMP1 as one of ten novel genes

affecting milk protein concentration. This was done using pair-end RNA sequencing of mammary tissue associated with cows that had low or high values for milk protein percentage.

A few of the genes discussed above were also responsible for mineral binding. *BMP1* is involved in the positive regulation of cartilage development and has metalloendopeptidase activity. It also has calcium and zinc ion binding activity, while the Z*GRF1* gene also has zinc ion binding activity.

RNA Binding Motif Protein 47 (*RBM47*) is heavily involved in translational processes in the form of mRNA and RNA binding as well as the conversion of cytidine to uridine. Weikard *et al.*, (2012) analysed tissue-specific mRNA expression patterns in crossbred cows. He concluded that the transcriptional gene expression levels in the mammary gland, skeletal muscle and liver of cows postpartum are modulated due to the genetic and phenotypic background that affects performance of milk ability. An analysis of milk production traits in a Chinese Holstein population linked this gene (Hu *et al.*, 2010). Further studies are required in order to understand this genes' potential role in milk production.

RTN4RL1 is involved in CNS regulation and has signalling receptor activity. It is involved in heparin and chondroitin sulfate binding. Heparin is an anti-coagulant while chondroitin sulphate is involved in the prevention of cartilage degeneration and aid in stabilizing neuronal structures (Dickendesher *et al.*, 2012). No literature was identified to have linked this gene with previous studies of fertility traits in beef cattle. This gene has been included in a prognostic test as a biomarker for breast cancer in humans (Marchionni *et al.*, 2013) and is linked to minimal lymphoma development.

Neurotrimin (*NTM*) is a protein coding gene associated with cellular adhesion through homophilic mechanisms and control the assimilation of various nitrogen sources (Jeter *et al.*, 1984). A GWA study on Holstein cattle (Freebern *et al.*, 2020) associated *NTM* with a displaced abomasum, which would subsequently affect milk production but has been rarely described in suckling beef cattle (Oman *et al.*, 2016). Another study, (De León *et al.*, 2019), associated this gene with AFC in Columbian beef cattle. The result of knocking *NTM* out in mice was a deficit in the learning of emotionally challenging tasks mainly involved with that of active avoidance (Mazitov *et al.*, 2017). These behaviours are linked with natural defensive response to aversive stimuli.

MACROD1 is involved with protein de-ADP-ribosylation and has been linked to hydrolase activity (Agnew *et al.*, 2018). A GWAS using genotyping-by-sequencing methodology identified *MACROD1* as a novel candidate gene that influences cow milk traits (Ibeagha-Awemu *et al.*,

2016), especially fat percentage of bovine milk. Three fatty acids, these being caproic, caprylic and tridecylic, were all linked to *MACROD1* gene region. These medium length carbon saturated fatty acids, or constituents of them, are all found in bovine milk; thus, this gene may influence the concentration of these acids. This gene was also associated with %KPH within the Simmental beef breed (Hieber *et al.*, 2018). %KPH refers to the percentage of total fat, specifically in the regions of the kidneys, pelvis, and heart, of a carcass.

MGLL, or monoglyceride lipase, is associated with the serine hydrolase superfamily, has lipase activity and a possible role in the fat digestion intracellularly (Senior & Isselbacher, 1963). This superfamily is linked to the conversion of monoacylglycerides into glycerol and their associated free fatty acids. These substrates are essential in ATP generation as well as the glycerol for either gluconeogenesis or hepatic lipogenesis processes. High *MGLL* activity is linked to the mobilization of lipid reserves (Viscarra & Ortiz, 2013) and may have a resultant cascading effect on the fat components in milk.

DSCAML1 mediates cell-cell adhesion activity, especially for homophilic cells, and is essential in neurological system development and maintenance (Fuerst *et al.*, 2009). The absence of this gene in previous discoveries in any cattle breed indicates that this may be a novel association within this study. Indel variants of down syndrome cell adhesion molecule 1 have been associated with poor sperm morphology in goats (Wang *et al.*, 2020), with Kosova *et al.*, (2014) also associating this gene in humans.

The large number of genes identified for WW_{MAT} in this study that are linked to the development of the mammary gland in conjunction with the production of milk, milk fat and protein concentrations and mobilization of body lipids is favourable. This once again illustrates our current understanding that lowly heritable traits are influenced by many genes with small polygenic effects. Studies (Buchanan *et al.*, 2003; Rasby *et al.*, 2014) have indicated that cows with the ability to better utilize available body reserves are able to produce amounts of milk suitable to enable a weaner calf to reach its maximum growth potential.

Both genes, *FILIP1* and Nucleoporin gp210 (*NUP210*) lack identified molecular functions and biological processes within the *Bos taurus* genome. According to research on human genes (Militello *et al.*, 2018), *FILIP1* is important during myogenesis and if silenced this inhibits the differentiation of myoblasts into myotubes. *NUP210* expression is critical for neuro-progenitor cell and myoblast differentiation (D'Angelo *et al.*, 2012) and up-regulates the expression of key differentiation genes.

LOC101902845, *LOC528282*, *NUP210* and *CEP164* have no identified biological or molecular role in the Bovidae family.

Age at first calving (AFC)

In this study, twenty significant ($p < 1x10^{-8}$) SNPs were associated with age at first calving and was expected as AFC is known to be a polygenic trait. Most studies on heifer fertility revolve around biological signals, like age at first corpus luteam, pregnant or not pregnant at 18 months of age as well as various ultrasound observations.

Significant SNPs associated with AFC were previously reported on BTA 4, 5, 6, 10, 13, 14, 22 and 23 in literature (Costa *et al.*, 2015; Nascimento *et al.*, 2016; Melo *et al.*, 2018; Fernández *et al.*, 2019) in *Bos indicus* and crossbred beef cattle. The results from this study found significant SNPs on BTA 4 and BTA 5. Additional SNPs that were identified in this study have not been previously reported in literature and may be unique to the SA Bonsmara composite breed or spurious associations that could be deemed false positives.

A total of eight genes were linked to significant SNPs associated with AFC. Cellular processes, with regards to organelle and compartment organization and/or cell biogenesis, occurs as the most common function at 45%. This could be linked to AFC, as it considers the heifers' ability to reach sexual maturity at an early age, become pregnant and maintain pregnancy for the first time. Immune system functions were the next primary functional group followed by binding, especially ATP binding which was associated with three different genes. Inhibitor activity only makes up 9% of identified functions but were all endopeptidase related. For biological processes, cellular processes once again make up most of the allocated processes. This may be related to the ability of the uterine tract to develop and prepare sufficiently for successful conception. Binding activity was the second most numerous process, relating to protein and lipid binding. Signalling pathways are the third most common process and may be linked to the onset of puberty and the cascade of gonadotrophin hormones activated at the onset of puberty.

AKT3 is involved in ATP binding, is essential in protein serine/threonine kinase activity and is a key regulator in the chemical signalling of the PI3K-AKT-mTOR pathway (Lee *et al.*, 2012). This signalling influences many critical cell functions, including the synthesis of new proteins, cell growth, cell proliferation and the longevity of cells (Cohen, 2013). This gene was one of four genes identified to be linked to the autosomal recessive disease known as familial partial lipodystrophy (Garg, 2011). Lipodystrophy is described as an abnormal distribution of fat in the body and can refer to both the abnormal accumulation and/or the loss of fat tissue. Marete *et al.*, (2018) identified

AKT3 as a candidate gene for the development of tissues associated with udder morphology in three French dairy breeds.

Metabotropic glutamate receptor 8 (*GRM8*) has various neurotransmitter receptor and signalling pathways, especially with regards to glutamate activity (Scherer *et al.*, 1997). A comprehensive study using whole-genome analysis associated this gene with body size in Brahman cattle (Chen *et al.*, 2020). It is known that heifers who reach an earlier AFC compared to their similar age counterparts have larger body sizes. This may have been resultant of indirect selection through incidences of dystocia, with larger heifers experiencing a lower rate of problematic births at their first calving. A study in humans linked this gene to endometriosis-related infertility (Galarneau *et al.*, 2018).

RDH16 is associated with oxidoreductase activity (Deng *et al.*, 2010) and has the highest activity in liver cells. No studies related to fertility traits have associated this gene in beef cattle. A recent study on mice identified the function of *RDH16* to be significantly linked as a tumour-suppressing gene, especially in hepatocellular membranes (Zhu *et al.*, 2020).

OVOS2, a gene similar to ovostatin has only one identified biological process in beef cattle, that being endopeptidase inhibitor activity involved in innate immune responses (Jacob *et al.*, 2009). A study performed by Makina *et al.*, (2015) in SA Bonsmara and other local Sanga breeds identified *OVOS2* as a selection signature related to reproductive performance, which AFC is a component of. González-Ruiz *et al.*, (2019) linked this gene to tuberculosis resistance in Mexican Holsteins.

A2M is involved in serine-type endopeptidase inhibitor activity and protease binding. This gene has been linked to up-regulation in bovine host tissues in response to pathogens and resulting inflammation responses (Ferreira *et al.*, 2013). A study in mice where this gene was knocked-out (Umans *et al.*, 1995) indicated that this gene has the highest level of expression around the partum period.

FER is a member of the FPS family of non-transmembrane receptor tyrosine kinases involved in several cellular processes including cell proliferation, differentiation and localization (Fan, 2020). Biological processes include ATP and lipid binding activities. A similar gene, *FER1L6*, belonging to the same family as *FER* was associated with AFC in Nellore cattle (Mota *et al.*, 2017).

Ubiquitin like modifier activating enzyme 2 (*UBAC2*) is involved in protein localization and regulation in the ER and is linked to the Wnt signalling pathway (Park *et al.*, 2012). No study has associated this gene to cattle with any traits of interest. This gene has been linked with Bechet's

disease (BD) in humans (Fei *et al.*, 2009) characterised by blood vessel inflammation throughout the entire body.

KIF1B has been described as having ATP and microtubule binding activities. It belongs to the kinesin family of proteins and is essential for the transport of mitochondria and other substrates intracellularly. This gene has frequently been identified as a strong selective sweep in multiple dairy breeds (Randhawa *et al.*, 2016). This is indicative of selective forces operating on the genetics that control the anatomical structure and physiological function of mammary glands, thus affecting the quantity and quality of milk constituents. Zhao *et al.*, (2001) identified muscle weakness and impaired transport of synaptic vesicle precursors in mice that were heterozygous for the *KIF1B* gene.

Scrotal circumference (SC)

A review of relevant literature indicated that most researchers, with regards to bull fertility, study the biological processes associated with sperm quality, motility and scrotal volume (Fortes *et al.*, 2012b; Lirón *et al.*, 2012; Taylor *et al.*, 2018) at a SNP marker or WGS level. More recent association studies that scrutinize male fertility revolve around sexual precocity, especially in tropical cattle (Fortes *et al.*, 2013; Buzanskas *et al.*, 2017; da Silva Romero *et al.*, 2018; Sweett *et al.*, 2018; de Melo *et al.*, 2020; Stafuzza *et al.*, 2020) located in Central and Southern America. These researchers analyzed the age an animal becomes sexually active. Buzanskas *et al.*, (2017) identified significant SNPs on BTA 5, 14, 20, and 28, with Sweett *et al.*, (2018) obtaining an association with BTA 16. None of these aforementioned results correlated with the results from this study. Four SNPs identified in this study have not been identified in literature and can be attributed to most studies being done on Taurine or Indicine breeds.

Three genes were identified for SC. Catalytic activity (75%), especially with regards to the serine/threonine protein pathway, dominated the list of molecular functions for SC with the remaining 25% being annotated as having a binding function. Metabolic processes were grouped into four sets with embryonic development (41%) and cellular processes (35%) contributing the most. The *WDPCP* gene accounts for 15 of the 17 biological processes, with most being embryological. These processes may indicate that the potential SC of a bull calf may be determined before birth.

WD repeat-containing planar cell polarity effector (*WDPCP*) has been attributed to embryological and developmental processes, as well as regulation of cell biogenesis. *WDPCP* regulates cell alignment required for collective cell movement during embryonic development (Cui

et al., 2013). The signalling pathway governs collective cell movements and proteins in the Wnt/PCP pathway control the morphogenesis of multiciliate epithelial cells during vertebrate embryogenesis (Ma *et al.*, 2017). Afonso *et al.*, (2019) described the *WDPCP* gene referring to its role in encoding a protein that inhibits Wnt activity, with the pathway involved in muscle regeneration.

PPP3CA belongs to the serine/threonine-protein phosphatase catalytic subunit gene family. *PPP3CA* has a calmodulin binding function, in that it helps mediate the transfer of calcium and has highest levels of gene expression during adipogenesis (López-Victorio *et al.*, 2013). This gene has been identified to have a significant effect on sexual precocity in Nellore cattle (Dias *et al.*, 2015) and then was further verified in five other cattle breeds (Dias *et al.*, 2017). A study on selection signatures in African cattle (Taye *et al.*, 2017), associated *PPP3CA* with thermotolerance which is an important trait with regards to maintaining the scrotum for optimal sperm production. SanGiovanni & Lee, (2013) associated this gene with metabolic signalling via MAPK. It has been reported by Miyata *et al.*, (2015) that infertility in male rats can be induced by the blocking of this gene's action.

TRPM6 is involved in protein serine/threonine kinase activity, similar to other genes identified in this study. It is primarily responsible for cation channel activity. Rondón *et al.*, (2008) investigated *TRMP6* expression in the kidney and large intestinal tissues of mice where *TRMP6* was identified as the initial component that becomes directly involved in active kidney and intestinal epithelial Mg^{2+} absorption as well as reabsorption when circulating blood Mg^{2+} levels were low. A review on the link between magnesium homeostasis in cattle and *TRMP6* (Martens *et al.*, 2018), states that the binding of Mg^{2+} with an enzyme or substrate is essential for enzymic reactions.

An interesting connection occurs between AFC and SC in this study, as they share molecular functions and biological processes between the different genes that were associated. Two genes, *AKT3* and *TRPM6*, both have protein serine/threonine kinase activity and another gene, *PPP3CA*, has protein serine/threonine phosphatase activity. Other genes, *OVOS2* and *A2M*, associated with AFC have endopeptidase inhibitor activity as well as serine-type endopeptidase inhibitor activity. Upon studying sexual hormone secretion patterns, one can conclude that the mechanisms that initiate puberty in both sexes are controlled by the same mechanisms.

In conclusion this study demonstrated the potential of uncovering the hidden genetic mechanism that contribute to the expression of fertility traits. This GWAS provided the first insight of these underlying forces in a local indigenous SA cattle breed.

Chapter 5 General conclusion and recommendations

The aim of this study was to conduct a genome-wide association study to investigate the underlying genetics of fertility and maternal traits within the SA Bonsmara breed. The traits of interest for this study were age at first calving (AFC), inter-calving period (ICP), scrotal circumference (SC) and weaning weight maternal (WW_{MAT}). Fertility traits are lowly heritable but are an essential component in selection programs for the genetic improvement of all livestock species. In order to gain a better understanding of the genetic basis of these traits and in turn in improve reproductive efficiency, a GWAS study as described in this project was required. Phenotypic data, given as estimated breeding values of 4 171 animals, in conjunction with a pedigree containing 2 020 248 animals dating back to 1949 was provided by SA Stud Book on behalf of the SA Bonsmara Breeders Society.

An original dataset of 3 291 SA Bonsmara animals, genotyped across three different commercial arrays underwent quality control (QC), an assessment of population stratification through principal component analysis (PCA), amalgamation of the three arrays via imputation to the highest density possible and finally a genome-wide association analysis using single SNP regression using the **--snap** runtime option (Meyer & Tier, 2012) in WOMBAT (Meyer, 2007). Each trait was statistically analysed with 128 793 SNP markers across all Bovine autosomes with 2 620 animals for AFC, 2 245 for ICP, 2 318 for SC and 2 730 for WW_{MAT} respectively. This above fulfilled the first objective set in Chapter 1 of this study.

The second objective was that of a detailed annotation of the biological processes and/or molecular functions associated with associated genes and was done using Panther (Mi *et al.*, 2017). This gene ontology identified no genes associated with ICP, three genes for SC, eight genes for AFC and a total of nineteen genes for WW_{MAT}. Genes could be grouped according into four main categories, these being cellular processes, embryological development processes, receptor binding and signalling pathways as well as catalytic functions.

In conclusion, the findings in this study improve our current understanding of the genetic mechanism affecting fertility and maternal traits. Chromosomes and SNPs that yielded novel associations with previously uncharacterised genes or that have never been previously reported in literature may possibly be unique to the SA Bonsmara beef breed and will require further investigation.

5.1 Challenges and limitations

Use of different models

The development of high throughput sequencing technology (Visscher *et al.*, 2012, 2017) has been coupled with the necessary development of mixed model equation (MME) methods. The appropriate models and software needed to process and identify significant associations through GWAS are still up for debate with multiple options available ranging from Bayesian analysis (Habier *et al.*, 2011), to pre-conjugated linear gradient (Lidauer *et al.*, 1999) and/or regression analysis (Meyer, 2007; Garrick *et al.*, 2009). Prior to model inclusion, the appropriate calculation and weightings of phenotypes and/or estimated breeding values must be considered for the input data be as linearly independent as possible and maximise blocking of dependant variables. Each is to their own and may be relevant for single trait or multi-trait analysis. The inclusion of dominance and epistatic genetic effects still needs to come to the forefront of this field, with some researchers (Bolormaa *et al.*, 2015; Dos Santos *et al.*, 2016) showing that these effects must not be discarded. WOMBAT was utilised as a plethora of recent literature (McGovern *et al.*, 2019; Ring *et al.*, 2019; Twomey *et al.*, 2019; Purfield *et al.*, 2019b) has shown favourable and sensical results.

Preliminary GWA studies utilized available lower-density platforms (Visscher *et al.*, 2012; Sharma *et al.*, 2015) and basic MME software (Meuwissen *et al.*, 2001), with advances made with regards to appropriate statistical approaches (Meyer, 2007; Habier *et al.*, 2011). The lightning speed progress of genomic technology has led to the use of whole-genome sequences (WGS) of thousands of animals across a wide range of livestock species (Bolormaa *et al.*, 2017; Pereira *et al.*, 2017; van Son *et al.*, 2017; Yoshida *et al.*, 2017; Do *et al.*, 2018). The use of WGS in conjunction with GWAS has allowed for the validation of previous studies on the same traits within populations and the comprehensive analyses of the underlying genetic architecture that affect traits of interest.

Chip bias, SA Studbook Chip in conjunction with Illumina

The development of the original bovine reference genome (Zimin *et al.*, 2009), and subsequent development of most commercially available genotyping platforms utilised the genetic variability of four taurine breeds and one indicus breed. This excluded the genetic variation present within local indigenous Sanga and composite breeds (Matukumalli *et al.*, 2009). This has caused ascertainment bias in that some SNPs are either not present or show low levels of segregation when used on *Bos taurus africanus*. This is detrimental as it may not capture the full amount of

variation in the population and leads to lower levels of genetic informativeness. This may lead to bias as the imputation software available (Sargolzaei *et al.*, 2014; Browning & Browning, 2016) relies on the genetic variation present within the reference population on the highest density array. This limitation translates to a scenario where younger animals that may be genotyped on lower density chips, for cost-saving reasons, may have genetic variation that is not present in the reference population. This variation may not be captured and thus is not able to be analysed in a GWAS study. This reiterates the need to consistently update a reference population with animals from newer generations in order to accurately describe the genetic variability present. This points to a need to develop a genotyping array that is more suitable for our Sanga breeds.

5.2 Recommendations

This study was the first GWA study performed in Bonsmara cattle and the first investigation of fertility in any indigenous SA beef breed using GWAS methodology. Although this is the largest breed in SA with the most phenotypic and genotypic data, certain limitations were experienced which could be addressed in future studies.

The number of animals that were excluded from the final analysis for this study, in my opinion, was too high. This was due to the threshold of requiring an ERC value of ≥ 0.5. This is dependent on the reliability the animal had for a specific trait of interest as well as the animals' number of descendants present within the population pedigree (Lidauer *et al.*, 2017b). This mainly affected younger animals or animals that were genotyped but were not used as breeding stock. With time, as these animals proceed to have progeny with accurate phenotypes it will conjunctively raise their reliability estimates and may allow them to reach the threshold for inclusion in future association analyses. A comparison of this analysis with different MME approaches, for instance Bayesian inference through GenSel (Fernando & Garrick, 2008) may indicate the repeatability or may yield different results.

A better methodology with regards to the blending of the first, second and third interval periods that will be more representative and result in a better reflection of the phenotypic variance within the population may need to be identified. This may result in the association of more significant SNPs and/or the identification of SNPs in gene regions. Reviews on fertility traits (Hawken *et al.*, 2012; Fortes *et al.*, 2013) indicated the need to compile a reference for fertility genes identified in other GWAS which is still lacking.

The high amount of noise observed in the Manhattan plot for weaning weight maternal indicates that an inclusion of a fixed effect in the WOMBAT analysis may be necessary. The

studies on bioregion for Bonsmara animals (Nephawe *et al.*, 2004; Webb *et al.*, 2017) indicate the need to account for bioregion in the model for EBV prediction, which SA Stud Book does account for. A further inclusion of this fixed effect may be necessary in the final step of GWAS in order to consolidate the statistical model and minimise false positives and unwanted bias. The use of whole genome sequences (WGS) will allow for the more conclusive identification of causal genes as it will be more accurate compared to the use of the 130 000 SNP markers used in this study.

References

Abdeen, S.K., Salah, Z., Maly, B., Smith, Y., Tufail, R., Abu-Odeh, M., Zanesi, N., Croce, C.M., Nawaz, Z. & Aqeilan, R.I., 2011. Wwox inactivation enhances mammary tumorigenesis. Oncogene 30, 3900–3906 https://doi.org/10.1038/onc.2011.115.

Afonso, J., Fortes, M.R.S., Reverter, A., da Silva Diniz, W.J., Cesar, A.S.M., de Lima, A.O., Petrini, J., de Souza, M.M., Coutinho, L.L., Mourão, G.B., Zerlotini, A., Gromboni, C.F., Nogueira, A.R.A. & de Almeida Regitano, L.C., 2019. Genetic regulators of mineral amount in Nelore cattle muscle predicted by a new co-expression and regulatory impact factor approach. bioRxiv, 804419 https://doi.org/10.1101/804419.

Agnew, T., Munnur, D., Crawford, K., Palazzo, L., Mikoc, A. & Ahel, I., 2018. MacroD1 is a promiscuous ADP-ribosyl hydrolase localized to mitochondria. Front. Microbiol. 9, 20 https://doi.org/10.3389/fmicb.2018.00020.

Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D., Legarra, A. & Misztal, I., 2018. BLUPF90 suite of programs for animal breeding with focus on genomics.in Proceedings of the 11th World Congress on genetics applied to livestock production. Auckland.

Allison, D.B., Fernandez, J.R., Heo, M., Zhu, S., Etzel, C., Beasley, T.M. & Amos, C.I., 2002. Bias in estimates of quantitative-trait-locus effect in genome scans: demonstration of the phenomenon and a method-of-moments procedure for reducing bias. Am. J. Hum. Genet. 70, 575–85.

Anderson, C.A., Pettersson, F.H., Barrett, J.C., Zhuang, J.J., Ragoussis, J., Cardon, L.R. & Morris, A.P., 2008. Evaluating the Effects of Imputation on the Power, Coverage, and Cost Efficiency of Genome-wide SNP Platforms. Am. J. Hum. Genet. 83, 112–119.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T., 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–73.

ARC., 2016. Agricultural Research Council Annual Report 2016/17. Accessed: http://www.arc.agric.za/Documents/Annual%20Reports/ARC%20Annual%20Report%202016-2017.pdf

Ardlie, K.G., Kruglyak, L. & Seielstad, M., 2002. Patterns of Linkage Disequilibrium in the Human Genome. Nat. Rev. Genet. 3, 299–309.

Asharani, P. V., Keupp, K., Semler, O., Wang, W., Li, Y., Thiele, H., Yigit, G., Pohl, E., Becker, J., Frommolt, P., Sonntag, C., Altmüller, J., Zimmermann, K., Greenspan, D.S., Akarsu, N.A., Netzer, C., Schönau, E., Wirth, R., Hammerschmidt, M., Nürnberg, P., Wollnik, B. & Carney, T.J., 2012. Attenuated BMP1 function compromises osteogenesis, leading to bone fragility in humans and Zebrafish. Am. J. Hum. Genet. 90, 661–674 https://doi.org/10.1016/j.ajhg.2012.02.026.

Bader, J.S., 2001. The relative power of SNPs and haplotype as genetic markers for association tests. Pharmacogenomics 2, 11–24 https://doi.org/10.1517/14622416.2.1.11.

Beffa, L.M., 2005. Genotype x environment interaction in afrikaner cattle. Genotype X Environ. Interact. Afrikaner Cattle https://doi.org/10.6100/IR155825.

Behjati, S. & Tarpey, P.S., 2013. What is next generation sequencing? Arch. Dis. Child. Educ. Pract. Ed. 98, 236.

Berry, D.P. & Evans, R.D., 2014. Genetics of reproductive performance in seasonal calving beef cows and its association with performance traits. J. Anim. Sci. 92, 1412–1422.

Berry, D.P. & Kearney, J.F., 2011. Imputation of genotypes from low-to high-density genotyping platforms and implications for genomic selection. Animal 5, 1162–1169.

Berry, D.P., Kearney, J.F., Twomey, K. & Evans, R.D., 2013. Genetics of reproductive performance in seasonal calving dairy cattle production. Irish J. Agric. Food Res. 52, 1–16.

Berry, D.P., Mcclure, M.C. & Mullen, M.P., 2014a. 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 https://doi.org/10.1111/jbg.12067.

Berry, D.P., Wall, E. & Pryce, J.E., 2014b. Genetics and genomics of reproductive performance in dairy and beef cattle. Animal 8, 105–121.

Binelli, M., Scolari, S.C., Pugliesi, G., Van Hoeck, V., Gonella-Diaza, A.M., Andrade, S.C.S., Gasparin, G.R. & Coutinho, L.L., 2015. The transcriptome signature of the receptive bovine uterus determined at early gestation. PLoS One 10 https://doi.org/10.1371/journal.pone.0122874.

Bolormaa, S., Hayes, B.J., Savin, K., Hawken, R., Barendse, W., Arthur, P.F., Herd, R.M. & Goddard, M.E., 2011. Genome-wide association studies for feedlot and growth traits in cattle 1. J. Anim. Sci 89, 1684–1697.

Bolormaa, S., Pryce, J.E., Reverter, A., Zhang, Y.D., Barendse, W., Kemper, K.E., Tier, B., Savin, K., Hayes, B.J. & Goddard, M.E., 2014. A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. PLoS Genet. 10, e1004198.

Bolormaa, S., Pryce, J.E., Zhang, Y., Reverter, A., Barendse, W., Hayes, B.J. & Goddard, M.E., 2015. Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. Genet. Sel. Evol. 47, 1–12.

Bolormaa, S., Swan, A.A., Brown, D.J., Hatcher, S., Moghaddar, N., Van Der Werf, J.H., Goddard, M.E. & Daetwyler, H.D., 2017. Multiple-trait QTL mapping and genomic prediction for wool traits in sheep. Genet. Sel. Evol. 49, 1–22.

Bonsma, J., 1980. Livestock production -- a global approach. 1st ed. Tafelberg Publishers Ltd, Cape Town, South Africa.

Bosman, L., van Marle-Köster, E., van der Westhuizen, R.R., Visser, C. & Berry, D.P., 2017. Short communication: Population structure of the South African Bonsmara beef breed using high density single nucleotide polymorphism genotypes. Livest. Sci. 197, 102–105 https://doi.org/10.1016/j.livsci.2017.01.012.

Bourdon, R.M., 2000. Understanding Animal Breeding. 2nd ed. Prentice Hall.

Bovine HapMap Consortium, T.B.H., Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., Dodds, K.G., *et al.*, 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science (80-.). 324, 528–32.

Brøndum, R.F., Guldbrandtsen, B., Sahana, G., Lund, M.S. & Su, G., 2014. Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. BMC Genomics 15, 1–8 https://doi.org/10.1186/1471-2164-15-728.

Browning, B.L. & Browning, S.R., 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am. J. Hum. Genet. 84, 210–223.

Browning, B.L. & Browning, S.R., 2016. Genotype Imputation with Millions of Reference Samples. Am. J. Hum. Genet. 98, 116–126.

Buchanan, F.C., Van Kessel, A.G., Waldner, C., Christensen, D.A., Laarveld, B. & Schmutz, S.M., 2003. Hot topic: An association between a leptin single nucleotide polymorphism and milk and protein yield. J. Dairy Sci. 86, 3164–3166 https://doi.org/10.3168/jds.S0022-0302(03)73918-6.

Burns, B.M., Fordyce, G. & Holroyd, R.G., 2010. A review of factors that impact on the capacity of beef cattle females to conceive, maintain a pregnancy and wean a calf—Implications for reproductive efficiency in northern Australia. Anim. Reprod. Sci. 122, 1–22.

Burrow, H.M., 2001. Variances and covariances between productive and adaptive traits and temperament in a composite breed of tropical beef cattle. Livest. Prod. Sci. 70, 213–233.

Buzanskas, M.E., Grossi, D. do A., Ventura, R.V., Schenkel, F.S., Chud, T.C.S., Stafuzza, N.B., Rola, L.D., Meirelles, S.L.C., Mokry, F.B., Mudadu, M. de A., Higa, R.H., da Silva, M.V.G.B., de Alencar, M.M., Regitano, L.C. de A. & Munari, D.P., 2017. Candidate genes for male and female reproductive traits in Canchim beef cattle. J. Anim. Sci. Biotechnol. 8 https://doi.org/10.1186/s40104-017-0199-8.

Cammack, K.M., Thomas, M.G. & Enns, R.M., 2009. Reproductive Traits and Their Heritabilities in Beef Cattle. Prof. Anim. Sci. 25, 517–528.

Casas, E., Lunstra, D.D. & Stone, R.T., 2004. Quantitative trait loci for male reproductive traits in beef cattle. Anim. Genet. 35, 451–453 https://doi.org/10.1111/j.1365-2052.2004.01190.x.

Cavani, L., Garcia, D.A., Carreno, L.O.D., Ono, R.K., Pires, M.P., Farah, M.M., Ventura, H.T., Millen, D.D. & Fonseca, R., 2015. Estimates of genetic parameters for reproductive traits in Brahman cattle breed. J. Anim. Sci. 93, 3287–3291.

Chan, E.K.F., Hawken, R. & Reverter, A., 2009. The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. Anim. Genet. 40, 149–56.

Chang, N.S., Schultz, L., Hsu, L.J., Lewis, J., Su, M. & Sze, C.I., 2005. 17β-estradiol upregulates and activates WOX1/WWOXv1 and WOX2/WWOXv2 in vitro: Potential role in cancerous progression of breast and prostate to a premetastatic state in vivo. Oncogene 24, 714–723 https://doi.org/10.1038/sj.onc.1208124.

Chen, Q., Huang, B., Zhan, J., Wang, J., Qu, K., Zhang, F., Shen, J., Jia, P., Ning, Q., Zhang, J., Chen, N., Chen, H. & Lei, C., 2020. Whole-genome analyses identify loci and selective signals associated with body size in cattle. J. Anim. Sci. 98 https://doi.org/10.1093/jas/skaa068.

Choudhury, S.R., Satishchandra, P., Sinha, S. & Anand, A., 2019. Gene variants in ZGRF1 implicated for a rare sensory reflex epilepsy. bioRxiv Genet., 728188 https://doi.org/10.1101/728188.

Chud, T.C.S., Ventura, R. V., Schenkel, F.S., Carvalheiro, R., Buzanskas, M.E., Rosa, J.O., de Alvarenga Mudadu, M., da Silva, M.V.G.B., Mokry, F.B., Marcondes, C.R., Regitano, L.C.A. & Munari, D.P., 2015. Strategies for genotype imputation in composite beef cattle. BMC Genet. 16, 1–10.

Cohen, M.M., 2013. The AKT genes and their roles in various disorders. Am. J. Med. Genet. Part A 161, 2931–2937 https://doi.org/10.1002/ajmg.a.36101.

Corbet, N.J., Allen, J.M., Laing, A.R., Fordyce, G., McGowan, M.R. & Burns, B.M., 2018. Using ultrasound to derive new reproductive traits in tropical beef breeds: implications for genetic evaluation. Anim. Prod. Sci. 58, 1735.

Corbet, N.J., Burns, B.M., Johnston, D.J., Wolcott, M.L., Corbet, D.H., Venus, B.K., Li, Y., McGowan, M.R. & Holroyd, R.G., 2013. Male traits and herd reproductive capability in tropical beef cattle. 2. Genetic parameters of bull traits. Anim. Prod. Sci. 53, 101–113 https://doi.org/10.1071/AN12163.

Corbet, N.J., Shepherd, R.K., Burrow, H.M., Prayaga, K.C., van der Westhuizen, J. & Bosman, D.J., 2006. Evaluation of Bonsmara and Belmont Red cattle breeds in South Africa. 2. Genetic parameters for growth and fertility. Aust. J. Exp. Agric. 46, 213–223.

Costa, R.B., Camargo, G.M.F., Diaz, I.D.P.S., Irano, N., Dias, M.M., Carvalheiro, R., Boligon, A.A., Baldi, F., Oliveira, H.N., Tonhati, H. & Albuquerque, L.G., 2015. Genome-wide association study of reproductive traits in Nellore heifers using Bayesian inference. Genet. Sel. Evol. 47, 1–9.

Cui, C., Chatterjee, B., Lozito, T.P., Zhang, Z., Francis, R.J., Yagi, H., Swanhart, L.M., Sanker, S., Francis, D., Yu, Q., San Agustin, J.T., Puligilla, C., Chatterjee, T., Tansey, T., Liu, X., Kelley, M.W., Spiliotis, E.T., Kwiatkowski, A. V., Tuan, R., Pazour, G.J., Hukriede, N.A. & Lo, C.W., 2013. Wdpcp, a PCP Protein Required for Ciliogenesis, Regulates Directional Cell Migration and Cell Polarity by Direct Modulation of the Actin Cytoskeleton. PLoS Biol. 11 https://doi.org/10.1371/journal.pbio.1001720.

da Silva Romero, A.R., Siqueira, F., Santiago, G.G., de Almeida Regitano, L.C., de Souza Júnior, M.D., de Almeida Torres Júnior, R.A., do Nascimento, A.V. & Grisolia, A.B., 2018. Prospecting genes associated with navel length, coat and scrotal circumference traits in Canchim cattle. Livest. Sci. 210, 33–38.

D'Angelo, M.A., Gomez-Cavazos, J.S., Mei, A., Lackner, D.H. & Hetzer, M.W., 2012. A Change in Nuclear Pore Complex Composition Regulates Cell Differentiation. Dev. Cell 22, 446–458 https://doi.org/10.1016/j.devcel.2011.11.021.

DAFF., 2017. Abstract of Agricultural statistics 2017, Department of Agriculture Forestry and Fisheries, South Africa. 51-67.

Dassonneville, R., Fritz, S., Ducrocq, V. & Boichard, D., 2012. Short communication: Imputation performances of 3 low-density marker panels in beef and dairy cattle. J. Dairy Sci. 95, 4136–4140.

Davis, G.P., 1993. Genetic parameters for tropical beef cattle in northern Australia: a review. Aust. J. Agric. Res. 44, 179–198.

De León, C., Manrique, C., Martínez, R. & Rocha, J.F., 2019. Genomic association study for adaptability traits in four colombian cattle breeds. Genet. Mol. Res. 18 https://doi.org/10.4238/gmr18373.

de Melo, T.P., Salinas Fortes, M.R., Hayes, B., de Albuquerque, L.G. & Carvalheiro, R., 2020. Across-breed validation study confirms and identifies new loci associated with sexual precocity in Brahman and Nellore cattle. J. Anim. Breed. Genet. 137, 139–154 https://doi.org/10.1111/jbg.12429.

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.

Dekkers, J.C.M., 2010. Use of high-density marker genotyping for genetic improvement of livestock by genomic selection. CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour. 5, 1–13.

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

Dempfle, L., 1977. Comparison of several sire evauluation methods in dairy cattle breeding. Livest. Prod. Sci. 4, 129–139.

Deng, Y., Johnson, D.R., Guan, X., Ang, C.Y., Ai, J. & Perkins, E.J., 2010. In vitro gene regulatory networks predict in vivo function of liver. BMC Syst. Biol. 4, 153 https://doi.org/10.1186/1752-0509-4-153.

Derecka, K., Ahmad, S., Hodgman, T.C., Hastings, N., Royal, M.D., Woolliams, J.A. & Flint, A.P.F., 2010. Sequence variants in the bovine gonadotrophin releasing hormone receptor gene and their associations with fertility. Anim. Genet. 41, 329–331.

Dias, M.M., Cánovas, A., Mantilla-Rojas, C., Riley, D.G., Luna-Nevarez, P., Coleman, S.J., Speidel, S.E., Enns, R.M., Islas-Trejo, A., Medrano, J.F., Moore, S.S., Fortes, M.R.S., Nguyen, L.T., Venus, B., Diaz, I.S.D.P., Souza, F.R.P., Fonseca, L.F.S., Baldi, F., Albuquerque, L.G., Thomas, M.G. & Oliveira, H.N., 2017. SNP detection using RNA-sequences of candidate genes associated with puberty in cattle. Genet. Mol. Res. 16, 16019522 https://doi.org/10.4238/gmr16019522.

Dias, M.M., Souza, F.R.P., Takada, L., Feitosa, F.L.B., Costa, R.B., Diaz, I.D.P.S., Cardoso, D.F., Tonussi, R.L., Baldi, F., Albuquerque, L.G. & Oliveira, H.N., 2015. Study of lipid metabolism-related genes as candidate genes of sexual precocity in nellore cattle. Genet. Mol. Res. 14, 234–243 https://doi.org/10.4238/2015.January.16.7.

Dickendesher, T.L., Baldwin, K.T., Mironova, Y.A., Koriyama, Y., Raiker, S.J., Askew, K.L., Wood, A., Geoffroy, C.G., Zheng, B., Liepmann, C.D., Katagiri, Y., Benowitz, L.I., Geller, H.M. & Giger, R.J., 2012. NgR1 and NgR3 are receptors for chondroitin sulfate proteoglycans. Nat. Neurosci. 15, 703–712 https://doi.org/10.1038/nn.3070.

Do, D.N., Ostersen, T., Strathe, A.B., Mark, T., Jensen, J. & Kadarmideen, H.N., 2014. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. BMC Genet. 15, 1–15 https://doi.org/10.1186/1471-2156-15-27.

Do, D.N., Schenkel, F.S., Miglior, F., Zhao, X. & Ibeagha-Awemu, E.M., 2018. Genome wide association study identifies novel potential candidate genes for bovine milk cholesterol content. Sci. Rep. 8, 13239.

Do Nascimento, A.V., Da Silva Romero, Â.R., Utsunomiya, Y.T., Utsunomiya, A.T.H., Cardoso, D.F., Neves, H.H.R., Carvalheiro, R., Garcia, J.F. & Grisolia, A.B., 2018. Genome-wide association study using haplotype alleles for the evaluation of reproductive traits in Nelore cattle. PLoS One 13, 1–15.

Dos Santos, J.P.R., De Castro Vasconcellos, R.C., Pires, L.P.M., Balestre, M. & Von Pinho, R.G., 2016. Inclusion of dominance effects in the multivariate GBLUP model (Q Zhang, Ed.). PLoS One 11, e0152045 https://doi.org/10.1371/journal.pone.0152045.

du Plessis, I., Hoffman, L.C. & Calitz, F.J., 2006. Influence of reproduction traits and pre-weaning growth rate on herd efficiciency of different beef breed types in an arid sub-tropical environment. S. Afr. J. Anim. Sci. 36, 89–98.

Edea, Z., Dadi, H., Kim, S.W., Park, J.H., Shin, G.H., Dessie, T. & Kim, K.S., 2014. Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in Ethiopia. J. Anim. Breed. Genet. 131, 358–366.

Elangovan, M., Chong, H.K., Park, J.H., Yeo, E.J. & Yoo, Y.J., 2017. The role of ubiquitin-conjugating enzyme Ube2j1 phosphorylation and its degradation by proteasome during endoplasmic stress recovery. J. Cell Commun. Signal. 11, 265–273 https://doi.org/10.1007/s12079-017-0386-6.

Eler, J.P., Ferraz, J.B.S., Baliero, J.C.C. & Mattos, E.C., 2008. Genetic analysis of average annual productivity of Nellore breeding cows (COWPROD). Genet. Mol. Res. 7, 234–242.

Elsik, C.G., Unni, D.R., Diesh, C.M., Tayal, A., Emery, M.L., Nguyen, H.N. & Hagen, D.E., 2016. Bovine Genome Database: new tools for gleaning function from the Bos taurus genome. Nucleic Acids Res. 44.

Erbe, M., Edel, C., Pimmental, E.C.G., Dodenhoff, J. & Gotz, K.-U., 2018. Approximation of Reliability in Single Step Models using the Interbull Standardized Genomic Reliability Method. INTERBULL Bul. 54.

Erdman, V. V., Karimov, D.D., Nasibullin, T.R., Timasheva, I.R., Tuktarova, I.A. & Mustafina, O.E., 2017. The role of Alu polymorphism of PLAT, PKHD1L1, STK38L, and TEAD1 genes in development of a longevity trait. Adv. Gerontol. 7, 107–113 https://doi.org/10.1134/S2079057017020059.

Espigolan, R., Baldi, F., Boligon, A.A., Souza, F.R., Gordo, D.G., Tonussi, R.L., Cardoso, D.F., Oliveira, H.N., Tonhati, H., Sargolzaei, M., Schenkel, F.S., Carvalheiro, R., Ferro, J.A. & Albuquerque, L.G., 2013. Study of whole genome linkage disequilibrium in Nellore cattle. BMC Genomics 14, 305.

Evans, J.L., Golden, B.L., Bourdon, R.M. & Long, K.L., 1999. Additive genetic relationships between heifer pregnancy and scrotal circumference in Hereford cattle. J. Anim. Sci. 77, 2621–8.

Fan, G., 2020. FER mediated HGF-independent regulation of HGFR/MET activates RAC1-PAK1 pathway to potentiate metastasis in ovarian cancer. Small GTPases 11, 155–159 https://doi.org/10.1080/21541248.2017.1379931.

Fei, Y., Webb, R., Cobb, B.L., Direskeneli, H., Saruhan-Direskeneli, G. & Sawalha, A.H., 2009. Identification of novel genetic susceptibility loci for Behçet's disease using a genome-wide association study. Arthritis Res. Ther. 11, 1–7 https://doi.org/10.1186/ar2695.

Fernández, J.C., Pérez, J.E., Herrera, N., Martínez, R., Bejarano, D. & Rocha, J.F., 2019. Research Article Genomic association study for age at first calving and calving interval in Romosinuano and Costeño con Cuernos cattle. Genet. Mol. Res. 18 https://doi.org/10.4238/gmr18258.

Fernando, R.L. & Garrick, D.J., 2008. GenSel - User manual for a portfolio of genomic selection related analyses (Z Liu, Ed.). , e61756.

Ferreira, A.M., Bislev, S.L., Bendixen, E. & Almeida, A.M., 2013. The mammary gland in domestic ruminants: A systems biology perspective. J. Proteomics 94, 110–123 https://doi.org/10.1016/j.jprot.2013.09.012.

Foote, R.H., 2003. Fertility estimation: a review of past experience and future prospects. Anim. Reprod. Sci. 75, 119–139.

Fortes, M.R.S., DeAtley, K.L., Lehnert, S.A., Burns, B.M., Reverter, A., Hawken, R.J., Boe-Hansen, G., Moore, S.S. & Thomas, M.G., 2013. Genomic regions associated with fertility traits in male and female cattle: Advances from microsatellites to high-density chips and beyond. Anim. Reprod. Sci. 141, 1–19.

Fortes, M.R.S., Lehnert, S.A., Bolormaa, S., Reich, D., Fordyce, G., Corbet, N.J., Whan, V., Hawken, R.J. & Reverter, A., 2012a. Finding genes for economically important traits: Brahman cattle puberty. Anim. Prod. Sci. 52, 143–150.

Fortes, M.R.S., Reverter, A., Hawken, R.J., Bolormaa, S. & Lehnert, S.A., 2012b. Candidate Genes Associated with Testicular Development, Sperm Quality, and Hormone Levels of Inhibin, Luteinizing Hormone, and Insulin-Like Growth Factor 1 in Brahman Bulls. Biol. Reprod. 87, 1–8.

Free Software Foundation., 2018. Ubuntu 18.04.2 LTS.

Freebern, E., Santos, D.J.A., Fang, L., Jiang, J., Parker Gaddis, K.L., Liu, G.E., Vanraden, P.M., Maltecca, C., Cole, J.B. & Ma, L., 2020. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC Genomics 21, 1–11 https://doi.org/10.1186/s12864-020-6461-z.

Frischknecht, M., Bapst, B., Seefried, F.R., Signer-Hasler, H., Garrick, D., Stricker, C., Fries, R., Russ, I., Sölkner, J., Bieber, A., Strillacci, M.G., Gredler-Grandl, B. & Flury, C., 2017. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. BMC Genomics 18, 1–13.

Fuchsberger, C., Abecasis, G.R. & Hinds, D.A., 2015. minimac2: faster genotype imputation. Bioinformatics 31, 782–784.

Fuerst, P.G., Bruce, F., Tian, M., Wei, W., Elstrott, J., Feller, M.B., Erskine, L., Singer, J.H. & Burgess, R.W., 2009. DSCAM and DSCAML1 Function in Self-Avoidance in Multiple Cell Types in the Developing Mouse Retina. Neuron 64, 484–497 https://doi.org/10.1016/j.neuron.2009.09.027.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J. & Altshuler, D., 2002. The structure of haplotype blocks in the human genome. Science 296, 2225–2229.

Galarneau, G., Fontanillas, P., Team, the C.R., Team, the 23andMe R., Clementi, C., Hu-Seliger, T., Parfitt, D.-E., Tung, J.Y. & Beim, P.Y., 2018. Genome-wide association studies on endometriosis and endometriosis-related infertility. bioRxiv, 401448 https://doi.org/10.1101/401448.

Garg, A., 2011. Lipodystrophies: Genetic and acquired body fat disorders. J. Clin. Endocrinol. Metab. 96, 3313–3325 https://doi.org/10.1210/jc.2011-1159.

Garrick, D.J., 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet. Sel. Evol. 43, 1–11.

Garrick, D.J., Taylor, J.F. & Fernando, R.L., 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41, 55.

Gobikrushanth, M., Purfield, D.C., Colazo, M.G., Wang, Z., Butler, S.T. & Ambrose, D.J., 2018. The relationship between serum insulin-like growth factor-1 (IGF-1) concentration and reproductive performance, and genome-wide associations for serum IGF-1 in Holstein cows. J. Dairy Sci. 101, 9154–9167 https://doi.org/10.3168/jds.2018-14535.

Goddard, M., 1985. A Method of Comparing Sires Evaluated in Different Countries. Livest. Prod. Sci. 13, 321–331.

Goddard, M.E. & Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Publ. Gr. 10.

Goldberger, A.S., 1962. Best linear unbiased prediction in the general linear regression model. J. Am. Stat. Assoc. 57, 369–375.

González-Ruiz, S., Strillacci, M.G., Durán-Aguilar, M., Cantó-Alarcón, G.J., Herrera-Rodríguez, S.E., Bagnato, A., Guzmán, L.F., Milián-Suazo, F. & Román-Ponce, S.I., 2019. Genome-wide association study in mexican holstein cattle reveals novel quantitative trait loci regions and confirms mapped loci for resistance to bovine tuberculosis. Animals 9 https://doi.org/10.3390/ani9090636.

Gorbach, D.M., Makgahlela, M.L., Reecy, J.M., Kemp, S.J., Baltenweck, I., Ouma, R., Mwai, O., Marshall, K., Murdoch, B., Moore, S. & Rothschild, M.F., 2010. Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. J. Anim. Breed. Genet. 127, 348–351.

Gororo, E., Makuza, S.M., Chatiza, F.P., Chidzwondo, F. & Sanyika, T.W., 2018. Genetic diversity in Zimbabwean Sanga cattle breeds using microsatellite markers. S. Afr. J. Anim. Sci. 48, 128.

Granleese, T., Clark, S.A., Swan, A.A. & van der Werf, J.H.J., 2015. Increased genetic gains in sheep, beef and dairy breeding programs from using female reproductive technologies combined with optimal contribution selection and genomic breeding values. Genet. Sel. Evol. 47, 70.

Grgurevic, L., Macek, B., Mercep, M., Jelic, M., Smoljanovic, T., Erjavec, I., Dumic-Cule, I., Prgomet, S., Durdevic, D., Vnuk, D., Lipar, M., Stejskal, M., Kufner, V., Brkljacic, J., Maticic, D. & Vukicevic, S., 2011. Bone morphogenetic protein (BMP)1-3 enhances bone repair. Biochem. Biophys. Res. Commun. 408, 25–31 https://doi.org/10.1016/j.bbrc.2011.03.109.

Grobler, S.M., Scholtz, M.M., Greyling, J.P.C. & Neser, F.W.C., 2014. Reproduction performance of beef cattle mated naturally following synchronization in the Central Bushveld bioregion of South Africa. S. Afr. J. Anim. Sci. 44, 70–74.

Grobler, R., Visser, C., Capitan, A. & van Marle-Köster, E., 2018. Validation of the POLLED Celtic variant in South African Bonsmara and Drakensberger beef cattle breeds. Livest. Sci. 217, 136–139.

Gurgul, A., Semik, E., Pawlina, K., Szmatoła, T., Jasielczuk, I. & Bugno-Poniewierska, M., 2014. The application of genome-wide SNP genotyping methods in studies on livestock genomes.

Gutiérrez, J.P., Alvarez, I., Fernández, I., Royo, L.J., Díez, J. & Goyache, F., 2002. Genetic relationships between calving date, calving interval, age at first calving and type traits in beef cattle. Livest. Prod. Sci. 78, 215–222.

Habier, D., Fernando, R.L., Kizilkaya, K. & Garrick, D.J., 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12.

Halperin, E., Kimmel, G. & Shamir, R., 2005. Tag SNP selection in genotype data for maximizing SNP prediction accuracy. Bioinformatics 21, i195-1203.

Harris, B. & Johnson, D., 1998. Approximate Reliability of Genetic Evaluations under an Animal Model. J. Dairy Sci. 81, 2723–2728 https://doi.org/10.3168/jds.S0022-0302(98)75829-1.

Hawken, R.J., Zhang, Y.D., Fortes, M.R.S., Collis, E., Barris, W.C., Corbet, N.J., Williams, P.J., Fordyce, G., Holroyd, R.G., Walkley, J.R.W., Barendse, W., Johnston, D.J., Prayaga, K.C., Tier, B., Reverter, A. & Lehnert, S.A., 2012. Genome-wide association studies of female reproduction in tropically adapted beef cattle. J. Anim. Sci. 90, 1398–1410.

Hay, E.H. & Roberts, A., 2018. Genome-wide association study for carcass traits in a composite beef cattle breed. Livest. Sci. 213, 35–43 https://doi.org/10.1016/j.livsci.2018.04.018.

Hay, E.H. & Roberts, A., 2019. Genomic evaluation of genotype by prenatal nutritional environment interaction for maternal traits in a composite beef cattle breed. Livest. Sci. 229, 118–125 https://doi.org/10.1016/j.livsci.2019.09.022.

Hayes, B. & Goddard, M., 2010. Genome-wide association and genomic selection in animal breeding. Genome 53, 876–883.

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, 206–213.

Hayes, B.J., Visscher, P.M., McPartlan, H.C. & Goddard, M.E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13, 635–43.

Henderson, C.R., 1950. Estimation of genetic parameters (abstract). Annu. Math. Stat. 21, 309–310.

Henderson, C.R., 1963. Selection index and expected genetic parameters. Washington, DC.

Hieber, J.K., Endecott, R.L. & Thomson, J.M., 2018. Identification of genetic markers and QTL for carcass quality traits within the American Simmental Association Carcass Merit Program. Transl. Anim. Sci. 2, S39–S43 https://doi.org/10.1093/tas/txy032.

Hirschhorn, J.N. & Daly, M.J., 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95–108.

Hlatshwayo, M., 2015. Beef Cattle Management: A Nutrtional Focus. Pretoria, RSA.

Hoch, R. V. & Soriano, P., 2006. Context-specific requirements for Fgfr 1 signaling through Frs2 and Frs3 during mouse development. Development 133, 663–673 https://doi.org/10.1242/dev.02242.

Holst, P.J., 1999. Recording and on-farm evaluations and monitoring: breeding and selection. Small Rumin. Res. 34, 197–202.

Howie, B.N., Donnelly, P. & Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529.

Hu, F., Liu, J.F., Zeng, Z.B., Ding, X.D., Yin, C.C., Gong, Y.Z. & Zhang, Q., 2010. QTL Identification using combined linkage and linkage disequilibrium mapping for milk production traits on BTA6 in Chinese Hstein population. Asian-Australasian J. Anim. Sci. 23, 1261–1267 https://doi.org/10.5713/ajas.2010.10011.

Ibeagha-Awemu, E.M., Peters, S.O., Akwanji, K.A., Imumorin, I.G. & Zhao, X., 2016. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. Sci. Rep. 6, 1–18 https://doi.org/10.1038/srep31109.

Iung, L.H. de S., Mulder, H.A., Neves, H.H. de R. & Carvalheiro, R., 2018. Genomic regions underlying uniformity of yearling weight in Nellore cattle evaluated under different response variables. BMC Genomics 19, 1–13.

Jacob, K., Albrecht, S., Sollier, C., Faury, D., Sader, E., Montpetit, A., Serre, D., Hauser, P., Garami, M., Bognar, L., Hanzely, Z., Montes, J.L., Atkinson, J., Farmer, J.P., Bouffet, E., Hawkins, C., Tabori, U. & Jabado, N., 2009. Duplication of 7q34 is specific to juvenile pilocytic astrocytomas and a hallmark of cerebellar and optic pathway tumours. Br. J. Cancer 101, 722–733 https://doi.org/10.1038/sj.bjc.6605179.

Jairath, L., Dekkers, J.C., Schaeffer, L.R., Liu, Z., Burnside, E.B. & Kolstad, B., 1998. Genetic herd evaluation for herd life in Canada. J. Dairy Sci. 81, 550–562.

James, J.W. & Roberts, E.M., 1979. Measurement in sheep and goat improvement. Kensington, N.S.W.

Jeter, R.M., Sias, S.R. & Ingraham, J.L., 1984. Chromosomal location and function of genes affecting Pseudomonas aeruginosa nitrate assimilation.

Johnston, D.J., Barwick, S.A., Fordyce, G., Holroyd, R.G., Williams, P.J., Corbet, N.J. & Grant, T., 2014a. Genetics of early and lifetime annual reproductive performance in cows of two tropical beef genotypes in northern Australia. Anim. Prod. Sci. 54, 1–15 https://doi.org/10.1071/AN13043.

Johnston, D.J., Corbet, N.J., Barwick, S.A., Wolcott, M.L. & Holroyd, R.G., 2014b. Genetic correlations of young bull reproductive traits and heifer puberty traits with female reproductive performance in two tropical beef genotypes in northern Australia. Anim. Prod. Sci. 54, 74–84 https://doi.org/10.1071/AN13044.

Judge, M.M., Kearney, J.F., McClure, M.C., Sleator, R.D. & Berry, D.P., 2016. Evaluation of developed low-density genotype panels for imputation to higher density in independent dairy and beef cattle populations. J. Anim. Sci. 94, 949–962.

Khatkar, M.S., Moser, G., Hayes, B.J. & Raadsma, H.W., 2012. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. BMC Genomics 13, 538.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A.L., Barris, W., Schnabel, R.D., Taylor, J.F. & Raadsma, H.W., 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics 9, 187.

Kinghorn, B.P., Banks, R.G. & Simm, G., 2014. Genetic improvement of beef cattle.Page pp 451-473 in The Genetics of Cattle. Garrick, D., Ruvinsky, A., eds. CABI.

Kirkpatrick, B.W., 2014. Genetics of reproduction in cattle.Pages 260–283 in The Genetics of Cattle. Garrick, D., Ruvinsky, A., eds. CABI.

Kluska, S., Olivieri, B.F., Bonamy, M., Chiaia, H.L.J., Feitosa, F.L.B., Berton, M.P., Peripolli, E., Lemos, M.V.A., Tonussi, R.L., Lôbo, R.B., Magnabosco, C. de U., Di Croce, F., Osterstock, J., Pereira, A.S.C., Munari, D.P., Bezerra, L.A., Lopes, F.B. & Baldi, F., 2018. Estimates of genetic parameters for growth, reproductive, and carcass traits in Nelore cattle using the single step genomic BLUP procedure. Livest. Sci. 216, 203–209.

Koivula, M., Strandén, I., Aamand, G.P. & Mantysaari, E.A., 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal 10, 1061–1066.

Kong, A., Masson, G., Frigge, M.L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P.I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F., Thorsteinsdottir, U., Gudbjartsson, D.F., Stefansson, H. & Stefansson, K., 2008. Detection of sharing by descent, long-range phasing and haplotype imputation. Nat. Genet. 40, 1068–1075.

Kosova, G., Hotaling, J.M., Ohlander, S., Niederberger, C., Prins, G.S. & Ober, C., 2014. Variants in DPF3 and DSCAML1 are associated with sperm morphology. J. Assist. Reprod. Genet. 31, 131–137 https://doi.org/10.1007/s10815-013-0140-9.

Kuehn, L.A., Keele, J.W., Bennett, G.L., McDaneld, T.G., Smith, T.P.L., Snelling, W.M., Sonstegard, T.S. & Thallman, R.M., 2011. Predicting breed composition using breed frequencies of 50,000 markers from the US Meat Animal Research Center. J. Anim. Sci. 89, 1742–1750.

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

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

Lashmar, S.F., Visser, C. & Muchadeyi, F.C., 2018b. Factors influencing imputation accuracy for the South African Drakensberger beef cattle breed.Page 11.472 in World Congress on Genetics Applied to Livestock Production.

Lee, J.H., Huynh, M., Silhavy, J.L., Kim, S., Dixon-Salazar, T., Heiberg, A., Scott, E., Bafna, V., Hill, K.J., Collazo, A., Funari, V., Russ, C., Gabriel, S.B., Mathern, G.W. & Gleeson, J.G., 2012. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. Nat. Genet. 44, 941–945 https://doi.org/10.1038/ng.2329.

Legarra, A., Christensen, O.F., Aguilar, I. & Misztal, I., 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166, 54–65 https://doi.org/10.1016/j.livsci.2014.04.029.

Li, C., Cai, W., Zhou, C., Yin, H., Zhang, Z., Loor, J.J., Sun, D., Zhang, Q., Liu, J. & Zhang, S., 2016. RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. Sci. Rep. 6, 1–11 https://doi.org/10.1038/srep26813.

Li, Y., Willer, C., Sanna, S. & Abecasis, G., 2009. Genotype imputation. Annu. Rev. Genomics Hum. Genet. 10, 387–406.

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M. & Strandén, I., 2017a. Techinal Reference Guide for MiX99 Solver. , 84.

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M. & Strandén, I., 2017b. Technical Reference Guide For MiX99 Pre-Processor. , 87.

Lidauer, M., Matilainen, K., Mäntysaari, E., Pitkänen, T., Taskinen, M. & Strandén, I., 2017c. Command Language Interface Manual (CLIM). , 84.

Lidauer, M., Strandén, I., Mantysaari, E.., Poso, J. & Kettunen, A., 1999. Solving Large Test-Day Models by Iteration on Data and Preconditioned Conjugate Gradient. J. Dairy Sci. 82, 2788–2796.

Lirón, J.P., Prando, A.J., Fernández, M.E., Ripoli, M. V, Rogberg-Muñoz, A., Goszczynski, D.E., Posik, D.M., Peral-García, P., Baldo, A. & Giovambattista, G., 2012. Association between GNRHR, LHR and IGF1 polymorphisms and timing of puberty in male Angus cattle. BMC Genet. 13, 26.

Lopes, F.B., Wu, X.L., Li, H., Xu, J., Perkins, T., Genho, J., Ferretti, R., Tait, R.G., Bauck, S. & Rosa, G.J.M., 2018. Improving accuracy of genomic prediction in Brangus cattle by adding animals with imputed low-density SNP genotypes. J. Anim. Breed. Genet. 135, 14–27.

López-Victorio, C.J., Velez-delValle, C., Beltrán-Langarica, A. & Kuri-Harcuch, W., 2013. EDF-1 downregulates the CaM/Cn/NFAT signaling pathway during adipogenesis. Biochem. Biophys. Res. Commun. 432, 146–151 https://doi.org/10.1016/j.bbrc.2013.01.069.

Ludes-Meyers, J.H., Kil, H., Nuñez, M.I., Conti, C.J., Parker-Thornburg, J., Bedford, M.T. & Aldaz, C.M., 2007. Wwox hypomorphic mice display a higher incidence of B-cell lymphomas and develop testicular atrophy. Genes Chromosom. Cancer 46, 1129–1136 https://doi.org/10.1002/gcc.20497.

Ma, Y., Sun, Y., Jiang, L., Zuo, K., Chen, H., Guo, J., Chen, F., Lai, Y. & Shi, J., 2017. WDPCP regulates the ciliogenesis of human sinonasal epithelial cells in chronic rhinosinusitis. Cytoskeleton 74, 82–90 https://doi.org/10.1002/cm.21351.

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

Makina, S.O., Muchadeyi, F.C., Van Marle-Köster, E., Taylor, J.F., Makgahlela, M.L. & Maiwashe, A., 2015. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genet. Sel. Evol. 47, 1–14 https://doi.org/10.1186/s12711-015-0173-x.

Makina, S.O., Whitacre, L.K., Decker, J.E., Taylor, J.F., MacNeil, M.D., Scholtz, M.M., van Marle-Köster, E., Muchadeyi, F.C., Makgahlela, M.L. & Maiwashe, A., 2016. Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide. Genet. Sel. Evol. 48, 88.

Mao, X., Johansson, A.M., Sahana, G., Guldbrandtsen, B. & De Koning, D.J., 2016. Short communication: Imputation of markers on the bovine X chromosome. J. Dairy Sci. 99, 7313–7318.

Mapholi, N.O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S.C., MacNeil, M.D., Banga, C., Taylor, J.F. & Dzama, K., 2016. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick. Borne. Dis. 7, 487–497.

Marchini, J. & Howie, B., 2010. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11, 499–511.

Marchionni, L., Afsari, B., Geman, D. & Leek, J.T., 2013. A simple and reproducible breast cancer prognostic test. BMC Genomics 14, 1–7 https://doi.org/10.1186/1471-2164-14-336.

Marete, A., Lund, M.S., Boichard, D. & Ramayo-Caldas, Y., 2018. A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. PLoS One 13 https://doi.org/10.1371/journal.pone.0199931.

Martens, H., Leonhard-Marek, S., Röntgen, M. & Stumpff, F., 2018. Magnesium homeostasis in cattle: absorption and excretion. Nutr. Res. Rev., 1–17 https://doi.org/10.1017/S0954422417000257.

Martinez-Velazquez, G., Gregory, K.E., Bennet, G.. & Van Vleck, L.D., 2003. Genetic relationships between scrotal circumference and female reproductive traits. J. Anim. Sci. 81, 395–401.

Martínez, R., Bejarano, D., Gómez, Y., Dasoneville, R., Jiménez, A., Even, G., Sölkner, J. & Mészáros, G., 2017. Genome-wide association study for birth, weaning and yearling weight in Colombian Brahman cattle. Genet. Mol. Biol. 40, 453–459.

Mateescu, R.G., 2020. Chapter 2 - Genetics and breeding of beef cattle. https://doi.org/10.1016/B978-0-12-817052-6.00002-1.

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.L., Sonstegard, T.S. & Van Tassell, C.P., 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle (AE Toland, Ed.). PLoS One 4, e5350.

Mazitov, T., Bregin, A., Philips, M.A., Innos, J. & Vasar, E., 2017. Deficit in emotional learning in neurotrimin knockout mice. Behav. Brain Res. 317, 311–318 https://doi.org/10.1016/j.bbr.2016.09.064.

McClure, M.C., Morsci, N.S., Schnabel, R.D., Kim, J.W., Yao, P., Rolf, M.M., McKay, S.D., Gregg, S.J., Chapple, R.H., Northcutt, S.L. & Taylor, J.F., 2010. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. Anim. Genet. 41, 597–607.

McDaneld, T.G., Kuehn, L.A., Thomas, M.G., Snelling, W.M., Smith, T.P.L., Pollak, E.J., Cole, J.B. & Keele, J.W., 2014. Genomewide association study of reproductive efficiency in female cattle. J. Anim. Sci. 92, 1945–1957.

McGovern, S.P., Purfield, D.C., Ring, S.., Carthy, T.R., Graham, D.A. & Berry, D.P., 2019. Candidate genes associated with the heritable humoral response to Mycobacterium avium ssp. paratuberculosis in dairy cows have factors in common with gastrointestinal diseases in humans. J. Dairy Sci., 4249–4263.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Neto, E., Gill, C.A., Gao, C., Mannen, H., Stothard, P., Wang, Z., Van Tassell, C.P., Williams, J.L., Taylor, J.F. & Moore, S.S., 2007. Whole genome linkage disequilibrium maps in cattle. BMC Genet. 8, 74.

McRae, J.F., Mainland, J.D., Jaeger, S.R., Adipietro, K.A., Matsunami, H. & Newcomb, R.D., 2012. Genetic Variation in the Odorant Receptor OR2J3 Is Associated with the Ability to Detect the “Grassy” Smelling Odor, cis-3-hexen-1-ol. Chem. Senses 37, 585–593.

Melo, T.P., Fortes, M.R.S., Bresolin, T., Mota, L.F.M., Albuquerque, L.G. & Carvalheiro, R., 2018. Multi-trait meta-analysis identified genomic regions associated with sexual precocity in tropical beef cattle. J. Anim. Sci., 1–9.

Messine, O., Schwalbach, L.M.J. & Greyling, J.P.C., 2004. The effects of restricted suckling and early weaning on cow reproduction and weaner production performance in Gudali cattle. S. Afr. J. Anim. Sci. 34, 119–121.

Meuwissen, T.H., Hayes, B.J. & Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–29.

Meyer, K., 1997. Estimates of genetic parameters for weaning weight of beef cattle accounting for direct-maternal environmental covariances. Livest. Prod. Sci., 187–199.

Meyer, K., 2007. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. Sci. B. 8, 815–821 https://doi.org/10.1631/jzus.2007.B0815.

Meyer, K., Hammond, K., Mackinnon, M.J. & Parnell, P.F., 1991. Estimates of covariances between reproduction and growth in Australian beef cattle. J. Anim. Sci. 69, 3533–43.

Meyer, K., Hammond, K., Parnell, P.F., MacKinnon, M.J. & Sivarajasingam, S., 1990. Estimates of heritability and repeatability for reproductive traits in Australian beef cattle. Livest. Prod. Sci. 25, 15–30.

Meyer, K. & Tier, B., 2012. ‘SNP Snappy’: A Strategy for Fast Genome-Wide Association Studies Fitting a Full Mixed Model. Genetics 190, 275–277.

Mi, H., Huang, X., Murunganujan, A., Tang, H., Mills, C., Kang, D. & Thomas, P.D., 2017. PANTHER version 11:expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183–D189.

Miar, Y., Plastow, G. & Wang, Z., 2015. Genomic Selection, a New Era for Pork Quality Improvement. Springer Sci. Rev. 3, 27–37.

Milanesi, M., Vicario, D., Stella, A., Valentini, A., Ajmone-Marsan, P., Biffani, S., Biscarini, F., Jansen, G. & Nicolazzi, E.L., 2015. Imputation accuracy is robust to cattle reference genome updates. Anim. Genet. 46, 69–72.

Militello, G., Hosen, M.R., Ponomareva, Y., Gellert, P., Weirick, T., John, D., Hindi, S.M., Mamchaoui, K., Mouly, V., Döring, C., Zhang, L., Nakamura, M., Kumar, A., Fukada, S.I., Dimmeler, S. & Uchida, S., 2018. A novel long non-coding RNA Myolinc regulates myogenesis through TDP-43 and Filip1. J. Mol. Cell Biol. 10, 102–117 https://doi.org/10.1093/jmcb/mjy025.

Miyata, H., Satouh, Y., Mashiko, D., Muto, M., Nozawa, K., Shiba, K., Fujihara, Y., Isotani, A., Inaba, K. & Ikawa, M., 2015. Sperm calcineurin inhibition prevents mouse fertility with implications for male contraceptive. Science (80-.). 350, 442–445 https://doi.org/10.1126/science.aad0836.

Mkhize, F.N., Webb, E.C. & Scholtz, M.M., 2019. Effects of non-genetic factors on the inter-calving period of Nguni cows in South Africa. S. Afr. J. Anim. Sci. 48.

Mokolobate, M.C., Scholtz, M.M., Jordaan, G.F. & Neser, F.W.C., 2018. Retrospect evaluation of cow productivity in the South African landrace breeds and its environmental impact. J. Anim. Sci. 96.

Moser, G., Khatkar, M.S. & Raadsma, H.W., 2009. Imputation of missing genotypes in high density SNP data. Proc. Assoc. Advmt. Anim. Breed. Genet., 612–615.

Mota, R.R., Guimarães, S.E.F., Fortes, M.R.S., Hayes, B., Silva, F.F., Verardo, L.L., Kelly, M.J., de Campos, C.F., Guimarães, J.D., Wenceslau, R.R., Penitente-Filho, J.M., Garcia, J.F. & Moore, S., 2017. Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle. J. Anim. Breed. Genet. 134, 484–492 https://doi.org/10.1111/jbg.12299.

Mukasa-Mugerwa, E., 1989. A Review of a Reproductive Performance of Female Bos Indicus (zebu) Cattle.

Musemwa, L., Mushunje, A., Chimonyo, M., Fraser, G., Mapiye, C. & Muchenje, V., 2008. Nguni cattle marketing constraints and opportunities in the communal areas of South Africa: Review. African J. Agric. Res. 3, 239–245.

Naderi, S., Bohlouli, M., Yin, T. & König, S., 2018. Genomic breeding values, SNP effects and gene identification for disease traits in cow training sets. Anim. Genet. 49, 178–192 https://doi.org/10.1111/age.12661.

Nascimento, A.V., Matos, M.C., Seno, L.O., Romero, A.R.S., Garcia, J.F. & Grisolia, A.B., 2016. Genome wide association study on early puberty in Bos indicus. Genet. Mol. Res. 15, 1–6.

National Center for Biotechnology Information., 2019.

Nephawe, K.A., Cundiff, L. V., Dikeman, M.E., Crouse, J.D. & Van Vleck, L.D., 2004. Genetic relationships between sex-specific traits in beef cattle: Mature weight, weight adjusted for body condition score, height and body condition score of cows, and carcass traits of their steer relatives. J. Anim. Sci. 82, 647–653 https://doi.org/10.1093/ansci/82.3.647.

Neser, F.W.., van Wyk, J.., Fair, M.. & Lubout, P., 2012. Genetic evaluation of growth traits in beef cattle using random regression models. S. Afr. J. Anim. Sci. 42, 474–477.

Nicolazzi, E.L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., Pirani, A., Soans, C., Brew, F., Jorjani, H., Evans, G., Simpson, B., Tosser-Klopp, G., Brauning, R., Williams, J.L. & Stella, A., 2015. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. BMC Genomics 16, 283.

Nicolazzi, E.L., Marras, G. & Stella, A., 2016. SNPConvert: SNP Array Standardization and Integration in Livestock Species. Microarrays 5.

Nielsen, R. & Slatkin, M., 2013. An Introduction to Population Genetics : Theory and Applications. 2nd ed. Sinauer Associates.

Nino-Soto, M.. & King, W.., 2004. Genetic Factors that Affect Normal Reproduction and Fertility in Domestic cattle.Pages 1–6 in 23rd World Buiactrics Congress. Quebec, Canada.

Norris, D., Banga, C., Benyi, K. & Sithole, B.., 2004. Estimation of genetic parameters and variance components for growth traits of Nguni cattle in Limpopo province, South Africa. Trop. Anim. Health Prod. 36, 801–806.

Oman, R.E., Streeter, R.N., Reppert, E.J. & Chako, C.Z., 2016. Left Displacement of the Abomasum in 4 Beef Calves. J. Vet. Intern. Med. 30, 1376–1380 https://doi.org/10.1111/jvim.14353.

Ong, S.H., Guy, G.R., Hadari, Y.R., Laks, S., Gotoh, N., Schlessinger, J. & Lax, I., 2000. FRS2 Proteins Recruit Intracellular Signaling Pathways by Binding to Diverse Targets on Fibroblast Growth Factor and Nerve Growth Factor Receptors. Mol. Cell. Biol. 20, 979–989 https://doi.org/10.1128/mcb.20.3.979-989.2000.

Organisasie, R.P. & RPO., 2017. Overview of the south african red meat industry. , 44.

Parish, G.A., Rhinehart, J.D. & Smith, T., 2011. Expected Progeny Differences and Selection Indices for Beef Cattle Selection.

Park, G., Kim, H.S., Choe, J.Y. & Kim, S.K., 2012. SUMO4 C438T polymorphism is associated with papulopustular skin lesion in Korean patients with Behçet's disease. Rheumatol. Int. 32, 3031–3037 https://doi.org/10.1007/s00296-011-2086-5.

Parkinson, T.J., 2004. Evaluation of fertility and infertility in natural service bulls. Vet. J. 168, 215–229.

Partipilo, G., D'Addabbo, P., Lacalandra, G.M., Liu, G.E. & Rocchi, M., 2011. Refinement of Bos taurus sequence assembly based on BAC-FISH experiments. BMC Genomics 12, 639.

Pausch, H., MacLeod, I.M., Fries, R., Emmerling, R., Bowman, P.J., Daetwyler, H.D. & Goddard, M.E., 2017. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. Genet. Sel. Evol. 49, 1–14.

Pereira, G.L., Chud, T.C.S., Bernardes, P.A., Venturini, G.C., Chardulo, L.A.L. & Curi, R.A., 2017. Genotype Imputation and Accuracy Evaluation in Racing Quarter Horses Genotyped Using Different Commercial SNP Panels. J. Equine Vet. Sci. 58, 89–96.

Phocas, F., 2009. Genetic analysis of breeding traits in a Charolais cattle population segregating an inactive myostatin allele. J. Anim. Sci. 87, 1865–1871.

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

Ponzoni, R.W., 1986. Economic Evaluation of Breeding Objectives in Sheep and Goats - Summary and Commentary.Pages 465–469 in 3rd World Congress on Genetics Applied to Livestock Production. Lincoln, Nebraska, USA.

Porto-Neto, L.R., Kijas, J.W. & Reverter, A., 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genet. Sel. Evol. 46 https://doi.org/10.1186/1297-9686-46-22.

Pritchard, J.K. & Przeworski, M., 2001. Linkage Disequilibrium in Humans: Models and Data. Am. J. Hum. Genet. 69, 1–14.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C., 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis.

Purfield, D.C., Evans, R.D. & Berry, D.P., 2019a. Reaffirmation of known major genes and the identification of novel candidate genes associated with carcass-related metrics based on whole genome sequence within a large multi-breed cattle population. BMC Genomics 20, 1–17 https://doi.org/10.1186/s12864-019-6071-9.

Purfield, D.C., Evans, R.D., Carthy, T.R. & Berry, D.P., 2019b. Genomic Regions Associated With Gestation Length Detected Using Whole-Genome Sequence Data Differ Between Dairy and Beef Cattle. Front. Genet. 10, 1068 https://doi.org/10.3389/fgene.2019.01068.

Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. & Simianer, H., 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Anim. Genet. 41, 377–89.

Qwabe, S.O., van Marle-Köster, E., Maiwashe, A. & Muchadeyi, F.C., 2013. Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. S. Afr. J. Anim. Sci. 43.

Randhawa, I.A.S., Khatkar, M.S., Thomson, P.C. & Raadsma, H.W., 2016. A meta-assembly of selection signatures in cattle. PLoS One 11 https://doi.org/10.1371/journal.pone.0153013.

Rasby, R.J., Stalker, A. & Funston, R.N., 2014. Body Condition Scoring Beef Cows: Body Condition Scores Reflect Body Fat. Lincoln.

Regatieri, I.C., Boligon, A.A., Costa, R.B., de Souza, F.R.P., Baldi, F., Takada, L., Venturini, G.C., de Camargo, G.M.F., Fernandes, G.A., Tonhati, H., de Oliveira, H.N. & de Albuquerque, L.G., 2017. Association between single nucleotide polymorphisms and sexual precocity in Nellore heifers. Anim. Reprod. Sci. 177, 88–96.

Ring, S.C., Purfield, D.C., Good, M., Breslin, P., Ryan, E., Blom, A., Evans, R.D., Doherty, M.L., Bradley, D.G. & Berry, D.P., 2019. Variance components for bovine tuberculosis infection and multi-breed genome-wide association analysis using imputed whole genome sequence data. PLoS One 14, 1–24 https://doi.org/10.1371/journal.pone.0212067.

Robinson, G.K., 1991. That BLUP Is a Good Thing: The Estimation of Random Effects. Stat. Sci. 6, 15–51.

Roche, J.R., Friggens, N.C., Kay, J.K., Fisher, M.W., Stafford, K.J. & Berry, D.P., 2009. Body condition score and its association with dairy cow productivity, health and welfare: A review. J. Dairy Sci. 92.

Rondón, L.J., Tiel Groenestege, W.M., Rayssiguier, Y. & Mazur, A., 2008. Relationship between low magnesium status and TRPM6 expression in the kidney and large intestine. Am. J. Physiol. - Regul. Integr. Comp. Physiol. 294, R2001–R2007 https://doi.org/10.1152/ajpregu.00153.2007.

RStudio Team. RStudio, I., 2019. RStudio: Integrated Developement for R. Version 1.2.5033.

Russell, P.J., 2010. iGenetic: a Molecular Approach (B Cummings, Ed.). 3rd ed. Pearson.

Rust, T. & Groeneveld, E., 2001. Variance component estimation on female fertility traits in beef cattle. S. Afr. J. Anim. Sci. 31, 131–141.

SA Stud Book., 2016. SA Stud Book Annual Report. The South African Stud Book and Animal Improvement Association. Bloemfontein.

SA Stud Book., 2019. SA Stud Book Report for Academic use only. The South African Stud Book and Animal Improvement Association. Bloemfontein.

Saatchi, M., Schnabel, R.D., Taylor, J.F. & Garrick, D.J., 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics 15, 1–17 https://doi.org/10.1186/1471-2164-15-442.

Sanarana, Y., Visser, C., Bosman, L., Nephawe, K., Maiwashe, A. & van Marle-Köster, E., 2016. Genetic diversity in South African Nguni cattle ecotypes based on microsatellite markers. Trop. Anim. Health Prod. 48, 379–385 https://doi.org/10.1007/s11250-015-0962-9.

SanGiovanni, J.P. & Lee, P.H., 2013. AMD-Associated Genes Encoding Stress-Activated MAPK Pathway Constituents Are Identified by Interval-Based Enrichment Analysis. PLoS One 8 https://doi.org/10.1371/journal.pone.0071239.

Saowaphak, P., Duangjinda, M., Plaengkaeo, S., Suwannasing, R. & Boonkum, W., 2017. Genetic correlation and genome-wide association study (GWAS) of the length of productive life, days open, and 305-days milk yield in crossbred holstein dairy cattle. Genet. Mol. Res. 16 https://doi.org/10.4238/gmr16029091.

Sargolzaei, M., Chesnais, J.P. & Schenkel, F.S., 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15.

Sartori, R., Bastos, M.R., Baruselli, P.S., Gimenes, L.U., Ereno, R.L. & Barros, C.M., 2010. Physiological differences and implications to reproductive management of Bos taurus and Bos indicus cattle in a tropical environment.

Schaeffer, L.R., 2001. Multiple tarit international bull comparisons. Livest. Prod. Sci. 69, 145–153.

Scheet, P. & Stephens, M., 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. Am. J. Hum. Genet. 78, 629–644.

Scherer, S.W., Soder, S., Duvoisin, R.M., Huizenga, J.J. & Tsui, L.C., 1997. The human metabotropic glutamate receptor 8 (GRM8) gene: A disproportionately large gene located at 7q31.3-q32.1. Genomics 44, 232–236 https://doi.org/10.1006/geno.1997.4842.

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

Scholtz, M.M., Gertenbach, W. & Hallowell, G., 2011. The Nguni Breed of Cattle:Past, Present and Future. Pretoria.

Sejrsen, K., Huber, J.T., Tucker, H.A. & Akers, R.M., 1982. Influence of Nutrition on Mammary Development in Pre- and Postpubertal Heifers. J. Dairy Sci. 65, 793–800 https://doi.org/10.3168/jds.S0022-0302(82)82268-6.

Sejrsen, K. & Purup, S., 1997. Influence of Prepubertal Feeding Level on Milk Yield Potential of Dairy Heifers: A Review. J. Anim. Sci. 75, 828–835 https://doi.org/10.2527/1997.753828x.

Selk, G.E., Wetteman, R.P., Lusby, K.S., Oltjen, J.W., Mobley, S.L., Rasby, R.J. & Garmendia, J.C., 1988. Relationships among weight change, body condition and reproductive performance of range beef cows. J. Anim. Sci. 66, 3153–3159.

Senior, J.R. & Isselbacher, K.J., 1963. Demonstration of an intestinal monoglyceride lipase: an enzyme with a possible role in the intracellular completion of fat digestion.

Sharma, A., Lee, J.S., Dang, C.G., Sudrajad, P., Kim, H.C., Yeon, S.H., Kang, H.S. & Lee, S.H., 2015. Stories and challenges of genome wide association studies in livestock - a review. Asian-Australasian J. Anim. Sci. 28, 1371–1379 https://doi.org/10.5713/ajas.14.0715.

Silva, L., Gasparino, E., Torres Júnior, R., Euclides Filho, K., Silva, L., Alencar, M., Júnior, S., Battistelli, J. & Silva, S., 2015. Repeatability and genotypic correlations of reproductive and productive traits of crossbred beef cattle dams. funpecrp.com.br Genet. Mol. Res. Genet. Mol. Res 14, 5310–5319.

Snelling, W.M., Allan, M.F., Keele, J.W., Kuehn, L.A., McDaneld, T., Smith, T.P.L., Sonstegard, T.S., Thallman, R.M. & Bennett, G.L., 2009. Genome-wide association study of growth in crossbred beef cattle. J. Anim Sci., 2009–2257.

Spencer, C., Su, Z., Donnelly, P. & Marchini, J., 2009. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 5.

Stafuzza, N.B., Costa e Silva, E.V. da, Silva, R.M. de O., Costa Filho, L.C.C. da, Barbosa, F.B., Macedo, G.G., Lobo, R.B. & Baldi, F., 2020. Genome-wide association study for age at puberty in young Nelore bulls. J. Anim. Breed. Genet. 137, 234–244 https://doi.org/10.1111/jbg.12438.

Strandén, I., 2014. RelaX2 program for pedigree analysis, User's guide for version 1.65.

Strandén, I. & Lidauer, M., 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. J. Dairy Sci. 82, 2779–2787 https://doi.org/10.3168/jds.S0022-0302(99)75535-9.

Strandén, I. & Mäntysaari, E.A., 2010. A recipe for multiple trait deregression. Interbull Bull., 21.

Stranger, B., Stahl, E. & Raj, T., 2011. Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. Genetics.

Su, W.-P., Chen, S.-H., Chen, S.-J., Chou, P.-Y., Huang, C.-C. & Chang, N.-S., 2012. WW Domain-Containing Oxidoreductase is a Potential Receptor for Sex Steroid Hormones. Sex Horm. https://doi.org/10.5772/26043.

Sweett, H., Miglior, F., Livernois, A., Fonseca, P., Id-Lahoucine, S., Troya, E., Suárez-Vega, A. & Cánovas, A., 2018. PSXIV-18 Genome-wide association study to identify genomic regions and single nucleotide polymorphisms functionally associated with bull fertility. J. Anim. Sci. 96, 138–139 https://doi.org/10.1093/jas/sky404.303.

Taskinen, M., Lidauer, M.H., Su, G. & Aamand, G.P., 2013. Comparison of Model Reliabilities from Single-Step and Bivariate Blending Methods. Interbull Bull.

Taye, M., Lee, W., Caetano-Anolles, K., Dessie, T., Hanotte, O., Mwai, O.A., Kemp, S., Cho, S., Oh, S.J., Lee, H.K. & Kim, H., 2017. Whole genome detection of signature of positive selection in African cattle reveals selection for thermotolerance. Anim. Sci. J. 88, 1889–1901 https://doi.org/10.1111/asj.12851.

Taylor, J.F., Schnabel, R.D. & Sutovsky, P., 2018. Identification of genomic variants causing sperm abnormalities and reduced male fertility. Anim. Reprod. Sci. 194, 57–62 https://doi.org/10.1016/j.anireprosci.2018.02.007.

Tellam, R., Lemay, D., Van Tassell, C., Lewin, H., Worley, K. & Elsik, C., 2009. Unlocking the bovine genome. BMC Genomics.

Thévenon, S., Dayo, G.K., Sylla, S., Sidibe, I., Berthier, D., Legros, H., Boichard, D., Eggen, A. & Gautier, M., 2007. The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. Anim. Genet. 38, 277–286.

Thomas, D.C., 2006. Are We Ready for Genome-wide Association Studies? Cancer Epidemiol. Biomarkers Prev. 15, 595–598.

Thomasen, J.R., Sorenson, A.C., Lund, M.S. & Guldbrandtsen, B., 2014. Adding cows to the reference population makes a small dairy population competitive. J. Dairy Sci. 97, 5822–5832.

Twomey, A.J., Berry, D.P., Evans, R.D., Doherty, M.L., Graham, D.A. & Purfield, D.C., 2019. Genome-wide association study of endo-parasite phenotypes using imputed whole-genome sequence data in dairy and beef cattle. Genet. Sel. Evol. 51, 1–17 https://doi.org/10.1186/s12711-019-0457-7.

Umans, L., Serneels, L., Overbergh, L., Lorent, K., Van Leuven, F. & Van den Berghe, H., 1995. Targeted inactivation of the mouse α2-macroglobulin gene. J. Biol. Chem. 270, 19778–19785 https://doi.org/10.1074/jbc.270.34.19778.

van der Westhuizen, L., MacNeil, M.D., Scholtz, M.M., Neser, F.W.C., Makgahlela, M.L. & van Wyk, J.B., 2020. Genetic variability and relationships in nine South African cattle breeds using microsatellite markers. Trop. Anim. Health Prod. 52, 177–184 https://doi.org/10.1007/s11250-019-02003-z.

van der Westhuizen, R.R., Schoeman, S.J., Jordaan, G.F. & van Wyk, J.B., 2001a. Heritability estimates derived from threshold analyses for reproduction and stayability traits in a beef cattle herd. S. Afr. J. Anim. Sci. 31, 25–32.

van der Westhuizen, R.R., Schoeman, S.J., Jordaan, G.F. & van Wyk, J.B., 2001b. Genetic parameters for reproductive traits in a beef cattle herd estimated using multitrait analysis. S. Afr. J. Anim. Sci. 31, 41–48.

van der Westhuizen, B., Theron, H., Burger, B. & Steyl, L., 2017. Mediavrystelling: Genomiese Teelwaardes vir Bonsmara.

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 https://doi.org/10.3389/fgene.2018.00331.

van Marle-Köster, E., Visser, C. & Berry, D.P., 2013. A review of genomic selection - Implications for the South African beef and dairy cattle industries. S. Afr. J. Anim. Sci. 43, 1–16.

van Marle, J., 1974. The breeding of beef cattle in South Africa: Past, Present and Future. S. Afr. I. Anim. Sci 4, 297–304.

van Son, M., Enger, E.G., Grove, H., Ros-Freixedes, R., Kent, M.P., Lien, S. & Grindflek, E., 2017. Genome-wide association study confirm major QTL for backfat fatty acid composition on SSC14 in Duroc pigs. BMC Genomics 18, 1–13.

VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci., 4414–4423.

VanRaden, P.M., Null, D.J., Sargolzaei, M., Wiggans, G.R., Tooker, M.E., Cole, J.B., Sonstegard, T.S., Connor, E.E., Winters, M., van Kaam, J.B.C.H.M., Valentini, A., Van Doormaal, B.J., Faust, M.A. & Doak, G.A., 2013. Genomic imputation and evaluation using high-density Holstein genotypes. J. Dairy Sci. 96, 668–678.

VanRaden, P.M., Wiggans, G.R., Tassell, C.P. Van, Sonstegard, T.S. & Schenkel, F., 2009. Benefits from Cooperation in Genomics. Interbull Bull. 39, 67–72.

Viscarra, J.A. & Ortiz, R.M., 2013. Cellular mechanisms regulating fuel metabolism in mammals: Role of adipose tissue and lipids during prolonged food deprivation. Metabolism. 62, 889–897 https://doi.org/10.1016/j.metabol.2012.12.014.

Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J., 2012. Five years of GWAS discovery. Am. J. Hum. Genet. 90, 7–24 https://doi.org/10.1016/j.ajhg.2011.11.029.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. & Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22 https://doi.org/10.1016/j.ajhg.2017.06.005.

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, 12–18 https://doi.org/10.1093/af/vfaa003.

Voorman, A., Lumley, T., McKnight, B. & Rice, K., 2011. Behavior of QQ-Plots and Genomic Control in Studies of Gene-Environment Interaction (S Cherny, Ed.). PLoS One 6, e19416.

Wall, L., Christiansen, T. & Orwant, J., 2000. The Perl programming language. Prentice Hall Software Series.

Wang, W.Y.S., Barratt, B.J., Clayton, D.G. & Todd, J.A., 2005. Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 6, 109–118.

Wang, Z., Huang, S., Fan, L., Zhang, T., Wang, L. & Wang, Y., 2018. Adaptive and dynamic RFID tag anti-collision based on secant iteration. PLoS One 13, e0206741.

Wang, K., Kang, Z., Jiang, E., Yan, H., Zhu, H., Liu, J., Qu, L., Lan, X. & Pan, C., 2020. Genetic effects of DSCAML1 identified in genome-wide association study revealing strong associations with litter size and semen quality in goat (Capra hircus). Theriogenology 146, 20–25 https://doi.org/10.1016/j.theriogenology.2020.01.079.

Wang, K.S., Liu, X., Zhang, Q., Pan, Y., Aragam, N. & Zeng, M., 2011. A meta-analysis of two genome-wide association studies identifies 3 new loci for alcohol dependence. J. Psychiatr. Res. 45, 1419–1425 https://doi.org/10.1016/j.jpsychires.2011.06.005.

Wang, H., Woodward, B., Bauck, S. & Rekaya, R., 2012. Imputation of missing SNP genotypes using low density panels. Livest. Sci. 146, 80–83.

Wathes, D.C., Pollott, G.E., Johnson, K.F., Richardson, H. & Cooke, J.S., 2014. Heifer fertility and carry over consequences for life time production in dairy and beef cattle. Animal 8, 91–104 https://doi.org/10.1017/S1751731114000755.

Webb, E.C., Visagie, P.C., van der Westhuizen, J. & Snyman, H.A., 2017. Influence of bioregion and environmental factors on the growth, size and reproduction of Bonsmara cows. South African J. Anim. Sci. 47, 542–552 https://doi.org/10.4314/sajas.v47i4.13.

Weigel, K.A., Van Tassell, C.P., O'Connell, J.R., VanRaden, P.M. & Wiggans, G.R., 2010. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using

reference panels and population-based imputation algorithms. J. Dairy Sci. 93, 2229–2238.

Weikard, R., Goldammer, T., Brunner, R.M. & Kuehn, C., 2012. Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. Physiol. Genomics 44, 728–739 https://doi.org/10.1152/physiolgenomics.00007.2012.

Whalen, A., Gorjanc, G., Ros-Freixedes, R. & Hickey, J.M., 2017. Assessment of the performance of different hidden Markov models for imputation in animal breeding. bioRxiv, 227157.

Wu, X., Ivanchenko, M. V., Al Jandal, H., Cicconet, M., Indzhykulian, A.A. & Corey, D.P., 2019. PKHD1L1 is a coat protein of hair-cell stereocilia and is required for normal hearing. Nat. Commun. 10, 1–15 https://doi.org/10.1038/s41467-019-11712-w.

www.teagasc.ie. Breeding and Genetics. Agric. Food Dev. Auth.

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

Yoshida, G.M., Lhorente, J.P., Carvalheiro, R. & Yáñez, J.M., 2017. Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (Salmo salar L.). Anim. Genet. 48, 698–703.

Zalesky, D.D., Day, M.L., Garcia-Winder, M., Imakawa, K., Kittok, R.J., D'Occhio, M.J. & Kinder, J.E., 1984. Influence of exposure to bulls on resumption of estrous cycles following parturition in beef cows. J. Anim. Sci. 59, 1135–1139 https://doi.org/10.2527/jas1984.5951135x.

Zhang, K., Chu, K., Wu, X., Gao, H., Wang, J., Yuan, Y.C., Loera, S., Ho, K., Wang, Y., Chow, W., Un, F., Chu, P. & Yen, Y., 2013a. Amplification of FRS2 and activation of FGFR/FRS2 signaling pathway in high-grade liposarcoma. Cancer Res. 73, 1298–1307 https://doi.org/10.1158/0008-5472.CAN-12-2086.

Zhang, Z. & Druet, T., 2010. Marker imputation with low-density marker panels in Dutch Holstein cattle. J. Dairy Sci. 93, 5487–5494.

Zhang, J., Zhang, J., Shao, Z., Yu, L., Zhang, J., Zhou, Y., Wu, Y.-J., Ma, L., Wang, R.-J., Huang, S.-Q., Gao, R.-R., Liu, L.-H., Shao, Z.-H., Shi, H.-J., Cheng, L.-M. & Yu, L., 2013b. Novel Cerebellum-Enriched miR-592 May Play a Role in Neural Progenitor Cell Differentiation and Neuronal Maturation Through Regulating Lrrc4c and Nfasc in Rat . Curr. Mol. Med. 13, 1432–1445 https://doi.org/10.2174/15665240113139990072.

Zhao, C., Takita, J., Tanaka, Y., Setou, M., Nakagawa, T., Takeda, S., Yang, H.W., Terada, S., Nakata, T., Takei, Y., Saito, M., Tsuji, S., Hayashi, Y. & Hirokawa, N., 2001. Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1Bβ. Cell 105, 587–597 https://doi.org/10.1016/S0092-8674(01)00363-4.

Zhou, X. & Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–4.

Zhu, Y.H., Li, J.B., Wu, R.Y., Yu, Y., Li, X., Li, Z.L., Zhang, H.L., Feng, G.K., Deng, R. & Zhu, X.F., 2020. Clinical significance and function of RDH16 as a tumor-suppressing gene in hepatocellular carcinoma. Hepatol. Res. 50, 110–120 https://doi.org/10.1111/hepr.13432.

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J.A. & Salzberg, S.L., 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10.

Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S., 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. U.S.A. 109, 1193–1198.

Zwane, A.A., Maiwashe, A., Makgahlela, M.L., Zwane, A.A., van Marle-Köster, E., Choudhury, A., Taylor, J.F. & Maiwashe, A., 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.

Addenda

Addendum A: Deregression of estimated breeding values (EBVs)

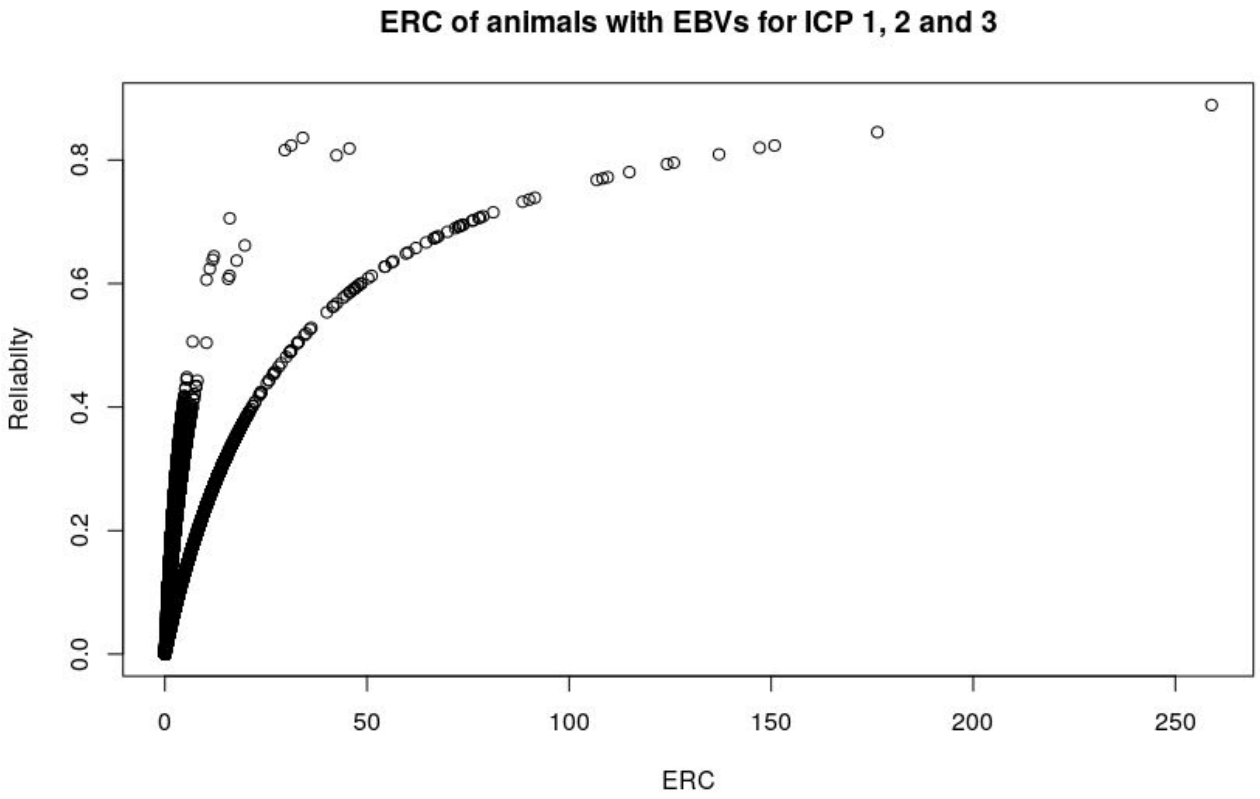


Figure A1 Plot of effective record contributions (ERC) against reliabilities for inter calving period

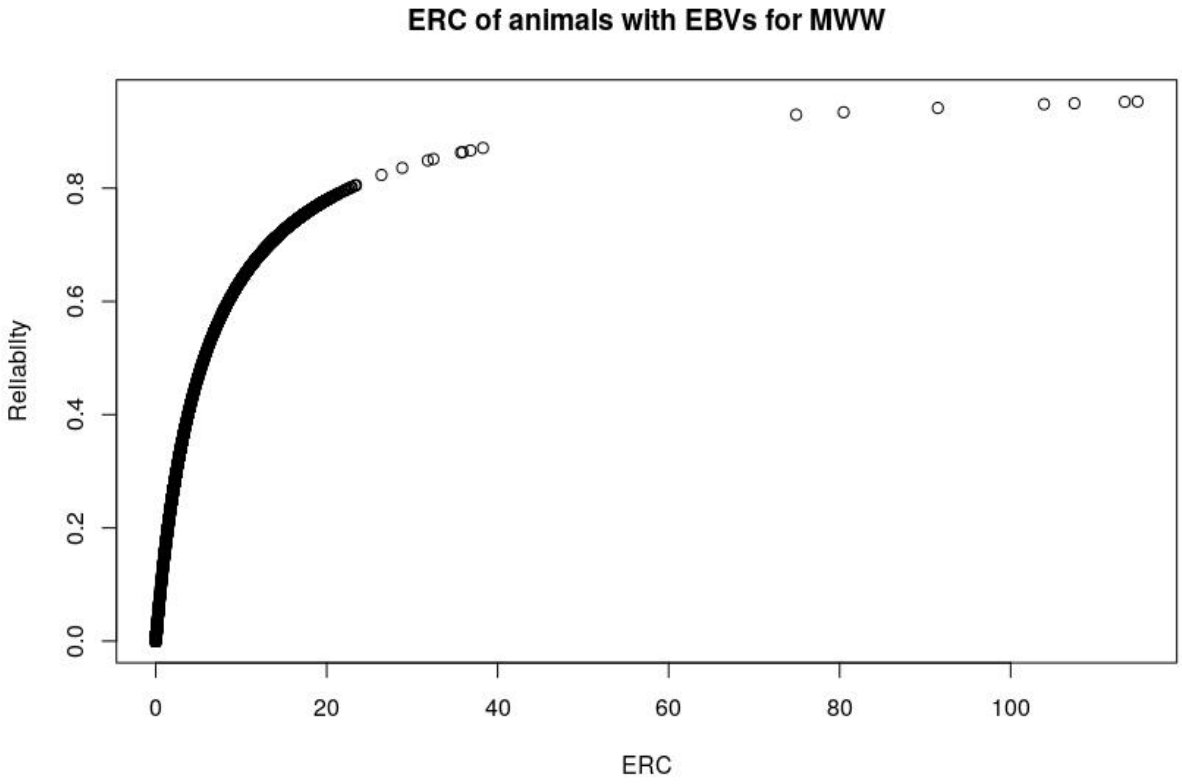


Figure A2 Plot of effective record contributions (ERCs) against reliabilities for weaning weight maternal

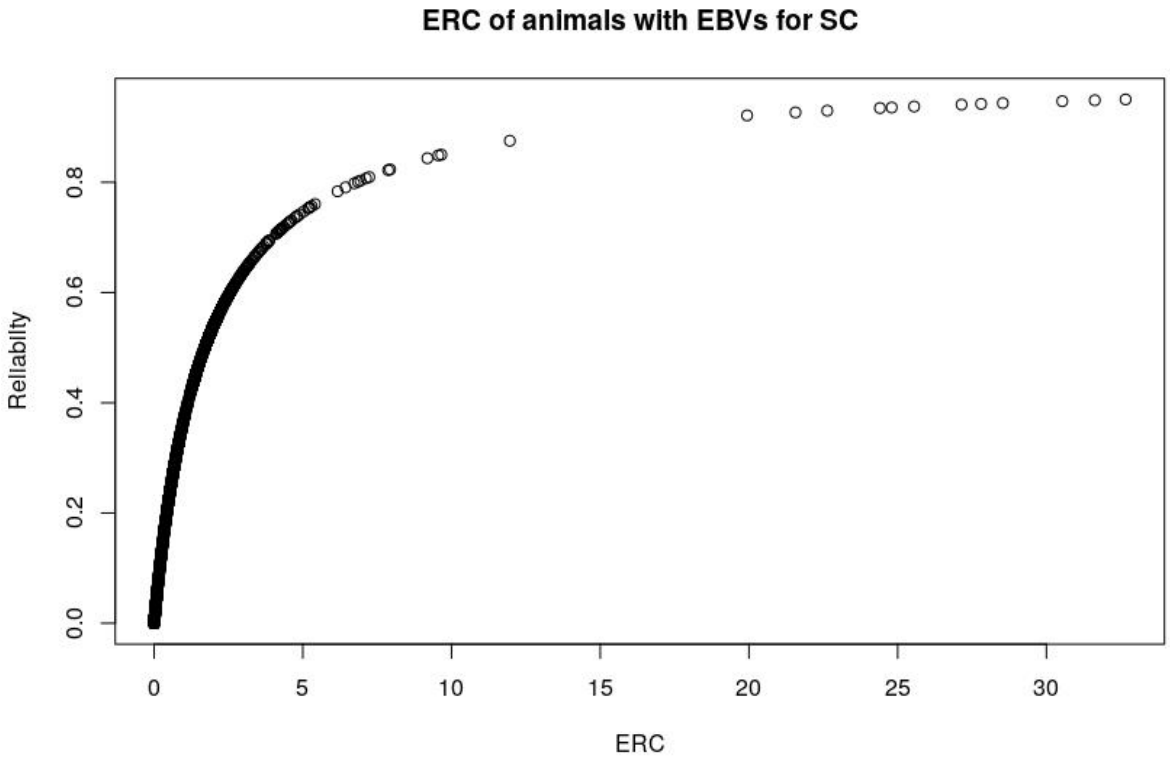


Figure A3 Plot of effective record contributions (ERCs) against reliabilities for scrotal circumference

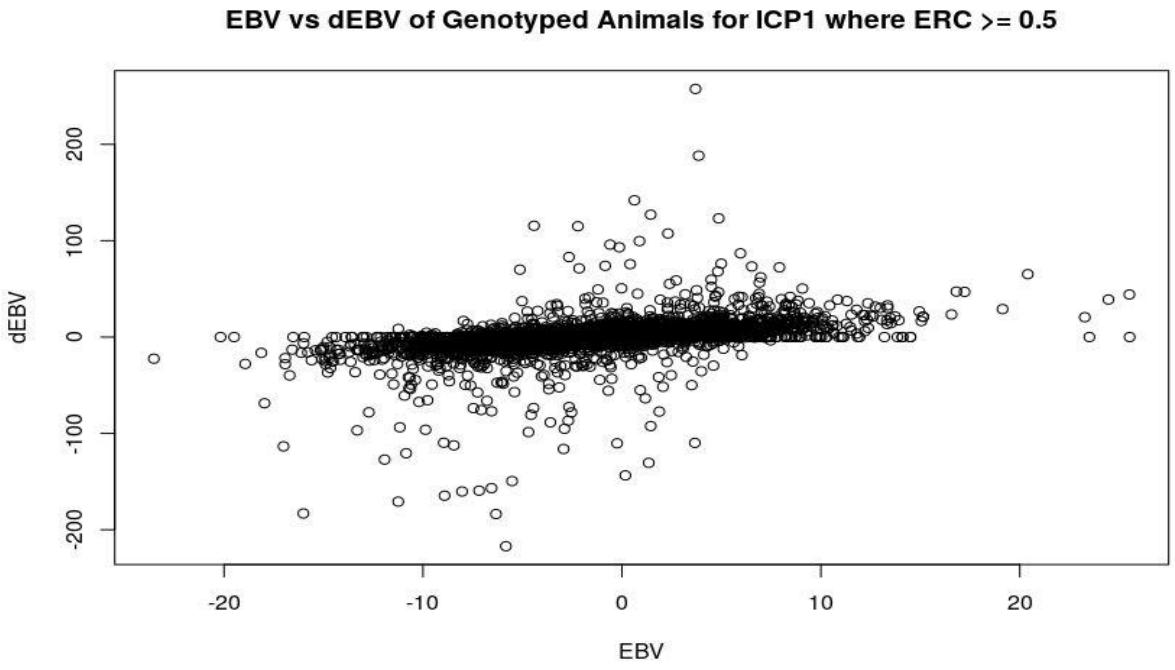


Figure A4 Plot of estimated breeding values against deregressed estimated breeding values for first inter calving period

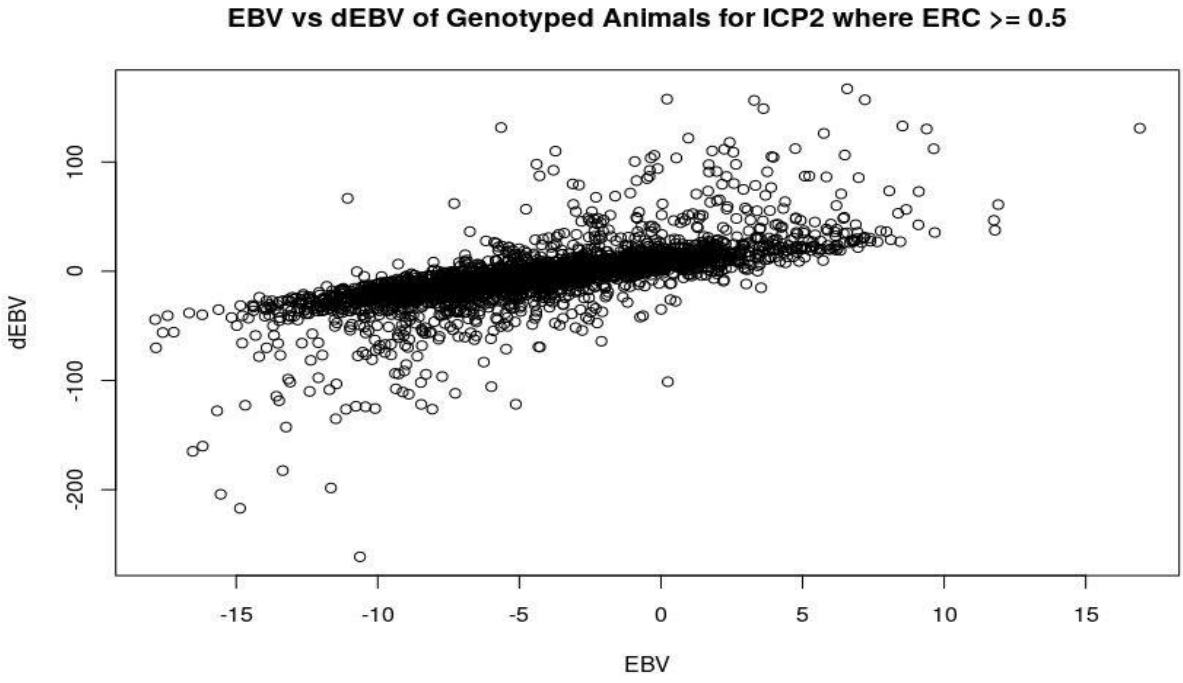


Figure A5 Plot of estimated breeding values against deregressed estimated breeding values for second inter calving period

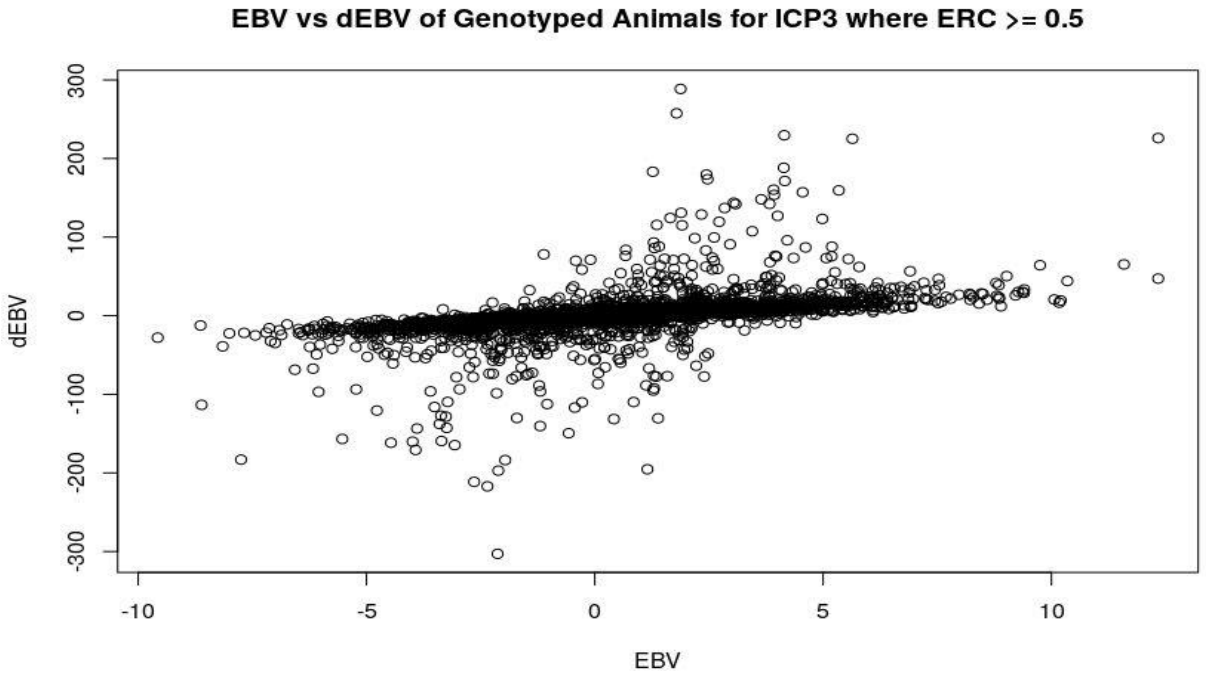


Figure A6 Plot of estimated breeding values against deregressed estimated breeding values for third inter calving period

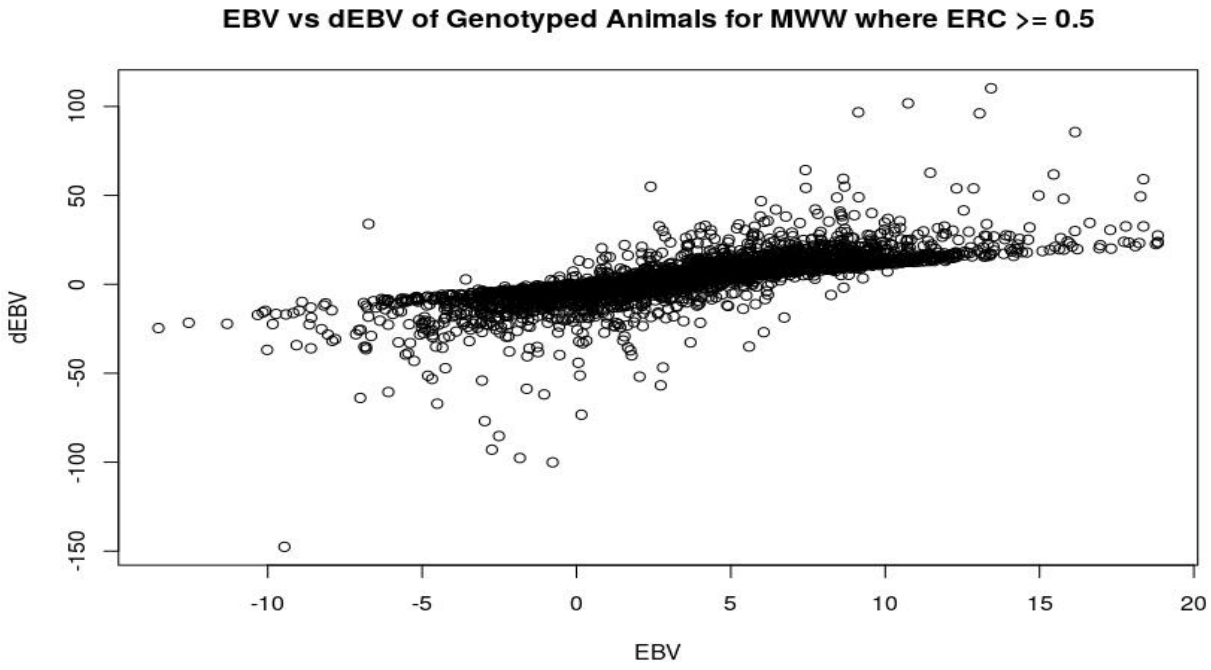


Figure A7 Plot of estimated breeding values against deregressed estimated breeding values for weaning weight maternal

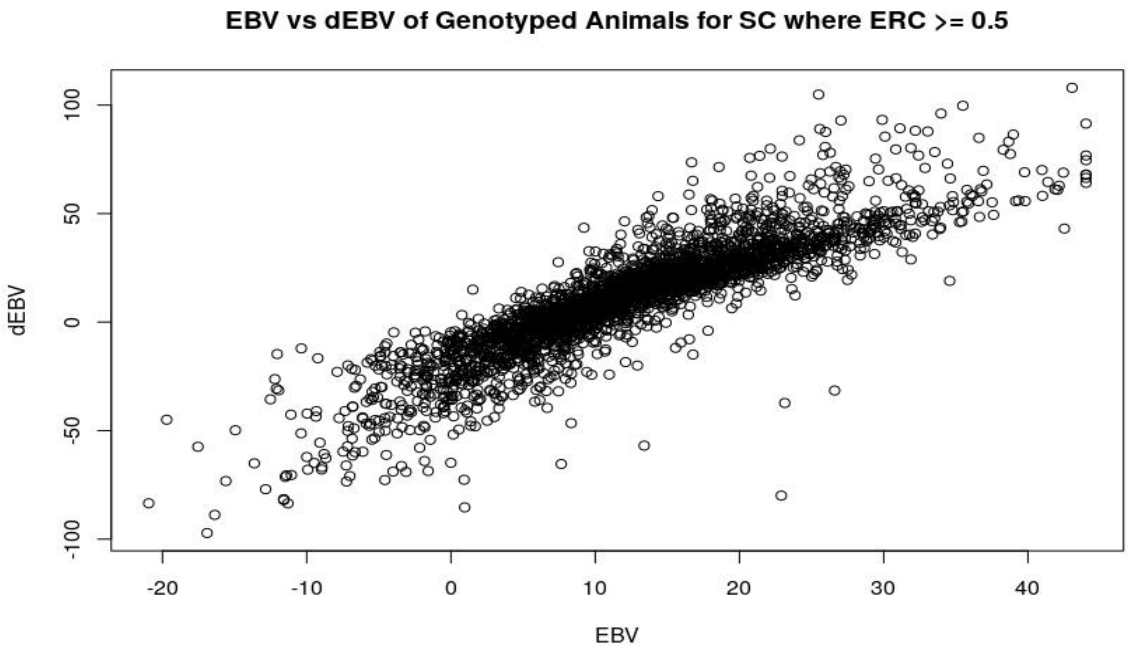


Figure A8 Plot of estimated breeding values against deregressed estimated breeding values for scrotal circumference

Addendum B: Principal component analysis (PCA)

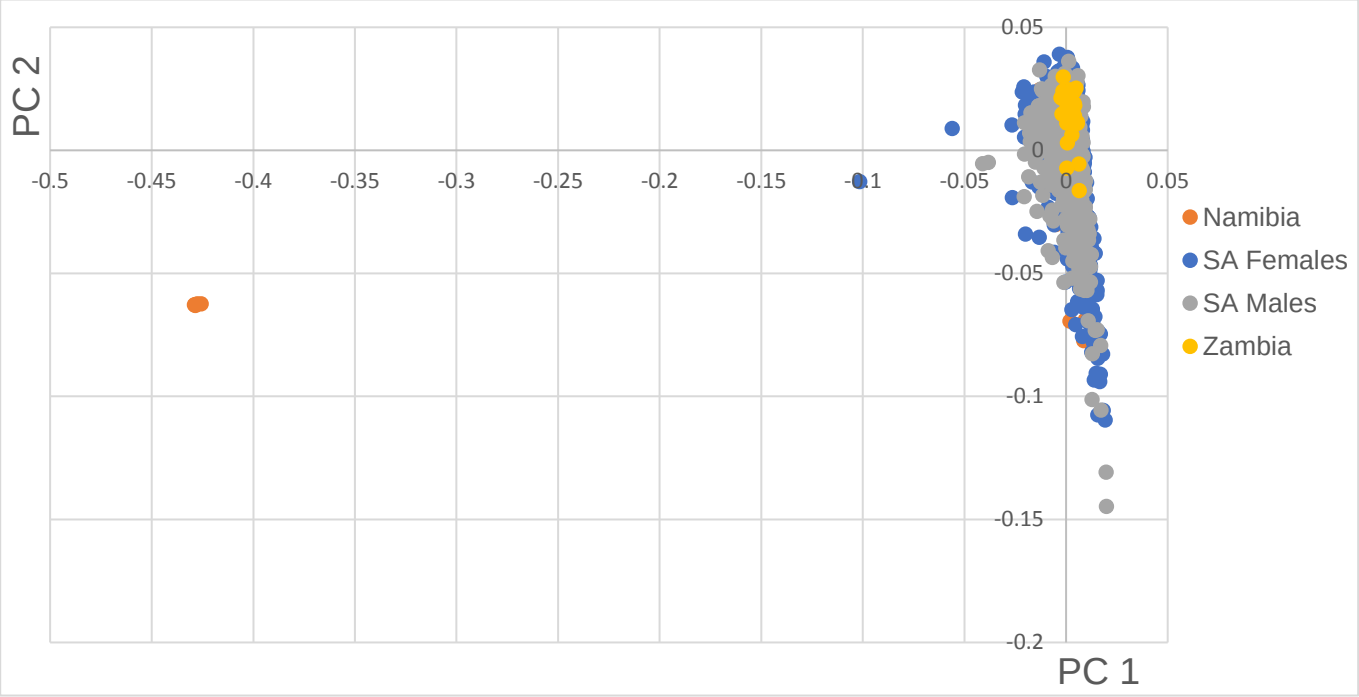


Figure B1 Genetic relationships among 1 937 Bonsmara animals and 128 793 SNPs genotyped on the GGP 150K HD array for the first and second principal components (PC 1 and PC 2)

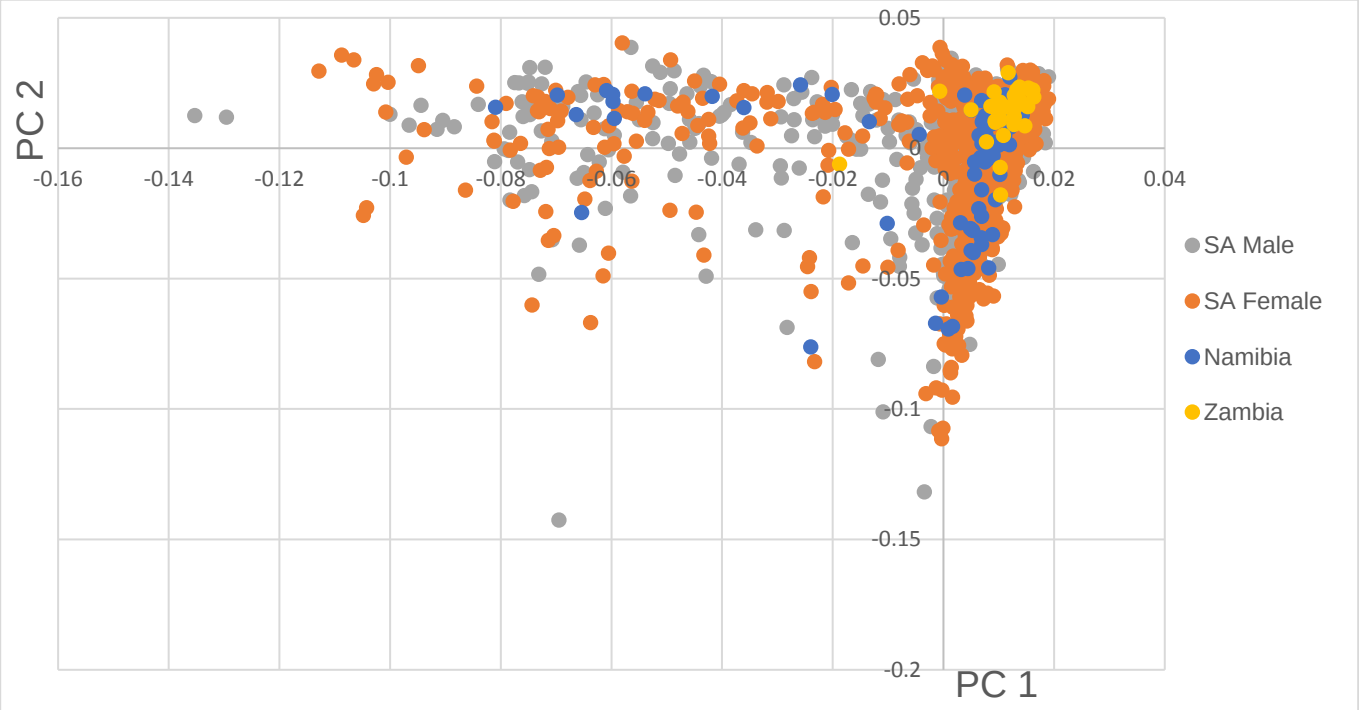


Figure B2 Genetic relationships between 1932 Bonsmara animals and 128 793 SNPs genotyped on the GGP 150K HD array for the first and second principal components (PC 1 and PC 2)

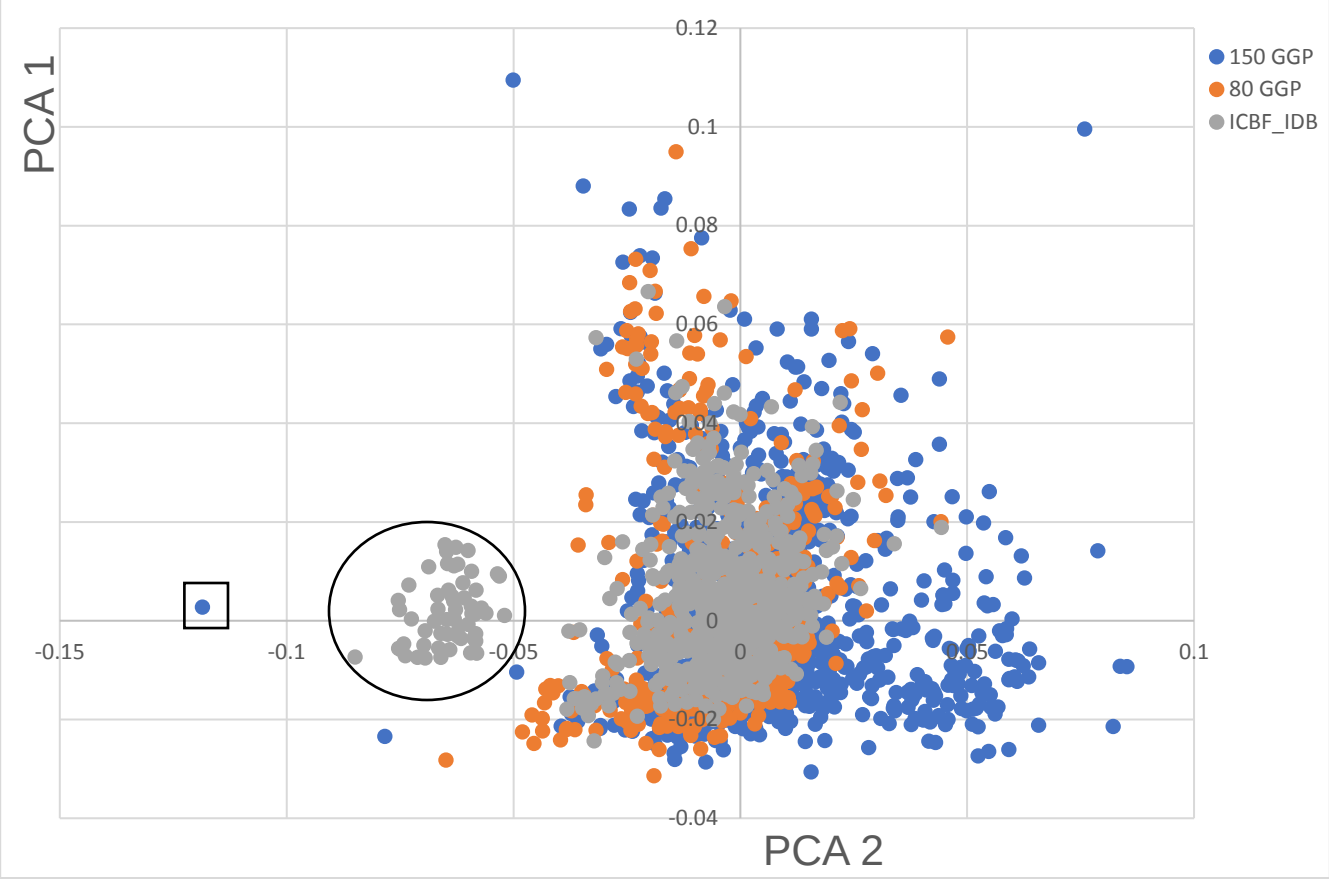


Figure B3 Genetic relationships between 3 253 Bonsmara animals and 128 793 SNPs genotyped on the GGP 150K HD array for the first and second principal components (PC 1 and PC 2) post-imputation