Genomic Diversity and Population Structure of the South African Hereford Population

Shenrae Carter

Submitted in partial fulfilment of the requirement for the degree

MSc Animal Breeding and Genetics

In the

Department of Animal and Wildlife Sciences

Faculty of Natural and Agricultural Sciences

University of Pretoria

October 2018

Declaration

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

Signature.....

Date.....

Acknowledgements

I would like to thank my parents, Wayne and Ahdel Carter, for providing me with the opportunity to partake of both under-graduate and post-graduate studies, which has afforded me the opportunity to pursue my desired future career. I would also like to thank my family for their support and guidance throughout this process.

I would like to thank my supervisors Prof Esté van Marle-Köster and Dr Carina Visser, for providing me with the opportunity to complete my MSc degree and many thanks for the guidance, knowledge and patience throughout this process.

I would like to thank my fellow students, Simon Lashmar and Rulien Grobler at the Animal Breeding and Genetics laboratory for their assistance and guidance throughout this process.

I would like to thank the Beef Genomics Programme and UNISTEL for providing me with the data to carry out my MSc degree.

Abstract

In this study both microsatellite and SNP data was evaluated, with the aim to quantify genetic variation, inbreeding and population structure at a genomic level to provide reference data for the South African (SA) Hereford breed. A total of 1799 microsatellite profiles were obtained from UNISTEL generated with 11 ISAG recommended microsatellite markers. Two data sets were evaluated, a complete data set of 1799 samples and a selected, representative subset of 500 samples. The mean number of microsatellite alleles was 10 and 8.5 for the complete and subset data sets, respectively across the 11 loci analysed. Heterozygosity and F_{IS} estimates across the loci were (0.663± 0.003) and (-0.014) for the complete data set and (0.661± 0.006) and (-0.009) for the subset data set, respectively, indicating relatively high genetic diversity present within the populations and low inbreeding. Population structure analysis assigned the animals to six inferred clusters, with three herds forming distinct herd clusters and the remaining herds assigning evenly across the inferred clusters. The SNP data set contained 184 SA and 316 Irish Herefords, which were analysed separately and then as a merged data set. The two data sets were subjected to quality control, utilising the following parameters, namely a sample call rate of <98%, a SNP call rate of <98%, MAF of <5% and deviation from HWE ($p<0.001$). The observed H_O for the SA and Irish Hereford in this study were similar, while inbreeding differed slightly, but was low in both cases (SA: -0.002± 0.051; Irish: -0.007± 0.034). Principal component analysis for the SA Hereford observed three genetically distinct herds, with the remainder of the herds sharing a close genetic relationship. The ADMIXTURE analysis assigned the SA Hereford to seven inferred clusters, supporting the PCA analysis. The Irish Hereford formed one distinct cluster in the PCA results, supported by the ADMIXTURE analysis, indicating one homogenous population. The merged data set population structure analysis, allocated the SA and Irish Herefords into two diverse clusters, with slight overlap, indicating the two populations are genetically distinct. The microsatellite and SNP data supported each other, indicating the SA Hereford are sufficiently genetically diverse for improvement and development of the breed. The results provide reference data on genomic diversity in the SA Hereford for application in the development of the current training population for genomic selection.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADG	Average daily gain
AFC	Age at first calving
AHA	American Hereford Association
AMOVA	Analysis of molecular variance
BGP	Beef Genomics Programme
Bp	Base pairs
CD	Calving date
CI	Calving interval
DMI	Dry matter intake
EBV	Estimated breeding value
FCR	Feed conversion ratio
FE	Feed efficiency
F_{IS}	Inbreeding coefficient of an individual relative to the subpopulation
F_{ST}	Fixation index
GDP	Gross domestic profit
GEBV	Genomic estimated breeding value
GS	Genomic selection
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HWE	Hardy-Weinberg Equilibrium
ISAG	International Society of Animal Genetics
K	Number of assumed populations
Kb	Kilo basepairs
LD	Linkage disequilibrium

MAF	Minor allele frequency
MAS	Marker assisted selection
MCMC	Markov Chain Monte Carlo
Ne	Effective population
PCA	Principle component analysis
PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
QC	Quality control
QTL	Quantitative trait loci
RFI	Residual feed intake
SA	South Africa
SNP	Single nucleotide polymorphism
UK	United Kingdom
USA	United States of America
WW	Weaning weight
YW	Yearling weight

CHAPTER 1

INTRODUCTION

The origin of the Hereford breed can be traced back to the western border of Herefordshire, in central England where Hereford cattle were present during the 1700's (Köster, 1992; Blott *et al.,* 1998; Leesburg, 2012). In the "History of Hereford Cattle", by MacDonald and Sinclair in 1909, some of the first records of the original Hereford were described. William Marshall was one of the first people who documented cattle breeds in Herefordshire with special reference to the Hereford breed (MacDonald & Sinclair, 1909). MacDonald and Sinclair (1909) also made special reference to Benjamin Tomkins, a renowned breeder, who implemented breeding strategies as early as 1769 which shaped the origins of the Hereford known today, focussing selection on animals that matured at an early age and fattened readily (MacDonald & Sinclair, 1909; Köster, 1992; Leesburg, 2012). Hereford cattle were represented by a solid colouring, varying from light to dark on the body with their distinguishing and characteristic white faces that have defined the original and modern breed of today (MacDonald & Sinclair, 1909; Köster, 1992). The modern Hereford breed after approximately 150 years of selection is known for its early maturation, high rate of gain in feedlot systems, good milk production and good temperament (American Hereford Association, 2016).

The production characteristics of the Hereford induced it's exportation around the world and as a result Hereford cattle are found on five continents, with more than five million pedigreed Herefords in over fifty countries (The Hereford Cattle Society UK, 2013). Hereford cattle are well distributed throughout the United Kingdom (UK), United States of America (USA), Canada, Uruguay, Argentina, New Zealand, Australia, and also growing in numbers in countries such as Kazakhstan, with populations exceeding 20 000 registered Hereford (Hereford Breeder.net, 2017).

The main documented exportations from the UK of Hereford, relevant to this study, include exports to Ireland, USA and South Africa (SA) in 1775; 1817 and 1892, respectively (MacDonald & Sinclair, 1909). The Hereford was favoured for its even temperament and adaptability and often imported to upgrade production performance and quality of local breeds (Grundy, 2002; Irish Hereford Breed Society, 2017). These favourable characteristics motivated the importation of Hereford into SA, in an attempt to improve the meat characteristics of the indigenous Afrikaner cattle. In addition the Hereford breed was also used as one of the base breeds for development of the composite Bonsmara beef cattle breed (Köster, 1992; Zwane *et al.,* 2016).

The SA Hereford Breeder's Society was established during 1917 and over a period of 100 years the breeders have imported semen from various countries, mostly from the UK, USA and Canada (Leesburg, 2012). In comparison to other countries in the world, for example the USA at 350 000 and Ireland at 21 500 registered Hereford cattle, the SA population is small with only 6 267 head of registered Hereford animals (SA Studbook, 2016). Registered Hereford cattle in SA are also

relatively small in number in comparison to other prominent breeds in South Africa, for example the Bonsmara, Beefmaster and Boran with 118 758, 47 517 and 36 048 registered head of cattle, respectively (SA Studbook, 2016). South Africa has a dynamic beef cattle industry with more than 28 different beef breeds delivering genetic material to the commercial market (SA Studbook, 2016). These breeds participate in official performance testing schemes, which is the systematic measurement of individual animal performance to provide reliable data for genetic evaluation, which is an essential step for genetic improvement of a breed (Bergh, 2008). The national beef cattle performance and progeny testing scheme in SA was established in 1959, with the aim of providing performance information on economically important traits to improve the efficiency of beef production (Mokoena *et al.,* 1999). The SA Hereford was one of the first breeds to participate in national performance testing since the establishment of the scheme in 1959 with a current participation of 75% in the Logix system scheme (SA Studbook, 2016).

Performance recording schemes provide a quantitative evaluation of genetic merit in the form of a single numerical value, known as an estimated breeding value (EBV) (Garrick & Golden, 2008). The use of EBV's, combined with that of molecular technologies has allowed for the identification of genetically superior individuals from a younger age and for the prediction of performance, increasing the rate of genetic improvement through more accurate and focussed selection pressure (Garrick & Golden, 2008; Bullock & Pollak, 2009). The SA Hereford has been applying EBV's since 1994 and genetic trends, for most traits, are favourable. In order to ensure further genetic progress, the seed stock farmers need to use all possible tools available that include animal recording, application of estimated breeding values and more recently genomic estimated breeding values (GEBV) (van Marle-Köster *et al.,* 2013).

The sequencing of the bovine genome led to the discovery of abundant bi-allelic single nucleotide polymorphisms (SNP), utilised to characterise the cattle genome (Williams *et al.,* 2009; Garrick, 2011). The sequenced genome was constructed from a single inbred female Hereford, to which additional sequence information was incorporated from six other cattle breeds, identifying putative SNPs (Tellam *et al.,* 2009). SNPs arrays have been utilised to study genetic diversity, inbreeding and population structure as early as 2008 (Singh et al., 2014). Some examples of cattle specific genetic diversity studies using SNPs include Makina *et al.,* 2014, Cañas-álvarez *et al.,* 2015 and Kasarda *et al.,* 2016, among others.

SNP discovery has opened up opportunity for commercialisation of genomic selection (GS). GS is based upon the computation of individual genetic merit using statistical analyses of prediction equations, derived from data of phenotyped and genotyped ancestral reference populations (Van Eenennaam *et al.,* 2014). GS holds benefits for estimating the genomic breeding value for low heritability traits; quantifying familial relationships and increasing the rate of genetic gain by shortening the generation (Goddard & Hayes, 2007; Miller, 2010; Henryon *et al.,* 2014).

Implementation of genomic selection is based on two prerequisites, namely the formation of a training population and the construction of a prediction equation (Goddard & Hayes, 2007). A reference population consist of animals phenotyped and genotyped, for economically important traits, for no less than 1000 high impact animals with high accuracy EBV's (Goddard & Hayes, 2007; van

Marle-Köster *et al.,* 2013; Theron *et al.,* 2014). To date a number of countries such as the USA, Canada and Ireland have successfully implemented genomic selection in beef and dairy cattle (Garrick, 2011; Berry *et al.,* 2016). The American Hereford Association (AHA), has established a reference population of approximately 2000 bulls, based on a 50K SNP chip in 2012 (Garrick , 2011; Berry *et al.,* 2016). The AHA has been working on the implementation of routine evaluations, including 20 000 - 50 000 genotyped Herefords (Garrick*,* 2011; Berry *et al.,* 2016).

In order for the South African Hereford to remain competitive, the breed has become part of the Beef Genomics Program (BGP) where 12 SA beef breeds are taking part with the primary aim to establish reference populations and engage in genomic selection for their breeds. The SA Hereford has strong genetic linkage with countries such as USA and Canada, due to semen imports, which can be exploited in building the reference population. The first step towards establishing a reference population is to gain an insight into the genetic variation of the SA Hereford with regard to diversity, inbreeding and population structure.

Aim of the Study

The aim of the study was to quantify genetic variation, inbreeding and population structure at the genomic level to provide reference data for the SA Hereford breed for potential application in genomic selection.

Objectives

1. Estimate heterozygosity, inbreeding and population structure using microsatellite data.
2. Estimate population structure and genomic diversity using genotypic data.
3. Compare genomic data from SA Hereford to a data set of Irish genomic data.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The methods implemented for livestock production maximise economic output but maximising production efficiency. Production efficiency can be improved through increasing the rate of genetic gain and through selection of genetically superior animals (Hill, 2016). The sequencing of the bovine genome, followed by the development and construction of extensive SNP panels, has provided the opportunity to implement tools such as genomic selection for improving production efficiency (Miller, 2010; Garrick, 2011). SNP panels can be used for quantification and characterisation of breed or population diversity, utilising the available genetic resources (Miller, 2010; Gamaniel & Gwaza, 2017).

Prior to these developments, genetic improvement of beef cattle has been limited to traits that were easy to measure at a young age, such as growth traits, with minimal improvement seen in traits related to reproduction or feed efficiency and tenderness (Garrick, 2011; Garrick & Saatchi, 2011). This chapter will provide a brief history of the Hereford breed and genetic improvement of beef cattle with special reference to the Hereford breed. A discussion will follow on genomic diversity based on microsatellite and SNP markers and the role thereof in genomic evaluations and their applications.

2.2 Development of the modern Hereford

The Hereford known today developed over a period of three centuries, with the original Hereford being a small frame, early maturing animal that readily fattened and naturally adapted to the climate in England (Leesburg, 2012). Hereford cattle were first imported into SA in 1892, in an attempt to upgrade the meat characteristics of the indigenous Afrikaner cattle (Köster, 1992; Hannote *et al.,* 2002; Leesburg *et al.,* 2013). The imported cattle were long haired and woolly, which left it poorly adapted to the sub-tropical climate, thus selection was focussed on smoother coats, thicker hides and pigmentation around the eyes. It took approximately 15 years to adapt the breed to the extreme climate and the prevalence of ticks (Köster, 1992; Leesburg *et al.,* 2013).

The modern South African Hereford as shown in Figure 2.1 has a solid red body, with white face, crest, dewlap and underbelly. Conformation characteristics vary from animal to animal, however they are generally well developed in the back, loin and hind-quarters (Franke *et al.,* 1975; The Hereford Cattle Society UK, 2016). The selection of the modern South African Hereford dates back to the 1940's at the Mara Research Station where Herefords were used in the development of the composite Bonsmara breed. Herefords were bred together with indigenous Afrikaner and Shorthorn breeds for the development of the Bonsmara (Maiwashe *et al.,* 2002; Leesburg *et al.,* 2013).

Figure 2.1 Hereford cow (Photo from http://www.hereford.co.za/, 2017)

In selection of Hereford cattle during the late 1950's and early 1970's in the USA two traits played a major role in selection, namely the selection for polled Hereford and selection for removal of the dwarf gene (Jones & Jolly, 1982; Blott et al., 1998; Leesburg, 2012). The original Hereford was horned but polled Hereford became more prominent due to the efforts of Warren Gammon, a Hereford breeder in USA in the 1890's, who had a preference for the polled phenotype (American Hereford Association, 2017). The polled Hereford was bred using introgression of the polled gene from polled animals (Schafberg & Swalve, 2015).). The polled and horned lines were separated through specific selection focus by individual breeders and breed associations with a specific preference for either polled or horned Hereford, with their registration in separate herd books (Blott *et al.*, 1998; Schafberg & Swalve, 2015).

The line 1 Hereford cattle herd were established by the Bureau of Animal Industry, (USDA) and the herd was initiated by the first purchase of Hereford in 1924 as part of a beef cattle experiment that is still running today and have been line-bred since 1934 at Fort Keogh, USA (Leesburg, 2012). The Line 1 Hereford played an important role in the elimination of the dwarf gene in the Hereford. Prior to 1940's, dwarfism was not considered a problem within the breed, however by the 1950-1960's the presence of the dwarf gene was an important limitation for the further development due to the mature weight of Hereford being approximately half that of other beef cattle (Jones & Jolly, 1982; Leesburg, 2012). The Line 1 Hereford did not carry the recessive allele for dwarfism and were thus utilised to eliminate the recessive allele, coupled with selection for weight and frame size, which focussed on the improvement of growth traits and maternal characteristics in 1977 (Köster, 1992; Jones & Jolly, 1982; Leesburg, 2012).

Line 1 Hereford contributed significantly to the USA Hereford populations with sales of cattle from Fort Keogh distributed to 34 states within the USA, with a significant proportion of the American Hereford sharing some lineage with the Line 1 Hereford (Leesburg, 2012; Leesburg *et al.,* 2013). Hereford semen and embryos are continually exported from the USA into South Africa, with South African Hereford estimated to have a 0.38 probability of sharing ancestry with the Line 1 Hereford (Leesburg *et al.,* 2013). The production characteristics of the Hereford, a lean, fast growing animal makes it a popular choice for the South African feedlot industry (SA Hereford, 2017; Hereford Namibia, 2017). The modern Hereford is known for efficient feed conversion ratio (FCR) and good

temperament, for ease of handling assisted, and is well adapted to South African feedlot systems and finishing on natural grazing (SA Hereford, 2017; Hereford Namibia, 2017).

2.3 Genetic improvement of beef cattle with special reference to Hereford cattle

The SA cattle industry is comprised of many breeds including indigenous, composite and of exotic origin (ARC Annual Beef Bulletin, 2016; Abin *et al.,* 2016). The beef cattle sector accounting for 80% and dairy cattle sector accounts for 20% of the total 13 million head of cattle in SA (ARC Annual Beef Bulletin, 2016; Abin *et al.,* 2016; DAFF, 2018). Popular beef breeds in SA include Nguni, Brahman, Bonsmara, Afrikaner, Simmentaler, Angus and Drakensberger (Scholtz *et al.,* 2008).

The beef production industry is comprised of commercial, seedstock and small holder producers (Grobler, 2016). Feedlot production accounts for 75% of beef production in SA (Strydom, 2008; DAFF, 2015; ARC Annual Beef Bulletin, 2016). The Sanga, British and European breeds dominate the feedlot industry with 82% of the cattle originating from these breeds, with the Bonsmara, Hereford and Simmentaler breeds being among the most popular (Scholtz *et al.,* 2008; DAFF, 2015).

In the past decade production of beef increased by 46% and consumption increased by 35.7%, with approximately 8.4 million tons of beef produced (DAFF, 2015). The beef sector is vertically integrated with many feedlots having their own abattoirs or selling directly to retailers or consumers (DAFF, 2015). The SA beef value chain can be seen below in Figure 2.2.

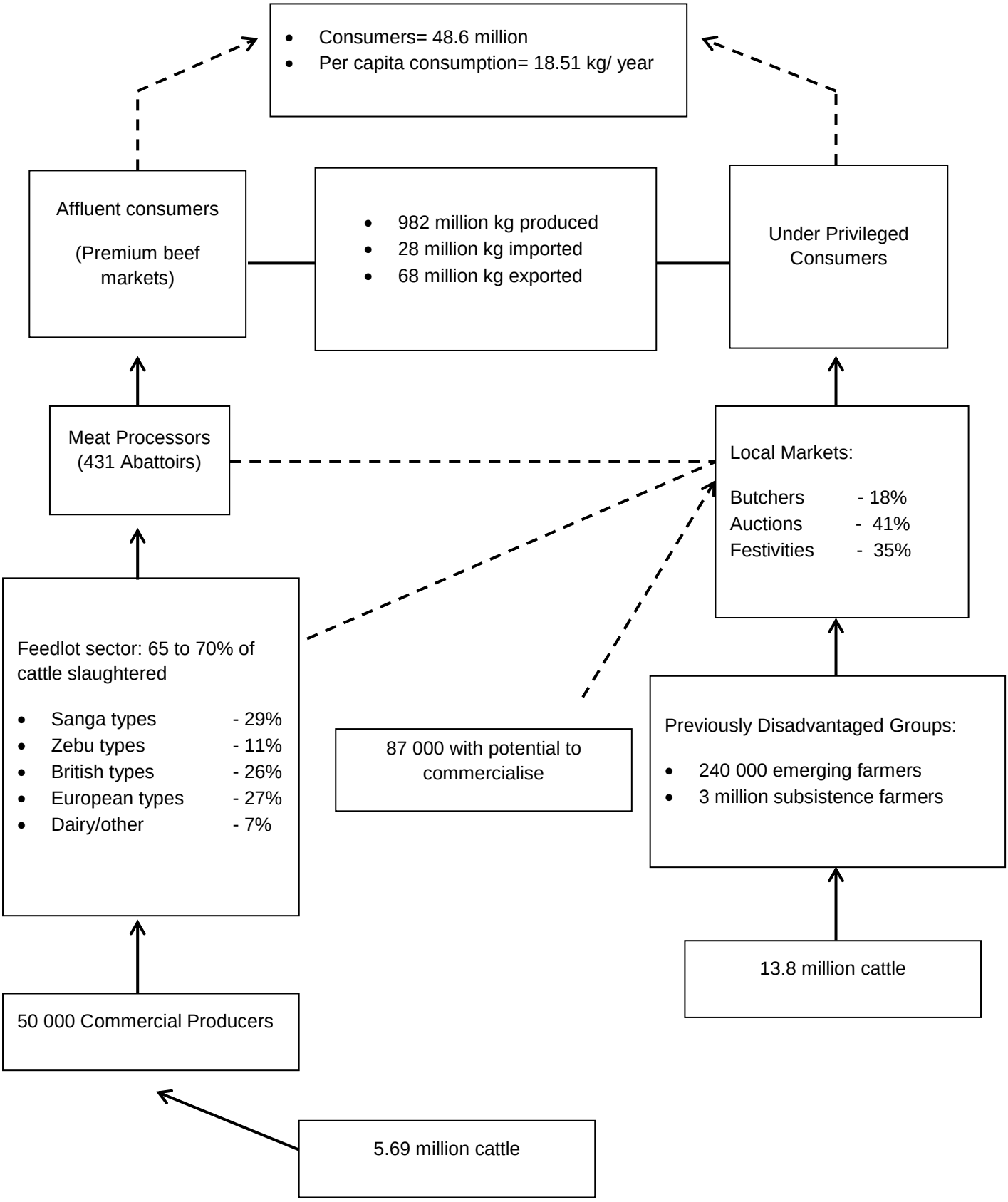

Figure 2.2 South African Beef Value Chain (Source: DAFF, 2015)

The beef enterprises have been driven to select for traits that influence market weight and carcass quality, thus the value of reproductive traits has often been underestimated (Glaze, 2011). Reproduction is a complex trait and is difficult to quantify and record accurate measures of reproductive efficiency (van der Westhuizen *et al.,* 2001; Glaze, 2011). Traits which are used to measure reproductive performance in beef cattle include calving interval (CI); age at first calf (AFC); calving date (CD) and pregnancy rate (Gargantini *et al.,* 2005; Glaze, 2011). Calving interval is the most commonly used measure of reproductive performance in cattle (van der Westhuizen *et al.,* 2001). Selection for AFC at two years was driven by the industry and led to an earlier AFC, which could results in earlier maturing animals, which could improve both the initial and re-conception rate post first calving but could also lead to calving difficulties (van der Westhuizen *et al.,* 2001; Gargantini *et al.,* 2005).

Studies have shown reproductive traits to be lowly heritable particularly for female cattle with lower potential rates of improvement (Table 2.1). Male fertility traits are found to have higher heritabilities and have been subjected to greater selection intensity and genetic gain (Table 2.1). Over the past two to three decades the beef industry has utilised tools to improve reproductive efficiency, mostly through the application of EBV's which have been the most effective methods for selection and genetic improvement (Glaze, 2011). A number of studies have been conducted on the reproductive characteristics of beef breeds and some recorded heritabilities for these traits are summarised below in Table 2.1.

Table 2.1 A summary of heritability estimates for reproductive traits in beef cattle

Trait	Heritability	Study
Age at puberty	0.4- 0.7	van der Westhuizen *et al.,* 2001; Cammack *et al.,* 2009; Spangler, 2016
Age at first calving	0.2- 0.3	van der Westhuizen *et al.,* 2001; Cammack *et al.,* 2009; Spangler, 2016
Calving date	0.03-0.21	van der Westhuizen *et al.,* 2001; Cammack *et al.,* 2009; Spangler, 2016
Calving interval	0.01-0.22	van der Westhuizen *et al.,* 2001; Cammack *et al.,* 2009; Glaze, 2011; Spangler, 2016
Pregnancy rate	0.10-0.21	Gargantini *et al.,* 2005; Cammack *et al.,* 2009; Spangler, 2016
Scrotal circumference at weaning	0.29- 0.78	van Marle-Köster *et al.*, 2000; Gargantini *et al.,* 2005; Cammack *et al.,* 2009
Scrotal circumference at yearling	0.25- 0.44	Nelsen *et al.,* 1986; van Marle-Köster *et al.*, 2000; Gargantini *et al.,* 2005

Selection for mature size at a younger age, survivability and weaning heavy calves has been a focal point of reproductive selection in Hereford breeding schemes (Baker *et al.*, 1991; Mwansa *et al.*, 2002). Early maturation and survivability will result in a shorter generation interval and improved longevity, with selection for birth and weaning weight to wean heavy calves potentially increasing the annual rate of genetic change (Baker *et al.*, 1991). Mwansa et al. (2002) has shown that direct and maternal birth weight, weaning weight and cow survivability are important characteristics included in selection index for Hereford cattle, where selection is focussed on weaning heavy calves. The genetic

trends for growth and production related traits in the South African Hereford populations indicating a gradual improvement in performance are shown below in Figure 2.3 (SA Studbook, 2016).

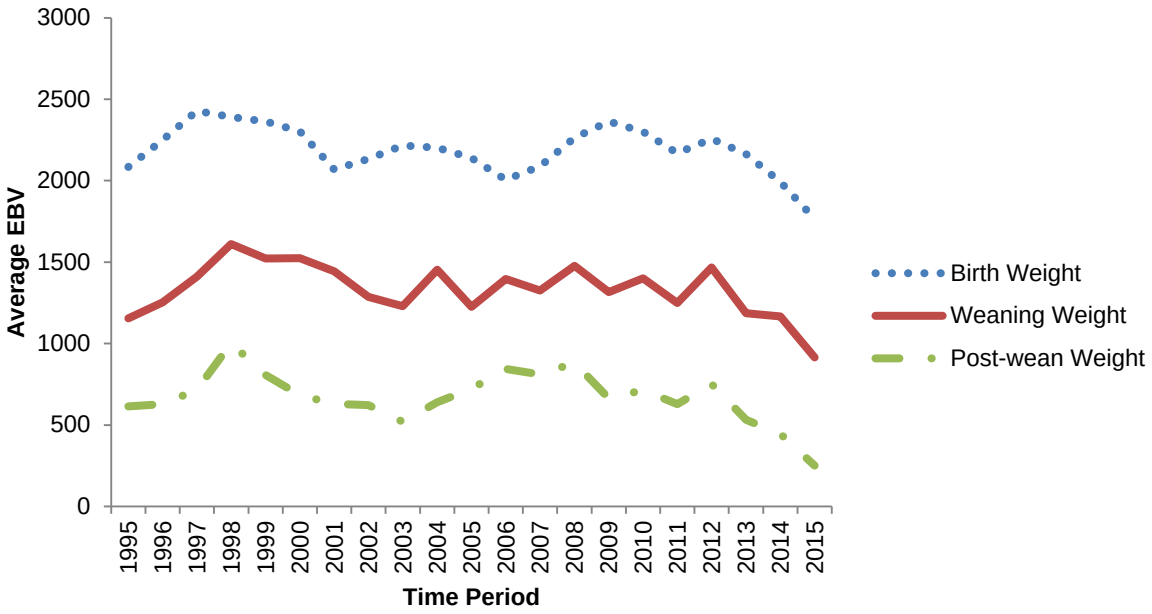

Figure 2.3 Genetic trends for birth weight; weaning weight and post-wean weight for the South African Hereford from 1987 to 2015 (Source: SA Studbook, 2016).

The major genetic advancements in beef cattle has been made in traits that are both easy and cost effective to measure, namely traits related to growth, such as birth, weaning and yearling weight (Miller, 2010; Garrick & Saatchi, 2011). Selection for growth rate has been a focus in most cattle breeds in SA and has been found to be positively associated with a desirable market weight (Bishop, 1992; Parnell, 1994; Abin *et al.,* 2016). A study performed by Koch et al. (2004) on the response to selection in Hereford cattle for weaning weight (WW), yearling weight (YW) and carcass traits over a 20 year period, revealed improved feed efficiency, improved market weight and a greater proportion of lean meat. A number of studies have been conducted on the growth characteristics of the Hereford breed and estimated heritabilities for these traits are summarised below in Table 2.2.

Table 2.2 A summary of heritability estimates for growth traits in Hereford cattle

Trait	Heritability	Study
Birth Weight	0.31- 0.54	Glaze & Schalles, 1994; van Marle-Köster *et al.*, 2000; Koch et al., 2004
Weaning Weight	0.13- 0.24	Glaze & Schalles, 1994; van Marle-Köster *et al.*, 2000; Koch et al., 2004
Yearling weight	0.11- 0.48	Glaze & Schalles, 1994; van Marle-Köster *et al.*, 2000; Koch et al., 2004
Cannon bone length	0.22- 0.32	van Marle-Köster *et al.*, 2000; Meyer, 2000
Hip height at weaning	0.11- 0.28	van Marle-Köster *et al.*, 2000; Rumph *et al.,* 2005
Hip height at yearling	0.27- 0.49	Nelsen *et al.,* 1986; Kriese *et al.,* 1991; van Marle-Köster *et al.*, 2000

One of the most important environmental influences on growth performance is nutrition and feed efficiency. The challenge for beef production is improving production without increasing the cost of production, its environmental impact or product quality (Ravagnolo *et al.,* 2018). Feed efficiency (FE) is a difficult trait to measure and has been characterised using feed conversion ratio (FCR) or residual feed intake (RFI), with FCR being the most common used measure in SA (Sherman *et al.,* 2009; Steyn *et al.,* 2014; Moore *et al.,* 2014). FCR as a trait for selection has the disadvantage of being unfavourably correlated with weight and mature size, resulting in increased mature size, thus higher maintenance requirements (Steyn *et al.,* 2014). RFI overcomes the disadvantage inherent to FCR being independent of mature weight and ADG, while strongly correlated to feed intake (Berry & Crowley, 2013; Steyn *et al.,* 2014; Moore *et al.,* 2014). Selection for feed efficiency using RFI, which has a moderate heritability (0.4- 0.7), can result in a lower DMI, manure production and methane emission, while maintaining desired body weights and production levels (Berry & Crowley, 2013; Steyn *et al.,* 2014; Moore *et al.,* 2014).

The development of genomic tools provides an opportunity for genetic improvement of FE (Bolormaa *et al.,* 2011; Moore *et al.,* 2014). Research into genetic markers associated with RFI has been carried out since 2003, identifying SNPs in association with RFI (Arthur & Herd, 2008; Moore *et al.,* 2014). Since 2014 the Canadian Hereford Association and 25% of the Uruguayan Hereford breeders, as part of a national project, have been implementing genomic tools to improve FE and collect individual FI records for Hereford herds (Navajas et al., 2014; Ravagnolo et al., 2018). Ravagnolo et al. (2018) observed a pooled RFI heritability estimate for the Canadian and Uruguayan Hereford herds of 0.26 ± 0.07, which was lower but in the same order of the pooled heritability of 0.33 for Hereford observed by Berry & Crowley (2013), studying the genetics of feed efficiency in dairy and beef cattle. Further studies evaluating the genetics of feed efficiency and related characteristics should be performed in order to improve within-breed selection for improved FE to minimise environmental impact and improve economic efficiency. Genetic improvement of all these traits simultaneously should be the goal of any beef production system to improve the overall genetic merit of the animal.

2.4 Genomic information for genetic improvement in cattle

2.4.1 Molecular Markers

Prior to the development of molecular tools, selection was based on phenotype, with limited understanding of the underlying genetic mechanism influencing traits of economic importance (Beuzen *et al.*, 2000; Singh *et al.,* 2014). The quantitative genetic approach to selection was based on population genetic parameters, such as heritability, genetic variance and correlations, which were evaluated and quantified (Dekkers & Hospital, 2002). The recorded phenotypic data and pedigrees were utilised for estimation of EBV, which was utilised to rank individuals for selection (Seidel, 2010; Singh *et al.,* 2014). Although quantitative genetic methods of selection were successfully applied for genetic improvement in livestock, the genetic gain remains limited for traits which are sex-limited, lowly-heritable or late-expressed (Dekkers & Hospital, 2002; Singh *et al.,* 2014).

The development of molecular technologies offered new solutions to study the genetics of various animal species (Dekkers, 2004; Mullen *et al.,* 2013). Research into molecular markers was initiated in the later 1980's (Singh *et al.,* 2014). Molecular genetics mitigated some of the limitations to quantitative genetics through identification of causal mutations and non-functional genetic markers that were linked to quantitative trait loci (QTL) (Fan *et al.,* 2010; Wakchaure *et al.,* 2015a). The information provided by OTL was not without its own limitations however (Dekkers, 2004; Moniruzzaman *et al.,* 2014). Selection based on marker-trait associations, is essentially selection for the marker as opposed to the trait (Singh *et al.,* 2014; Moniruzzaman *et al.,* 2014). Therefore, selection only remains effective as long as the marker and gene of interest remain in linkage disequilibrium (LD) (Dekkers, 2004; Singh *et al.,* 2014; Moniruzzaman *et al.,* 2014). Selection for a trait using microsatellites or SNP markers required validation from phenotypic expression of the selected trait (Singh *et al.,* 2014; Wakchaure *et al.,* 2015b). This led to the integration of both phenotypic and genetic evaluation of a trait, which improved the accuracy of breeding value estimations (Singh *et al.,* 2014; Wakchaure *et al.,* 2015b).

The development of the Polymerase Chain Reaction (PCR) and sequencing technologies, led to DNA-based polymorphisms becoming the markers of choice for evaluation of genetic variation, of which microsatellites and SNPs are both examples (Hanotte & Jianlin, 2005; Singh *et al.,* 2014). These DNA based markers have a range of broad applications, including gene mapping, parentage verification, population genetics and MAS (Bruford *et al.,* 2003; Schlötterer, 2004; Singh *et al.,* 2014). The choice of marker utilized is dependent on the ease of genotyping and associated cost, as well as the information content, the neutrality and independence of the marker (Vignal et al., 2002).

Copy number variations (CNV's) include duplications, insertions or deletions within the genome larger than 50bp (Wang *et al.,* 2015; Pierce *et al.,* 2018). These markers have been found in regions of the cattle genome responsible for adaptation and can be a useful tool to evaluate genetic diversity, providing supplementary information to that gained using SNPs (Wang *et al.,* 2015; Pierce *et al.,* 2018).

Microsatellites markers, which are short repetitive elements in genomic DNA, have been popular markers of choice due to easy application using PCR and electrophoresis (Hanonotte & Jianlin, 2005; Chistiakov, 2006). The advantages of microsatellites markers include high information content; dense genome coverage; high level polymorphism; locus-specific nature and codominant transmission (Bruford *et al.,* 2003). They have been used to quantify genetic diversity within and among breeds; characterise population admixture and determine population assignment, among that of parentage verification and estimation of inbreeding (Bruford *et al.,* 2003; Hanonotte & Jianlin, 2005; Singh *et al.,* 2014).

Autosomal microsatellite markers were isolated for most livestock species and the International Society for Animal Genetics (ISAG) constructed lists of markers to be utilised for genetic characterisation and parentage verification in 2001/2002 (Hanotte & Jianlin, 2005; ISAG website: http://www.isag.org.uk/journal/comparisonguide.asp; http://www.isag.org.uk/ISAG/all02_PVpanels_LPCGH.doc). The ISAG-FAO panels were designed to ensure the results are relevant on a regional and international level, providing a view of genetic diversity (Balloux, 2002; FAO, 2011; Brenig & Schütz, 2016). The recommended ISAG microsatellite marker panel for parentage verification was constructed based on markers with high heterozygosity and polymorphism and has been utilized worldwide since the 1990s (Kios, 2011; Brenig & Schütz, 2016). The ISAG panel of microsatellite markers for cattle parentage testing is shown below in Table 2.3.

Table 2.3 ISAG panel of microsatellite markers applied in cattle parentage testing

Locus		Primer Sequence (5'- 3')	Reference
BM1824	Forward	GAG CAA GGT GTT TTT CCA ATC	Barendse et al., 1994
	Reverse	CAT TCT CCA ACT GCT TCC TTG	
BM2113	Forward	GCT GCC TTC TAC CAA ATA CCC	Sunden et al., 1993
	Reverse	CTT CCT GAG AGA AGC AAC ACC	
ETH10	Forward	GTT CAG GAC TGG CCC TGC TAA CA	Toldo et al., 1993
	Reverse	CCT CCA GCC CAC TTT CTC TTC TC	
ETH225	Forward	GAT CAC CTT GCC ACT ATT TCC T	Steffen et al., 1993
	Reverse	ACA TGA CAG CCA GCT GCT ACT	
INRA23	Forward	GAG TAG AGC TAC AAG ATA AAC TTC	Vaiman et al., 1994
	Reverse	TAA CTA CAG GGT GTT AGA TGA ACT C	
SPS115	Forward	AAA GTG ACA CAA CAG CTT CTC CAG	Moore & Byrne, 1993
	Reverse	AAC GAG TGT CCT AGT TTG GCT GTG	
TGLA122	Forward	CCC TCC TCC AGG TAA ATC AGC	Georges & Massey, 1992
	Reverse	AAT CAC ATG GCA AAT AAG TAC ATA C	
TGLA126	Forward	CTA ATT TAG AAT GAG AGA GGC TTC T	Georges & Massey, 1992
	Reverse	TTG GTC TCT ATT CTC TGA ATA TTC C	
TGLA227	Forward	CGA ATT CCA AAT CTG TTA ATT TGC T	Georges & Massey, 1992
	Reverse	ACA GAC AGA AAC TCA ATG AAA GCA	

Parentage verification has received much attention as it improves the accuracy of pedigree records and thus the derived genetic evaluations, which is critical for accurate EBV estimation (Radko *et al.,* 2010; McClure *et al.,* 2012; McClure *et al.,* 2013). There is some debate with regard to the use of microsatellites versus SNPs for parentage verification, quantifying genetic diversity and characterising population structure (McClure *et al.,* 2012; Strucken *et al.,* 2014; Lyons *et al.,* 2015). Accuracy for use in parentage testing became an issue with microsatellites due to discrepancies between laboratories (Vignal et al., 2002; Fernández et al., 2013). Common discrepancies related to genotyping errors, where results were not always comparable due to differences in allele calling, size determination and the creation of false alleles (Vignal et al., 2002; Fernández et al., 2013). Multiple factors, such as the duration of analysis; high rate of genotyping error and the advent of high-throughput DNA sequencing and bioinformatics, resulted in SNPs becoming the marker of choice for most genomic applications (Vignal et al., 2002; Fernández et al., 2013; Brenig & Schütz, 2016). Microsatellites markers remain suitable for application in parentage verification, due to ease of testing and abundance of breeding database results (Fernández et al., 2013; Brenig & Schütz, 2016), however SNPs are quickly becoming a more popular option due to improved parentage accuracy and availability due to routine genotyping in breeds using genomic selection (McClure *et al.,* 2018).

SNPs are biallelic markers which have gained much popularity in the last decade due to their abundance, stability and suitability for high-throughput automated analysis (Heaton *et al.,* 2002; Fan *et al.,* 2010; Singh *et al.,* 2014). The popularity of these markers was accelerated with the bovine genome sequencing project which produced a SNP database composed of two million bovine SNPs, the majority of which were identified in a single Hereford cow (Eck *et al.,* 2009; Williams *et al.,* 2009). SNP validation studies led to the incorporation of an additional 23 000 SNPs into the database, sequenced from Holstein and seven prominent beef breeds including the Hereford (Eck *et al.,* 2009). Of the total SNPs within the database, 123 000 were identified from regions of deep coverage and were selected for the generation of the initial SNP chips by both the International Bovine Hapmap Consortium, which produced the HapMap 30K SNP chip which has been validated across 18 breeds of both *Bos Indicus* and *Bos Taurus* (Williams *et al.,* 2009; Eck *et al.,* 2009; Seidel, 2010). Two companies Illumina, based in San Diego, California, USA and Affymetrix, based in Santa Clara, California, USA have produced multiple bovine SNP arrays, the details of which are summarised in Table 2.4.

Table 2.4 A list of the available Affymetrix and Illumina bovine genotyping arrays (www.illumina.com; www.thermofisher.com/za/en/home/life-science/microarray-analysis/affymetrix.html)

Platform	SNP Chip	Number of SNPs	
		Fixed	Custom
Illumina	Bovine SNP50	53 714	
	BovineLD BeadChip	7 931	80 000
	BovineHD BeadChip	> 777 000	-
	GGP Bovine 150K Array	>134 000	-
	GGP Bos Indicus HD Array	> 74 000	-
	GGP Bovine LD Array	>26 000	-
	TruSeq Bovine Parentage Sequencing Panel	200 ISAG-recommended parentage SNPs and 66 recessive disease associated SNPs	-
	Infinium iSelect		≤700 000
Affymetrix	GeneChip Bovine Mapping 10K	10 000	-
	Targeted Genotyping Bovine 25K	~ 25 000	-
	Axiom Genome-wide BOS 1 Array Plate	> 640 000	-

These SNP arrays have been a crucial step to implementing genomic selection and the combination of EBV's, matched with a respective SNP profile, can be integrated to produce a genomic EBV to improve selection accuracy two-fold (Seidel, 2010; Meuwissen *et al.,* 2018). The benefits of genomic selection include a simplified selection process and improvement in the rate of genetic gain (Seidel, 2010; Boichard *et al.,* 2016). Both microsatellites and SNPs have been successfully utilised to estimate genetic diversity parameters and characterise population structure for population genetics, which is important to evaluate prior to implementing genomic selection in a population.

2.4.2 Genetic diversity

The processes of natural and artificial selection have crafted the phenotypic and production characteristics in livestock breeds today, with livestock animals selected for the purpose of high meat and milk yields; docility for ease of handling and fertility for population longevity (Engelsma *et al.,* 2010; Magee *et al.,* 2014). Commercial production market drivers are one of the threats to genetic diversity, due to narrow selection focus, which is broadly implemented as part of the breeding objectives (Hoffmann & Baumung, 2013; Tixier-Boichard *et al.,* 2015). The genetic diversity of a population dictates its ability to respond to selection, with long-term survival depending on allelic variation (Medugorac *et al.,* 2011; Kristensen *et al.,* 2015). Future food security can be ensured through maintenance of genetically diverse livestock populations, where adaptive and neutral diversity is encouraged, allowing for effective response to selection for changing environmental conditions,

disease threats and changes in market drives (Hoffman, 2010; Medugorac et al., 2011; Hoffmann & Baumung, 2013).

Population diversity studies require the quantification of various parameters, in order to characterise the population structure, namely Hardy-Weinberg equilibrium (HWE); heterozygosity values; inbreeding coefficient; FST and LD. HWE evaluation allows one to compare the actual versus expected population structure, which assumes the population complies with Hardy-Weinberg principles (Falconer & Mackay, 1996; Bourdon, 2000).The test for HWE evaluates the relationship between gene and genotype frequencies and a population is said to be in HWE when these frequencies remain constant through generations (Falconer & Mackay, 1996). Possible causes for digression from HWE in a population can be indicative of selection, inbreeding or population substructure (Namipashaki *et al.,* 2015; Chen et al., 2017).

Heterozygosity is a measure of diversity within a population and is defined as the probability that two alleles chosen at random from a population are different (Toro & Caballero, 2005; Engelsma *et al.,* 2010; Makina *et al.,* 2014). High estimates could reflect natural selection and admixture within a population, whereas low estimates could reflect population isolation (Ojango *et al.,* 2011). The observed and expected heterozygosity values can be utilised to estimate the degree of inbreeding, numerically represented as the inbreeding coefficient (F) which is the probability that two alleles are identical by descent (Falconer & Mackay, 1996; Bourdon, 2000; Slate *et al.,* 2004). A high estimate is indicative of a small number of heterozygote genotypes and a large number of homozygote genotypes, with the opposite being true for a low estimate (Ojango *et al.,* 2011).

Wright's F-statistics, with special reference to F_{ST} have important application for population history and identifying genomic regions under selection (Holsinger, 2009; Jakobsson *et al.,* 2012). F_{ST} is a parameter of genetic differentiation between populations and can be defined as the correlation between randomly selected alleles from a population in comparison to its ancestral population (Frkonja *et al.,* 2012; Bhatia *et al.,* 2013). F_{ST} is one of three interrelated parameters namely, F_{IT} which is the correlation of gametes in an individual compared to that of the whole population and F_{IS} which is the correlation between gametes in an individual compared to that of the subpopulation from which it originates (Holsinger, 2009; Jakobsson *et al.,* 2012). F_{ST} can be calculated by using the variance of allele frequencies, which is divided into inter and intra-locus variance, with inter-locus variance arising from mutation, random drift and selection (Nei, 1978; Frkonja *et al.,* 2012). This causes an increase in expected heterozygosity (H_e) between loci and intra-locus variance a result of sampling a finite number of individuals (Nei, 1978; Frkonja *et al.,* 2012). The F_{ST} statistic and its associated components, F_{IS} and F_{IT} are a useful tool to characterise population structure (Weir & Cockerham, 1984). F_{ST} is a parameter of both the population and the selected SNPs for computation (Bhatia *et al.,* 2013). Hudson's estimate of F_{ST} is recommended, as it does not overestimate, nor is it affected by the ratio of sample sizes in the population (Bhatia *et al.,* 2013). F_{ST} ranges of 0- 0.05 imply low genetic variation, 0.05- 0.15 imply moderate genetic variation and 0.15- 0.25 imply large genetic variation (Hartl & Clark, 1996).

Linkage disequilibrium (LD) throughout the genome facilitates the determination of the genetic relationship between specific regions in the genome that have been a focus of selection and estimate

the success of the specific selection criteria (Lee *et al.,* 2011; Khanyile *et al.,* 2015; Makina *et al.,* 2015). LD is defined as the non-random association between alleles at different loci within a population, and is population specific (Hartl & Clark, 1997). The genome is a mosaic of haplotype blocks, which are regions of the genome with high marker-marker LD and low haplotype diversity, separated by shorter segments of low LD (Gautier *et al.,* 2007). The LD phases between two populations is likely to be different for a few reasons, namely if the populations have been separate for many generations or if the effective population (Ne) is small (Roos *et al.,* 2008). Common measures of LD include D' and r^2, both derived from Lewontin's D (Lee *et al.,* 2011; Makina *et al.,* 2015; Kasarda *et al.,* 2016). The D' value indicates whether recombination between two loci has occurred, whereas the r^2 value denotes the correlation between two loci (Mustafa et al., 2018). The r^2 measure is generally considered the more robust and reliable parameter to determine LD, as the D' statistic tends to be inflated with small populations or low haplotype frequencies (Lee *et al.,* 2011; Makina *et al.,* 2015; Kasarda *et al.,* 2016). The r^2 value is the squared correlation coefficient between two variables, namely alleles at two different SNP loci (Matukumalli *et al.,* 2009; Visser *et al.,* 2016). LD is an important parameter due to its association with evolutionary genetics (Slatkin, 2008). LD extent can be evaluated as an indication of the degree of inbreeding, selection or admixture present, with a high degree of LD being indicative of potential inbreeding or selection, whereas long-range LD is indicative of recent population admixture (Ai *et al.,* 2013; Wientjes *et al.,* 2013; O'Brien *et al.,* 2014).

Population diversity and structure is quantified and characterised using software designed to evaluate genomic data. A major motivation for developing more advanced and computationally adept software is due to the complex nature of traits and polygenetic effects (Purcell *et al.,* 2007). The advent of SNP genotyping has allowed for whole genome analysis and with the application of large numbers of markers, allows for accurate assessment of population association and stratification studies (Purcell *et al.,* 2007). There are broad ranges of software available to evaluate genomic data, among others PLINK (Purcell *et al.,* 2007); GCTA (Yang *et al.,* 2011); ADMIXTURE (Alexander *et al.,* 2009); STRUCTURE (Pritchard *et al.,* 2000) and GenAlEx (Peakall & Smouse, 2012). A summary of the software, parameters and applications of said parameter estimates can be seen in Table 2.5.

Table 2.5 A summary of relevant software for genetic diversity and population structure analysis used in the current study

Software	Main Parameters	Application
GenAlEx (Peakall & Smouse, 2012)	Observed & Expected heterozygosity Inbreeding coefficient Chi-square Polymorphic information content (PIC) Wright's F-statistics AMOVA	Inference and evaluation of genetic diversity.
STRUCTURE (Pritchard et al., 2000)	LnP(K) K value	Evaluation and characterisation of population structure.
PLINK (Purcell et a., 2007)	Minor allele frequency (MAF) Hardy-Weinberg equilibrium (HWE) Observed & Expected heterozygosity Inbreeding coefficient Linkage disequilibrium (r2)	Inference and evaluation of genetic diversity.
GCTA (Yang et al., 2011)	Eigenvectors Eigenvalues	Inference of genetic relationships among individuals.
ADMIXTURE (Alexander et al., 2009)	Cross validation error K value	Evaluation and characterisation of population structure.

PLINK and GCTA are used to analyse genetic diversity parameters and infer population relationships. PLINK software has the capability to handle large data sets with computational efficiency for data management, summary statistics, population stratification, association analysis, and identity-by-descent estimation (Purcell *et al.,* 2007). GCTA was designed to estimate the proportion of genome- or chromosome-wide SNPS phenotypic variance for complex traits (Yang *et al.,* 2011). The software is capable of data management; estimating the genetic relationship among individuals; the proportion of variance in phenotype explained by SNPs; inbreeding coefficients for individuals; FST values; computing LD scores and PCA analysis, among other things (Yang *et al.,* 2011).

ADMIXTURE (Alexander *et al.,* 2009) and STRUCTURE (Pritchard *et al.,* 2000) software use similar statistical modules to infer population admixture and structure (Frkonja *et al.,* 2012). STRUCTURE is a free software package for analysing multi-locus genotype data in order to characterise population structure, using both microsatellite and SNP data (Pritchard *et al.,* 2000). The software is capable of inferring population distinction; population assignment; identifying migrants and admixed individuals and estimating population allele frequencies, utilising a Bayesian approach which employs the Markov Chain Monte Carlo (MCMC) algorithm to estimate posterior distribution (Pritchard *et al.,* 2000). Admixture uses a maximum likelihood framework for estimation of individual ancestries from multi-locus SNP genotypes and although it utilises the same likelihood model as that of STRUCTURE (Pritchard *et al.,* 2000), this software maximizes the likelihood instead of estimating the

posterior and the estimation process uses a faster numerical optimisation algorithm, which is beneficial for larger data sets (Alexander *et al.,* 2009).

GenAIEx (Peakall & Smouse, 2012) is a cross-platform package, that runs in Microsoft excel, capable of analysing population data utilising microsatellites or SNPs (Peakall & Smouse, 2012). The package is capable of computing both frequency-based and distance-based analyses, namely Hardy-Weinberg Equilibrium (HWE), Heterozygosity estimates and F-statistics and Analysis of Molecular Variance (AMOVA), respectively (Peakall & Smouse, 2012).

Evaluating genetic diversity and population structure can be done using these various software programs. Genome-wide data is time-consuming and computationally expensive to analyse, thus it's ideal to reduce the number of final markers analysed, by pruning data according to its information content (Wilkinson *et al.,* 2011). Most methodologies employ the process of SNP quality control (QC), to ensure reliable results. QC parameters reported in diversity studies include, removing SNPs with less than 95% call rate, SNPs with less than 0.02 minor allele frequency (MAF) and samples with more than 10% missing genotypes (Decker *et al.,* 2014; Mbole-Kariuki *et al.,* 2014; Zwane et al., 2016). In addition, SNPs that were in high LD may be removed as they have been shown to counter the effect of ascertainment bias and improve estimates of genetic relatedness between populations (Kijas *et al.,* 2012; Makina *et al.,* 2014). Parameters such as HWE, heterozygosity, inbreeding coefficient, FST and LD estimates are determined using various programmes. ADMIXTURE (Alexander *et al.,* 2009) and STRUCTURE (Pritchard *et al.,* 2000) analyses can be used to evaluate ancestry proportions for predefined ancestral populations, which can be presented graphically for easy inference of population structure (Decker *et al.,* 2014).

2.4.3 Genetic diversity and population structure in beef cattle

The development of high throughput genomic technologies in combination with data analysis, allows the study of the genetic diversity of livestock populations (Zhang & Plastow, 2011; Edea *et al.,* 2014; Tixier-Boichard *et al.,* 2015). The results of analysis are applied to characterise genetic history; population structure; parentage determination and traceability (Negrini *et al.,* 2008; Engelsma *et al.,* 2010). Population structure is important to future selection drives, with the use of SNP arrays, allowing for the detection of informative markers that can discriminate between populations (Makina *et al.,* 2015; Kasarda *et al.,* 2016; Zwane *et al.,* 2016). There has been a number of genetic diversity studies performed on various beef breeds using both microsatellite and SNP markers. A summary of some studies is shown below in Table 2.6.

Table 2.6 Genetic diversity studies performed on various beef breeds using both microsatellite and SNP markers

Study title	Marker type	Marker number or SNP chip	Authors
Admixture and diversity in West African cattle	Microsatellite	20 markers	Freeman *et al.* (2004)
Genetic diversity in selected stud and commercial herds of the Afrikaner cattle breed	Microsatellite	11 markers	Pienaar *et al.* (2014)
Genetic diversity between and within Sudanese Zebu cattle breeds using microsatellite markers	Microsatellite	9 markers	Hussein *et al.* (2015)
Genetic diversity in South African Nguni cattle ecotypes based on microsatellite markers	Microsatellite	22 markers	Sanarana *et al.* (2016)
Genetic diversity in Zimbabwean Sanga cattle breeds using microsatellite markers	Microsatellite	16 markers	Gororo *et al.* (2018)
Prediction of breed composition in an admixed cattle population	SNP	Illumina BovineSNP50 Beadchip	Frkonja *et al.* (2012)
Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel	SNP	Illumina BovineSNP50 Beadchip	Makina *et al.* (2014)
Genetic diversity and divergence among Spanish beef cattle assessed by a bovine high-density SNP chip	SNP	Illumina BovineHD Beadchip	Cañas-álvarez *et al.* (2015)
Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide	SNP	Illumina BovineSNP50 Beadchip	Makina *et al.* (2016)
Genetic divergence of cattle populations based on genomic information	SNP	Illumina BovineSNP50 Beadchip	Kasarda *et al.* (2016)

Genetic diversity studies using microsatellites have often utilised the recommended ISAG markers, among others (Hussein *et al.,* 2015; Sanarana *et al.,* 2016; Gororo *et al.,* 2018). Pienaar *et al.* (2014) evaluated genetic diversity in the Afrikaner breed finding that a moderate to high degree of variation was present within the breed despite recent population decline. Hussein *et al.* (2015) evaluated the genetic diversity and relationship among three Sudanese cattle breeds, finding that the breeds originated from a common ancestor; no inbreeding between or within the breeds and observing a good degree of genetic variability across breeds, promoting their use in breeding and conservation programmes as a source of genetic diversity. Sanarana *et al.* (2016) evaluated the genetic diversity of the SA Nguni ecotypes finding that gene flow between the ecotypes has led to low to moderate genetic differentiation among populations, however despite the admixture among populations there was distinct separation where selection had been practiced and the indigenous nature of the breed promotes its conservation.

Genetic diversity studies utilising SNP markers require a small number of samples and SNPs (Frkonja *et al.,* 2012; Cañas-álvarez *et al.,* 2015; Kasarda *et al.,* 2016). Studies can be carried out with as little as 50 samples and 4000 SNPs to accurately quantify genetic variation (Baumung *et al.,* 2004; Frkonja *et al.,* 2012). Makina *et al.* (2014) evaluated the genetic diversity and population structure of six cattle breeds in South Africa. A low to moderate degree of genetic diversity was found across all six breeds, observing a closer genetic relationship among the indigenous breeds and clear genetic distinction between the indigenous breeds and the imported *Bos taurus* breeds (Makina *et al.,* 2014). Cañas-álvarez *et al.* (2015) evaluated the genetic diversity and divergence among seven important indigenous Spanish cattle breeds using the high-density beadchip, finding a low degree of divergence among breeds but high degree of admixture within the breeds. Makina *et al.* (2016) evaluated the SA Sanga cattle finding that the cattle originated from African taurine and Bos indicus cattle, with the Bonsmara breed confirming the hybrid history of these cattle.

Genetic diversity studies conducted using similar methodologies allow for comparisons between studies and potentially provides data to be utilised for genetic improvement and conservation. Pienaar et al. (2014) performed a genetic diversity study on an indigenous South African cattle breed utilising the recommended ISAG microsatellite markers. Sanarana et al. (2016) emphasised the importance of the Nguni breed in communal and commercial systems and recommends further studies be performed to determine the breeds broader application with regard to ecotypes. Makina et al. (2014) evaluated indigenous and exotic breeds within South Africa, observing a distinct genetic divergence between the two groups, identifying distinct genetic resources which could be exploited in order to adapt to future environmental conditions.

2.4.4 Genomic selection

Genomic selection can be applied for within-breed genetic improvement, utilising genetic markers to improve EBV accuracy (Garrick, 2011; van Marle-Köster *et al.,* 2015). The training population should contain more than 1000 impactful animals, phenotyped and genotyped for important traits, used to derive a prediction equation (Goddard & Hayes, 2007; van Marle-Köster *et al.,* 2013; Theron *et al.,* 2014). The prediction equation facilitates the implementation of genomic selection in beef breeding schemes, which have a high proportion of cross-bred animals (Goddard & Hayes, 2007). SNP arrays used in combination with imputation can utilise limited genomic data within beef breeding systems (Georges *et al.,* 2018). The cost associated with genomic selection can be minimised by using low density SNP panels, however there is an associated risk that causative SNPs are not directly interrogated on the arrays (Van Binsbergen *et al.,* 2015; Georges *et al.,* 2018). Imputation of low density SNP panels to infer higher density genotypes, can be used to improve selection accuracy, through incorporation of this more inclusive genomic information into the EBV without increasing cost (Saatchi *et al.*, 2013; Georges *et al.,* 2018). The genomic selection approach combined with predictions of breeding values for beef cattle, has observed an improvement in genetic gains of 15- 44% (Ibtisham *et al.,* 2017).

Applying genomic selection in practice is achieved through the inclusion of genomic predictions into traditional genetic evaluations (Miller, 2010; Garrick & Saatchi, 2011; Saatchi *et al.*, 2013). The beef industry had the benefit of learning from the success of genomic selection within the dairy industry (Miller, 2010). Some of the challenges the beef industry has faced with regard to implementing genomic selection and its impact on performance include the lack of artificial insemination, high degree of breed heterogeneity and small effective population sizes (Berry *et al.,* 2016; Ibtisham *et al.,* 2017). The beef industry can overcome this constraint through implementation of multi-breed genomic prediction models and the utilisation of lower-cost genotyping technologies (Berry *et al.,* 2016). The benefits of genomic selection include a shorter generation interval through selection of younger animals and refinement of selection focus for desirable traits such as feed efficiency, carcass and reproductive traits through selection of a commercial reference population, extensively recorded for traits of interest (Goddard & Hayes, 2007; Ibtisham *et al.,* 2017). Genomic data has become an integral part of the selection process for breeders due to its many benefits (Saatchi *et al.*, 2013).

2.5 Conclusion

The South African Hereford was selected for medium frame with early maturation; increased growth rate with maximum lean deposition and fertility traits and is well established in SA. Traditional breeding strategies laid the foundation of animal evaluation and selection, which still form an integral part of any selection scheme. The role of genetic diversity studies in breed characterisation, genetic variation and the relationship among local populations is imperative for the management, improvement and conservation of livestock genetic resources.

CHAPTER 3

MATERIALS AND METHODS

3.1 Introduction

The aim of the study was to quantify genetic variation, inbreeding and population structure at a genomic level to provide reference data for the SA Hereford breed for potential application in genomic selection, using both microsatellite and genome-wide SNP data.

The microsatellite profiles of 1945 animals were provided by UNISTEL (Unistel Medical Laboratories (Pty) Ltd, SA, http://www.unistelanimalservices.co.za/) with the consent of the South African Hereford Breed Society. Furthermore, a total of 500 SNP-genotyped animals were available for genomic data analysis that included 184 South African Hereford genotypes generated within the Beef Genomic Program (BGP) and 316 Irish Hereford genotypes from TEAGASC Ireland by Prof D Berry (https://www.teagasc.ie/). The SA Hereford Society provided consent for the use of the data and approval for use of genotypes were provided by the BGP. Ethics approval (EC 160802-063) was obtained for all data analysed in this study from the Ethics Committee of the Faculty of Natural and Agricultural Sciences at the University of Pretoria.

3.2 Materials

3.2.1 Microsatellite data

The 1945 microsatellite profiles included the name of the breeder; the laboratory number; animal name and tag number; registration number; date of birth; gender and the alleles for each of the indicated markers. These microsatellite profiles were edited for absent or incomplete marker data, which resulted in the removal of 146 animals, leaving 1799 animals that represented 36 of the original 58 herds. A subset of 500 animals was then selected from the 1799 animals, to create a population representative data set. Only the 22 herds who contributed at least 15 animals with complete profiles were included in the subset. A maximum of 25 animals per herd were included in the subset, to avoid over-representation by a small number of herds. The number of herds and the number of animals included in the complete and subset data sets are indicated below in Table 3.1.

Table 3.1 The number of herds and animals included in the complete and subset data set

Complete data set		**Subset data set**	
Herd label	**Number of animals**	**Herd label**	**Number of animals**
A	39	A	25
B	49	B	25
C	6	F	18
D	6	G	25
E	6	H	25
F	18	I	25
G	66	N	19
H	45	O	25
I	28	P	25
J	5	Q	25
K	12	R	20
L	8	U	18
M	11	V	25
N	19	W	25
O	42	Y	18
P	70	AA	16
Q	88	BB	25
R	20	CC	25
S	5	DD	25
T	6	FF	25
U	19	GG	16
V	58	HH	25
W	83		
X	9		
Y	19		
Z	14		
AA	17		
BB	109		
CC	34		
DD	706		
EE	11		
FF	66		
GG	17		
HH	63		
II	14		
JJ	11		
Total	1799	**Total**	500

The complete and subset data was analysed for eleven common microsatellite markers namely, BM1824, BM2113, ETH10, ETH225, ETH3, INRA23, SPS115, TGLA122, TGLA126, TGLA227 and TGLA53. These ISAG-recommended markers are commonly used for genetic characterisation and parentage verification.

3.2.2 SNP data

The 184 South African SNP genotypes were generated with the GGP HD Bovine 150K chip as part of the BGP programme. The genotypes generated by the BGP were for influential herd sires and these genotypes will form part of the reference population for the Hereford breed for implementation of genomic selection. The 316 Irish genotypes were generated with the BovineHD BeadChip array by TEAGASC. The GGP Bovine array (https://www.illumina.com/products/ggp-whole-genome-genotyping-arrays.html) features approximately 134 000 SNPs for *Bos taurus* breeds specifically, whereas the BovineHD BeadChip (https://www.illumina.com/products/by-type/microarray-kits/bovinehd.html) contains over 777 000 SNPs for both *Bos taurus* and *indicus* breeds.

The South African data set consisted of 184 individuals, provided by 19 stud breeders. The herd labels and number of animals contributing to each to herd are indicated in Table 3.2.

Table 3.2 The number of herds and animals included in the SA SNP data set

Herd Label	**Number of Animals**
B	9
H	31
N	4
P	1
R	23
U	2
W	28
X	7
Z	4
CC	27
FF	20
JJ	1
KK	3
LL	10
MM	6
NN	1
PP	1
QQ	3
RR	3
Total	184

The Hereford herds in SA that contributed either microsatellite and/or genotypic data for analyses in this study is distributed over six provinces in South Africa as shown in Figure 3.1.

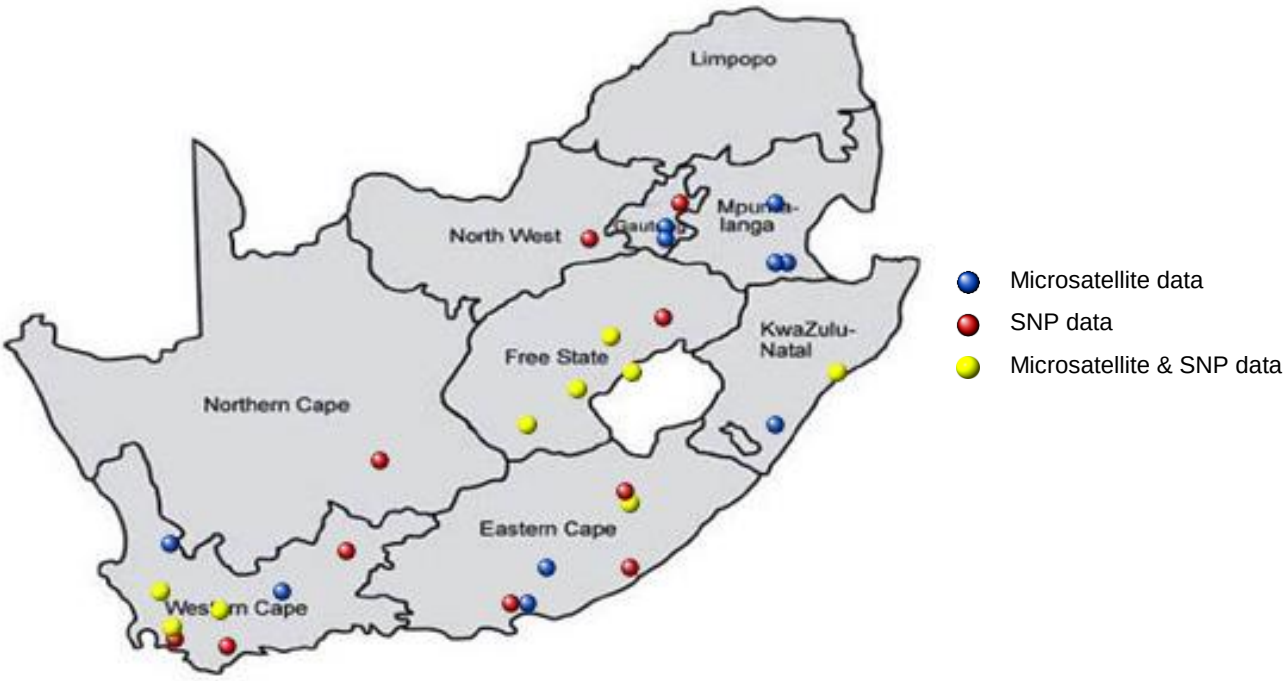

Figure 3.1 Distribution of the Hereford herds contributing microsatellite data, genotypic data or both in South Africa

3.3 Methods

3.3.1 Statistical analysis of microsatellite data

Microsatellite analyses were performed on both the complete and subset data sets. The data was analysed for the following population parameters using various software. Allele frequency, polymorphic information content (PIC) and heterozygosity (observed and expected) values were performed using Excel MS toolkit v3.1 (Park, 2001). Hardy-Weinberg equilibrium (HWE) deviations were performed using GenePop v4.7 (Raymond & Rousset, 1995). Wright's F-statistics for each locus; including θ and f which are analogous to Wright's F_{ST} and F_{IS} were performed using FSTAT v2.9.3 (Goudet, 1995) and AMOVA was performed using GenAIEx v6.503 (Peakall & Smouse 2006, 2012).

The Hardy-Weinberg model is based on the assumption that gene and genotypic frequencies remain constant in a population (Bourdon, 1997). Forces such as mutation or selection can cause changes in these frequencies and affect the degree of genetic variation (Bourdon, 1997; Hartl & Clark, 1997). The mathematical relation between the allele and genotype frequencies and their relationship is shown below, where $p2$, $2pq$ and $q2$ are the allele frequencies of A and a in gametes from the previous generation (Hartl & Clark, 1997).

$$\text{AA: } p^2 \qquad \text{Aa: } 2pq \qquad \text{aa: } q^2$$

$$(p + q)^2 = p^2 + 2pq + q^2$$

Heterozygosity is a measure used to quantify the amount of genetic variation due to polymorphic loci (Falconer & Mackay, 1996). Heterozygosity is expressed as an observed or expected value, if the two values are not equal, the population is not in HWE (Falconer & Mackay, 1996). Heterozygosity can be calculated for single loci, where *P*i is the frequency of the *i*th of *k* alleles, shown below (Nei, 1973).

$$H = 1 - \sum_{i=1}^{k} Pi^2$$

Heterozygosity can be calculated over several loci, where the first summation is for the *l*th of *m* loci and the second summation, where *P*i is the frequency of the *i*th of *k* alleles, shown below (Nei, 1973).

$$H = 1 - \frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{k} Pi^2$$

Wright's F-statistics, F_{IS} and F_{ST}, represent the proportion of genetic variance in the subpopulation contained in an individual and the proportion of genetic variance contained in the population, respectively. F_{IS}, the inbreeding coefficient, represents the individual (I) relative to the subpopulation (S), with F_{ST}, representing the subpopulation (S) relative to the total population (T), which can be calculated using the sample variance s^2 and sample mean p, with the formula below (Weir, 2012).

$$F_{ST} = \frac{s^2}{\bar{p}(1-\bar{p})}$$

Population structure was estimated and visualised using STRUCTURE (Pritchard *et al.,* 2000). The Bayesian-based assignment test was utilised to infer the estimated number of population (K clusters). The data was analysed using the admixture ancestry model, with correlated allele frequencies. To estimate the ideal K, the LnPr (X|K) was applied, the formula for calculating delta K can be seen below, where Ln P(D) is the mean posterior probability value per cluster (Pritchard *et al.,* 2000).

$$L(K) = Mean\ of\ the\ Ln\ P(D) values$$
$$L'(K) = L(K)_n - L(K)_{n-1}$$
$$L''(K) = L'(K)_n - L'(K)_{n-1}$$
$$\Delta K = \frac{L''(K)}{Standard\ deviation}$$

The peak value for delta K at each specified cluster is assumed to be the ideal population number. Each K from 1 to 10 was run for twelve repetitions, thus producing twelve run results for K=

1, etc. Each of the twelve runs for each K value was carried out with a burn-in period of 20 000 steps and MCM of 100 000 iterations.

3.3.2 Statistical analysis for SNP data

The South African and Irish data sets were first analysed separately, subsequently the two data sets were merged and the analyses were repeated. PLINK (Purcell *et al.,* 2007) software was utilised for quality control and estimation of summary statistics for the evaluation of genetic diversity of the population. GCTA (Yang *et al.,* 2011) software was run to produce eigenvectors and eigenvalues, which were used to perform PCA and ADMIXTURE (Alexander *et al.,* 2009). Genesis 0.2.3 (Buchmann, R & Hazelhurst, S., University of the Witwatersrand, Johannesburg, SA, http://www.bioinf.wit.ac.za/oftware/genesis) software was used to visualise population structure.

3.3.2.1 South African and Irish data sets

Quality control

The South African and Irish data sets were received from BGP and TEAGASC in the .MAP and .PED file format which were used for analyses. The .MAP and .PED files for the South African and Irish data sets were converted to binary .BED and .BIM and .FAM format prior to quality control and summary statistic estimation. PLINK (Purcell *et al.,* 2007) was used to identify and remove individuals and variants which failed to satisfy quality control parameters.

Individual based quality control was performed to remove individuals that violated a sample call rate of 98%. SNP-based quality control was performed to remove all SNPs that violated a call rate of 98%; minor allele frequency (MAF) of less than 5% and digressed from Hardy-Weinberg equilibrium (HWE).

Summary statistics

Marker-based summary statistics namely observed and expected heterozygosity, individual inbreeding coefficients (F), MAF and LD estimates were generated with PLINK using the commands shown below in Table 3.3 (Purcell *et al.,* 2007).

Table 3.3 Summary of the parameters, commands and software used for evaluation of genetic diversity in the current study

Parameter	Parameter description	Command	Program
HO & HE	Observed & expected heterozygosity	- -hardy	Plink (Purcell *et al.,* 2007)
F	Inbreeding coefficient per individual	- -het	
MAF	Minor allele frequency per SNP	- -freq	
LD	Linkage disequilibrium between SNPs on the same chromosome	- -r^2	
LD window	LD for different mapping distances, namely 10, 10-20kb, 20-40kb, 40-60kb, 60-100kb, 100-200kb, 200-500kb and 500kb- 1Mb	- -r^2 - -ld-window	
GRM	Genetic relationship matrix generation	- -make-grm	GCTA (Yang *et al.,* 2011)
PCA	Computing eigenvalues and eigenvectors for Principal Component Analysis (PCA), for three principal components	- -pca 3	

The eigenvectors file was utilised to graph the principle component one and two against each other to characterise the genetic relationship among the individuals in the analysed population in Microsoft Excel (2010).

Population structure

Population structure was estimated using the maximum likelihood estimation of ancestry for each animal. The software was run for a K value range of 1 to 10 for the South African and 1 to 20 for the Irish data set, respectively. Each run produced a cross-validation error estimate for each of the individual K values and the ideal K was identified by the corresponding, lowest cross-validation error estimate. Once the ideal K was identified, the software Genesis 0.2.3 (Buchmann, R & Hazelhurst, S., University of the Witwatersrand, Johannesburg, SA, http://www.bioinf.wit.ac.za/oftware/genesis) was used to create a visualisation of population structure in bar plot format, which was then manually annotated from the .Q and .FAM file.

3.3.2.2 Merged data set

The .MAP files for the South African and Irish data set were used to identify the common SNPs between the two data sets. A total of 127 935 SNPs were identified as common and were listed in a .txt file. Using PLINK (Purcell *et al.,* 2007), the SNP list text file was used to extract the specific SNPs from each data set, to produce a combined .MAP file which contained only the common SNPs. The two new .Map files for each data set were merged using the - - merge command to produce one

merged .MAP file with 500 individuals and 127 935 SNPs. The merged file was updated for the SNP chromosome and physical positions using SNPchiMp V.3 (Nicolazzi *et al.,* 2015). The updated merged .MAP file was then converted to the binary file formats .BED, .BIM and .FAM for quality control and estimation of summary statistics.

The process of data evaluation outlined in 3.3.2.1 for quality control, summary statistic estimations, genetic relatedness and population structure analyses that was performed for each of the individual data sets, was similarly carried out for the merged data sets.

CHAPTER 4

RESULTS

4.1 Microsatellite Data

4.1.1 Genetic diversity

Eleven ISAG microsatellite markers routinely used for parentage and genetic diversity were analysed in this study based on 1799 microsatellite profiles. The complete and subset datasets were evaluated for PIC and HWE, the results of which are shown below in Table 4.1 and 4.2, respectively.

The PIC values of the markers tested in both data sets were moderate to high, with the average PIC estimates exceeding 0.5. The markers ETH3 and INRA23 had PIC estimates below 0.5, in both the complete and subset data set.

Table 4.1 PIC estimates per locus for the complete and subset data

Chromosome	**Locus**	**PIC**	
		Complete data set	**Subset data set**
1	BM1824	0.603	0.575
2	BM2113	0.759	0.769
5	ETH10	0.609	0.582
9	ETH225	0.726	0.723
19	ETH3	0.434	0.468
3	INRA23	0.381	0.429
15	SPS115	0.769	0.776
21	TGLA122	0.745	0.719
22	TGLA126	0.621	0.608
18	TGLA227	0.739	0.755
16	TGLA53	0.648	0.665
	Mean	0.64	0.643

* Mean PIC estimated per loci across all contributing herds

The microsatellite markers were tested for deviations from HWE for both data sets. The complete data observed that all loci, except for ETH10 and ETH3, deviated from HWE. The results for the subset observed all loci, except one (TGLA227), in HWE (P>0.05).

Table 4.2 Hardy-Weinberg Equilibrium exact test results per locus for the complete and subset data

Locus	Complete data set		Subset data set	
	P-value	**S.E.**	**P-value**	**S.E.**
BM1824	**0.029**	0.015	0.255	0.017
BM2113	**0.000**	0.000	0.124	0.016
ETH10	0.111	0.026	0.371	0.042
ETH225	**0.000**	0.000	0.154	0.028
ETH3	0.069	0.012	0.240	0.017
INRA23	**0.000**	0.000	0.51	0.052
SPS115	**0.000**	0.000	0.758	0.031
TGLA122	**0.009**	0.006	0.612	0.037
TGLA126	**0.032**	0.008	0.089	0.013
TGLA227	**0.000**	0.000	**0.003**	0.001
TGLA53	**0.000**	0.000	0.582	0.026

* P-Values in bold did not adhere to HWE (P<0.05)

Following marker evaluation, both data sets were evaluated for inbreeding and population differentiation per locus, attached as Addendum A. Both data sets were evaluated for observed and expected heterozygosity, inbreeding and population differentiation presented in Table 4.3.

.

Table 4.3 Estimates for heterozygosity, inbreeding and population differentiation for the complete and subset data

Data set	Number of herds	Average Number of alleles	H_E	H_O	F_{IS} (*f*)	F_{ST} (*θ*)
Complete data set	36	10	0.684 ± 0.035	0.663 ± 0.003	-0.014	0.053
Subset data set	22	8.45	0.687 ± 0.032	0.661 ± 0.006	-0.009	0.049

*H_E: expected heterozygosity, H_O: observed heterozygosity, F_{IS}: inbreeding coefficient and F_{ST}: population differentiation index

In both data sets the mean heterozygosity estimates exceeded 0.5, indicating relatively high genetic diversity, for the Herefords included in this study. Wright's F-statistics for both data sets indicated no reduction in heterozygosity, with both the average F_{IS} values low and negative indicative of little to no inbreeding.

An AMOVA analyses was performed to further define the genetic variation partitioning in the complete data set (Table 4.4) indicating 95% of the genetic diversity observed within the herds and only 5% observed between herds.

Table 4.4 AMOVA output analyses for the complete data set

Parameters	**Complete data**			
	df	**SS**	**Est Var**	**%**
Among Pops	35	729.785	0.204	5
Within Indiv	1798	6560.500	3.649	95
Total	1833	7290.285	3.853	100

*df: degrees of freedom, SS: sum of squares, Est Var: variance components and %: percentage variation

4.1.2 Population Structure

Population structure and admixture was analysed for the subset data set and are summarised in Table 4.5. It is clear that Cluster 2 consists mainly of cattle from Herd R, while Herds F, U and V contributed significantly to Cluster 4.

Table 4.5 The subset data's proportion of membership in each of the six inferred clusters in the structure program

Predefined Herds	**Number of Individuals**	**Inferred Clusters**					
		1	**2**	**3**	**4**	**5**	**6**
Herd A	25	0.131	0.153	0.188	0.075	0.205	0.248
Herd B	25	0.245	0.107	0.199	0.043	0.13	0.277
Herd F	18	0.073	0.152	0.133	**0.508**	0.061	0.072
Herd G	25	0.281	0.098	0.194	0.07	0.115	0.242
Herd H	25	0.143	0.067	0.206	0.067	**0.312**	0.204
Herd I	25	**0.316**	0.094	0.131	0.064	0.176	0.219
Herd N	19	0.161	0.147	0.234	0.054	0.262	0.142
Herd O	25	0.121	0.151	**0.341**	0.165	0.118	0.103
Herd P	25	0.12	0.092	0.246	0.249	0.173	0.12
Herd Q	25	**0.317**	0.053	0.119	0.044	0.195	0.273
Herd R	20	0.039	**0.757**	0.065	0.052	0.024	0.063
Herd U	18	0.087	0.187	0.111	**0.464**	0.054	0.098
Herd V	25	0.063	0.076	0.07	**0.668**	0.038	0.085
Herd W	25	0.16	0.122	0.212	0.121	0.179	0.206
Herd Y	18	0.179	0.131	0.144	0.086	0.276	0.185
Herd AA	16	0.156	0.18	0.128	0.198	0.041	0.297
Herd BB	25	0.074	0.254	0.221	0.257	0.098	0.097
Herd CC	25	0.151	0.147	0.232	0.116	0.2	0.153
Herd DD	25	**0.44**	0.102	0.125	0.06	0.082	0.191
Herd FF	25	0.148	0.073	0.139	0.062	**0.498**	0.08
Herd GG	16	0.197	0.149	0.103	0.096	0.106	**0.349**
Herd HH	25	0.147	0.115	0.228	0.114	0.182	0.213

*Bold values indicate a Herd with ≥30% proportion of membership in a specific cluster

In Figure 4.1 the estimated probabilities (Ln Pr) of the number of true populations (K) was shown.

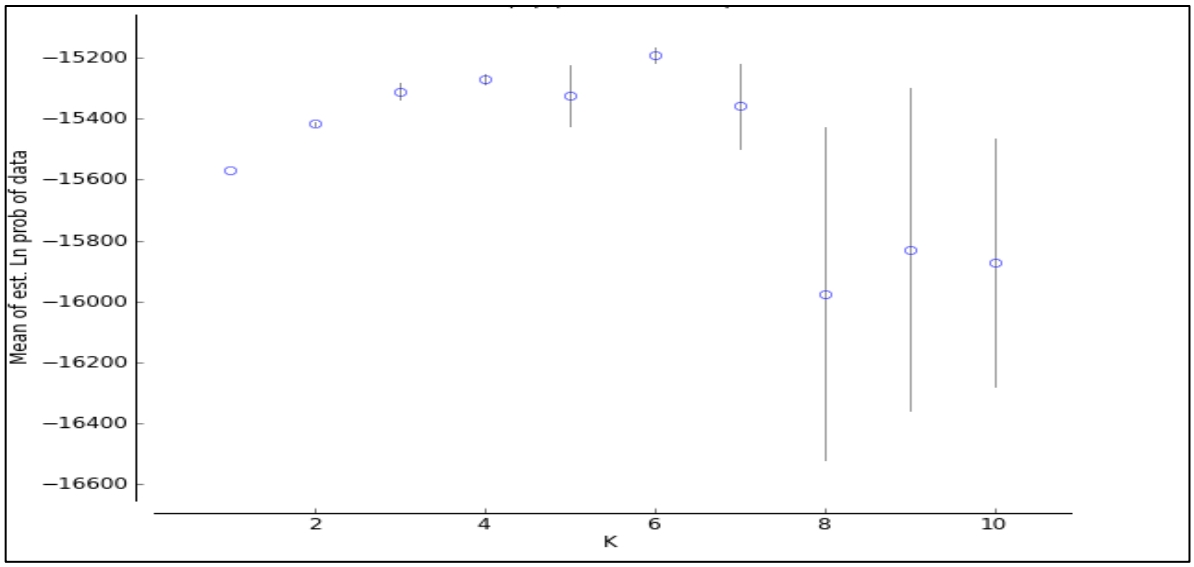

Figure 4.1 Plot of estimated probabilities of the data for different numbers of inferred clusters (K=2 to 9), with representation of probabilities for the mean of 12 runs at each K

In Figure 4.2 Ln Pr (XlK) graph showed gradual decline from K=1 to K=5, after which it peaks at K=6, then once again steadily declines from K=6 to K=9. Therefore K=6 was identified as the most probable number of inferred populations

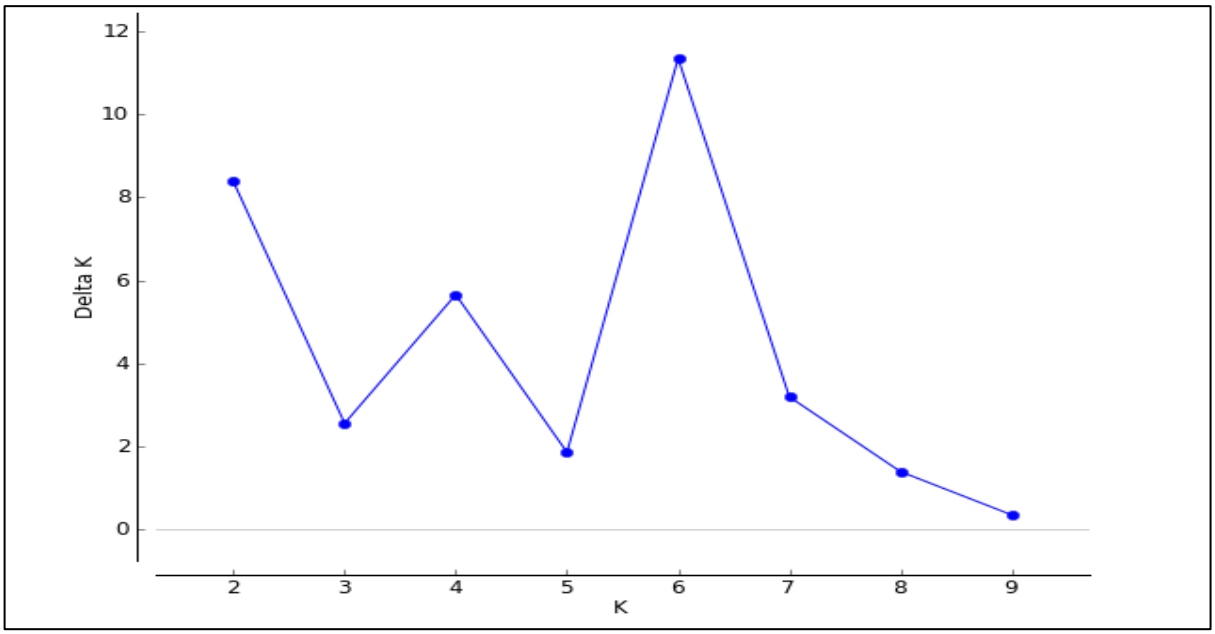

Figure 4.2 A delta K plot, indicating the choice of the appropriate K-value for the subset data

The bar graph for the subset population structure illustrates population stratification for 6 assumed ancestral populations (Figure 4.3). The structure plot elucidates the population structure of the 500 selected individuals, grouped into six clusters. Each of the individual clusters is distinct and describes the genetic variability within each cluster.

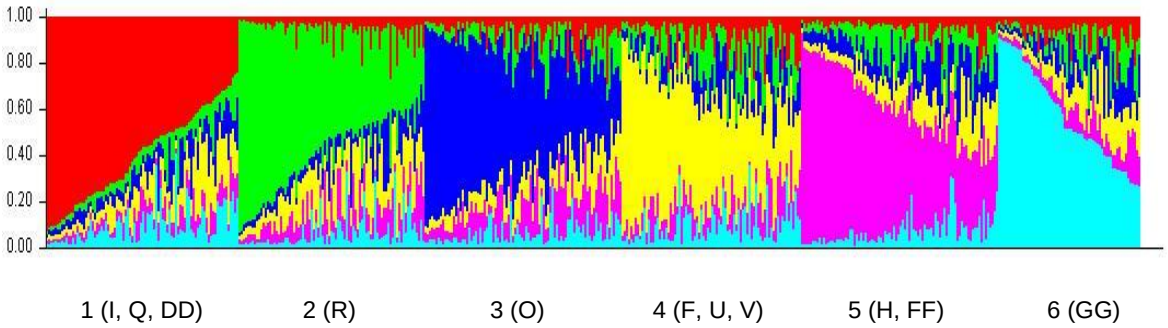

Figure 4.3 Population structure plot showing the proportions of contributing herds for each individual for k=6 for the subset data set

4.2 SNP Data

4.2.1 South African Data

4.2.1.1 Genetic diversity analysis

Sample-based quality control was implemented to exclude samples with a high level of missing genotypes and marker-based quality control to ensure the maximum number of informative markers is utilized for downstream analysis. No individuals were removed as a result of sample based QC, using a sample call rate of <98% as the threshold. A total of 15 755 SNPs were removed as a result of marker-based QC. The following is a breakdown of the results of marker-based QC: 0 SNPs removed due to low call rate (<98%); 15 037 removed due to low MAF (<5%) and 718 SNPs removed due to HWE (P<0.001) violation. A total of 116 498 SNPs passed QC and were validated for downstream analysis.

The mean expected (H_E) and observed (H_O) heterozygosity and individual inbreeding coefficient (F) were 0.386± 0.112, 0.387± 0.117 and -0.002± 0.051, respectively. The summary statistics on a per-chromosome basis for the SA data set are summarized in Table 4.6.

Table 4.6 A summary of statistics for autosomal SNPs for the SA data set

Chromosome	Number of SNPs	Average interval between SNPs (bp) per chromosome	Mean MAF	Mean Heterozygosity
1	7 184	91 382	0.299	0.389
2	6 183	94 915	0.297	0.390
3	5 489	91 400	0.292	0.381
4	5 351	95 736	0.302	0.399
5	5 832	81 617	0.308	0.398
6	6 430	78 737	0.293	0.381
7	5 434	84 013	0.293	0.383
8	4 874	95 597	0.296	0.384
9	4 838	92 818	0.308	0.392
10	4 780	90 196	0.303	0.387
11	4 925	89 628	0.302	0.388
12	4 074	87 427	0.296	0.385
13	3 705	91 137	0.292	0.386
14	4 216	76 658	0.285	0.374
15	3 874	89 965	0.301	0.396
16	3 659	89 923	0.296	0.386
17	3 375	96 152	0.299	0.383
18	2 963	90 903	0.291	0.382
19	2 906	89 866	0.294	0.384
20	3 727	73 841	0.302	0.390
21	3 258	81 293	0.292	0.384
22	2 805	90 466	0.296	0.386
23	2 415	91 910	0.299	0.392
24	3 225	73 936	0.295	0.382
25	2 016	86 213	0.307	0.396
26	2 426	87 004	0.288	0.376
27	2 092	85 631	0.304	0.393
28	2 109	89 015	0.295	0.389
29	2 333	90 613	0.298	0.396

The mean number of SNPs observed per chromosome was 4017, observing an average interval of 87 862bp between SNPs on each chromosome. The MAF and H_O per chromosome was 0.298 and 0.387 and LD was estimated at 0.487± 0.238 for the SA data set.

4.2.1.2 Genetic relatedness between individuals and Population structure

Principle component analysis

In order to evaluate the genetic relatedness of animals in the South African data set, a principle component analysis (PCA) was performed on the sample set of 184 Hereford animals. The principle

components, PCA 1 and PCA 2 were graphed against each other, with the results in Figure 4.4.

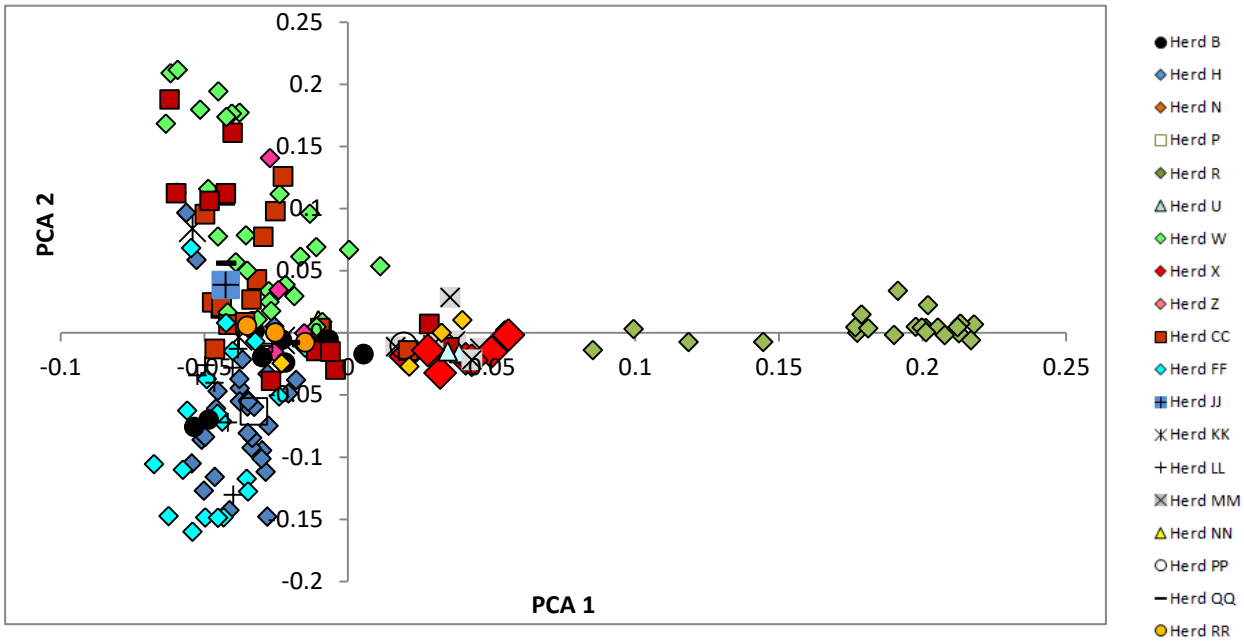

Figure 4.4 PCA illustrating the relationships among the 184 SA Hereford cattle

Figure 4.4 illustrates that the SA data set had a disperse data spread. The majority of samples fall to the left of PC1, representing 15 of the 19 herds. Herd H, R and FF form very distinct clusters, whereas Herds W and CC form a looser cluster. The remainder of the herds do not form clearly observable clusters. The majority of the individuals, with the exception of the indicated herds, are genetically related and a moderate degree of homogeneity is observed within the SA data set.

Population Structure

Admixture was utilized to evaluate population structure based on the shared familial SNP genotype percentages. It utilizes cross-validation, in order to identify the ideal K value, which is representative of the true quantity of genetic familial populations. The lowest cross-validation (CV) value is indicative of the most appropriate K value. Cross-validation error values for K values ranging from 1-10 were run and the ideal K value was identified at 7. The proportion of membership each herd allocated, to each of the seven clusters, is shown in Table 4.7.

Table 4.7 The proportion of herds allocated to each of the seven inferred clusters

Predefined Herds	Number of animals	Inferred Clusters						
		1	2	3	4	5	6	7
B	9	0.072	0.075	0.041	0.265	0.037	**0.413**	0.096
H	31	0.043	**0.397**	0.085	0.184	0.016	0.178	0.097
N	4	0.025	0.159	0.125	0.116	0.199	0.218	0.158
P	1	0.040	**0.355**	0.043	0.075	0.010	0.276	0.201
R	23	0.007	0.008	0.011	0.045	**0.818**	0.020	0.091
U	2	0.011	0.024	0.006	0.109	0.085	0.037	**0.728**
W	28	0.059	0.099	**0.480**	0.046	0.031	0.089	0.196
X	7	0.004	0.031	0.018	0.029	0.064	0.059	**0.796**
Z	4	0.148	0.058	0.195	0.042	0.056	**0.367**	0.134
CC	27	**0.337**	0.087	0.097	0.073	0.055	0.270	0.081
FF	20	0.016	**0.461**	0.114	0.183	0.009	0.180	0.038
JJ	1	0.000	0.000	0.194	0.000	0.000	**0.806**	0.000
KK	3	0.017	**0.401**	0.133	0.157	0.010	0.243	0.039
LL	10	0.070	0.223	0.066	0.207	0.019	**0.359**	0.056
MM	6	0.063	0.039	0.039	0.064	0.138	0.060	**0.595**
NN	1	0.112	0.001	0.029	0.055	0.085	**0.553**	0.166
PP	1	0.007	0.048	0.050	0.015	0.054	0.118	**0.709**
QQ	3	0.000	0.000	0.115	0.027	0.021	**0.812**	0.025
RR	3	0.059	0.174	0.114	0.028	0.057	**0.333**	0.234

*Bold values indicate a Herd with ≥30% proportion of membership to a specific cluster

Population structure assuming 7 inferred populations is depicted in Figure 4.5.

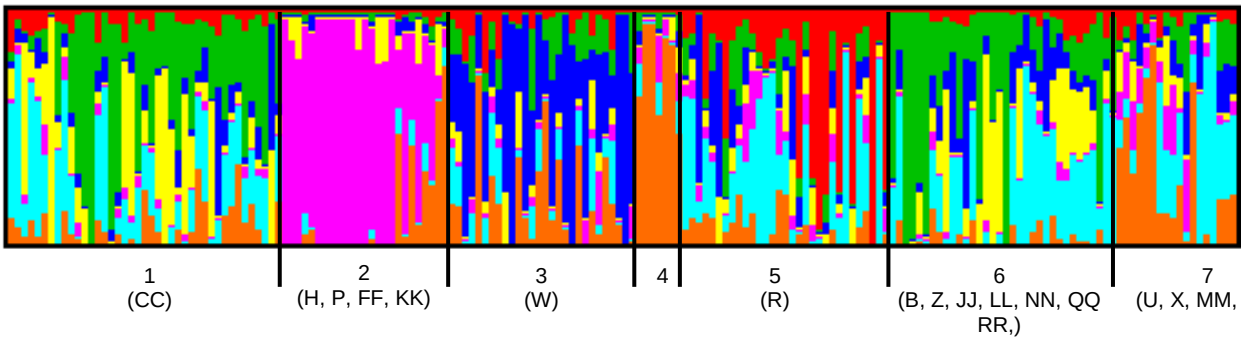

Figure 4.5 Population structure plot showing the proportion of inferred populations for each individual, for k=7, organised according to cluster for the SA data set

Figure 4.5 elucidates the South African population structure organized according to K=7 ancestral populations. The structure plot organized according to herd supports the PCA plot, with herd R forming a distinct cluster and Herd H, W, CC and FF forming more diverse clusters. The cluster

formation for herd H, W, CC and FF is expected given the overlap of these herds which can be seen in the PCA analysis. The structure plot organized according to K=7 clusters, observes that although the SA data set forms distinct clusters, there is a moderate degree of diversity within each of the specified clusters.

4.2.2 Irish data

4.2.2.1 Genetic diversity analysis

A sample call rate of <98% was used as the threshold for sample-based QC, resulting in the removal of 11 individuals. Marker-based QC removed (1) 38 527 SNPs due to low call rate (<98%); (2) 162 216 due to low MAF (<5%) and (3) 14 133 SNPs due to HWE (P<0.001) violation, resulting in a total of 214 876 SNPs removed. A total of 563 032 SNPs passed QC and were validated for downstream analysis. There were 11 missing genotypes in the Irish data set and 72% of the SNPs passed marker-based QC for downstream analysis.

The mean expected (H_E) and observed (H_O) heterozygosity and individual inbreeding coefficient (F) are 0.384± 0.113, 0.387± 0.116 and -0.007± 0.034, respectively. The inbreeding coefficient indicated no presence of inbreeding, with the heterozygosity estimates implying a modest degree of genetic diversity within the Irish data set.

A per-chromosome analysis was performed to compute summary statistics for the Irish data set in Table 4.8.

Table 4.8 A summary of statistics for autosomal SNPs for the Irish data set

Chromosome	Number of SNPs	Average interval between SNPs (bp) per chromosome	Mean MAF	Mean Heterozygosity
1	34 994	19 002	0.2930	0.3840
2	29 539	20 011	0.3038	0.3926
3	27 297	19 162	0.2992	0.3921
4	26 133	19 551	0.2938	0.3855
5	26 157	19 491	0.2949	0.3826
6	26 919	19 075	0.2929	0.3866
7	24 972	19 251	0.3015	0.3922
8	20 531	22 969	0.2887	0.3803
9	24 130	19 028	0.3062	0.3974
10	24 412	18 152	0.3020	0.3937
11	25 863	17 254	0.2985	0.3889
12	19 248	19 124	0.2896	0.3855
13	16 047	20 983	0.2855	0.3784
14	16 263	21 391	0.2817	0.3729
15	18 975	19 133	0.2899	0.3842
16	18 522	18 418	0.2958	0.3869
17	17 518	18 550	0.3037	0.3998
18	15 637	18 029	0.2855	0.3792
19	15 281	17 642	0.2868	0.3789
20	17 254	17 876	0.2957	0.3911
21	16 258	18 056	0.2872	0.3787
22	15 066	17 295	0.2924	0.3899
23	12 475	18 235	0.2976	0.3870
24	14 564	18 327	0.2949	0.3845
25	10 572	17 525	0.3034	0.3935
26	12 374	17 655	0.2870	0.3840
27	10 957	17 738	0.2983	0.3904
28	10 444	17 995	0.2977	0.3906
29	11 828	18 280	0.2928	0.3882

The mean number of SNPs observed per chromosome was 19 318 with an average interval of 87 862bp between SNPs on each chromosome. The average MAF and H_O was 0.295 and 0.387, respectively and the LD was estimated at (0.380 ± 0.166) for the Irish data set.

4.2.2.2 Genetic relatedness between individuals and Population structure

Principle component analysis

The genetic relatedness among the 305 Irish Hereford was established using principle component analysis (PCA). The principle components, PCA 1 and PCA 2 were graphed against each other, with the results in Figure 4.7.

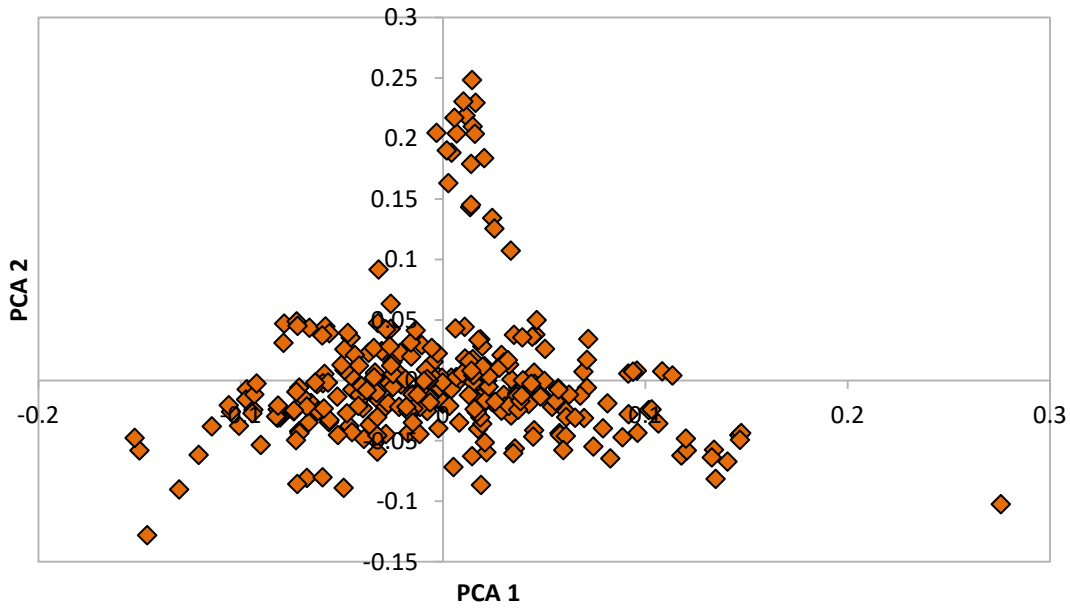

Figure 4.7 PCA illustrating the relationships among the 305 Irish Hereford cattle

A high degree of homogeneity can be observed with the majority of the individuals forming a loose cluster and only a few outlying individuals.

Population Structure

Admixture analysis was performed for the Irish data set. The CV error was not minimised within a range of K=1 to 20, indicating that the Irish data set represents a single population with minimal differentiation, which is supported by the PCA results.

4.2.3 Merged Data

The merged data set comprised of 500 samples and 127 568 common SNPs. The sample-based QC, removed eight individuals, with a sample call rate of <98% as the threshold. Marker-based

QC removed 20 054 SNPs as follows: 4 437 SNPs removed due to low call rate (<98%); 14 740 removed due to low MAF (<5%) and 877 SNPs removed due to HWE (P<0.001) violation. 107 514 SNPs satisfied the thresholds for marker-based QC and were validated for downstream analysis.

The results for genetic diversity evaluation were 0.387± 0.112, 0.385± 0.113 and 0.005± 0.04 for the expected (H_E) and observed (H_O) heterozygosity and individual inbreeding coefficient (F), respectively. The merged data set observed little to no inbreeding, with a moderate degree of genetic diversity.

The merged data set was analysed for summary statistics on a per-chromosome basis and the chromosome estimates for summary statistics can be seen in Addendum B. The mean number of SNPs observed per chromosome was 3 707, with observed averages of 0.299 and 0.385 for MAF and H_O, respectively. The LD for the merged data set was estimated at 0.503± 0.252.

4.2.3.1 Genetic relatedness between individuals and Population structure

Principle component analysis

PCA was computed for the 492 Hereford, in order to determine the genetic relationship between the two data sets. The principle components, PCA 1 and PCA 2 were graphed against each other, with the results in Figure 4.8.

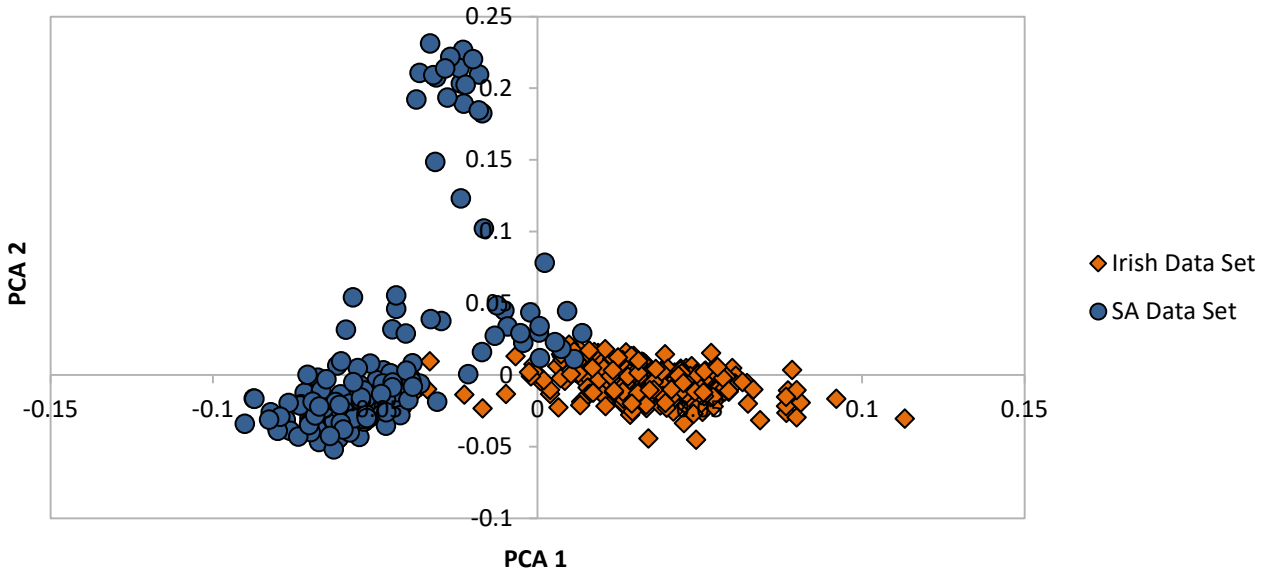

Figure 4.8 PCA illustrating the relationships among the 492 SA and Irish Hereford cattle

Figure 4.8 illustrates the two populations separating into two distinct clusters, with minimal overlap, indicating that the two populations are genetically distinct.

Population Structure

The results of the Admixture and PCA analyses, indicated that the merged data set would be best represented using K=2, the results are shown below in Table 4.9.

Table 4.9 The SA data sets proportion of membership in each of the seven inferred clusters in the admixture program

Data set	Number of animals	Inferred clusters	
		1	**2**
South Africa	184	0.210	**0.790**
Irish	308	**0.730**	0.270

*Bold values indicate a Herd with ≥30% proportion of membership to a specific cluster

The structure plot observes a clear separation between the Irish and SA Hereford data sets, which is shown in Figure 4.9.

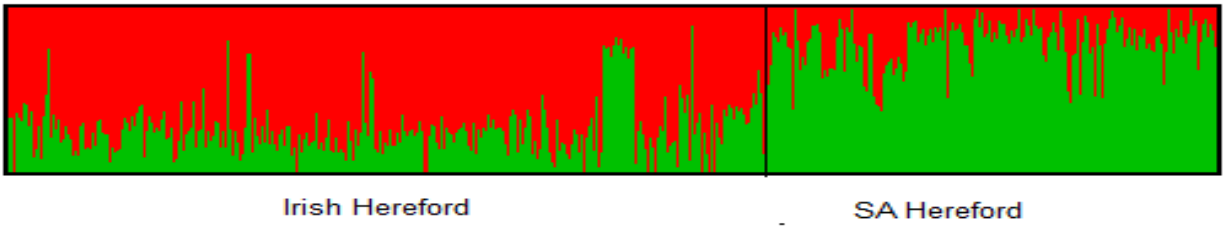

Figure 4.9 Population structure plot showing the proportion of inferred populations for each individual, for k=2, for the merged data set.

CHAPTER 5

DISCUSSION

5.1 Introduction

The survivability of a breed depends on its ability to compete and remain relevant in the current production environment and be able to adapt to changing markets (Groeneveld *et al.,* 2014; Biscarini *et al.,* 2015). Genetic diversity studies utilise phenotypic and genomic information to quantify and characterise the genetic variability present within a breed, which is important for optimal response to selection and critical for breed security (Ruane, 1999; Baumung *et al.,* 2004; Groeneveld *et al.,* 2014). Breeding and production of beef cattle can be effectively performed using the information garnered from genetic diversity studies to direct selection focus to achieve optimal production outcomes (Amer *et al.,* 2015; Kristensen *et al.,* 2015). Microsatellite markers were the marker of choice for molecular genetics, genetic diversity and parentage applications; however SNPs have rapidly superseded microsatellites for animal genetics applications (Yang *et al.,* 2013; Yaro *et al.,* 2016).

The formation of breed associations in the 19th century led to the initiation of performance recording schemes with the aim of genetic improvement of breeds (FAO, 2011). The South African Hereford has actively participated in animal recording since the establishment of national recording schemes in South Africa in the sixties (SA Studbook, 2016). Currently there are 56 registered Hereford herds listed in the SA Studbook Logix system, of which 42 are actively taking part and contributes performance data (SA Studbook, 2016). Pedigree completeness for the participating Hereford herds over the last 10 years ranged from 94.6% to 87.8% completeness for 1 to 6 generations deep, respectively, which shows a positive trend for pedigree recording in the breed (SA Studbook, 2016). An overall positive genetic trend for performance traits can be observed over a period of 36 years, with the population remaining stable for birth and mature weight performance and increasing for weaning weight, AFC and milk production performance (SA Studbook, 2016).

Since the development of DNA marker technology the genetic variation at the DNA level between different populations and individuals can be promptly and reliably identified (Yang *et al.,* 2013; Shalaby *et al.,* 2016). The use of DNA marker technologies in the selection process has allowed for the development of traits in animals in a more precise and time efficient manner (Ahmed & Khosa, 2010; Yaro *et al.,* 2016). Molecular information applied with care in selection programs has the potential to increase productivity, enhance environmental adaptation and maintain genetic diversity (Gündüz *et al.,* 2016; Shalaby *et al.,* 2016). DNA markers may realise these benefits with the opportunity to select at an early age; select for a wide range of traits and improve the prediction accuracy of mature phenotypes (Gündüz *et al.,* 2016).

The South African Hereford is one of 30 breeds used for beef production in South Africa. In this

study the aim was to quantify the genomic diversity, degree of inbreeding and population structure of the SA Hereford based on available microsatellite parentage profiles and whole genome genotypes generated by the BGP.

5.2 Microsatellite data

Microsatellite markers have been the preferred DNA markers for parentage testing and studies of genetic diversity in livestock species for the past two decades (Yang *et al.,* 2013; Yaro *et al.,* 2016). The polymorphic information content (PIC) of a marker determines the usefulness of the individual markers, with PIC values above 0.5 being highly informative (Qwabe, 2011; Suh *et al.,* 2014). In this study the average PIC value across the 11 microsatellite markers for the complete and subset data sets was 0.640 and 0.643, respectively. An average PIC of 0.765 was observed for Creole & commercial Brazilian cattle based on 22 microsatellite markers, similar to those achieved for this study (Egito et al., 2007) which had 10 markers in common with the current study. The markers used in this study were chosen from the ISAG list of recommend microsatellite for diversity studies, which were generated for parentage testing and therefore it was expected they be moderately to highly polymorphic, as this is a prerequisite for markers applied in parentage testing (Zhang et al., 2010).

The study revealed that the subset data set markers were in HWE equilibrium ($P>0.05$), with only one marker (TGLA227) deviating from HWE. The complete data set only had two markers which satisfied HWE parameters, namely ETH10 and ETH3, however the remainder of the markers deviated from HWE ($P<0.05$). The complete data set was generated for parentage testing, thus it contained related animals. The data set also had herd sizes ranging from 5 to 706 individuals; these factors could explain the deviations from HWE in the complete data set. HWE estimations are indicative of the evolutionary forces a population is exposed to; deviations from the estimate can be indications of possible selection, inbreeding and population stratification (Hartl & Clark, 1997; Qwabe, 2011).

The genetic diversity in the Hereford population in this study was relatively high with expected heterozygosity estimates for the complete and subset data sets of 0.684 ± 0.035 and 0.687 ± 0.032, respectively. Kantanen et al. (2000) used 10 microsatellites to genotype 20 North European cattle breeds, and found that the average observed heterozygosity was equal to 0.599, which was slightly lower than observed for this study. Similar results were observed in 8 British cattle breeds of which the Hereford was one, genotyped using 30 microsatellite markers (Wiener et al., 2004) where the average across breeds was a lower expected H_O (0.626), compared to this study.

Moderate to high heterozygosity estimates are indicative of natural selection and an exchange of genetic material between different populations. The SA Hereford breeders import semen from bulls from USA, Canada, Australia and Europe (SA Hereford Cattle Society, 2018). Migration by introduction of new alleles into the population is one of the approaches important for maintaining genetic diversity (Decker *et al.,* 2014). Genetic improvement of a breed should be carried out with a balance between rapid genetic progress and maintenance of genetic diversity (Khatib, 2015). The advent of technologies that allowed for animal germplasm exchange between countries has impacted the degree of genetic diversity present in cattle populations worldwide, influencing the degree of inbreeding and ability to adapt to changing environments (Cronin & Leesburg, 2016; Krehbiel, 2017).

The complete data set had an F_{IS} and F_{ST} of -0.014 and 0.053, respectively, and the subset data set had an F_{IS} and F_{ST} of -0.009 and 0.049, respectively. A low F_{ST} estimate (<0.05) is suggestive of a reduction in heterozygosity and a positive F_{IS} estimate is indicative of inbreeding (Hartl & Clark, 1997). The F_{IS} estimates for the complete and subset data set, observed low, negative F_{IS} nearing zero. The F_{ST} estimates were above the threshold that is indicative of heterozygosity loss. The F_{IS} and F_{ST} estimates for this study are indicative of limited inbreeding and moderate levels of genetic differentiation within the population. The AMOVA analysis indicated that 95% of the genetic diversity was within the population with the remaining 5% among the population. Similar results were observed in 11 Indian cattle breeds, genotyped using 21 microsatellite markers (Sharma *et al.*, 2015) with a reported F_{IS} at -0.091 and F_{ST} at 0.084, comparative with this study. Maretto & Cassandro (2016) used 24 microsatellite markers to genotype 279 Burlina cattle, reporting F_{IS} at 0.049 and F_{ST} at 0.133, observing a lesser degree of genetic differentiation and higher degree of inbreeding, compared to this study.

The analyses of the population structure of the herds in this study based on the microsatellite data indicated a moderate level of admixture where 10 of the herds (F, H, I, O, Q, R, U, V, FF and GG) assigned to specific clusters. The remainder of the herds was admixed and did not assign specifically to any one cluster. The pattern of clustering could be attributed to the close relationships among the animals within herds as parentage profiles were used. Furthermore, common ancestry could also have contributed to the clusters observed, as influential sires are often used across stud herds.

The markers used in the study were proposed by ISAG, which were developed for European cattle breeds, thus they were informative for the breed evaluated in this study (Friedrich, 2009; Pienaar *et al.,* 2014). The information gained from microsatellite and SNP markers, is analogous and comparative, and despite SNP based genotypes becoming more available, microsatellites still provide a viable and informative option for use in genetic studies (Coates *et al.,* 2009; Wakchaure *et al.,* 2015b). In developing countries like SA, the use of SNP genotyping is still relatively low, thus microsatellites remain a useful DNA technology for generating individual profiles and parentage testing (Ducrocq *et al.,* 2018). The microsatellite data indicated that the SA Hereford is genetically diverse, which is important to allow for adaptive differentiation between populations and control of inbreeding.

5.3 SNP data

Since the development of commercial SNP arrays, single nucleotide polymorphisms (SNPs) have gained popularity as a marker also used for genetic diversity studies (Negrini *et al.,* 2008; Lin *et al.,* 2010; Edea *et al.,* 2013). The application in genetic diversity studies in cattle has been demonstrated in a number of studies (Negrini *et al.,* 2008; Lin *et al.,* 2010; Edea *et al.,* 2013).

The SA beef genomics programme (BGP) was established to encourage and expand accurate measurements of phenotypes in SA beef cattle and to combine these phenotypes with genomic data to enhance EBV accuracies. The programme provides the opportunity to establish reference populations with genomic data for all the participating breeds (http://www.livestockgenomics.co.za/).

During the first three years, 13 breeds have participated including the SA Hereford (http://www.livestockgenomics.co.za/).

The establishment of reference populations is a key step in implementing genomic selection (GS), from which the prediction equation can be formulated, with which to select (Boichard *et al.,* 2016; Georges *et al.,* 2018; Meuwissen *et al.,* 2018). The benefit of GS is that candidates can be evaluated and selected at a young age, resulting in reduced generation interval (Boichard *et al.,* 2016; Georges *et al.,* 2018; Meuwissen *et al.,* 2018). Genomic selection can be utilised to balance progress and diversity simultaneously (Heslot *et al.,* 2013).

In this study the SA Hereford had 184 genomic genotypes available for diversity analyses. The average expected and observed heterozygosity estimates for the SA data were 0.386 and 0.387, respectively. The estimates obtained were comparable with estimates of 0.25 and 0.30 for expected heterozygosity by Makina *et al.* (2014) and higher than the study by Zwane et al. (2016) of 0.24 for H_E and 0.20 for H_O for the SA Hereford. The heterozygosity estimates obtained for the SNP data supported that obtained for the microsatellite data, indicating a moderate degree of heterozygosity in the SA Hereford.

The SA Hereford diversity was comparable with the Irish Hereford with observed average expected and observed heterozygosity estimates of 0.384 and 387, respectively. The inbreeding coefficient for the SA population supported the observed heterozygosity estimates, with a low, negative estimate indicating minimal to no population inbreeding. The inbreeding coefficient for the South African data set was -0.002. The inbreeding estimate obtained for this study compared favourably with the estimate F range: -0.026- 0.005 for six South African cattle breeds by Makina *et al.* (2014) and inbreeding coefficient of -0.02 for the SA Hereford by Zwane *et al.* (2016). The SA Hereford compared comparatively with the Irish Hereford with an inbreeding coefficient of -0.007.The discrepancies between the H_E and H_O estimates are small but both data sets observe $H_O > H_E$ and the individual estimates are comparatively high to the indicated studies, this coupled with the inbreeding coefficient estimates, suggests genetic diversity within the breed is favourable. The SA Hereford is genetically diverse and maintaining levels of diversity comparative with international herds, indicative of sound selection programmes (Hedrick, 2015; Hill, 2016). The level of genetic diversity is an asset in terms of promoting the SA Hereford as a source of germplasm for export (Heslot *et al.,* 2013; Ibtisham *et al.,* 2017). The animals genotyped in this study are also recommended as candidates for inclusion in the reference population (Heslot *et al.,* 2013; Ibtisham *et al.,* 2017).

The average MAF estimate for the SA data set was 0.298. Similar estimates were obtained of 0.27 for indigenous South Africa breeds and 0.30 for the SA Hereford by Zwane *et al.* (2016). The Irish data set observed a similar MAF estimate of 0.295. The SA and Irish data sets both had the highest proportion of SNPs in the MAF interval of 0.4-0.5, which is indicative that the two populations are highly polymorphic. The minor allele frequency is an indicator of the presence of rare alleles which are important for conservation of breed genetic diversity (Panagiotou *et al.,* 2010; Eynard, 2018). Evaluation and characterisation of MAF can direct selection focus for inclusion of rare variants, which are necessary to ensure long term response to genomic selection (Engelsma, 2012; Eynard, 2018).

The average LD r^2 estimate was 0.487 for the SA data set. The LD estimates in comparative studies are considerably lower compared to the estimates obtained in this study with an r^2 estimate of 0.20, 0.27 and 0.19 for Pinzgau, Brown Swiss and Nelore cattle by Kasarda *et al.* (2016) and r^2 estimates of 0.21 and 0.16 for Hereford and Braford by Biegelmeyer *et al.* (2016). The Irish Hereford r^2 estimate observed was higher than that of the SA Hereford at 0.503. LD is a measure of the correlation of genetic variation at two or more loci (Khatib, 2015). Linkage disequilibrium can arise from interbreeding of populations with different allele frequencies or selection increasing specific allele frequencies (Lashmar *et al.,* 2015). High LD around selected alleles can results in haplotype fixation, ultimately reducing genetic variation (Engelsma, 2012; Sudrajad *et al.,* 2017). Ideal r^2 estimates will be closer to 0 than 1, with higher LD estimates indicative of a lesser degree of genetic variation present and is cause for concern with regard to conservation of genetic diversity (Lashmar *et al.,* 2015; Biegelmeyer *et al.,* 2016; Sudrajad *et al.,* 2017). However, a low to moderate degree of LD is desirable to ensure long term response to selection, with estimates of 0.2 considered sufficient for genomic selection to be successful, assuming desirable association of marker with QTL (Lashmar *et al.,* 2015). The LD estimate for the SA Hereford was moderate; this could be attributed to LD decay occurring at a slower rate in Taurine breeds, such as the Hereford, due to recent intense selection for production (O'Brien *et al.,* 2014; Biegelmeyer *et al.,* 2016). The r^2 estimates imply that greater accuracy of marker predictions can be achieved and that sufficient LD is present for implementation of genomic selection. .

Principle component analysis identifies genetically related samples with the principle components representing uncorrelated samples produced from potentially correlated SNPs (Yang *et al.,* 2011; Ojango *et al.,* 2011; Lashmar *et al.,* 2015). The SA Hereford formed a loose cluster, with the samples partitioning distinctly according to PC 1. The herds R, X, MM and PP all come from horned Hereford breeder. These herds separated distinctly from the remainder of the 15 herds, all falling to the right of PCA 1. The herds formed distinct clusters, with herds X, MM and PP showing a greater association with each other than that of herd R. Herds H, P, FF and KK showed a strong genetic association, with these herds showing some overlap. Herds W and CC were significant contributors of genetic material to SA Hereford breeders, which explains the wide spread overlap of these herds with the data set. The clustering and degree of overlap seen in the PCA, suggests a low to moderate degree of homogeneity in the SA Hereford. The Irish Hereford formed a distinct cluster, with one outlying cluster. The Irish Hereford has a higher degree of homogeneity compared to the SA Hereford. The merged data set showed the SA and Irish Hereford form distinct clusters with some overlapping samples. The SA and Irish Hereford, although forming distinct clusters, do share admixture with some overlap, indicating a small degree of genetic relatedness, but shared genes as a breed.

ADMIXTURE is a model-based approach which estimates the most probable number of ancestral populations (Alexander *et al.,* 2009). The PCA and admixture analysis for the SA Hereford supported each other. The admixture plot showed a similar clustering pattern, with herd R forming a distinct cluster and herds X, MM and PP assigning collectively to one cluster, which represent the horned herds. The herds H, P, FF and KK clustered together loosely in the PCA, similarly assigned

collectively to one cluster in the Admixture. The herds W and CC assigned specifically to one cluster and did not share admixture with other herds. The admixture clusters where multiple herds assigned, had more diverse composition, which is to be expected given the remaining herds, did not form strong associative clusters as seen in herd R, but had more disperse data spreads and overlapped with at least one other herd. The admixture plot showed that the majority of the individuals had diverse compositions in terms of contributory populations, except for herd R, W and CC, expected given the moderate degree of genetic relatedness and homogeneity among the analysed herds, apart from the horned herds. The admixture analysis for the Irish Hereford indicated that the data set was one homogenous population, supporting the PCA analysis. The admixture analysis for the merged data supported the PCA analysis, with the SA and Irish Hereford clustered separately, with each data set assigned to a specific inferred cluster. The data set included influential sires, which confirms the relatedness observed in the admixture analyses; additionally some herds contained at least 48 individuals explaining the clustering shown for the distinct herds. The genetic linkage present in the data set is important for development of a reference population.

The microsatellite and SNP data have 12 Hereford herds in common. The microsatellite and SNP population structure analyses support each other. The number of inferred clusters was similar across marker analyses with six and seven inferred clusters for the microsatellite and SNP data, respectively. The structure analyses across the two markers showed that herds H, R and FF formed distinct herd clusters, which did not share a close genetic relationship with the remainder of the herds. However, the remaining Hereford herds shared a close genetic relationship, assigning evenly across the inferred clusters. The population structure results across marker analyses, indicates that the SA Hereford has a low to moderate degree of homogeneity and that conservation of genetic diversity within the breed can be achieved through introduction of germplasm, from local genetically distinct Hereford herds in SA or international imports (Khatib, 2015; Eynard, 2018).

5.4 Conclusion

This study provides a reference of genomic information for the SA Hereford, which can be utilised for setting up the reference population for genomic selection. The animals genotyped as part of this study show sufficient genetic diversity and compare comparatively with international herds. The exchange of genetic material between SA and international Hereford herds can advance population genetic diversity and initiate the development of genetic linkage between populations. Genetic diversity can be exploited for further progress in genetic improvement for the SA Hereford population and combat the potential for inbreeding.

CHAPTER 6

CONCLUSION & RECOMMENDATION

6.1 Conclusion

In this study, the South African Hereford breed was characterised by analysing microsatellite and genotypic data. Eleven microsatellite markers were utilised to evaluate genetic diversity and population structure. The results revealed a moderate degree of heterozygosity and limited inbreeding. The FST and AMOVA analysis indicated that the majority of genetic variation occurred within the population, and the population structure analysis revealed the breed differentiated into six sub-populations.

The genotypic data set comprised of 184 Hereford, genotyped as part of the beef genomics programme, with the aim of compiling breed specific reference populations. The results of the SNP analysis indicated a moderate degree of heterozygosity, with no cause for concern with regard to inbreeding. The LD estimate was moderate; the small number of genotypes available with a high degree of relatedness could have contributed to the moderate estimate. PCA analysis showed a close genetic relationship among the 184 animals; however five clusters tended to be more genetically distinct. Population structure results revealed the breed differentiated into seven sub-populations. The results supported the PCA analysis with the three genetically distinct herds, forming distinct clusters.

The Irish Hereford genotypic data was evaluated for comparison purposes. The genetic diversity analysis results confirmed that the SA Hereford is comparable in diversity on an international level. PCA and population structure analysis revealed that the two Hereford populations formed distinct clusters, indicating the two populations were genetically diverse, with shared genomic regions.

The microsatellite and SNP evaluation results supported each other. The Hereford breeds ability to respond to selection for genetic improvement and prevention of inbreeding is dependent on the degree of genetic diversity within the population. This study serves as a reference for further genetic studies on the South African Hereford.

6.2 Recommendations

Breed improvement depends on the accuracy of genetic diversity estimation (Groeneveld *et al.,* 2014; Biscarini *et al.,* 2015). Improved accuracy can be achieved through analysis of larger sample sizes from various geographic locations, in order to be truly representative of population diversity. The SA Hereford, as mentioned, has a relatively small productive population in comparison to other prominent breeds in SA and could benefit from the establishment of larger production herds. Larger production herds mean a larger available genetic pool from which to select samples for scientific studies, with the additional benefit of a higher level of genetic variation experiencing greater gains, imperative for the future production efficiency of the breed.

Encouraging international collaboration is important for breed improvement and competitiveness. The breeding focus in SA is centred on calf crop and the growth performance of the calf crop, with important traits including low BW for ease of calving; weaning weight for rate of gain; post-wean rate of gain; feed conversion efficiency; mature mass and mothering ability, thus reproduction and growth should be the main focus of diversity evaluation and selection. Imported purebred bulls and cows are the stock utilised to establish most beef herds in SA, with the combined utilisation for upgrading leading to an increase in the available genetic resource base compared to that of foreign countries.

The utilisation of SNP markers should be encouraged, with its comparative benefits to other molecular markers (Fernández *et al.,* 2013; Brenig & Schütz, 2016). Genetic selection tools allow for directional selection aimed at optimising heterozygosity, to mitigate the effects of shorter generation intervals and allowing for the management of genetic resources, to determine the degree of genetic variation or inbreeding present in populations. SNP markers can be utilised to select for increased heterozygosity in genes for adaptation to climate change or breed for a reduction in gene diversity loss through evaluation of shared ancestral genes.

EBV systems have been successful thus far but the estimates are an indirect measure of animal performance. Genomic information allows for individual animal characterization, which can produce an animal specific SNP profile through genotyping. The inclusion of genomic information to that of complete pedigree records can be utilized to reveal the genetic structure of the SA Hereford. Special reference can be made for traits that as of yet have had little characterization or undergone minimal genetic improvement and breed genetically diverse individuals, which match the purpose of commercial production.

In order to better understand breed diversity, it would be ideal to include reference samples, evaluated with a standardized marker set, allowing for the combination or comparison of individual studies (Erhardt & Weimann, 2007). The data can ultimately be utilised to preserve the genetic resources that the breed offers for application in genomic selection. The construction of reference populations should include a minimum of 1000 high impact animals, preferably using breeds that align with local agro-climatic environments and markets (Theron *et al.,* 2014; Ducrocq *et al.,* 2018)

Further genetic diversity evaluations of the SA Hereford should be performed at regular intervals to keep track of genetic diversity levels within the population, allowing for the implementation of breeding or management strategies to mitigate the potential effects of a loss in genetic diversity.

REFERENCES

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

Ahmed, S., & Khosa, A.N., 2010. An introduction to DNA and their role in livestock production: A review. J Anim Plant Sci. 20 (4), 305- 314.

Ai, H., Huang, L., & Ren, J., 2013. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. PLOS. 8 (2), 1-11.

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

Amer, P., Allain, D., Avendano, S., Baslega, M & Boettcher, P., 2015, Breeding strategies and programmes. Animal Science White Papers. Technical Reports and Fact Sheets. 6.

American Hereford Association, 2016. http://hereford.org/node/47 accessed on the 8 July, 2016.

American Hereford Association, 2017. http://hereford.org/node/47 accessed on the 24 January, 2017.

ARC Annual Beef Bulletin, 2016. The Growth Prospects of the SA Beef Industry. 22- 24.

Arthur, J.P.F. & Herd, R.M., 2008. Residual feed intake in beef cattle. R Bras Zootec. 37, 269- 279. (In Brazilian, English abstract)

Baker, R.L., Morris, C.A., Johnson, D.L., Hunter, J.C. & Hickey, S.M., 1991. Results of selection for yearling or 18-month weight in Angus and Hereford cattle. Livest Prod Sci. 29, 277-296.

Balloux, F. & Lugon-Moulin, N., 2002. The estimation of population differentiation with microsatellite markers. Mol Ecol. 11, 155- 165.

Barendse, W., Armitage, S.M., Kossarek, L.M., Shalom, A., Kirkpatrick, B.W., Ryan, A.M., Clayton, D., Li, L., Neibergs, H.L., & Zhang, N., 1994. A genetic linkage map of the bovine genome. Nat Genet. 6 (3), 227- 235.

Baumung, R., Simianer, H & Hoffmann, I., 2004. Genetic diversity studies in farm animals – a survey. J Anim Breed Genet. 121, 361-373.

Bergh, L., 2008. Management and recording guidelines for accurate and reliable breeding values. Appl Anim Husb Rural Dev. 1, 10-15.

Berry, D.P., & Crowley, J.J., 2013. CELL BIOLOGY SYMPOSIUM: Genetics of feed efficiency in dairy and beef cattle. J Anim Sci. 91, 1594- 1613.

Berry, D.P., Garcia, J.F. & Garrick, D.J., 2016. Development and implementation of genomic predictions in beef cattle. Anim Front. 6 (1), 32- 38.

Beuzen, N.D., Stear, M.J. & Chang, K.C., 2000. Molecular markers and their use in animal breeding. Vet J. 160 (1), 42-52.

Bhatia, G., Patterson, N., Sankararaman, S. & Price, A.L., 2013. Estimating and interpreting FST: The impact of rare variants. Genome Res. 23 (9), 1514- 1521.

Biegelmeyer, P., Gulias-Gomes, C.G., Caetano, A.R., Steibel, J.P., & Cardoso, F.F., 2016. Linage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. BMC Genet. 17 (32), 1- 12.

Biscarini, F., Nicolazzi, E.L., Stella, A., Boettcher, P.J & Gandini, G., 2015. Challenges and opportunities in genetic improvement of local livestock breeds. Front Genet. 6 (33), 1- 7.

Bishop, S.C., 1992. Phenotypic and genetic variation in body weight, food intake and energy utilisation in Hereford cattle I. Performance test results. Livest Prod Sci. 30, 1-18.

Blott, S.C., William, J. & Haley, C., 1998. Genetic variation within the Hereford breed of cattle. Anim Genet. 29, 202-211.

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

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. J Anim Sci. 89 (6), 1684- 1697.

Bourdon, R.M., 2000. Chapter 4. In: Understanding Animal Breeding. Upper Saddle River, New Jersey. 56-65.

Bullock, K.D & Pollak, E.J., 2009. Beef Symposium: Evolution of Beef Cattle Genetic Evaluation. J Anim Sci. 87 (14), E1-E2.

Brenig, B. & Schütz, E., 2016. Recent development of allele frequencies and exclusion probabilities of microsatellites used for parentage control in the German Holstein Friesian cattle population. BMC Genet. 17 (18), 1- 9.

Bruford, M.W., Bradley, D.G. & Luikart, G., 2003. DNA markers reveal the complexity of livestock domestication. Nat Rev Genet. 4 (11), 900- 910.

Cammack, K.M., Thomas, M.G. & Enns, R.M., 2009. Reproductive traits and their heritabilities in Beef Cattle. PAS. 25 (5). 517- 528.

Cañas-álvarez, J.J., González-Rodriguez, A., Munilla, S., Varona, L., Diaz, C., Baro, J.A., Molina, A. & Piedrafita, J., 2015. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip. J Anim Sci. 93 (11), 5164- 5174.

Chen, B., Cole, J.W. & Grond-Ginsbach, C., 2017. Departure from Hardy Weinberg Equilibrium and Genotyping Error. Front Genet. 8 (167).

Chistiakov, D.A., Hellemans, B. & Volckaert, F.A.M., 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquacult. 255, 1-29.

Cronin, M.A., & Leesburg, V.L.R., 2016. Genetic variation and differentiation in parent-descendant cattle and bison populations. J Anim Sci. 94, 4491- 4497.

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

DAFF, 2015. A Profile of the SA Beef Market Value Chain. http://www.nda.agric.za/doaDev/sideMenu/Marketing/Annual%20Publications/Commodity%20Profiles/field%20crops/Beef%20market%20value%20chain%20profile%202015.pdf

DAFF, 2018. ABSTRACT of Agricultural Statistics. http://www.daff.gov.za/Daffweb3/Portals/0/Statistics%20and%20Economic%20Analysis/Statistical%20Information/Abstract%202018.pdf accessed on the 16 July 2018.

Decker, J.E., 2014. Decreasing generation interval to increase genetic progress. extension. www.eBEEF.org, accessed on the 30 January 2017.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Alcalá, A.M., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., Babar, M.E., de Almeida Regitano, L.C., Yildiz, M.A., Heaton, M.P., Liu, W., Lei, C., Reecy, J.M., Saif-Ur-Rehman, M., Schnabel, R.D. & Taylor, J.F., 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLOS. 10 (3), 1-14.

Dekkers, J.C.M. & Hospital, F., 2002. The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet. 3 (1), 22- 32.

Dekkers, J.C.M., 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lesson. J Anim Sci. 82, 313- 328.

Ducrocq, V., Laloe, D., Swaminathan, M., Rognon, X., Tixier-Boichard,M., & Zerjal, T., 2018. Genomics for Ruminants in Developing Countries: From Principles to Practice. Front Genet. 9 (251), 1- 7.

Eck, S.H., Benet-Pagès, A., Flisikowski, K., Meitinger, T., Fries, R. & Strom, T.M., 2009. Whole genome sequencing of a single Bos taurus animal for single nucleotide polymorphism discovery. Genome Biol. 10 (8).

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

Edea, Z., Bhuiyan, M.S.A., Dessie, T., Rothschild, M.F., Dadi, H. & Kim, K.S., 2014. Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. Anim. 9 (2), 218-226.

Egito, A.A., Paiva, S.R., do Socorro M Albuquerque, M., Mariante, A.S., Almeida, L.D., Castro, S.R. & Grattapaglia, D., 2007. Microsatellite based genetic diversity and relationships among ten Creole and commercial cattle breeds raised in Brazil. BMC Genet. 8 (83).

Engelsma, K.A., Calus, M.P.L., Bijma, P. & Windig, J.J., 2010. Estimating genetic diversity across the neutral genome with the use of dense marker maps. Genet Select Evol. 42 (12).

Engelsma, K.A., 2012. Use of SNP markers to conserve genome-wide genetic diversity in livestock. PhD Thesis. Wageningen University, Netherlands.

Erhardt, G., & Weimann, C., 2007. Use of molecular markers for evaluation of genetic diversity and in animal production. Arch Latinoam Prod Anim. 15 (1), 63- 66.

Eynard, S.E., 2018. Using genomic information to conserve genetic diversity in livestock. PhD Thesis. Wageningen University, Netherlands.

Falconer, D.S., & Mackay, T.F.C., 1996. Chapter 1 & 2. In: Introduction to Quantitative Genetics. 90 Tottenham Court Road, London. 5- 11 & 42-45.

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

FAO, 2011. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. 9.

Fernández, M.E., Goszczynski, D.E., Lirón, J.P., Villegas-Castagnasso, E.E., Carino, M.H., Ripoli, M.V., Rogberg-Muñoz, A., Posik, D.M., Peral-García, P. & Giovambattista, G., 2013. Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Genet Mol Biol. 36 (2), 185- 191.

Franke, D.E., Burns, W.C & Koger, M., 1975. Variation in the Coat-Color Pattern of Hereford Catte. J. Hered. 66 (3), 147- 150.

Freeman, A.R., Meghen, C.M., MacHugh, D.E., Loftus, T., Achukwi, M.D., Bado, A., Sauveroche, B. & Bradley, D.G., 2004. Admixture and diversity in West African cattle populations. Molec Ecol. 13, 3477-3487.

Friedrich, H., 2009. Evaluation of microsatellite markers for parentage verification in South African Angora goats. MSc (Agric) Thesis. University of Pretoria, South Africa.

Frkonja, A., Gredler, B., Schnyder, U., Curik, I. & Sölkner, J., 2012. Prediction of breed composition in an admixed cattle population. Anim Genet. 43 (6), 696- 703.

Gamaniel, I.B. & Gwaza, D.S., 2017. Molecular Characterization of Animal Genetics Resources, its Potential for use in Developing Countries. J Genet Genet Eng. 1 (1), 43-57.

Gargantini, Gl., Cundiff, L.V., Lunstra, D.D. & Van Vleck, L.D., 2005. Genetic relationships between male and female reproductive traits in beef cattle. PAS. 21, 195- 199.

Garrick, D.J. & Golden, B.L., 2008. Producing and using genetic evaluations in today's U.S. beef industry. J Anim Sci. 87 (14).

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

Garrick, D.J. & Saatchi, M., 2011. Opportunities and challenges for genomic selection of beef cattle. J Anim Sci. 40, 310-316.

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J., Boichard, D., Lathrop, G.M., Gut, I.G. & Eggen, A., 2007. Genetic and Haplotypic Structure in 14 European and African Cattle Breeds. Genet. 177 (2), 1059-1070.

Georges, M., & Massey, J. M., 1992. Polymorphic DNA markers in Bovidae. Patent WO 92/13102 1992.

Georges, M., Charlier, C., & Hayes, B., 2018. Genomic selection of livestock and beyond. GIGA.

Glaze, J.B. & Schalles, R.R., 1994. Heritabilities and genetic correlations for birth weight, weaning weight, and yearling weight in polled hereford cattle. Cattleman's day 1994.

Glaze, J.B., 2011. Genetic selection for fertility and performance. Proceedings: Applied Reproductive Strategies in Beef Cattle. 255- 262.

Goddard, M.E. & Hayes, B.J., 2007. Genomic Selection. J Anim Breed Genet. 124 (6), 323- 330.

Gororo, E., Makuza, S.M., Chatiza, F.P., Chidzwondo, F. & Sanyika, T.W., 2018. Genetic diversity in Zimbabwean Sanga cattle breeds using microsatellite markers. SA J Anim Sci. 48 (1), 128- 141.

Goudet,J., 1995. FSTAT (version 2.9.3): A computer programme to calculate F-statistics. J Hered. 8, 485-486.

Grobler, S.M., 2016. Alternative management systems to increase beef production under extensive conditions. PhD Thesis. University of the Free State, South Africa.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E., Weigend, S. & The GLOBALDIV consortium., 2010. Genetic diversity in farm animals – a review. Anim Genet. 41 (1), 6- 31.

Grundy, J.E., 2002. The Hereford Bull: His Contribution to New World and Domestic Beef Supplies. Agric Hist Rev. 52 (1), 69-88.

Gündüz, Z., Yilmaz, O., Cemal, I., & Biçer, O., 2016. A comparison of old and modern type DNA marker technologies and their impact of animal breeding programs. Univ J Inst Sci & Tech. 6 (2), 175- 180.

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

Hanotte, O. & Jianlin, H., 2005. Genetic characterization of livestock populations and its use in conservation decision making. FAO. 131- 136.

Hartl, D.L., & Clark, A.G., 1997. Chapter 3 & 4. In: Principles of Population Genetics. Sunderland, Masachusetts.95-107 & 117-120.

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

Hedrick, P.W., 2015. Heterozygote Advantage: The Effect of Artificial Selection in Livestock and Pets. J Hered. 106 (2), 141- 154.

Henryon, M., Berg, P. & Sørensen, A.C., 2014. Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains. Livest Sci. 166, 38-47.

Hereford Breeder.net, 2017. https://www.herefordbreeder.net/ accessed on the 25 January, 2017.

Hereford Namibia, 2017. http://www.hereford-namibia.com/info.html, accessed on the 26 January, 2017.

Heslot, N., Rutkoski, J., Poland, J., Jannik,J., & Sorrels,M.E., 2013. Impact of Marker Ascertainment Bias on Genomic Selection Accuracy and Estimates of Genetic Diversity. PLoS ONE. 8 (9), 1- 8.

Hill, W.G, 2016. Is continued genetic improvement of livestock sustainable? Genet. 202, 877-881.

Hoffmann, I., 2010. Livestock biodiversity. Rev Sci Tech Off Int Epiz. 29 (1), 73- 86.

Hoffmann, I. & Baumung, R., 2013. Chapter 3. In: Diversifying food and diets using agricultural biodiversity to improve nutrition and health. p68-78.

Holsinger, K.E., 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet. 10 (9), 639- 650.

Hussein, I.H., Alam, S.S., Makkawi, A.A.A., Sid-Ahmed, S.A., Abdoon, A.S. & Hassanane, M.S., 2015. Genetic diversity between and within Sudanese Zebu cattle breeds using microsatellite markers. Res Genet. 2015.

Ibtisham, F., Zhang, L., Xiao, M., An, L., Ramzan, M.B., Nawab, A., Zhao, Y., Li, G., & Xu, Y., 2017. Genomic selection and its application in animal breeding. Thai J Vet Med. 47 (3), 301- 310.

Irish Hereford Breed Society, 2017. https://irishhereford.com/, accessed 2017.

Jakobsson, M., Edge, M.D. & Rosenberg, N.A., 2012. The relationship between FST and the frequency of the most frequent allele. Genet. 193 (2), 515- 528.

Jones, J.M. & Jolly, R.D., 1982. Dwarfism in Hereford cattle: a genetic morphological and biochemical study. NZ Vet J. 30 (12), 185- 189.

Kantanen, J., Olsaker, I., Holm, L., Lien, S., Vilkki, J., Brusgaard, K., Eythorsdottir, E., Danell, B. & Adalsteinsson, S., 2000. Genetic diversity and population structure of 20 North European cattle breeds. J Hered. 91 (6), 446- 457.

Kasarda, R., Moravčiková, N., Mészáros, G., Trakovická, A. & Kadlečik, O., 2016. Genetic divergence of cattle populations based on genomic information. Sci Agric Biol. 47 (3), 113- 117.

Kasinathan, P., Wei, H., Xiang,T., Molina, J.A., Metzger, J., Broek, D., Kasinathan, S., Faber, D.C. & Allan, M.F., 2015. Acceleration of genetic gain in cattle by reduction of generation interval. Scientific Reports. 5 (8674).

Khanyile, K.S., Dzomba, E.F. & Muchadeyi, F.C., 2015. Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations in Southern Africa. Front Genet. 6 (13).

Khatib, H., 2015. Chapter 2 & 16. In: Molecular and Quantitative Animal Genetics. Hoboken, New Jersey. 15-23, 155- 160.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R., Cristobal, M.S., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J. & Dairymple, B., 2012. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. PLoS ONE. 10 (2).

Kios, D.K., 2011. Application of DNA markers in parentage verification of Boran cattle in Kenya. MSc (Agric) Thesis. University of Pretoria, South Africa.

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

Köster, E., 1992. Genetic Studies in a Hereford Stud. MSc (Agric) Thesis. University of the Orange Free State, South Africa.

Krehbiel, B.C., 2017. Evaluation of Population Genetic Structure in Two British Bos taurus Breeds Across Five U.S. Climate Zones. MSc Thesis. Colorado State University, USA.

Kriese, L.A., Bertrand, J.K. & Benyshek, L.L., 1991. Age adjustment factors, heritabilities and genetic correlations for scrotal circumference and related growth traits in hereford and brangus bulls. J Anim Sci. 69, 478- 489.

Kristensen, T.N., Hoffmann, A.A., Pertoldi, C. & Stronen, A.V., 2015. What can livestock breeders learn from conservation genetics and vice versa? Front Genet. 6 (38).

Lashmar, S.F., Visser, C. & Köster, E., 2015. Validation of the 50k Illumina goat SNP chip in the South African Angora goat. S Afr J Anim Sci. 45 (1).

Lee, S.H., Cho, Y.M., Lim, D., Kim, H.C., Choi, B.H., Park, H.S., Kim, O.H., Kim, S., Kim, T.H., Yoon, D. & Hong, S.K., 2011. Linkage disequilibrium and Effective Population size in Hanwoo Korean Cattle. Asian-Australas J Anim Sci. 24 (12), 1660- 1665.

Leesburg, V.L.R., 2012. The Influence of Line 1 Hereford on the Global Hereford Population. Animal, Wildlife and Grassland Sciences Thesis. University of the Free State, South Africa.

Leesburg, V.L.R., MacNeil, M.D., Köster, E., Mapholi, O. & Neser, F.W.C., 2013. Impact of Line 1 germplasm on South African Hereford cattle. S Afr J Anim Sci. 43 (2).

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

Lyons, R.E., Buttsworth, S., Waine, D & Kelly, M., 2015. SNP-based parentage in an Australian cattle industries context: does one size fit all? Anim Breed Genet. 21, 249- 252.

MacDonald, J. & Sinclair, J., 1909. Chapter 1, 3, 6 & 10. In: History of Hereford Cattle. London, Vinton. 1-31, 49- 79, 120- 147 & 356- 373.

Magee, D.A., MacHugh, D.E. & Edwards, C.J., 2014. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. Anim Front. 4 (3).

Maiwashe, A.N., Bradfield, M.J., Theron, H.E. & van Wyk, J.B., 2002. Genetic parameter estimates for body measurements and growth traits in South African Bonsmara cattle. Livest Prod Sci. 75 (3), 293- 300.

Makina, S.O., Muchadeyi, F.C., 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., Taylor, J.F., Köster, E., Muchadeyi, F.C., Makgahlela, M.L., MacNeil, M.D. & Maiwashe, A., 2015. Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. Front Genet.6 (337).

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

Maretto, F. and Cassandro, M., 2016. Temporal variation in genetic diversity and population structure of Burlina cattle breed. Ital J Anim Sci. 13 (3091).

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O' Vonnell, J., Moore, S.S., Smith, T.P.L., Sonstegrad, T.S. & Tassell, C.P.V., 2009. Development and Characterisation of a SNP Genotyping Assay for Cattle. PLoS ONE. 4 (4).

Mbole-Kariuki, M.N., Sonstegard, T., Orth, A., Thumbi, S.M., Bronsvoort, B., Kiara, H., Toye, P., Conradie, I., Jennings, A., Coetzer, K., Woolhouse, M., Hanotte, O. & Tapio, M., 2014. Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. Heredity. 113, 297- 305.

McClure, M., Sonstegard, T., Wiggans, G. & Van Tassell, C.P., 2012. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. Front Genet. 3 (140).

McClure, M.C., Sonstegrad, T.S., Wiggans, G.R., Van Eenennaam, A.L., Weber, K.L., Penedo, C.T., Berry, D.P., Flynn, J., Garcia, J.F., Camo, A.S., RRegitano, L.C.A., Albuquuerque, M., Silva, M.V.G.B., Machado, M.A., Coffey, M., Moore, K., Boscher, M., Genestout, L., Mazza, R., Taylor, J.F., Schnabel, R.D., Simpson, B., Marques, E., McEwan, J.C., Cromie, A., Coutinho, L.L., Kuehn, L.A., Keele, J.W., Piper, E.K., Cook, J., Williams, R., Bovine HapMap Consortium. & Van Tassell, C.P., 2013, Imputation of microsatellite alleles from dense SNP genotypes for parentage verification across multiple *Bos taurus* and *Bos indicus* breeds. Front Genet. 4 (176).

McClure, M.C., McCarthy, J., Flynn, P., McClure, J.C., Dair, E., O'Connell, D.K., & Kearney, J.F., 2018. SNP Data Quality Control in a National Beef and Diary Cattle System and Highly Accurate SNP Based Parentage Verification and Identification. Front Genet. 9 (84), 1- 14.

Medugorac, I., Veit-Kensch, C.E., Ramljak, J., Brka, M., Marković, B., Stojanović, S., Bytyqi, H., Kochoski, L., Kume, K., Grūnenfelder, H., Bennewitz, J. & Förster, M., 2011. Conservation priorities of genetic diversity in domestic metapopulations: a study in taurine cattle breeds. Ecol Evol. 1 (3), 408- 420.

Meissner, H.H., Scholtz, M.M. & Palmer, A.R., 2013. Sustainability of the South African Livestock Sector towards 2050 Part 1: Worth and impact of the sector. S Afr J Anim Sci. 43 (3), 282- 297.

Meuwissen, T., Hayes, B., & Goddard, M., 2018. Genomic selection: A paradigm shift in animal breeding. Anim Front. 6 (1), 6- 14.

Meyer, K., 2000. Estimates of genetic parameters for cannon bone length in beef cattle. Anim Genet. 11.

Miller, S., 2010. Genetic improvement of beef cattle through opportunities in genomics. Braz J Anim Sci. 39, 247-255.

Mokoena, M.R., Townsend, R.F & Kirsten, J.F., 1999. Cattle Improvement Schemes in South Africa: Measuring the Returns to Research Investments. Agrekon. 38 (1), 1-12.

Moniruzzaman, M., Khatun, R., & Mintoo, A.A., 2014. Application of marker assisted selection for livestock improvement in Bangladesh. Bangl vet. 31 (1), 1-11.

Moore, S.S., & Byrne, K., 1993. Dinucleotide polymorphism at the bovine calmodulin independent adenylcyclase locus. Anim Genet. 24 (150).

Moore, S.S., Mujibi, F.D. & Sherman, E.L., 2014. Molecular basis for residual feed intake in beef cattle. J Anim Sci. 87, 41- 47.

Mullen, M.P., McClure, M.C., Kearney, J.F., Waters, S.M., Weld, R., Flynn, P., Creevy, C.J., Cromie, A.R., & Berry, D.P., 2013. Development of a custom SNP chip for dairy and beef cattle breeding, parentage and research. Interbull Bulletin. (47).

Mustafa, H., Ahmad, N., Heather, H.J., Eui-soo, K., Khan, W.A., Ajmal, A., Javed, K., Pasha, T.N., Ali, A., Kim, J.J., & Sonstegard, T.S., 2018. Whole genome study of linkage disequilibrium in Sahiwal cattle. S Afr J Anim Sci. 48 (2), 353- 360.

Mwansa, P.B., Crews, D.H., Wilton, J.W. & Kemp, R.A., 2002. Multiple trait selection for maternal productivity in beef cattle. J Anim Breed Genet. 119, 391-399.

Namipashaki, A., Razaghi-Moghadam, Z. & Ansari-Pour, N., 2015. The essentiality of reporting Hardy-Weinberg equilibrium calculations in population-based genetic association studies. Cell J. 17 (2), 187-192.

Navajas, E. A., Pravia, M. I., Lema, M., Clariget, J., Aguilar, I., Ravagnolo, O., Brito, G., Peraza, P., Rizza, M. D., & Montossi, F., 2014. Genetic improvement of feed efficiency and carcass and meat quality of hereford cattle by genomics. ALPA. 22 (5), 13- 16.

Negrini, R., Nicoloso, L., Crepaldi, P., Milanesi, E., Colli, L., Chegdani, F., Pariset, L., Dunner, S., Leveziel, H., Willams, J.L. & Marsan, P.A., 2008. Assessing SNP markers for assigning individuals to cattle populations. Anim Genet. 40, 18- 26.

Nei, M., 1973. Analysis of Gene Diversity in Subdivided Populations. Proc Nat Acad Sci USA. 70 (12), 3321-3323.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genet. 89, 583-590.

Nelsen, T.C., Short, R.E., Urick, J.J. & Reynolds, W.L., 1986. Heritabilities and genetic correlations of growth and reproductive measurements in hereford bulls. J Anim Sci. 63, 409-417.

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- 289.

O'Brien, A.M.P., Utsunomiya, Y.T., Mészáros, G., Bickhart, D.M., Liu, G.E., Van Tassell, C.P., Sonstegard, T.S., Da Silva, M.V.B., Garcia, J.F., & Sölkner, J., 2014. Assessing signatures of selection thrugh variation in linkage disequilibrium between taurine and indicine cattle. Genet Sel Evol. 46 (19), 1- 14.

Ojango, J.M., Mpofu, N., Marshall, K. & Andersson-Eklund, L., 2011. Quantitative methods to improve the understanding and utilisation of animal genetic resources. Animal Genetics Training Resource. Module 4.

Panagiotou, O.A., Evangeou, E., & Ioannidis, J.P.A., 2010. Genome-wide Significant Association for Variants with Minor Allele Frequency of 5% or Less- An Overview: A HuGE Review. Am J Epidemiol. 172 (8), 869- 889.

Park, S.D.E., 2001. Trypanotolerance in West African Cattle and the population genetic effects of selection. PhD Thesis. University of Dublin, Ireland.

Parnell, P.F., 1994. The consequences of selection for growth rate in beef cattle. Aus Soci Anim Prod. 20.

Peakall, R. & Smouse, P.E., 2012. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teching ad research- an update. Bioinformatics. 28 (19), 2537- 2539.

Pienaar, L., Grobler, J.P., Neser, F.W.C., Scholtz, M.M., Swart, H., Ehlers, K. & Marx, M., 2014. Genetic diversity in selected stud and commercial herds of the Afrikaner cattle breed. S Afr J Anim Sci. 44 (5).

Pierce, M.D., Dzama, K & Muchadeyi, F.C., 2018. Genetic Diversity of Seven Cattle Breeds Inferred Using Copy Number Variations. Front Genet. 9 (163), 1- 14.

Pritchard, J.K ., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics. 155, 945- 959.

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 tool set for whole-genome association and population-based linkage analysis. Am J Hum Genet. 81 (3), 559- 575.

Qwabe, S.O., 2011. Genetic and phenotypic characterisation of the South African Namaqua Afrikaner sheep breed. MSc (Agric) Animal Science Thesis. University of Pretoria, South Africa.

Radko, A., Słota, E. & Marczyńska, J., 2010. Usefulness of a supplementary set of microsatellite DNA markers for parentage testing in cattle. Pol J Vet Sci. 13 (2), 313- 318.

Ravagnolo, O., Aguilar, I., Crowley, J.J., Pravia, M.I., Lema, M., Macedo, F.L., Scott, S., & Navajas, E.A., 2018. Accuracy of genomic predictions of residual feed intake in Hereford with Uruguayan and Canadian training populations. WCGALP. 1, 723- 751.

Raymond, M., & Rousset, F., 1995. GENEPOP (version 4.7): population genetics software for exact tests and ecumenicism. J Hered. 86, 248- 249.

Ruane, J., 1999. A critical review of the value of genetic distance studies in conservation of animal genetic resources. J Anim Breed Genet. 116, 317- 323.

Rumph, J.M., Kress, D.D., Davis, K.C., Anderson, D.C., Van Wagoner, H.C. & Boss, D.L., 2005. Characterization of the montana line 4 inbred Hereford herd. AM Soci Anim Sci. 56, 116-120.

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. Genet. 179, 1503- 1512.

SA Hereford, 2017. http://www.hereford.co.za/, accessed on the 24 January, 2017.

SA Hereford, 2018. http://www.hereford.co.za/p14/the-breed/hereford-breed-history.html, accessed on the 23 June 2018.

S.A. Studbook Association, 2016. Annual Report, 2016.

Saatchi, M., Ward, J. & Garrick, D.J., 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. J Anim Sci. 91 (4), 1538- 1551.

Sanarana, Y., Visser, C., Bosman, L., Nephawe, K., Maiwashe, A. & Köster, E., 2016. Genetic diversity in South African Nguni cattle ecotypes based on microsatellite markers. Trop Anim Healh Prod. 48 (2), 379- 385.

Schafberg, R. & Swalve, H.H., 2015. The history of breeding for polled cattle. Livest Sci. 179, 54-70.

Schlötterer, C., 2004. The evolution of molecular markers- just a matter of fashion. Nat Rev. 5 (1), 63-69.

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 (1), 1- 9.

Seidel, G.E., 2010. Brief introduction to whole-genome selection in cattle using nucleotide polymorphisms. Reprod Fertil Dev. 22, 138- 144.

Singh, U., Deb, R., Alyethodi, R.R., Alex, R., Kumar, S., Chakraborty, S., Dhama, K. & Sharma, A., 2014. Molecular markers and their applications in cattle genetic research: A review. BGM. 6, 49- 58.

Shalaby, S.I.A., Hozyen, H.F., & Awasthi, D.K., 2016. Molecular markers in farm animals: Advanced tools for evaluation of genetic diversity. Int J Curr Innov Res. 2 (9), 508- 516.

Sharma, R., Kishore, A., Mukesh, M., Ahlawat, S., Maitra, A., Pandey, A.K. & Tantia, M.S., 2015. Genetic diversity and relationship of indian cattle inferred from microsatellites and mitochondrial DNA markers. BMC Genet. 16 (73).

Sherman, E.L., Nkrumah, J.D. & Moore, S.S., 2009. Whole genome single nucleotide polymorphism associations with feed intake and feed efficiency in beef cattle. J Anim Sci. 88 (1), 16- 22.

Slate, J., David, P., Dodds, K.G., Veenvliet, B.A., Glass, B.C., Broad, T.E. & McEwan, J.C., 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity. 93, 255- 265.

Slatkin, M., 2008. Linkage disequlibrium- understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 9 (6), 477- 485.

Spangler, M., 2016. Traditional genetic selection for fertility: Indicator traits and potential antagonisms. Applied Reproductive Strategies in Beef Cattle. Des Moines, Iowa.

Steffen, P., Eggen, A., Dietz, A.B., Womack, J.E., & Stranzinger, G., 1993. Isolation and mapping of polymorphic microsatellites in cattle. Animal Genet. 24 (121).

Steyn, Y., Köster, E. & Theron, H.E., 2014. Residual feed intake as selection tool in South Africa Bonsmara cattle. Livest Sci. 164, 35- 38.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H & Gondro, C., 2014. Performance of different SP panels for parentage testing in two East Asian cattle breeds. Anim Genet. 45 (4), 572- 575.

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

Sudrajad, P., Seo, D.W., Choi, T.J., Park, B.H., Roh, S.H., Jung,W.Y., Lee, S.S., Lee, J.H., Kim, S., & Lee, S.H., 2017. Genome-wide linkage disequilibrium and past effective population size in three Korean cattle breeds. Anim Genet. 48 (1), 85- 89.

Suh, S., Kim, Y., Cho, C., Byun, M., Choi, S., Ko, Y., Lee, C.W., Jung, K., Bae, K.H. & Kim, J., 2014. Assessment of genetic diversity, relationships and structure among Korean Native cattle breeds using microsatellite markers. Asian Australas J Anim Sci. 27 (11), 1548- 1553.

Sunden, S.L.F., Stone, R.T., Bishop, M.D., & Kappes, S.M., 1993. A highly polymorphic bovine microsatellite locus: BM2113. Anim Genet. 24 (69).

Tellam, R.L., Lemay, D.G., Van Tassell, C.P., Lewin, H.A., Worley, K.C. & Elsik, C.G., 2009. Unlocking the bovine genome. BMC Genomics. 10 (193).

The Hereford Cattle Society UK, 2013. http://www.herefordcattle.org/about-us/history/ accessed on the 8 July, 2016.

The Hereford Cattle Society UK, 2016. The Hereford Management Guide.

Theron, H.E., van der Westhuizen, R.R., Köster, E. & van der Westhuizen, J., 2014. Reference Population for South African Bonsmara Cattle. Conference: 10th World Congress on Genetics Applied to Livestock Production.

Tixier-Boichard, M., Verrier, E., Rognon, X. & Zerjal, T., 2015. Farm animal genetic and genomic resources from an agroecological perspective. Front Genet. 6 (153).

Toldo, S., Fries, S.R., Steffen, P., Neibergs, H.L., & Barendse, W., 1993. Physically mapped cosmid-derived microsatellite markers as anchor loci on the bovine chromosome. Mamm Genome. 4, 720- 727.

Toro, M.A. & Caballero, A., 2005. Characterisation and Conservation of genetic diversity in subdivided populations. Philos Trans R Soc Lond B Biol Sci. 360, 1367-1378.

Vaiman, D., Mercier, D., Moazami-Goudarzi, K., Eggen, A., Ciampolini, R., Lepingle, A., Velmala, R., Kaukinen, J., Varvio, S.L., Martin, P., Levéziel, H., & Guérin, G., 1994. A set of 99 cattle microsatellite: characterization, synteny mapping and polymorphism. Mamm Genome. 5, 288- 297.

Van Binsbergen, R., Calus, M.P.L., Bink, M.C.A.M., van Eeuwijk, F.A., Schrooten, C., & Veerkamp, R.F., 2015. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. Genet Sel Evol. 47 (71), 1-13.

van der Westhuizen, R.R., Schoeman, S.J., Jordaan, G.F. & van Wyk, J.B., 2001. Genetic parameters for reproductive traits in a beef cattle herd estimated using multitrait analysis. S Afr J Anim Sci. 31 (1), 41- 48.

Van Eenennaam , A.L., Weigel, K.A., Young, A.E., Cleveland, M.A. & Dekkers, J.C.M., 2014. Applied Animal Genomics: Results from the Field. Annu Rev Anim Biosci. 2, 105- 139.

van Marle-Köster, E., Mostert, B.E. & Van der Westhuizen, J., 2000. Body measurements as selection criteria for growth in South African hereford cattle. S Afr J Anim Sci. 1, 5-15.

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), 1-17.

van Marle-Köster, E., Visser, C., Makgahlela, M. & Cloete, S.WP., 2015. Genomic technologies for food security: A review of challenges and opportunities in Southern Africa. Food Res Int. 76 (4), 971- 979.

Vignal, A., Milan, D., SanCristobal, M. & Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol. 34, 275- 305.

Visser, C., Lashmar, S.F., Köster, E., Poli, M.A. & Allain, D., 2016. Genetic Diversity and Population Structure in South African, French and Argentinian Angora Gotas from Genome-Wide SNP Data. PLoS ONE. 11 (5).

Wang, M.D., Dzama, K., Hefer, C.A., Muchadeyi, F.C., 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genomics. 16 (894), 1- 16.

Wakchaure, R., Ganguly, S., Praveen, P.K., Kumar, A., Sharma, S., & Mahajan, T., 2015a. Marker Assisted Selection (MAS) in Animal Breeding: A Review. J Drug Metab Toxicol. 6 (5), 1-5.

Wakchaure, R., Ganguly, S., Para, P.A., Praveen, P.K., & Qadri K., 2015b. Molecular Markers and their Applications in Farm Animals: A review. Int J Rec Boitech. 3 (30, 23- 29.

Weir, B.S. & Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evol. 38 (6), 1358- 1370.

Weir, B.S., 2012. Estimating F-statistics: A historical view. PMC. 79 (5), 637- 643.

Wiener, P., Burton, B. & Williams, J.L., 2004. Breed relationship and definition in British cattle: a genetic analysis. J Hered. 93, 597- 602.

Wientjes, Y.C.J., Veerkamp, R.F., & Calus, M.P.L., 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genet. 193, 621- 631.

Wilkinson, S., Wiener, P., Archibald, A.L., Law, A., Schnabel, R.D., McKay, S.D., Taylor, J.F. & Ogden, R., 2011. Evaluation of approaches for identifying population informative markers from high density SNP Chips. BMC Genet. 12 (45).

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

Yang, H., Lee, S.H., Goddard, M.E. & Visscher, P.M., 2011. SAQC: SNP array quality control. BMC Bioinformatics. 12 (100).

Yang, W., Kang, X., Yang, Q., in, T., & Fang, M., 2013. Review on the development of genotyping methods for assessing farm animal diversity. J Anim Sci Biotechnol. 4 (2), 1- 6.

Yaro, M., Munyard, K.A., Stear, M.J., & Groth, D.M., 2016. Molecular identification of livestock breeds: A tool for modern conservation biology. Bio Rev. 92 (2), 993- 1010.

Zhang, Y., Wang, Y., Sun, Yu, Y., & Zhang, Y., 2010. Validation of 17 microsatellite markers for parentage verification and identity test in Chinese Holstein Cattle. Asian-Aust J Anim Sci. 23(4), 425- 429.

Zhang, C. & Plastow, G., 2011. Genomic Diversity in Pig (Sus scrofa) and its Comparison with Humans and other Livestock. Curr Genom. 12, 138-146.

Zwane, A.A., Maiwashe, A., Makgahlela, M.L., Choudhury, A., Taylor, J.F. & Köster, E., 2016. Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South Africa indigenous cattle breeds. S Afr J Anim Sci. 46 (3), 303- 312.

Addendum A

Wright's F-statistical for 11 microsatellite loci (F_{IS} and F_{ST} values) for each locus for the complete and subset data

Locus	**Wright's F-Statistics**			
	Complete Data set		**Subset Data set**	
	F_{IS} (f)	F_{ST} (θ)	F_{IS} (f)	F_{ST} (θ)
BM1824	-0.046	0.045	-0.021	0.047
BM2113	-0.007	0.029	-0.009	0.029
ETH10	-0.013	0.045	0.032	0.053
ETH225	0.003	0.04	0.011	0.057
ETH3	-0.023	0.025	-0.008	0.036
INRA23	-0.011	0.048	0.009	0.046
SPS115	-0.006	0.071	-0.027	0.049
TGLA122	-0.034	0.052	-0.065	0.042
TGLA126	-0.021	0.066	0.007	0.069
TGLA227	-0.003	0.097	-0.004	0.075
TGLA53	0	0.049	-0.011	0.031

* Mean F_{IS} and F_{ST} estimated per loci for both data sets

Addendum B

A summary of statistics for autosomal SNPs that passed quality control for the merged data set

Chromosome	Number of SNPs	Average interval between SNPs (bp) per chromosome	Mean MAF	Mean Heterozygosity
1	6560	100 669	0.2988	0.3849
2	5672	102 187	0.2984	0.3851
3	5058	98 992	0.2990	0.3862
4	4928	105 410	0.2999	0.3894
5	5368	90 994	0.3019	0.3861
6	5914	86 693	0.2968	0.3835
7	5038	91 613	0.2996	0.3864
8	4495	103 797	0.2990	0.3837
9	4491	101 017	0.3084	0.3918
10	4426	97 437	0.3039	0.3888
11	4576	95 242	0.3024	0.3877
12	3722	96 996	0.2964	0.3851
13	3440	94 720	0.2943	0.3834
14	3987	79 346	0.2884	0.3731
15	3551	99 381	0.2928	0.3831
16	3356	97 322	0.2983	0.3847
17	3076	105 991	0.3028	0.3885
18	2742	97 762	0.2916	0.3800
19	2663	100 637	0.2961	0.3833
20	3442	81 803	0.3021	0.3890
21	2975	91 312	0.2926	0.3809
22	2625	96 741	0.2961	0.3866
23	2232	96 193	0.3021	0.3887
24	3069	76 159	0.2941	0.3779
25	1847	96 326	0.3027	0.3869
26	2251	96 368	0.2897	0.3778
27	1949	95 315	0.3060	0.3907
28	1924	98 999	0.3000	0.3885
29	2137	102 332	0.2991	0.3894