Table of contents

List of tables

List of figures

List of abbreviations

ARC	Agricultural Research Council
ARC−API	Agricultural Research Council – Animal Production Institute
ARC−BTP	Agricultural Research Council – Biotechnology Platform
ARS	Agricultural Research Service
API	Animal improvement Institute
BAC	Bacterial artificial chromosome
BLUP	Best Linear Unbiased Prediction
BTP	Biotechnology Platform
CNV	Copy number variants
CV	Cross validation
DALR	Department of Agriculture and Land Reform
DST	Identity by state distance
EBV	Estimated breeding value
EHH	Extended haplotype homozygosity
FAO	Food and Agriculture Organisation
F_{ST}	Fixation indices
GLM	Generalized Linear Model
HWE	Hardy-Weinberg Equilibrium
H_e	Heterozygosity
iHS	Integrated Haplotype Score
LD	Linkage disequilibrium
LE	Linkage equilibrium
LRH	Long-range haplotype
LSMEANS	Least square means
N_e	Effective population size
N_T	Past effective population size at generation t
MAF	Minor allele frequency
MAS	Marker-assisted selection
NJ	Neighbor Joining
QTL	Quantitative Trait Loci
QC	Quality control
REHH	Relative extended haplotype homozygous
SA	South Africa

SNP	Single nucleotide polymorphism
SVS	SNP Variation Suite
UCSC	University of California, Santa Cruz
USDA	United States Department of Agriculture
WGS	Whole-genome shotgun sequencing

Chapter 1

1.1 Background and motivation

Throughout evolution, forces such as mutation, selection, migration and genetic drift have led to differential survival of individuals and the continual development of new cattle breeds with characteristics that improve their ability to grow and reproduce in particular environments (Blakely & Bade, 1982). Today, there are more than 800 cattle breeds in the world, subdivided into two major categories of taurine and indicine that are morphologically and genetically distinct from each other (MacHugh *et al*., 1997; Qanbari *et al*., 2010). Taurine cattle (*Bos taurus*) account for most of the herds in the temperate regions of Europe, Western Africa and Northern Africa. Indicine cattle are subdivided into pure zebu and zebu crossbred-types and are better adapted and dominant in arid conditions across Africa and the Indian sub-continent (Gautier *et al*., 2009). Zebu cattle are thought to have been introduced into Africa about 4000 years ago, however they only became widespread about 700 AD during the time of Arabic migrations into North and East Africa (MacHugh *et al.,* 1997). Currently, the African continent is dominated by the taurine, Zebu and its derived forms known as Sanga cattle (MacHugh *et al.,* 1997). Sanga cattle were brought to Southern Africa by the Khoi-Khoi people between 600 and 700 AD. The southward migration through some of the harshest countries on the subcontinent led to the evolution of hardy cattle breeds that are heat, parasite and drought tolerant (Scholtz, 2010).

Cattle in South Africa could be classified into four different groups, namely *B. taurus* (e.g. Angus, Hereford and Holstein), *B. indicus* (e.g. Brahman), *Sanga* and *Sanga derived types* (e.g. Afrikaner and Nguni), those of unclear origins (e.g. Drakensberger) and locally developed breeds (e.g. Bonsmara and Brangus). These cattle breeds have played an important role in the social, cultural and economic development of the country. Furthermore, Sanga cattle (indigenous and locally developed cattle) possess adaptive traits that are usually associated with tolerance to various diseases, extreme temperatures and humidity and to change in the availability of feed. These breeds are also adapted to low input management systems and have shown the ability to survive, produce and reproduce under harsh environments (Scholtz, 2010). Therefore, in the harsh and fluctuating South African environments, these cattle breeds hold potential with regard to their adaptation to the pressures of the specific local environment (Scholtz, 2010). Despite their large numbers and not endangered status, their adaptive traits are of importance and there is a worldwide drive for effective management of indigenous genetic resources, as they could be most valuable in selection and breeding programs in times of biological stress such as famine, drought or disease epidemics (FAO, 2010).

Genetic diversity is a prerequisite for genetic improvement and environmental adaptation of livestock population (FAO, 2007). It allows farmers to select stocks or develop new breeds in response to changing conditions, including climate change, new or resurgent disease threats, new knowledge of human nutritional requirements, and changing market conditions or societal needs (FAO, 2010). It is also important for determining breeding behaviour of species, individual reproductive success and the existence of gene flow (Groeneveld *et al.*, 2010). Furthermore, information such as data on the population size, structure, and geographic distribution and production environments is a prerequisite for control of inbreeding and effective utilization of breed specific characteristics (Boettcher *et al*., 2010). In South Africa a number of studies have focused on the characterization of small stock such as goats: Visser *et al.* (2004) and sheep: Soma *et al.* (2012); Qwabe *et al*. (2012). Limited studies have focused on the genetic characterization of South African cattle breeds and therefore there is a need for genetic characterization of these breeds as a means towards their management.

Unlike other demographic processes that affect the entire genome, selection affects specific functionally important sites in the genome. Selection acts on a mutation and reduces variation at linked sites and leaves its signature in adjacent chromosomal region of the genome (Nielsen, 2005). Therefore, selection is thought to have shaped the patterns of variation among cattle breeds in the world by creating selective sweeps or selection signatures in specific regions of genome associated with traits of importance to farming communities (The Bovine HapMap Consortium *et al*. 2009). Given that South African cattle breeds have also been subjected to both natural and artificial selection it can be postulated that these breeds also carry some selection signatures in their genome. Selection signatures are regions in the genome that have favorably increased in frequency and get fixed in a population because of their functional importance in biological processes (Akey *et al*., 2002). Scanning for signatures of selection in the genome is important as it could assist in detecting regions of the genome that are, or have been, functionally important and have thus been targeted by either natural or artificial selection. It can also assists with providing a genomic understanding of how and where natural and artificial selection have shaped the patterns of variation in the genome, thereby giving important insights with regard to mechanisms of evolution (Otto, 2000), selection of loci for breeding and selection programs (Vitalis *et al*., 2001) and annotation of functionally important genomic regions (Sabeti *et al*., 2002).

Molecular based methods such as microsatellite markers have been used for capturing of information to estimate genetic diversity among farm animal genetic resources (Visser *et al*., 2004; Soma *et al.*, 2012; Qwabe *et al*., 2012). These markers have provided insights into breed history and information regarding both the distinctiveness (across-breeds) and the (within-breeds) diversity of population (Boettcher *et al.*, 2010). They have also been used to identify genes and traits that has been targeted by

selection (The Bovine HapMap Consortium *et al*., 2009). Single nucleotide polymorphisms (SNPs) are rapidly becoming the marker of choice and could also be used to provide information on selected regions as well as those that portray neutral variation (Gautier *et al*., 2009; Boettcher *et al*., 2010; Porto-Neto *et al*., 2013). The recently developed genome wide SNP arrays (e.g. 7K, 50K and 700K) (Illumina, San Diego, CA), allow for simultaneous high throughput examination of thousand and hundreds of thousands of loci with high accuracy (Matukumalli *et al*., 2009) and thus offer the opportunity to characterize South African cattle breeds at the genomic level and unravel the genetic history of these breeds, that will be relevant for managing the present and future genetic diversity. In addition, the distribution and density of the SNPs and the linkage disequilibrium (LD) within the populations could reveal statistical associations with phenotypes even if the SNPs are not causative mutations (Hayes *et al*., 2009). Therefore these SNP arrays offer an opportunity to reveal genomic regions underlying phenotypic variation among South African cattle breeds.

One possible limitation with regard to the use of SNP chip is an ascertainment bias associated with the design of the SNP chip (Matukumalli *et al*., 2009). Due to the bias associated with not sampling rare SNPs, e.g. high minor allele frequency (MAF) SNP in small sampled populations, the estimates of nucleotide diversity, population size, demographic changes, linkage disequilibrium, selection signatures and inference of population structure will be affected (Nielsen, 2005). Measures of breed-wise genetic variability will be biased and diversity in local breeds will be underestimated (The Bovine HapMap Consortium, 2009; FAO, 2010). Therefore, before these SNP arrays could be used for genome wide studies it is essential that their utility is investigated in order to determine the number of polymorphic SNPs available for application among any given cattle breed (The Bovine HapMap Consortium, 2009).

To date, studies with regard to the adaptability, ticks and parasite tolerance, disease resistance, production potential, production efficiency and meat quality of South African cattle breeds (e.g. Afrikaner, Nguni, Drakensberger and Bonsmara) have been undertaken (e.g. Muchenje *et al*., 2008; Muchenje *et al*., 2009; Marufu *et al*., 2011). However, comprehensive research about the nature or extent of the genetic variation underlying these breeds is limited. This emphasizes the need for a genome wide characterization of Afrikaner, Nguni, Drakensberg and Bonsmara cattle as a first step towards understanding their population structure and selection pressures that shaped the patterns of these breeds.

1.2 Aim and Objectives

The aim of this study was to conduct a genome wide scan for signatures of selection among Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein cattle breeds of South Africa cattle using data generated from the Bovine SNP50 BeadChip as a first step towards unraveling the nature or extent of the genetic variation underlying these breeds. The Angus and Holstein breeds were included in this study as reference breeds since they have been extensively characterized using similar tools in other studies (Prasad *et al*., 2008; Qanbari *et al*., 2010; Ramey *et al*., 2013).

To achieve this aim a set objectives were as follows:

The first objective was to investigate the usefulness of the BovineSNP50 in local breeds and also to evaluate its application in cattle breeds that are widely used in South Africa including Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein.

The second objective was to investigate genetic diversity and population structure within and between the six mentioned cattle breeds using genome wide SNP data generated from the Illumina Bovine SNP50 BeadChip in order to unravel possible historic events during domestication and breed formation.

The third objective was to quantify the extent of genome-wide LD and estimate effective population sizes within the above mentioned six cattle breeds. LD pattern across the genome can influence the number of markers that must be screened to achieve reasonable power for detecting selection signatures.

The fourth objective was to perform a genome-wide scan for signatures of selection within and between the above mentioned six cattle breeds in South Africa using data generated from the Illumina BovineSNP50 BeadChip. This was done in order to understand selection forces that has shaped the pattern of genetic variation among these breeds.

These objectives have been achieved in phases and are reported in this thesis by way of scientific articles. Chapter two presents the literature review related to the objectives of this study. Chapter three contains the manuscript “Evaluation of the BovineSNP50 genotyping array in four South African cattle populations” that describes the evaluation of BovineSNP50 BeadChip to determine its utility for genome wide studies among South African cattle breeds. Chapter four describes the genetic diversity and population structure of the six cattle breeds included in this study. The extent of linkage disequilibrium and effective population size is described in chapter five. Finally, chapter six focuses on detection of selection signatures in the genome of South African cattle breeds. Chapter seven gives a general discussion, critical review of research findings and recommendations for future studies.

References

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

Blakely, J. & Bade D.H. (1982). The science of animal husbandry, 3rd edition. Virginia, Reston Publishing House. ISBN 0-83596978-9.

Boettcher, P.J., Tixier-Boichard, M., Toro, M.A., Simianer, H., Eding, H., Gandini, G., Joost, S., Garcia, D., Colli, L., Ajmone-Marsan, P. & the GLOBALDIV Consortium. (2010). Objectives, criteria, and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics* 41(1), 64-77.

FAO. (2007), The State of the World's Animal Genetic Resources for Food and Agriculture, Rome, Italy.

FAO. (2010). Progress report of the FAO on selected activities related to agricultural biodiversity. http://www.fao.org/docrep/meeting/018/k5700e.pdf

Gautier, M., Flori, L., Reibler, A., Jaffrézic, F., Laloe, D., Gut, I., Moazimi-Goudarzi, K. & Foulley, J.L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10, 550.

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. *Animal Genetics* 41, (1) 631.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92, 433–443.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P. & Bradley, D.G. (1997). Microsatellite DNA variation and the evolution, domestications and phylogeography of taurine and zebu cattle. *Genetics* 146, 1071-1086.

Marufu, M.C., Qokweni, L., Chimonyo, M. & Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. *Ticks and Tick Borne Diseases* 2, 172-177.

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

Muchenje, V., Dzama, K., Chimonyo, M., Strydom, P.E. & Raats, J.G. (2009). Relationship between stress responsiveness and meat quality in three cattle breeds. *Meat Science* 81, 653-657.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G. & Strydom, P.E. (2008). Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. *South African Journal of Animal Science* 2, 298-304.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Reviews Genetics* 39, 197-218.

Otto, S.P. (2000). Detecting the form of selection from DNA sequence data. *Trends Genetics* 16, 526-529.

Porto-Neto, L.R., Lee, S.H., Sonstegard, T.S., Van Tassell, C.P., Lee, H.K., Gibson, J.P. & Gondro, C. (2013). Genome-wide detection of signatures of selection in Koren Hanwoo cattle. *Animal Genetics* 45, 180-190.

Prasad, A., Schnabel, R.D., McKay, S.D., Murdoch, B., Stothard, P., Kolbehdar, D., Wang, Z., Taylor, J.F. & Moore, S.S. (2008). Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal Genetics* 39, 597–605.

Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Litchtner, P., Sharifi, A.R. & Simianer, H. (2010). A genome-wide scan for signatures of recent selection in Holstein cattle. *Animal Genetics* 41, 377-389.

Qwabe, S.O., van Marle-Köster, E. and Visser, C. (2012). Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. *Tropical Animal Health and Production* 45 (2), 511-516.

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D. & Taylor, J.F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14, 382-.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P, Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platkon, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R. & Lander, E.S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832-837.

Scholtz, M.M. (2010). Beef breeding in South Africa. 2nd Edition, (Pretoria: Agricultural Research Council, South Africa), ISBN-13: 978-1-86849-391-3.

Soma, P., Kotze, A., Grobler, J.P. & van Wyk, J.B. (2012). South African sheep breeds: Population genetic structure and conservation implications. *Small Ruminant Research* 103, 528-532 DOI: 10.1126/science.1167936.

The Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528-532.

Visser, C., Hefer, C.A., van Marle-Köster, E. & Kotze, A. (2004). Genetic variation of three commercial and three indigenous goat populations in South Africa. *South African Journal of Animal Science* 34 (1), 24-27.

Vitalis, R., Dawson, K. & Boursot, P. (2001). Interpretation of variation across marker loci as evidence of selection. *Genetics* 158, 1811-1823.

Chapter 2

Literature review

2.1 Introduction

Farm animal genetic resources include all species, breeds and strains of animals, particularly those of economic, scientific and cultural interest to mankind used for agriculture either at present or potentially in the future (FAO, 2005; Alderson, 2010). A breed is a homogenous group of domestic livestock with definable and identifiable external characteristics that allow it to be separated by visual assessment from other similarly distinct groups within the same species (FAO, 2005). Breeds may share a large fraction of their genome with other breeds, but each holds distinctive combination of genes that may confer unique traits particularly for adaptation to specific environments (Scherf, 2000; FAO, 2010). Breeds are also linked to their origin with regard to tradition and history or a geographical region (Alderson, 2010). Indigenous and local developed breeds possess adaptive traits that are usually associated with tolerance to various diseases, extreme temperatures and humidity, changes in the availability of feed, adaptation to low capacity management and ability to survive, produce and reproduce for extended period of time (Scherf, 2000). These specific breeds hold potential in times of biological stress such as famine, drought or disease epidemics due to years of adaption to the pressures of the specific local environments. This review was aimed at unveiling the characterisitics of South African indigenous and locally developed cattle breeds (Afrikaner, Nguni, Drakensberger and Bonsmara), their genetic potential and to discuss opportunities for application of genomics tools.

2.2 Historical overview of South African indigenous and locally developed breeds

Afrikaner breed

The Afrikaner cattle is one of the oldest indigenous cattle breeds in South Africa. This breed was brought to South Africa by the Khoi-Khoi people who migrated along the western side of Southern Africa to areas of the Western Cape Province in 1652 (Scholtz, 2010). This Southward migration through some of the harshest country of the subcontinent led to the evolution of a hardy, heat, parasite and drought tolerant breed. The first selection of this breed was carried out by the transporters who carried goods around the country in ox wagons before the advent of railways; and selected for cattle of uniformity in build and color (Scholtz, 2010). At the end of the 18th century, the Afrikaner was well-defined as a cattle breed and was greatly appreciated by the settlers for the power, speed, stamina, and hardness and also for their meat and milk (Bergh *et al*., 2010)

The Afrikaner (Figure 2.1), which belongs to the Sanga type of cattle, was the first indigenous South African breed to form a breed society in 1912 (Bergh *et al*., 2010).

Figure 2.1 The Afrikaner cattle (www.arc.agric.za)

The modern Afrikaner is medium-framed, yellow to red coloured with lateral horns with a typical twist or sometimes dehorned. This breed has the characteristics of heat, parasite and tick tolerance, disease resistance (e.g. redwater, heartwater and gallsickness), good walking and grazing ability, easy calving and exceptional mothering ability, ideal mothering line for cross breeding and development of hardy composite among others. The Afrikaner has mature weight of 820 - 1090 kg and 550 - 730 kg, average birth weight of 34 kg and 32 kg and average weaning weight of 210 kg and 195 kg for males and females respectively. Their age at first calving is between 36 to 41 months and they have an average inter-calving period of 445 days. This breed has an average carcass weight and dressing percentage of 181.7 kg and 54 % respectively (Bergh *et al*., 2010). It is known for good meat quality being tender, tasty and succulent. In crossbreeding programs the Afrikaner has been shown to improve the quality of the meat of the breed with which it is crossed, especially with regard to tenderness (Strydom, 2008). The Afrikaner breeders have over the past two decade focused on economically important traits in the modern beef production environment, namely fertility, all aspects of functional efficiency and the implementation of performance recording information in the breeding programs (Bergh *et al*., 2010).

Nguni breed

The Nguni cattle breed was brought to South Africa by the Khoi-Khoi people who migrated southwards from the central lake of Africa between 600 and 700 AD. These people have their lives centered around these cattle i.e. their ceremonies, their wealth and dowry, their links with other tribes,

their food, clothes and welfare (Scholtz *et al*., 2011). The Nguni cattle are still found wherever the descendants of the original groups of the Nguni tribe settled, namely in Swaziland, Zululand and Mozambique. This breed is classified as a Sanga type (with low cervico-thoracic humps, in front legs, instead of the high thoracic humps of pure Zebu) that is native to Africa. The protein analyses indicated that they have characteristics of both *Bos Taurus* and *Bos Indicus* cattle (Meyer, 1984). The Nguni cattle are small to medium sized with bulls weighing between 500 kg and 600 kg, and cows weighing between 300 kg and 500 kg. They are characterized by their multi-coloured skin, which can present many different patterns (white, brown, golden yellow, black, dappled, or spotty), but their noses are always black-tipped and they present a variety of horn shapes. Different ecotypes of Nguni cattle were developed in the different agro-ecological regions and these are still maintained within the breed (Scholtz *et al.*, 2011). Bothma (1993) recorded that they are about fourteen ecotypes of Nguni, however, Nortier *et al*. (2002) reported that the genetic differentiation determined with the aid of microsatellite markers, did not reflect differences between various ecotypes in Namibia.

Data from the South African national performance scheme shows that Nguni cattle have an average birth and weaning weight of 25 kg and 155 kg respectively (Bergh *et al*., 2010). Their average age at first calving is 31 months and their average inter calving period is 404 days (Muchenje *et al*., 2007). Scholtz *et al*. (1990) reported that Nguni cows are the lightest and they weaned light calves, however their growth rate and feed conversion ratio compare well with other breeds. Nguni are also the most fertile beef breed in South Africa and this might be the result of natural selection to which these animals were subjected to under the management system of Nguni people (Scholtz, *et al*., 2011). Strydom (2008) reported that the Nguni showed no difference in terms of meat tenderness compared to Bonsmara, Afrikaner and Brown Swiss. In addition Gertenbach *et al*. (2011) reported that the Nguni steers provided a satisfactory carcass in terms of classification and in their subsequent trial concluded that the carcasses of both yearling and two year old Nguni were comparable to the carcasses of Sussex steers in terms of grading and dressing percentages. The Nguni cattle has the ability to produce and reproduce under harsh environment conditions (Scholtz *et al*., 1991). Research has revealed that the Nguni has the most resistant to ticks in South Africa and has shown to be tolerant to tick borne diseases (Muchenje *et al*., 2008; Marufu *et al*., 2011), which are considered to be the most important problem for livestock farmers in most parts of Africa. In the study by Spickett *et al*. (1989), Nguni's productivity as measured by the weaning weight of their calves was shown to be least affected by tick infestation due to its natural resistance to ticks. The Nguni is also an excellent dam line for crossbreeding, with little calving difficulties. It has an increased ability to recycle nitrogen back to the rumen, which improves microbial growth and organic matter digestion, reducing the requirement on low quality pastures and therefore need little or no supplementation during winter (Scholtz *et al.*, 2011).

Figure 2.2 The Nguni cattle (www.arc.agric.za)

The Nguni breed was regarded as inferior in the past due to the lack of performance recording during the colonization. This perception was the results of African man living in a symbiotic relationship with his animals and also due to the fact that the Nguni had variety of colours amongst animals of the breed, which was in contrast to the general tendency in the stud breeding industry that emphasize uniformity (Scholtz & Ramsay, 2011). Thus, the breeding industry was unable to identify the much emphasized antiquated breed standard (Bonsma, 1980) and regarded these animals as an indiscriminate mixture of breeds (Scholtz, 1988). As a result little attention was paid to the improvement of this indigenous breed (Scholtz & Ramsay, 2011). Only in 1959 was the potential of the Nguni recognized following the introduction of the beef cattle recording scheme and the publication of results on the Nguni in the early 1980's. Thus in 1983 the Nguni was recognized as a developing breed under the livestock Improvement Act (1977) of South Africa and in 1986 a breed society was established for this breed (Scholtz *et al*., 2011). Nguni cattle have survived over several decades in the dry, hot and harsh environment which was usually affected by tick and parasite and thus they were forced by natural selection to retain those characteristics that ensure theirs survival. Today Nguni cattle are very popular as a beef breed and also for their skin, with its Society being the second largest seed stock beef breed in South Africa (Scholtz & Ramsey, 2011).

Drakensberger breed

The Drakensberger cattle was bred and developed in South Africa. Its origin can be related back to the indigenous cattle of the Khoi-Khoi and other indigenous groups of the Cape and the adjacent areas (Dreyer, 1982). In 1837 several "Voortrekker" families left Cape Province to travel north with herds of similar black oxen, by then referred as the "Vaderland" cattle. Most of these trekkers settled along the

Drakensberg Mountain range, among them were the Uys family who began farming in Volksrust area in the Mpumalanga province. This family played a significant role in the development of the Drakensberger, with strong selection within a closed herd. They bred exceptional animals in the district of Wakkerstroom and Utrecht, where they chose to settle. Their breeding program resulted in a definitive breed type, which was later known as "Uys-cattle" (Scholtz, 2010). Thirteen years later, the Department of Agriculture recommended that the "Uys-cattle" be acknowledge as a breed in terms of the Stock Improvement Act of 1934. Since the "Uys-cattle" were more prevalent in the pastures of the Drakensberg Mountains, it was recommended that the name of the "Uys-cattle" be changed to Drakensberger (Scholtz, 2010). The breeding society for this breed was recognized in 1969 and in 1980 the Society made performance recording compulsory (Dreyer, 1982). Today this breed has spread throughout the country, from Humansdorp in the south, throughout the eastern Free State, KwaZulu-Natal and eastern Mpumalanga to Messina in the Northern Province.

Figure 2.3 The Drakensberger cattle (www.studbook.co.za)

The Drakensberger breed date back to history, where there were no dipping fluid or proven medicine, thus it was able to withstand harsh conditions, flies, mosquitoes, ticks and parasite-borne diseases. Therefore, this breed has an edge over other breeds regarding adaptability, hardiness and tick and disease resistance (Scholtz, 2010). The Drakensberger cattle are black in color with a smooth coat, long and deep bodied medium to large framed cattle. Mature bulls weigh from 820 kg to 1100 kg and cows from 550 kg to 720 kg, calves weigh approximately 35 kg at birth. Cows remain in production for up to 20 years. The Drakensberger cows have good milk production; 240 kg weaning weight is common in stud. Cows have up to 90 % fertility and low incidence of abortion (Bergh *et al*., 2010). In a study by Strydom (2002), Drakensberger were shown to have the juiciest and most tasty meat with the best cut

ability compared to British and Europe breeds. Drakensberger have also been shown to be resistance to heat, this could be attributed to a loose skin as well as short and shiny blue-black hair color which reflect sunlight (Scholtz, 2010). Over the last decade the emphasis for beef cattle breeding moved to meat production and thus animals with a distinctive long and well balanced, well-muscled beef carcass were developed. The modern Drakensberger is the product of such development brought by strong selection based on scientific norms. Drakensberger has continuously been selected for economically important traits such as adaptability, fertility, milk production traits, longevity, growth ability, feed turnover conversion and carcass quality (Scholtz *et al.*, 2010).

Bonsmara breed

Bonsmara was bred and strictly selected for economic production at Mara and Messina Research Station from 1937 to 1963 by scientists under the guidance of late Prof Jan C Bonsma (Bonsma, 1980). It is the only beef breed in South Africa created through a well-documented crossbreeding programme with the aid of objectively recorded performance data combined with visual evaluation according to norms of functional efficiency (Bonsma, 1980). It was bred to perform in sub-tropical environment, where British cattle had proved to be unsuccessful and Afrikaner struggled with calving regularly. Initially five bulls of British beef breed were used on Afrikaner cows after which the progeny were performance tested. After pilot trails it was decided to continue only with the better performing Hereford and Shorthorn cross breeds. Cross-breeding trails for 5/8 Afrikaner and 3/8 Hereford or Shorthorn blood provided suggestions that the calving percentage and weaning weights were higher, while calf mortality dropped. Through strict selection of breeding animals rapid genetic progress was made. Twenty years later a superior cattle breed, performing better than other breeds in the bushveld of the Northern Transvaal was established (Bergh *et al.*, 2010). This breed was tested successfully in about 20 commercial herds in different areas of South Africa. In 1964 the breed society was formed by twelve breeders, since then this breed has expanded within the borders of South Africa to be the most prominent beef breed with over 100 000 registered Bonsmara cattle (Bonsma, 1980).

Figure 2.4. The Bonsmara cattle (www.studbook.co.za)

The Bonsmara is a medium framed, smooth coated, heat and tick tolerant beef breed. It is uniform red-brown to light brown in colour, with slight cervico-thoracic hump in the bull and good beef conformation (Bosman & Scholtz, 2010). The average adult live weights for males and females range from 544 - 950 kg and 300 - 700 kg respectively (Scholtz *et al*., 2010). Bonsmara cattle are bred and selected through a system that effectively promotes cow efficiency. Thus they have excellent mothering ability under all environmental conditions. Bonsmara cows wean calves that are very suitable for finishing off on pasture or in feedlots. Under pasture condition the average weight at birth, at weaning and at yearling age for females were 36, 220 and 317 kg, respectively (Scholtz *et al*., 2010). On the other hand under feedlot condition, the respective weights for males were 39, 240 and 388 kg, respectively (Bosman & Scholtz, 2010). In a survey conducted at ARC, Bonsmara cattle had the highest percentages intake in feedlots of all breeds in South Africa (Scholtz, 2010). The Bonsmara breed has excellent meat qualities (tender, tasty and succulent) and in crossbreeding programs, the Bonsmara have shown to improve the quality of the meat of the breed with which it is crossed, especially as regards tenderness (Bosman & Scholtz, 2010). They are well adapted to warm bushveld and subtropical areas. Bonsmara are very fertile with an average of 414 days interval period (Scholtz, 2010).

2.3 Genomic tools

To date several investigations into the adaptability, ticks and parasite tolerance, disease resistance, production potential, production efficiency and meat quality of Afrikaner, Nguni, Drakensberger and Bonsmara breeds have been undertaken (e.g. Muchenje *et al*., 2008; Muchenje *et al*., 2009; Marufu *et al*., 2011). However, comprehensive research on the genetic potential regarding these characteristics

are limited. This therefore emphasized the need for a genome wide scan of these breeds as a first step toward understanding their genetic merit.

The recent development in molecular genetics and bioinformatics such as whole genome sequencing technology has enabled the development of genome wide SNP DNA arrays for many livestock including cattle (The Bovine Genome Sequencing and Analysis Consortium *et al.*, 2009). This has identified more than ten million single nucleotide polymorphism (SNP), which could explain high percentage of the phenotypic variation in cattle. The availability of these massive on ten million SNP markers has resulted in the development of the Bovine SNP7K, SNP50K and SNP770K BeadChips by Illumina in collaboration with the USDA-ARS, University of Missouri, and the University of Alberta (Illumina, San Diego, CA). These chips allow simultaneous high throughput interrogation of large number of loci with high measurement precision (Matukumalli *et al*., 2009) and thus they present opportunities to study South African cattle breeds in order to unravel population structure as well as the genetic potential of these breeds.

2.4 SNP discovery and SNP array development

The whole-genome sequencing for most livestock included both whole-genome shotgun sequencing (WGS) and BAC-to-BAC sequencing (Green, 2001). This revealed a large number of genetic variants across the genomes of livestock species which mostly consisted of SNPs, for example in cattle, ~2.2 million draft SNPs were detected with one SNP per kb (The Bovine HapMap Consortium, 2009). Before application, these SNPs required further validation as the process of SNP discovery was also prone to errors associated with SNP prediction software, sequencing errors and false discovery of SNP (Fan *et al*., 2010). As a results high quality control (QC) criteria were set up after the whole-genome sequencing and HapMap projects were completed to filter the draft SNPs and to select candidate SNPs for placement on the SNP arrays. Quality filters implemented stipulated that (i) each allele of the SNP is included in at least two sequence reads; ii) there were no repetitive elements surrounding the SNP (within 100 bp); iii) the SNP must be predicted by a minimum of six sequence reads; and iv) the predicted SNPs did not overlap with complex regions (e.g. duplicated sequences) (Matukumalli *et al*., 2009). Furthermore, the physical distribution of SNPs evenly across the genome and reasonable intervals between neighbouring SNPs (except Y) were prioritized. The Illumina (Illumina, San Diego, CA) and Affymetrix (Affymetrix.com) platforms developed the currently utilised commercial SNP array. Table 2.1 shows differences in SNP array developed by the Illumina and Affymetrix companies. However, only Illumina products will be discussed in this review.

Table 2.1 Number of SNPs contained in the different chips produced by the Illumina (Illumina, San Diego, CA) and Affymetrix companies (Affymetrix.com)

	Illumina			Affymetrix	
Species	Lower density	Medium density	High density	Medium density	High density
Chicken		60K			600K
Dog		50K	170K	50K	127K
Cattle	7K	50K	770K		640K
Sheep	5K	50K	600K		
Goat		50K			
Horse		50K			
Porcine		60k	510K		

Illumina in collaboration with the USDA ARS, University of Missouri, and the University of Albert (Illumina, San Diego, CA) were involved in the design of the commercially available Bovine SNP7K, SNP50K and SNP700K Bead arrays (Matukumalli *et al*., 2009). The design of Bovine SNP arrays required that each SNP is processed and assigned an Illumina Infinium design score related to assay performance. The design score for each SNP included on the assay was evaluated based on the SNP performance on the genotypes of 576 animals (Matukumalli *et al*., 2009). Infinium I assays (which require two beads to interrogate a SNP) was found to have lower performance than did the Infinium II assays (which require only one bead), regardless of SNP source. Furthermore the utility of the SNPs set were examined by characterizing SNPs allele frequencies among multi-breed panels (Table 2.2). Matukumalli *et al*. (2009) revealed that 51,383 (95%) of the 54,001 called SNPs from the Bovine SNP50 BeadChip v1 were polymorphic among the 558 cattle belonging to 21 cattle breeds, with an average minor allele frequency of 0.26 across the entire set.

Table 2.2 Summary statistics on minor allele frequency for different cattle breeds involved in the validation of the Bovine SNP50 BeadChip (Matukumalli *et al*., 2009)

Breed	**Average MAF**[1]	**Informative**[2]	**Heterozygous**[3]	**MAF**				
				0.1	**0.2**	**0.3**	**0.4**	**0.5**
Taurine								
Hereford	0.27	0.89	0.29	0.15	0.17	0.15	0.21	0.21
Charolais	0.26	0.91	0.31	0.15	0.2	0.18	0.2	0.19
Holstein	0.26	0.90	0.31	0.16	0.17	0.18	0.20	0.20
Piedmontese	0.26	0.89	0.31	0.15	0.18	0.19	0.17	0.20
Norwegian Red	0.26	0.88	0.31	0.13	0.20	0.17	0.20	0.18
Limousin	0.25	0.90	0.30	0.17	0.19	0.17	0.19	0.19
Romagnola	0.25	0.84	0.28	0.16	0.18	0.17	0.16	0.18
Angus	0.25	0.89	0.30	0.16	0.18	0.16	0.19	0.19
Red Angus	0.26	0.84	0.30	0.12	0.19	0.15	0.2	0.18
Guernsey	0.25	0.80	0.27	0.13	0.19	0.15	0.17	0.16
Jersey	0.24	0.78	0.26	0.17	0.16	0.14	0.16	0.15
Brown Swiss	0.25	0.80	0.27	0.16	0.16	0.16	0.15	0.16
Simmental	0.30	0.62	0.30	0.00	0.24	0.00	0.25	0.12
Gelbvieh	0.30	0.65	0.30	0.00	0.25	0.00	0.26	0.13
Taurine x Indicine								
Beefmaster	0.26	0.92	0.32	0.16	0.19	0.19	0.17	0.20
Santa Gertrudis	0.25	0.91	0.30	0.18	0.19	0.18	0.17	0.19
African								
Sheko	0.24	0.75	0.25	0.15	0.17	0.13	0.16	0.14
N'dama	0.24	0.64	0.20	0.14	0.15	0.11	0.13	0.11
Indicine								
Brahman	0.18	0.76	0.19	0.28	0.19	0.11	0.10	0.08
Gir	0.19	0.59	0.16	0.20	0.14	0.10	0.07	0.08
Nelore	0.19	0.59	0.15	0.21	0.13	0.09	0.08	0.07

[1]Average MAF calculated across all loci including the monomorphic SNP within a given breed.

[2]The fraction of informative SNP with MAF $\geq$ 0.01.

[3]The fraction of heterozygous SNP averaged across all animals within a breed.

The evaluation of the BovineSNP50 array by Matukumalli *et al*. (2009) suggested that this array enables simultaneous high throughput interrogation of tens of thousands of loci with high measurement precision to investigate genetic variation in any cattle breed. However it is important to point out that the design of these SNP arrays was associated with ascertainment bias (Matukumalli *et al*., 2009; Helyar *et al*., 2011). Ascertainment bias occurs when the markers are generally identified in a small panel of individuals from part of the specie's genomic range. For example the inclusion of SNPs only occurring more than the predefined number of times (MAF) among European breeds during the design of the BovineSNP50 array resulted in a higher proportion of the markers within indicine and African breeds having lower MAF. The ascertainment bias of the SNP50 BeadChip made the assay more useful amongst common beef and dairy breeds in the U.S. and Europe, and reduced its power for genome wide scan studies within African and indicine breeds (Matukumalli *et al*., 2009).

Ascertainment bias is applicable to all analyses that relies on allele frequency. The bias associated with not sampling rare SNPs results in over estimation of the average diversity of polymorphic sites and under estimation of the average diversity across all sites (Helyar *et al*., 2011). Ascertainment bias will therefore affect the estimates of nucleotide diversity, population size demographic changes, linkage disequilibrium, selection sweep and inference of population structure (Nielsen *et al*., 2005).

2.5 Applications of genomic tools in population genetics and diversity studies

Genetic differentiation and population structure

Previous studies on genetic diversity and structure analyses have used low density microsatellite markers, mitochondrial or Y- chromosomes (Qwabe *et al*., 2012; Soma *et al*., 2012). However in recent years SNPs have become the markers of choice in genetic diversity studies probably due to the fact that they are plentiful in the genome, more stable and adjustable to high throughput automation analysis and also due to the availability of genome wide SNP arrays. (Edea *et al*., 2013; Dekkers *et al*., 2014).

Several statistical approaches are available to study genetic differentiation and population structure of different breeds and populations. Genetic diversity is usually expressed as the frequencies of genotypes and alleles, the proportion of polymorphic loci and the observed and expected heterozygosity. To measure diversity within populations, the expected heterozygosity (H_e) or gene diversity is the most widely used parameter (Nei, 1973). Alternatively the genetic diversity can be measured by the allelic diversity (number of alleles segregating in the population); this parameter is of key relevance in

conservation programs (Toro *et al.*, 2009). A high number of alleles imply more genetic variation (Nei, 1978). The mean number of alleles detected depends on sample size of the population because of the potential presence of unique alleles in a population that may occur at low frequencies. The number of detected alleles may increase with an increase in population size. Therefore it is important to sample population sizes that are more or less equal for comparison. Allelic diversity is also important from a long-term perspective, as the limit of selection response is determined by the initial number of alleles. It is more sensitive to bottlenecks than expected heterozygosity as it reflects past variations in population size more accurately (Toro *et al.*, 2009).

In a structured population with *n* breeds/populations, the total gene diversity is partitioned into a component within breeds/populations and another between breeds/populations (Toro *et al.*, 2009). The analysis of molecular variance (AMOVA) can be used (Excoffier *et al*., 2005) to illustrate the partition of gene diversity into components.

The population subdivision involve an inbreeding-like effect, therefore its effect is usually measured in terms of decrease in the proportion of heterozygous genotypes (Hartl, 1988). The effects of population subdivision are measured by a quantity called the fixation index (Wright, 1978). These include: F_{ST} which measures the reduction in heterozygosity of a subpopulation due to random genetic drift; F_{IS} which is the inbreeding coefficient concerned with inbreeding in individual (I) relative to the total subpopulation (S) to which they belong and F_{IT} which measures the reduction in heterozygosity of an individual relative to the total population (T) (Hartl, 1988).

The genetic relationship between populations can be measured by estimating the genetic distance between populations. The difference measured between two populations offers a good estimate of how different they are genetically (Nie's, 1978). Nie's, (1978) unbiased genetic distance estimate is one of the common measurements of genetic distance.

Population structure can be described using ADMIXURE computer software (Alexander *et al.*, 2009), which implements a model-based clustering method for inferring population structure using genotypic data. This software is suitable for the assignment of populations and assumes a model in which there are K populations (where K may be unknown), each of which is characterized by a set of allele frequencies at each locus. Individuals are assigned to populations according to their membership confidence for each cluster which is interpreted as a probability of membership. This program may be used to assign individuals correctly to a population or a breed, especially when the phenotypic differentiation between breeds/ populations is difficult to detect or when genealogical information is absent. Molecular markers can detect whether introgression or crossbreeding occurred (Alexander *et al.*, 2009).

Using the above mentioned statistical procedures genome wide characterization studies can unravel possible historical events that occurred and shaped population during domestication and breed formation. Such information can assist in preserving the genetic diversity within endangered indigenous breeds and also with scientific conservation of adaptation traits (Fan *et al*., 2010). Using SNP data Mackay *et al*. (2008) showed that the genetic diversity of breeds is linked to their areas of origin, suggesting that breeds that have diverged more recently were generally closer together geographically. They also showed larger difference between taurine and indicine breeds due to greater divergence time. In another study Hayes *et al.* (2009) revealed significant differences between beef and dairy cattle and this was attributed to divergent selection pressure across these breeds Dekkers *et al.* (2014) analysed the phylogenetic relationships among 372 animals from 48 cattle breeds using the BovineSNP50 array. Their results were consistent with the biogeography of breeds but also indicated the presence of admixed populations and revealed pedigree relationships between individuals. Furthermore the genetic analyses of 403 individuals from 23 sheep breeds and 210 individuals from two wild sheep species with 1,536 SNPs revealed that the genetic variability within both African and Asian sheep breeds were lower than those between European breeds (Kijas *et al.*, 2009). This study by Kijas *et al*. (2009) also revealed that the genetic distances between individuals from African and Asian breeds was smaller than those of European breeds. The genetic relationships among breeds were also consistent with the geographical distribution and history of breed formation (Kijas *et al.*, 2009).

Whole-genome LD patterns

Linkage disequilibrium (LD) is the non-random association of alleles at different loci (Hayes *et al*., 2003). It usually arises due to migration, mutation, selection, small finite population size or other genetic events which the population has been subjected to. In other cases LD is deliberately created through F2 QTL design by crossing two inbred lines or backcrossing (Hayes *et al*., 2009). Linkage disequilibrium have been used to investigate genes underlying genetic variation in different cattle breeds and populations. However, linkage disequilibrium is population specific and has some degree of heterogeneity between populations depending on the demographic histories of animals under investigation (Hayes *et al*., 2008). Thus, it is essential to estimate the extent of LD in a given population in order to predict the best array size for future genome wide association studies, genomic selection studies or genome wide scan for selection signatures (Goddard & Hayes, 2007). For example, for population with longer range LD, there is less value in moving to a higher density SNP array due to most QTL already being in LD with markers on a smaller array. However, if a population has a relatively short LD range, then not all QTL may be in LD with markers on a smaller array, making it necessary to use larger SNP array. Figure 2.1 shows the variations in the extent of LD (r^2) between populations (Tenesa *et al*., 2007).

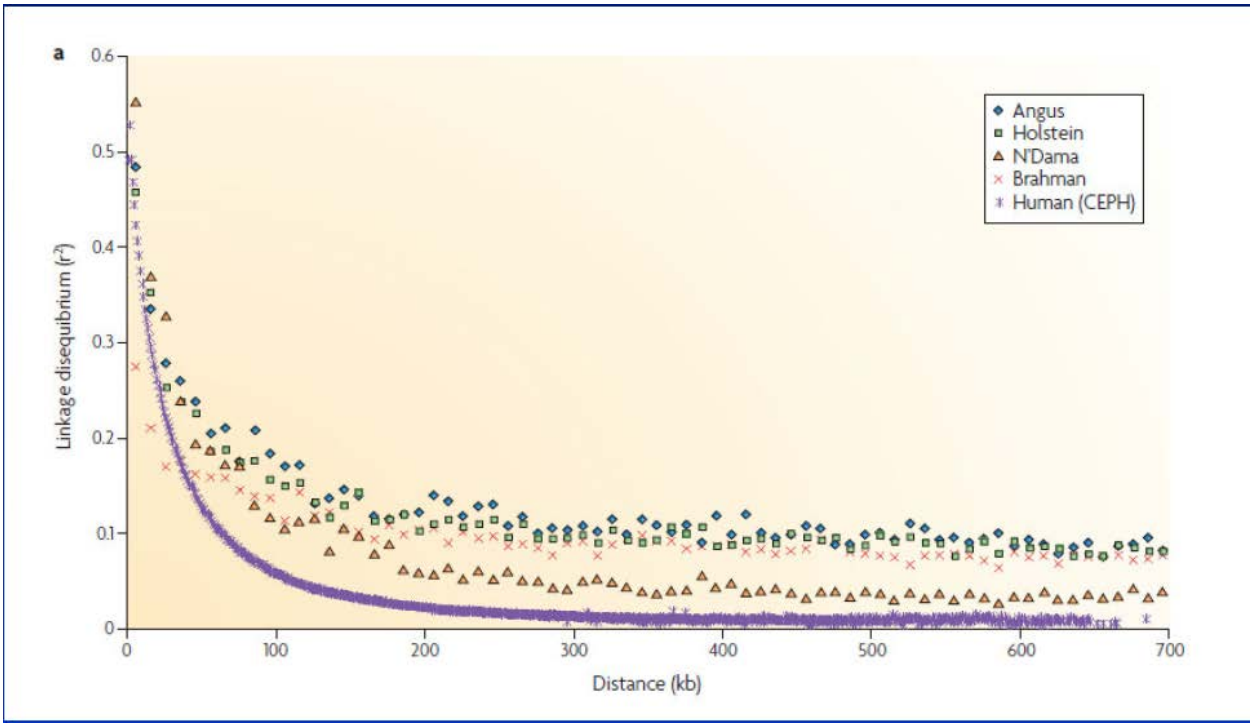


Figure 2.5 Extent of LD in humans and livestock (Tenesa *et al*., 2007).

Evolutionary and molecular forces acting on different breeds or populations have been shown to be responsible for the variation on the extent of LD (Hayes *et al*., 2008). Selection is considered to be an important source of LD, however, its effect is likely to be localized around specific genes and thus has moderate effects on the amount of average LD over the whole genome. Small finite population size is generally implicated as the key source of extensive LD in livestock populations that are characterized by small effective population size (N_e). Thus, the extent of LD in the genome can be used to infer ancestral effective population size. Effective population size is the number of individuals in an idealized population that would give rise to similar rate of inbreeding as observed in the actual breeding population and is an important population parameter that can assists to explain how different populations have evolved (Hayes *et al*., 2003).

Previous studies have shown that moderate LD ($r^2 \geq 0.2$) extends up to 100 kb in cattle while its can only extends up to 5 kb in Humans (Tenesa *et al*., 2007). However, very high LD ($r^2 \geq 0.8$) only extends a very short distance (1 kb) in both humans and cattle (Tenesa *et al*., 2007). Results from MacKay *et al*. (2007) revealed that moderate LD extended up to between 40 and 60 kb in cattle, which suggested that about 50 000 SNPs would capture most of the LD information necessary for whole genome wide scan studies in *Bos taurus* cattle populations. Similarly, García-Gámez *et al*. (2012) observed similar trends of LD decay in sheep, where moderate LD extended up to between 20-40 kb using SNP data. On the other hand, Du *et al*. (2007) revealed that pig may have considerable more LD

than cattle as average r^2 values of 0.20 were observed even for SNPs that were separated by more than 0.2 Mb.

Detection of signatures of selection

Unlike other demographic processes that affect the entire genome, selection affects specific functionally important sites in the genome (Nielsen, 2005). When selection acts on a mutation it will affect linked sites and leave its signature in the adjacent chromosomal regions thereby creating the signals close to the selected genes (Simianer & Quanbari *et al*., 2014). These signals can be: (i) allele frequency spectrum that will be shifted towards extreme (high or low) frequencies, (ii) an excess of homozygous genotypes and (iii) long haplotypes with high frequency. Different approaches uses one or a combination of these signals to detect signatures of selection. Briefly, whole genome scan is conducted in which each single locus (e.g. SNP) is given a value based on a chosen test statistics. However in some cases, point-wise statistic are discouraged as they are usually associated with random noise. Thus a moving window or a creeping window approach are used to smoothen the picture and remove the noise (Simianer & Quanbari *et al*., 2014).

Signature of selection studies differ from classical genetics approach that starts with a phenotype to the identification of underlying causal genes and mutations. They follow a reverse approach that start with a signature of selection and then endeavor to infer selected mutation and its associated phenotype (Qanbari *et al*., 2011). Detection for signatures of selection however remains a challenge, because the effect of selection can be confused with the effect of demographic factors (e.g. the size, structure, matting patterning of a population (Nielsen, 2005). For example, both adaptive hitching and population expansion can result in the excess of rare variants observed in DNA sequence data compared to what is known under a standard neutral model (Tajima, 1989). In addition most of the available data consist of SNP genotypes that were discovered using the ascertainment process (Ramey *et al*., 2013) which will affect the level of variability, distribution of allele frequencies, and level of linkage disequilibrium (Nielsen *et al*., 2005). However, despite these challenges, detection for signatures of selection has been the focus of theoretical (simulated) and empirical (observed) studies (Sabeti *et al*., 2002; Nielsen, 2005; Ramey *et al*., 2013). Different methods are available for genome wide detection of selection signatures using molecular data. These methods are classified based on the main variable that affect the pattern of molecular variation left by selection (Nielsen, 2005). Table 2.3 shows general approaches of detecting selection in genome wide selection studies.

Table 2.3 General approaches for detecting selection in genome wide selection studies

Test	**Pattern**	**Signatures**
Tajima's D and related	Frequency spectrum	A relative increase in the proportion of either low- or high-frequency mutations in the selected region (Tajima 1989)
dN/ dS ratio tests	Non-synonymous and synonymous substitution	Elevated ratio of nonsynonymous (N) to synonymous(S) changes (dN/dS) in coding regions of selected genes evolving under the assumed neutrality (Nielsen& Yang 1998).
Test based on reduction in variation	Local reduction in genetic variation	A signature decrease in genetic variation (often measured as heterozygosity) around the selected site relative to its chromosomal neighbour or genome-wide (Olekskyk *et al*. 2008).
F_{ST} based and related tests	amount of population subdivision	An increase or decrease in the population differentiation in genomic regions under selection relative to the rest of the genome (Akey *et al*. 2002)
Tests based on LD	LD and/or haplotype structure	Extended LD producing remarkably long haplotypes around the selected region Sabeti *et al*. 2002; Voight *et al*. 2006)

Test based on the frequency spectrum: detecting changes in the shape of the frequency distribution of genetic variation

The Tajima test was developed to detect selection based on excess of rare allele (Tajima, 1989). This approach uses Tajima's test to measure the differences between two estimators of the population mutation rate, θ_w and π_w (Tajima, 1989). Under neutrality, the means of θ_w and π_w should be equal to

one another and the expected value of Tajima's D (corresponds to the standardized difference θ_w and π_w) between populations conforming to neutral model should be zero. Any significant deviation from zero suggests a skew in allele frequency distribution relative to neutral expectations and thus a signature of selection. However, the results of this test do not always have a clear interpretation as in some cases it is impossible to differentiate between positive and negative selection and also because this test is sensitive to demography (Tajima, 1989).

Test based on synonymous and non-synonymous substitution rates: detection for increased rates of functional mutations

It is expected that when the coding sequence of orthologous genes of interest are compared with what occurred under neutral evolution, the rate of mutation as expressed as the number of substitutions per non-synonymous site (d_n)/ the number of substitutions per synonymous site (d_s) is equal to one. However when positive selection is in effect $d_n/d_s<1$, and under negative selection $d_n/d_s >1$ (Nielsen & Yang, 1998). This approach identifies genomic regions that has been the target of past selection by comparing the ratio of non-synonymous to synonymous changes (d_n/d_s test). Signatures of selection are observed by an elevated ratio of non-synonymous to synonymous changes (d_n/d_s test) in coding regions of selected genes compared with other genes evolving under the assumed neutrality. The non-synonymous to synonymous ratio gives information about the evolutionary forces operating on a particular gene. (Nielsen & Yang, 1998).

Test based on an excess of homozygous genotypes: local reduction in genetic variation

A local reduction in variation within a selected gene and adjacent SNP variants is considered a genomic signature of selection (Maynard Smith & Haugh, 1974; Ramey *et al.,* 2013). This local reduction in genetic diversity can persist over a long genomic region and suggest a long term selection (Maynard Smith & Haigh, 1974). While it may be easy to detect regions of the genome where haplotype have been driven to complete fixation, their interpretation can be influenced by several limitations. Firstly, the observed selection sweep may not be easily distinguished from the effects of demographic history because population bottlenecks or recent founder effects can reduce polymorphism across the genome of the derivative population (Ramey *et al*., 2013). To distinguish between these effects, Hayes *et al*. (2009) suggested looking at the location of the loci. For example, demographic events are expected to change the patterns of allele frequencies across the entire genome

while selection is expected to change allele frequencies only at the loci which are in close proximity to the selected mutation (Hayes *et al*., 2009).

Test based on population differentiation: differentiation between populations (F_{st})

Several tests that are aimed at detection of selection signatures are based on the estimation of F_{st} over multiple loci (Akey *et al*., 2002). This statistic assumes that geographically variable selection pressures favor different variants in different genomic regions. Thus, between-population allele frequency differences may be more extreme in genome regions harboring such variants (Qanbari *et al*., 2011). By comparing F_{st} among loci and between populations one can get an estimate of how much genetic variability exists between, rather than within populations. Since natural selection is locus specific, environmental forces can cause systematic deviations in F_{st} values for a selected gene and nearby genetic markers resulting in highly differentiated regions between populations. Selection signatures will then be observed if genetic differentiation in the genomic region is greater than the level expected under neutrality (Akey *et al*., 2002). This method is however sensitive to demographic factors including migration and genetic drift (Oleksyk *et al*., 2008). To bypass this limitation, researchers have taken advantage of large-population datasets and compared outlier loci with the empirical distribution of population differentiation across the genomes of the compared populations (Oleksyk *et al*., 2008).

Test based on linkage disequilibrium: Extended linkage disequilibrium segments

Tests based on the exploitation of LD have been widely used in the detection of signatures of selection (Sabeti *et al*., 2002). However these signatures are likely to be temporary since recombination tend to quickly break down this LD as soon as the selected locus reach fixation (Nielsen, *et al*., 2005). To capture these signatures Sabeti *et al*. (2002) developed an approach known as the long-range haplotype (*LRH*) method to identify selection signatures by examining the relationship between the allele frequency and the LD level surrounding it. Signatures of selection are observed if one core haplotype has a combination of high relatively extended haplotype homozygosity ((*REHH*) and occur at high frequency in population (Sabeti *et al*., 2002). Voight *et al*. (2006) developed the Integrated Haplotype Score (iHS) which is an extension of extended haplotype homozygosity (*EHH*) based on the comparison of *EHH* between alleles within a population, and the detected wide-spread signals of positive selection using the human genome data. However, this approach lacks the power to detect selection signals that have resulted in near or compete fixation of an allele in a population and may fail to detect a significant fraction of loci that have experienced localized positive selection (Tang *et al*., 2007).

Signatures of selection studies in cattle breeds

A number of studies using different analytical concepts have been conducted to detect signatures of selection on a genome wide scale in cattle using SNP data (Prasad *et al*., 2008; Qanbari *et al*., 2010; Chan *et al*., 2010; Ramey *et al*., 2013; Porto-Neto *et al*., 2013). These studies have detected interesting genomic regions which are important to the understanding the role of natural and artificial selection in shaping the pattern of genetic variation in cattle.

Hayes *et al*. (2009) identified regions of the genome that have been differentially selected between production systems by comparing minor allele frequencies of SNPs in sliding window across the genome between beef and dairy cattle breeds. For example, by comparing MAF between beef and dairy cattle they observed the largest differences on chromosome 20 in the region of the *GHR* gene. This region carries a mutation known to have a large effect on milk production traits in a number of dairy populations (Blott *et al*., 2003). Another gene observed was the *ABCG2* between Holstein and Angus that harbours a mutation affecting protein percentage in milk (Cohen-Zinder *et al*., 2005).

Ramey *et al*. (2013) identified 28 putative selected sweep regions within 14 diverse cattle breeds by analyzing localized reduction in genetic diversity. These selective sweeps were in the genomic regions harboring genes for horned vs polled, coat color, stature and ear morphology. Using the population differentiation approach Gautier *et al*. (2009) identified 53 genomic regions under selection in West African cattle. Most of these were related to immune response, nervous system, skin and hair properties. Using similar approach Chan *et al*. (2010) detected 14 genomic regions with significantly different allele frequencies between Zebu and Taurine cattle. The functional genomic analysis of these regions pointed towards signatures of selection on tropical adaptation genes, including keratins, heat shock and heat resistance genes. In addition genomic regions relevant to behaviour, immune response and feed efficiency were discovered based on F_{st} estimates (The Bovine HapMap Consortium, 2009a).

Using EHH and iHS approaches Barendse *et al*. (2009) revealed a total of 12 putative selective sweeps associated with residual feed efficiency, beef yield and intra-muscular fatness. In addition, a set of genes including *GHR*, *MC1R*, *FABP3*, *CLPN3*, *SPERT*, *HTR2A5*, *ABCE1*, *BMP4* and *PTGER2* were subjected to selection in cattle (Flori *et al*., 2009; Qanbari *et al*., 2010). Another study by Qanbari *et al*. (2011) identified a total of 236 genomic regions potentially under selection in Holstein cattle using both population differentiation (F_{st}) and Integrated Haplotype Score (iHS) approaches. Both approaches suggested selection in vicinity of the Sialic acid binding Ig-lecitn 5 gene on BTA18, a region which was shown to contain a major QTL with strong effect on productive life and fertility traits in Holstein cattle.

There are several approaches for detection of selection signatures each approach captures specific pattern of molecular variation. Combining alternative approaches for detection of selection signatures has been suggested as a means of increasing the reliability of these studies. However, the success of one test and the failure of another does not exclude the possibility of a selection signature as different tests focus on different signatures left by or selected at different times (Simianer *et al*., 2014). In this thesis signatures of selection were identified using two approaches namely: detection for local reduction in genetic variation and identification of highly differentiated genomic regions between populations (F_{st}). These methods were chosen as they have been shown to possess high statistical power in identification of signature of selection within (Ramey *et al*., 2013) and among cattle breeds respectively (Porto-Neto 2013; Gautier *et al*., 2009).

2.6 Conclusion

South African cattle breeds possess a wide range of variation which could be the result of a combination of various processes, including domestication, migration, genetic isolation, environmental adaptation, selective breeding, introgression and admixture of subpopulations. The availability of genomic tools present opportunities to study South African cattle breeds at the genomic level in order to discovery their unique genetic structure and to unravel their genetic potential with regard to production, reproduction, diseases resistance and adaptation. Identifying signatures of selection among South African cattle breeds will be the first step towards discovering their unique genetic structure and can provide insight into genome requirement for survival in the African environment.

References

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

Alderson, L. (2010). Breeds at risk: criteria and classification. Report from a seminar held in London 16-17 February 2010.

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

Barendse, W., Harrison, B.E., Bunch, R.J., Thomas, M.B. & Turner, L.B. (2009). Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* 10, 178.

Bergh, L., Gerhard, R., Scholtz, M.M. & Mamabolo, M.J. (2010). Introduction to the information on beef and dual purpose breeds in South Africa. Part two in: Beef breeding in South Africa. 2nd ed. Ed.

Scholtz, M.M. Agricultural Research Council, Animal Production Institute, Irene. Published by Agricultural Research Council. pp. 150-291.

Blott, S., Kim, J.J. & Moisio, S. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163, 253-266.

Bothma, A.J. (1993). Comparison of the Government Nguni studs in Venda. Minst. Agrar. Thesis. University of Pretoria, South Africa

Bosman, D.J & Scholtz, M.M. (2010). Selection cattle for functional efficiency. In: Beef breeding in South Africa. Ed. Scholtz MM 2nd Edition, Pretoria, Agricultural Research Council.

Bonsma, J.C. (1980). Livestock Production – A global approach. Tafelberg, Cape Town.

Chan, E.K.F., Nagaraj, S.H. & Reverter, A. (2010). The evolution of tropical adaptation: comparing taurine and zebu cattle. *Animal Genetics* 41, 467-477.

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J., Everts-Van Der Wind, A., Lee, J.H., Drackley, J.K., Band, M.R., Hernandez, A.G., Shani, M., Lewin, H.A., Weller, J.I. & Ron, M. (2005). Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research* 15, 936-944.

Dekkers, 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., Regitano, L.C.A., 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 Genetics* 10 (3), e1004254. http://dx.doi.org/10.1371/journal.pgen.1004254

Dreyer, C.J. (1982). The breed structure of the Drakensberger cattle breed and factors that influence the efficiency of production. DSc (Agric) Thesis, University of Pretoria, South Africa.

Du, F-X., Clutter, A.C. & Lohuis, M.M. (2007). Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Science* 3, 166-178.

Edea, Z., Dadi, H., Kim, S., Dessie, T., Lee, T. & Kim, H. (2013). Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Frontiers in Genetics*, doi:10.3389/fgene.2013.00035.

Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin version 3.0: An integrated Software Package for Population Genetics Data Analysis. Computational and Molecular Population Genetic Laboratory (CMPG), Institute of Zoology, University of Berne, Switzerland.

Fan, B., Du, Z-Q., Gorbach, D.M. & Rothschild, M.F. (2010). Development and application of high-density SNP arrays in genomic studies of domestic animals. *Asian - Australasian Journal of Animal Sciences* 7, 833-847.

FAO. (2005). Genetic characterization of livestock populations and its use in conservation decision making, by O. Hannotte & H. Jianlin. In J. Ruane & A. Sonnino, eds. The role of biotechnology in exploring and protecting agricultural genetic resources, pg. 89–96. Rome 1.

FAO. (2007). The State of the World's Animal Genetic Resources for Food and Agriculture, Rome, Italy.

FAO. (2010). Progress report of the FAO on selected activities related to agricultural biodiversity. http://www.fao.org/docrep/meeting/018/k5700e.pdf

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M. & Gu, I. (2009). The genome response to artificial selection: A case study in dairy cattle. *PLoS ONE* 4, e6595.

García-Gámez, E., Sahana, G., Gutiérrez-Gil1, B. & Arranz, J.J. (2012). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* 13, 43 http://www.biomecentral.com/1471-2156/13/43

Gautier, M., Flori, L., Reibler, A., Jaffrézic, F., Laloe, D., Gut, I., Moazimi-Goudarzi, K. & Foulley, J.L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10, 550.

Gertenbach, W.D., Dugmore, T. J., Schotlz, M.M & Kirkman, K. (2011). The production potential of Nguni cattle. In: The Nguni breed of cattle - past, present and future. Nguni Cattle Breeders' Society of South Africa. The Nguni Cattle Breeders' Society 2011. ISBN: 978-0-620-51723-2. South Africa.

Goddard, M.E., & Hayes, B.J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics* 124, 323–330.

Green, E.D. (2001). Strategies for the systematic sequencing of complex genomes. *Nature Reviews Genetics* 2, 573-583.

Hartl, D.L. (1988). A primer of population Genetics. 2nd edition. Sinauer Associates, Inc. Publisher, Sunderland, MA, USA. pg 78.

Hayes, B.J., Visscher, P.M., McPartlan, H.C., & Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* 13, 635–643.

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

Hayes, B.J., Chamberlain, A.J., Maceachern, S., Savin, K., McPartlan, H., Macleod, I., Sethuraman, L. & Goddard, M.E. (2009). A genome map of divergent artificial selection between *Bos taurus* dairy and Bos taurus beef cattle. *Animal Genetics* 40, 176-184.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., Maes, G.E., Diopere, E., Carvalho, G.R. & Nielsen, E.E. (2011). Application of SNPs for population genetics of non-model organisms: new opportunities and challenges. *Molecular Ecology Resources* 11, 1-14.

Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., Mc William, S., Tang, D. & McEwan, J. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE*, 4(3), e4668.

Marufu, M.C., Qokweni, L., Chimonyo, M. & Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. *Ticks and Tick Borne Diseases* 2, 172-177.

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

Maynard-Smith, J. & Haugh, J. (1974). The hitch-hiking effect of a favourable gene. *Genome Research* 23, 23-25.

McKay, S.D., Schnabel, D., Murdoch, M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Dias Neto, E., Gill, C.A., Gao, C., Mannen, H., Stothard, P., Wang, Z., Van Tassel, C.P., Williams, J.L., Taylor, J.F. & Moore, S.S. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* 8, 74-85.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W. (2008). An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genetics* 9, 37. doi: 10.1186/1471-2156-9-37.

Meyer, E.H.H. (1984). Chromosomal and biochemical genetic markers of cattle breeds in Southern Africa. In Proceedings of the 2nd World Congress on Sheep and Beef Cattle Breeding (pp. 328-339), Pretoria, South Africa.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G. & Strydom, P.E. (2007). Meat quality of Nguni, Bonsmara and Angus steers raised on natural pasture in the Eastern Cape, South Africa. *Meat Agricultural Research* 4, 239-245.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G. & Strydom, P.E. (2008). Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. *South African Journal of Animal Science* 2, 298-304.

Muchenje, V., Dzama, K., Chimonyo, M., Strydom, P.E. & Raats, J.G. (2009). Relationship between stress responsiveness and meat quality in three cattle breeds. *Meat Science* 81, 653-657.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proceedings of Natural Academic Science USA 70, 3321-3323.

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

Nielsen, R. & Yang, Z. (1998). Likelihood models for detecting positively selected amino acids sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929-936.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Reviews Genetics* 39, 197-218.

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

Nortier, C.L., Els, J.F., Kortze, A. & van der Bank, F.H. (2002). Proceedings of the 7th World Congress on Genetic Applied to Livestock Production. Montpellier, France, August: Communication 26-07.

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

Prasad, A., Schnabel, R.D., McKay, S.D., Murdoch, B., Stothard, P., Kolbehdar, D., Wang, Z., Taylor, J.F. & Moore, S.S. (2008). Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal Genetics* 39, 597–605.

Porto-Neto, L.R., Lee, S.H., Sonstegard, T.S., Van Tassell, C.P., Lee, H.K., Gibson, J.P. & Gondro, C. (2013). Genome-wide detection of signatures of selection in Koren Hanwoo cattle. *Animal Genetics* 45, 180-190.

Qanbari, S., Pimentel, E.C.G, Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. & Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics* 41, 346-356.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G. & Simianer, H. (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* 12, 318.

Qwabe, S.O., van Marle-Köster, E. and Visser, C. (2012). Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. *Tropical Animal Health and Production* 45 (2), 511-516.

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D. & Taylor, J.F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14, 382.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P, Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platkon, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R. & Lander, E.S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832-837.

Scherf, B.D. (2000). World Watch List for domestic animal diversity. 3rd edition. FAO, Rome.

Scholtz, M.M. (1988). Selection possibilities of hardy beef breeds in Africa: the Nguni example. Proceedings of the 3rd World Congress on Sheep & Beef Cattle Breed. Paris, 19–23 June 1988.

Scholtz, M.M., Roux, C.Z. & Lombard, P.E. (1990). Breeding strategies for beef cattle in the subtropics and tropics: Terminal crossbreeding. Proceedings of the 4th World Congress on Genetics Applied to Livestock Production. 15, 361-364.

Scholtz, M.M., Lombard, P.E. & Enslin, C.B. (1991). A note on the early mating of beef heifers. *South African Journal of Animal Science* 21, 55-61-.

Scholtz, M.M. (2010). Beef breeding in South Africa. 2nd Edition, ARC Pretoria, South Africa (ISBN-13: 978-1-86849-391-3).

Scholtz, M.M., Bosman, D.J., Erasmus, G.J. & Maiwashe, A. (2010). Selection as the base of improvement in beef cattle, in Beef Breeding in South Africa, 2nd edition., Ed MM Scholtz (Pretoria: Agricultural Research Council), 2-10.

Scholtz, M.M., Gertenbach, W.D. & Hallowell, G. (2011). The Nguni breed of cattle: past, present and future. The Nguni Cattle Breeders' Society 2011. ISBN: 978-0-620-51723-2.

Scholtz, M.M. & Ramsey, K.A. (2011). The Nguni Breed of Cattle—Past, Present and Future, Nguni Cattle Breeders' Society, Bloemfontein, pp. 14-17 (Chapter 2).

Simianer, H. Ma, Y. & Qanbari, S. (2014). Statistical Problems in Livestock Population Genomics. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. Vancouver, BC, Canada, 17-22 August 2014.

Soma, P., Kotze, A., Grobler, J.P. and van Wyk, J.B. (2012). South African sheep breeds: Population genetic structure and conservation implications. *Small Ruminant Research* 103, 5926 pp. 528-532 DOI: 10.1126/science.1167936.

Spickett, A.M., de Klerk, D., Enslin, C.B. & Scholtz, M.M. (1989). Resistance of Nguni, Bonsmara and Hereford cattle to tick in a bushveld region of South Africa. Onderstepoort Journal of Veterinary Research 56, 245-250

Strydom, P.E. (2002). Factors influencing growth, carcass cattle yield and carcass composition of cattle. In: Feedlot management. Ed. Leeuw, K-J. Agricultural Research Council, Animal Production Institute, Irene. pp. 40.

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

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585-595.

Tang, H., Choudhry, S., Mei, R., Morga, M., Rodriguez-Cintron, W., Burchard, E., Gonz, I. & Risch, N.J. (2007). Recent genetic selection in the ancestral admixture of Puerto Ricans. *American Journal of Human Genetics* 81, 626-633.

Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E. & Visscher, P.M. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Research* 17, 520-526.

The Bovine Genome Sequencing and Analysis Consortium (2009). The genome sequence of Taurine cattle: A window to ruminant biology and evolution. *Science* 324, 522-528.

The Bovine HapMap Consortium, (2009a). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528-532.

Toro, M. A., Fernandez, J. & Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science* 120, 174-194.

Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *PLoS Biology* 4, e72.

Wright, S. (1978). Evolution and the Genetics of populations. Variability within and among Natural Populations. Volume 14. University of Chicago Press, Chicago.

Chapter 3

Evaluation of the BovineSNP50 genotyping array in four South African cattle populations

S.O. Qwabe[1,2#], E. van Marle-Köster[2], A. Maiwashe[1] & F.C. Muchadeyi[3]

[1] Agricultural Research Council, Private Bag X2, Irene 0062, South Africa

[2] Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

[3] Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort, 0110, South Africa

Published in South African Journal of Animal Science 2013, 43 (No. 1)

Short communication

Evaluation of the BovineSNP50 genotyping array in four South African cattle populations

S.O. Qwabe[1,2#], E. van Marle-Köster[2], A. Maiwashe[1] & F.C. Muchadeyi[3]

[1] Agricultural Research Council, Private Bag X2, Irene 0062, South Africa

[2] Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

[3] Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort, 0110, South Africa

Abstract

The BovineSNP50 genotyping array is a product with a wide range of applications in cattle such as genome-wide association studies, identification of copy number variation and investigation of genetic relationships among cattle breeds. It also holds potential for genomic selection, especially for traits that are expensive and difficult to measure. The usefulness of this chip for any of these applications depends on the degree of polymorphisms in the cattle breeds. The SNP50 array has not been validated in any South African cattle population and this could lead to overestimating the number of polymorphic SNPs available for application. This study is a first attempt to evaluate the BovineSNP50 genotyping array in the South African cattle population. A total of 96 bovine samples, consisting of 45 Holstein, 29 Nguni, 12 Angus and 10 Nguni x Angus crossbred animals, were genotyped with the BovineSNP50 infinium assay. The results of this study demonstrated that 40 555 SNPs were polymorphic (MAF >0.05) in these breeds and indicated potential for application in South African cattle populations. Genomic information generated from the BovineSNP50 can now be applied in genetic prediction, genetic characterization and genome-wide association studies

Keywords: Call rate, minor allele frequency

[#]Corresponding author: qwabes@arc.agric.za

The BovineSNP50 genotyping array (Illumina Inc., San Diego, CA), which features approximately 54 609 highly informative SNP probes uniformly distributed across the genome of the bovine, first became available in 2007 (Matukumalli *et al*., 2009). Since then the array has gained wide acceptance for use in genome-wide association studies to identify genomic regions contributing to natural variation in phenotypic traits in cattle. Several genomic regions associated with traits of economic

interest have been identified. These include genomic regions associated with feed intake and feed efficiency traits (Sherman *et al*., 2009); milk production traits (Mai *et al*., 2010); growth and feedlot traits (Bolormaa *et al*., 2011) and carcass traits (Lee *et al*., 2010). Furthermore, the BovineSNP50 genotyping array has been used to demonstrate the value of genomic selection where genomic data has been used to supplement extensive sets of performance data to predict genetic merit for application in selection programmes (Meuwissen *et al*., 2001; Hayes *et al*., 2009; VanRaden *et al*., 2009). The success of genomic selection is based on exploitation of the linkage disequilibrium (LD) relationship between SNPs and quantitative traits loci (QTL) affecting a trait (Hayes *et al*., 2009). Genomic information has already been incorporated into the genetic evaluation of dairy and beef cattle in the United States and Canada (VanRaden *et al*., 2009; MacNeil *et al*., 2010, respectively). In addition, the BovineSNP50 array has been used to detect copy number variants (CNV), which have been implicated in both disease phenotype and normal phenotypic variation associated with QTL (Hou *et al*., 2011). Furthermore, the BovineSNP array has been used to investigate genetic relationships among cattle breeds (Decker *et al*., 2009). The information has been applied in studies to detect selection signatures in Holstein cattle (Qanbari *et al*., 2010) and other dairy and beef breeds found on various continents (Ramey *et al*., 2013).

Genetic improvement of cattle for traits of economic importance has primarily been performed using conventional selection based on quantitative genetics for many decades. Phenotypic and pedigree data collected by stud breeders participating in animal recording schemes have been used in genetic evaluations to provide breeders with estimated breeding values. However, the genetic architecture of the trait was treated as a black box, with no knowledge of the number of genes that affect the trait, let alone the effects of each gene or their locations in the genome (Dekkers, 2007). Genomic selection has the potential to increase genetic progress for traits that are difficult and expensive to measure (Meuwissen *et al*., 2001). To date, no local or indigenous South African cattle breeds have been included in either the production or the validation of the BovineSNP50 array. It is therefore necessary to investigate the usefulness of the BovineSNP50 in local breeds and also to evaluate the application in cattle breeds that are widely farmed in South Africa to establish the necessary reference populations. Thus, the aim of the study was to determine the level of polymorphism and allele frequency distribution of the BovineSNP50 in South African populations of Nguni, Angus, and Holstein cattle.

Blood samples were obtained from 29 Nguni and 12 Angus animals and 10 Nguni x Angus F1 cross steers from the beef cattle management and system development project of the Northern Cape Department of Agriculture and Land Reform (DALR) and Agricultural Research Council–Animal Production Institute (ARC-API). Blood samples (10 mL) were collected using EDTA VACUETTE®

tubes. The blood samples were transported on ice to the ARC-API laboratory where they were refrigerated at – 20 ºC until extraction of DNA was performed. For Holstein, 45 semen samples were obtained from an artificial insemination company (Taurus, South Africa). Genomic DNA was extracted at the ARC-Biotechnology Platform laboratory from the whole blood and semen samples using the Qiagen DNeasy extraction kit (Qiagen, South Africa) according to the manufacturer's protocol. The protocol was adapted for the semen samples where Dithiothreitol (DTT) was added with proteinase K in the first step. Genomic DNA for all samples was quantified using a Qubit® 2.0 Fluorometer and the Nano drop Spectrophotometer (Nanodrop ND-1000). In addition, gel electrophoresis was performed to quantify the DNA. DNA samples extracted from blood had 260:280 ratio ranging from 1.8 to 1.94, while eleven of the 45 DNA samples from semen extraction had 260:280 ratio of less than 1.8. DNA concentrations for all samples ranged from 55 to 80 ng/µL.

Genotyping was conducted at the ARC-Biotechnology Platform with the Illumina BovineSNP50 BeadChip v2 which features 54 609 SNP probes distributed across the whole bovine genome with an average spacing of 49.9 kb (Matukumalli *et al*., 2009). Approximately 200 ng (12 µL of DNA loaded in each well of a BeadChip) of genomic DNA was used to genotype each sample. Samples were processed according to the Illumina Infinium–II assay protocol (Illumina, Inc. San Diego, CA 92122 USA). Briefly, each sample was whole-genome amplified for 20 hours at 37 °C. The samples were then fragmented, precipitated and re-suspended in an appropriate hybridization buffer. The samples were hybridized on the prepared BovineSNP50 BeadChip for 20 hours at 48 ºC. Following the hybridization, non-specifically hybridized samples were removed by washing, while the remaining specifically hybridized loci were processed for the single-base extension reaction, stained and imaged on an Illumina iScan Reader.

Genotype data generated from the iScan system were loaded into Illumina Genome Studio version 1.9.0 software, which uses algorithms to perform primary data analysis, including raw data normalization, clustering and genotype calling. A final custom report was created from the genome studio using Plink Input Report 2.1.1, which created a Ped (Pedigree file) and Map (SNP panel file) file to use for downstream analyses.

Basic genotype statistics for each marker, including call rate, minor allele frequency (MAF), Hardy-Weinberg Equilibrium (HWE), allele and genotype counts were calculated using the Quality Assurance Module from the SNP Variation Suite version 7 (SVS; Golden Helix Inc., Bozeman, Montana: www.goldenhelix.com). The following quality control criteria (filters) were used to remove from further analysis any SNPs with less than 95% call rate, and SNPs with less than 0.05 MAF. SNP were

tested for HWE (P <0.001) to identify possible typing error. Samples with more than 10% missing genotypes were removed from the study.

For Nguni, Angus and Nguni x Angus populations all samples were successfully genotyped (>98% of SNP were genotyped). Five Holstein samples were removed because they had more than 10% missing genotype; these samples were part of the eleven samples with less than 1.8 260:280 ratio. Their 260:280 ratio ranged from 1.62 to 1.69, thus it was concluded that these failed to genotype owing to sample contamination. 77, 256 and 139 SNPs were removed in the Nguni, Angus and Holstein breeds, respectively, for violating HWE (P <0.001).

The average call rate across four breeds was 98%. Over 95% of the SNPs had a call rate of greater than 95% in this study. This was comparable with an average call rate of 99.7% reported by the manufacturer (Illumina Inc., San Diego, CA) and also with the average call rate of greater than 97.9% reported by Matukumalli *et al*. (2009) across 21 different cattle breeds. Average call rate for individual breeds ranged from 96.9% (Holstein) to 99.7% (Angus). The results indicated that the BovineSNP50 array provides a useful tool for interrogating bovine genotype in numerous bovine breeds. Therefore, this array provides a robust resource for genome-wide association studies, genomic selection applications, investigating genetic relationship and detecting signatures of selection on South African Nguni, Angus, Nguni x Angus cross and Holstein cattle.

The successful application of the BovineSNP50 array depends largely on their degree of polymorphisms in the various cattle breeds (Fan *et al*., 2010). Therefore, SNPs with less than 5% MAF were removed from this study. Across all the populations, 40 555 or 74% of the 54 609 called SNPs were polymorphic, with an average minor allele frequency of 0.23 (Table 3.1). This indicated that the Bovine SNP50 array is informative among South African Nguni, Angus, Nguni x Angus and Holstein breeds to determine genetic variation underling these breeds.

The average MAF ranged from 0.17 (Nguni) to 0.22 (Holstein) (Table 3.1). This was in agreement with the previous observation that reveals considerable variations in MAF between breeds (Matukumalli *et al*., 2009). Holstein (41 078) and Angus (40 146), which are common breeds in the USA and Europe, had higher proportions of polymorphic SNP compared with the Nguni (35 843), an indigenous breed to South Africa. Matukumalli *et al*. (2009) observed a similar trend in studies with European and African breeds with 42 849 and 41 073 polymorphic SNPs in Holstein and Angus cattle, respectively, and only 28 869 and 35 084 SNPs in the African N'Dama and Sheko breeds. It was encouraging to confirm that the BovineSNP50 array will be equally informative for use in South

African Sanga compared with the other African breeds that were included during the validation of the BovineSNP50 array.

Table 3.1 Number of polymorphic loci in South African breeds

Populations	Samples	Polymorphic loci*	Mean MAF***	Median MAF***
Holstein	40	41 078	0.22	0.21
Angus	12	40 146	0.21	0.20
Nguni x Angus	10	38 979	0.19	0.20
Nguni	29	35 843	0.17	0.13
All breeds combined	91	40 555	0.23	0.23

*Minor allele frequency >0.05

***Across all 54 609 loci

It can be concluded from this study that the BovineSNP50 array will be applicable to the South African cattle populations, provided that the quality of DNA used meets the required quality for infinium assay. It was observed that sample contamination reduces SNP call rate for individual animals which, in this proof-of-concept study, then reduced the average call rate for Holstein. Overall, the results of this study demonstrate that the BovineSNP50 array will be useful for genomic studies across three breeds that are widely used by South African farmers for dairy and beef production.

Acknowledgments

The authors would like to thank the ARC-Biotechnology Platform for making their laboratory resources available for genotyping samples. Financial support from the ARC is greatly appreciated. Provision of semen on Holstein bulls by Taurus Co-operative is also acknowledged.

References

Bolormaa, S., Hayes, B.J., Savin, K., Hawken, R., Barendse, W., Arthur, P.F., Herd, M.R. & Goddard, M.E. (2011). Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science* 89, 1684-1697.

Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilkki, J., Seabury, C.M., Caetano, A.R., Johnson, G.S., Brenneman, R.A., Hanotte, O., Eggert, L.S., Wiener, P., Kim, J.J., Kim, K.S., Sonstegard, T.S., Tassell, C.P., Neibergs, H.L., McEwan, J.C., Brauning, R.,

Coutinho, L.L., Babar, M.E., Wilson, G.A., McClure, M.C., Rolf, M.M., Kim, J., Schnabel, R.D. & Taylor, J.F. (2009). Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proceedings of the National Academy of Sciences of the United States of America 106, 18644-18649.

Dekkers, J.C.M. (2007). Prediction of response from marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* 124, 331-341.

Fan, B., Zhi-Qiang, D., Danielle, M., Gorbach, M. & Rothschild, F. (2010). Development and application of high-density SNP arrays in genomic studies of domestic animals. *Asian-Australasian Journal of Animal Sciences* 23, 833-847.

Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K. & Goddard, M.E. (2009). Accuracy of Genomic breeding values in multi-breed dairy cattle population. *Genetics Selection Evolution* 41, 51-56.

Hou, Y., Liu, G.E., Bickhart, D.M., Cardone, M.F., Wang, K., Kim, E.S., Matukumalli, L.K., Ventura, M., Song, J. & VanRaden, P.M. (2011). Genomic characteristics of cattle copy number variations. *BMC Genomics* 12, 127-135.

Lee, Y.M., Han, C.M., Li, Y., Lee, J.J., Kim, L.H., Kim, J.H., Kim, D.I., Lee, S.S., Park, B.L., Shin, H.D., Kim, K.S., Kim, N.S. & Kim, J.J. (2010). A whole genome association study to detect single nucleotide polymorphism for carcass traits in Hanwoo populations. *Asian-Australasian Journal of Animal Sciences* 23, 417-424.

MacNeil, M.D., Nkrumah, J.D., Woodward, B.W. & Northcutt, S.L. (2010). Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. *Journal of Animal Science* 88, 517-522.

Mai, M.D., Sahana, G., Chistiansen, F.B. & Guldbrandtsen, B. (2010). A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *Journal of Animal Science* 88, 352-3528.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S. & Van Tassell, C.P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE 4*, e5350-5063.

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

Qanbari S., Pimentel, E.C.G, Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. & Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet. 41, 346-356. Schaeffer, L.R. (2006). Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and *Genetics* 12, 218-223.

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D. & Taylor, J.F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14, 382.

Sherman, E.L., Nkrumah, J. D., Bartusiak, C.L.R., Murdoch, B. & Moore, S.S. (2009). Fine mapping quantitative trait loci (QTL) for feed intake and efficiency in beef cattle. *Journal of Animal Science* 87, 37-45.

Van Raden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92, 16-24.

Chapter 4

Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel

S. O. Makina[1, 3*], F.C. Muchadeyi [2], E. van Marle-Köster[3], M. D. MacNeil[1, 4, 5] & A. Maiwashe[1, 4]

[1]Agricultural Research Council-Animal Production Institute, Private Bag X 2, Irene 0062, South Africa

[2]Agricultural Research Council-Biotechnology Platform, Private Bag X 5, Onderstepoort, 0110, South Africa

[3]Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

[4]Dept. Animal, Wildlife and Grassland Sciences, University of Free State, Bloemfontein 9300, South Africa

[5]Delta G, Miles City, Montana 59301 USA

Published in Frontiers In Genetics: published: 22 September 2014 doi: 10.3389/fgene.2014.00333

Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel

S.O. Makina[1, 3*], F.C. Muchadeyi [2], E. van Marle-Köster[3], M. D. MacNeil[1, 4, 5] & A. Maiwashe[1, 4]

[1]Agricultural Research Council-Animal Production Institute, Private Bag X 2, Irene 0062, South Africa

[2]Agricultural Research Council-Biotechnology Platform, Private Bag X 5, Onderstepoort, 0110, South Africa

[3]Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

[4]Dept. Animal, Wildlife and Grassland Sciences, University of Free State, Bloemfontein 9300, South Africa

[5]Delta G, Miles City, Montana 59301 USA

Abstract

Information about genetic diversity and population structure among cattle breeds is essential for genetic improvement, understanding of environmental adaptation as well as utilization and conservation of cattle breeds. This study investigated genetic diversity and the population structure among six cattle breeds in South Africa (SA) including Afrikaner (n=44), Nguni (n=54), Drakensberger (n=47), Bonsmara (n=44), Angus (n=31) and Holstein (n=29). Genetic diversity within cattle breeds was analyzed using three measures of genetic diversity namely allelic richness (A_R), expected heterozygosity (H_e) and inbreeding coefficient (*f*). Genetic distances between breed pairs were evaluated using Nei's genetic distance. Population structure was assessed using model-based clustering (ADMIXTURE). Results of this study revealed that the allelic richness ranged from 1.88 (Afrikaner) to 1.73 (Nguni). Afrikaner cattle had the lowest level of genetic diversity (H_e=0.24) and the Drakensberger cattle (H_e=0.30) had the highest level of genetic variation among indigenous and locally-developed cattle breeds. The level of inbreeding was relatively low across the studied cattle breeds. As expected the average genetic distance was the greatest between indigenous cattle breeds and *Bos taurus* cattle breeds but the lowest among indigenous and locally-developed breeds. Model-based clustering revealed some level of admixture among indigenous and locally-developed breeds and supported the clustering of the breeds according to their history of origin. The results of this study provided useful insight regarding genetic structure of South African cattle breeds.

Key words: South Africa, cattle breeds, genetic resources, genetic diversity, population structure

[#]Corresponding author: qwabes@arc.agric.za

Background

African cattle breeds can be divided into two major categories, namely Taurine cattle (*Bos taurus*) and Indicine cattle (*Bos indicus*). *Bos indicus* is subdivided into zebu proper and zebu crossbred-types and is phenotypically identifiable by the presence of a substantial cerciothoracic hump (Rege, 1999). The position of the hump on the animal's back is used to classify the zebu proper and zebu crossbred types into cervico thoracic-humped and thoracic-humped stocks (Epstein 1971). Cervico-thoracic-humped cattle occur in or are derived from contact areas of thoracic-humped Zebu and humpless cattle. In crossbreds of humped and thoracic-humped Zebu cattle, the hump is usually cervico-thoracic and these cattle are referred to as Sanga. However the Sanga is nowadays considered a separate group of cattle. Thus African cattle can be classified into four distinctive groups namely *B. taurus, B. indicus, Sanga* and *Sanga zebu types* (Rege, 1999). Afrikaner and Nguni cattle are classified under the Sanga group and indigenous to South Africa. Drakensberger and Bonsmara cattle are also classified under Sanga types, however, the origin of the Drakensberger cattle is unclear with a history dating back to the early settlers in the late 1700's (Scholtz *et al.,* 2010). The Bonsmara cattle was developed at Mara and Messina Research Station from 1937 to 1963 using Milk Short Horn, Hereford and Afrikaner cattle with the aim to produce a locally adapted beef breed (Bonsma, 1980). Angus and Holstein belong to *Bos taurus* group and these originate from Britain and Europe respectively.

The Afrikaner is one of the oldest breeds with a medium–frame, yellow to red coloured with lateral horns with a typical twist. It has exceptional good quality meat and is the ideal minimum care and maximum profit breed (Strydom *et al*., 2000). Nguni cattle are characterized by their multi-coloured coats, which can present many different patterns (white, brown, golden yellow, black, dappled, or spotty), but their noses are always black-tipped and they present a variety of horn shapes. This small framed breed has been kept in rural areas for centuries and often used as dam lines in crossbreeding systems (Scholtz *et al.,* 2011) Drakensberger is a medium to large frame breed and has a black smooth coat. A study by Strydom (2008) has shown that the Drakensberger compare well to British and Europe breeds with regard to meat quality. Bonsmara is medium to large framed, smooth coated with heat and tick tolerance and currently the breed with the largest number of registered females in South Africa (Muchenje *et al*., 2008).

Bos indicus are known to be adapted to the sub-tropical areas in Africa and have a higher tolerance to various diseases (Muchenje *et al*., 2008; Marufu *et al*., 2011). These breeds are also suited to low input systems with lower maintenance and management requirements. In a changing South African environment breeds such as the Afrikaner, Nguni, Drakensberger and Bonsmara holds potential. Despite their large numbers and not endangered, breeds genetic diversity information is essential for

control of inbreeding and effective utilization of breed specific characteristics. The adaptive traits are of importance and there is worldwide a drive for effective management of indigenous genetic resources as they could be most valuable in selection and breeding programs in times of biological stress such as famine, drought or disease epidemics (FAO, 2010). In order to effectively manage these cattle breeds comprehensive knowledge of their characteristics is required. These include population size and structure as well as knowledge of within and between breeds' divergence (Groeneveld *et al.*, 2010; Boettcher *et al*., 2010). In South Africa a number of studies have focused on the characterization of small stock such as goats: Visser *et al*. (2004) and sheep: Soma *et al.* (2012); Qwabe *et al*. (2012). Limited studies have focused on the genetic characterization of South African cattle breeds and this thus emphasised the need for a genetic characterization of these breeds as genetic resources.

Worldwide genetic markers have been used to assess the genetic variation among many cattle breeds relative to their area of origin (Blott *et al*., 1998; Hanotte *et al*., 2002; Gautier *et al*., 2007; Edea *et al*., 2013). Results have shown that genetic diversity of breeds is directly linked to their areas of origin, indicating that breeds which have diverged more recently were generally closer together geographically. These studies have also demonstrated larger differences between taurine and indicine breeds due to a greater time since their divergence (McKay *et al*., 2008; Edea *et al*., 2013). In addition, significant differences were reported between beef and dairy cattle compared to within beef or dairy; this was attributed to different selection pressure across these contemporary groups (Hayes *et al*., 2003).

This study therefore investigated genetic diversity and population structure within and between six cattle breeds in South Africa including Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein using genome wide single nucleotide polymorphism (SNP) generated from the Illumina Bovine SNP50BeadChip.

Materials and methods

Animal resources

A total of 249 animals including three indigenous breeds (Afrikaner=44, Nguni=54, Drakensberger=47), one composite (locally-developed) (Bonsmara=44), and two *Bos taurus* (Angus=31 and Holstein=29) cattle breeds were included in this study. Breeders and Research Stations which keep pure breeds of the populations included in this study were identified and requested to provide animals for blood sampling. All animal handling and sample collection were done according to the regulations of the Animal Ethics Committee of the University of Pretoria (E087-12). To maximize

the genetic diversity within each sampled population, pedigree data were used to select against full and half sib animals. Figure 4.1 show the map of South Africa indicating the location of farms and research station where populations under study were sampled. The sampling of these animals included collection of 10ml whole blood using EDTA VACUETTE® tubes. Holstein (48) semen samples were obtained with permission from an artificial insemination company (Taurus, South Africa). However, to maximize the genetic diversity within Holstein samples, identity by descent analysis was performed using data generated from the Bovine SNP50 BeadChip to select the least related bulls. In which a total of 29 least related bulls were selected for the purpose of this study.

Genotyping and quality control

Genomic DNA was extracted at the ARC-Biotechnology Platform from whole blood and semen samples using the Qiagen DNeasy extraction kit (Qiagen, South Africa) according to the manufacturer's protocol. The protocol was adapted for the semen samples where Dithiothreitol (DTT) was added with proteinase K in the first step. Genomic DNA for all samples was quantified using a Qubit® 2.0 Fluorometer and the Nanodrop Spectrophotometer (Nanodrop ND-1000). In addition, gel electrophoresis was performed to quantify the DNA.

Genotyping was conducted at the ARC-Biotechnology Platform with the Illumina BovineSNP50 BeadChip v2 which features 54 609 SNP probes distributed across the whole bovine genome with an average spacing of 49.9 kb (Matukumalli *et al.*, 2009). Approximately 12 µL of DNA loaded in each well of a BeadChip of genomic DNA was used to genotype each sample. Samples were processed according to the Illumina Infinium–II assay protocol (Illumina, Inc. San Diego, CA, 92122, USA). Quality control criteria were performed across six cattle breeds to remove from further analysis any SNPs with less than 95% call rate, SNPs with less than 0.02 MAF and samples with more than 10% missing genotypes (Purcell *et al*., 2007). This left about 46 236 SNPs across the breeds. Furthermore SNPs that were in high LD were pruned using the following parameter; --indep 50 5 2 in plink (Purcell *et al*., 2007); this left about 21 290 SNPs for further analysis. Pruning of SNPs that are in high LD have been shown to counter the effect of ascertainment bias and to generate meaningful comparison between breeds (Kijas *et al*., 2009).

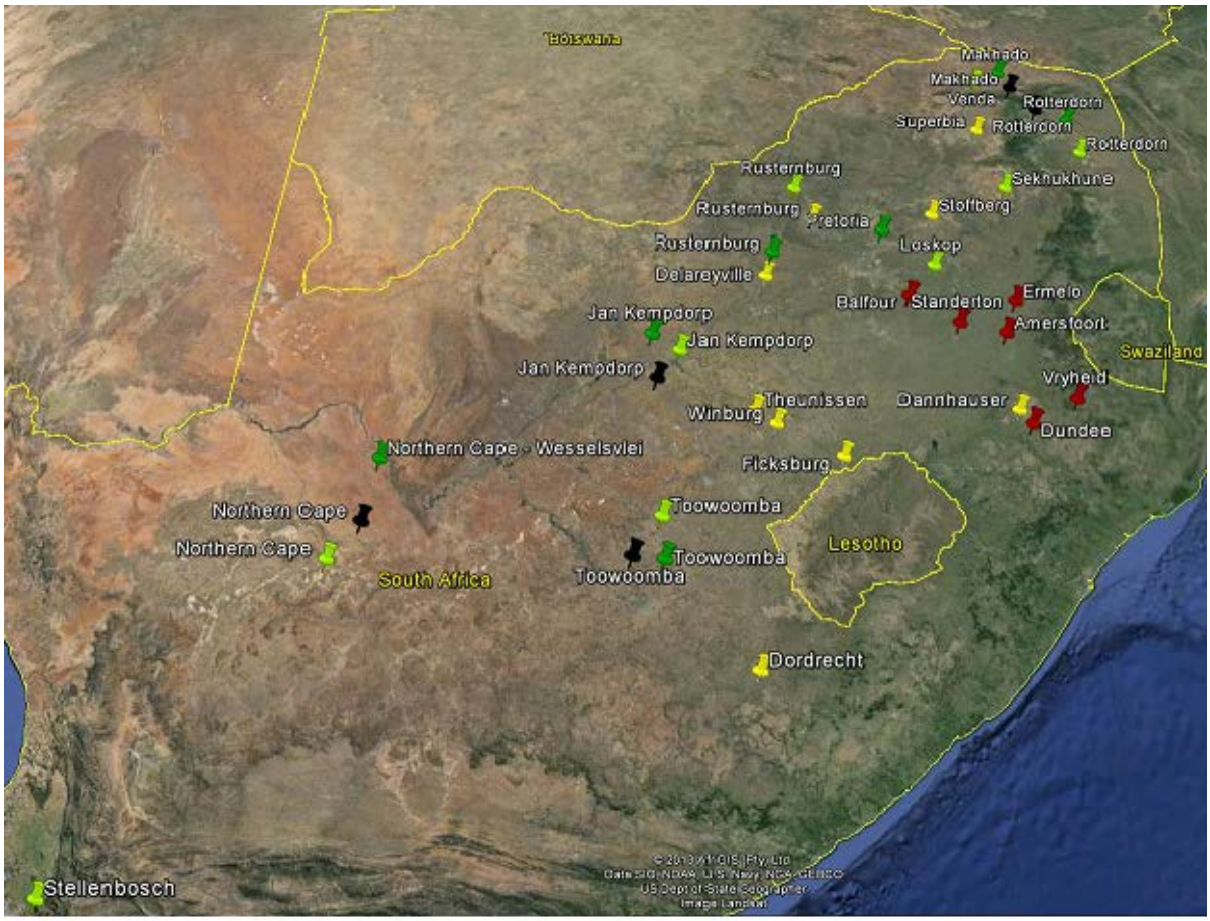


Figure 4.1 Geographic origin of five cattle breeds in South Africa sampled in the current study. Afrikaner (yellow) (44), Nguni (light green) (56), Drakensberger (red) (47), Bonsmara (dark green) and Angus (black) (31)

Estimates of within breed genetic diversity

Three measures of genetic variability were used to compare the levels of heterogeneity within the cattle breeds (allelic richness, expected heterozygosity and inbreeding coefficient). Allelic richness (A_R) was determined within each population using ADZE v 1.07 (Szpiech *et al*., 2008), while expected heterozygosity (H_e) and Inbreeding coefficient (*f*) was calculated using Plink v1.07 (Purcell *et al*., 2007) under the default setting.

Analyses of molecular variance (AMOVA) and population differentiation

Analyses of molecular variance to determine the partitioning of genetic diversity was first performed among indigenous and locally-developed cattle breeds and then amongst all six cattle breeds with the program ARLEQUIN 3.1 version (Excoffier *et al*., 2005).

Populations differentiation was evaluated using pairwise F_{ST} estimates according to Weir and Cockerham (1984) using Golden Helix SNP Variation Suite (SVS) Version 8.1(Golden Helix Inc., Bozeman, Montana, 2012).

Allele sharing and genetic distance

Genetic distance between all pairwise combination of individuals (*D*) was estimated as one minus the average proportion of allele shared (Purcell *et al*., 2007) where the average proportion of allele shared was calculated as Dst using Plink v1.07 (Purcell *et al*., 2007) as:

$$\text{Dst} = \frac{\text{IBS2} + 0.5 * \text{IBS1}}{\text{N}}$$

Where IBS1 and IBS2 are the number of loci which are shared either 1 or 2 alleles identical-by-state (IBS), respectively and N is the number of loci tested.

Pairwise genetic distance among cattle breeds was estimated based on Nei's unbiased genetic (Nei, 1987) distance using Phylip v 3.695 genetic software (Felsenstein, 1989), in which a Neighbor-joining (NJ) relationship tree was then constructed using DrawTree application within Phylip v 3.695 software (Felsenstein, 1989).

Structure analysis

To investigate the population structure of the studied cattle breeds, ADMIXTURE 1.2.3 Software (Alexander *et al*., 2009) was used. In order to infer the true number of genetic populations (clusters or K) between the six cattle breeds. Prior population information was ignored before testing and identifying distinct genetic populations, and assigning individuals to populations. ADMIXTURE uses cross validation (CV) procedure to estimate most preferable *K*. Most preferable *K* exhibit a low cross-

validation error compared to other *K* values. In the current study CV error estimates were plotted (Figure 4.2) for comparison of *K* and K = 6 exhibited low cross validation error values thus K= 6 was taken as the most probable number of inferred populations.

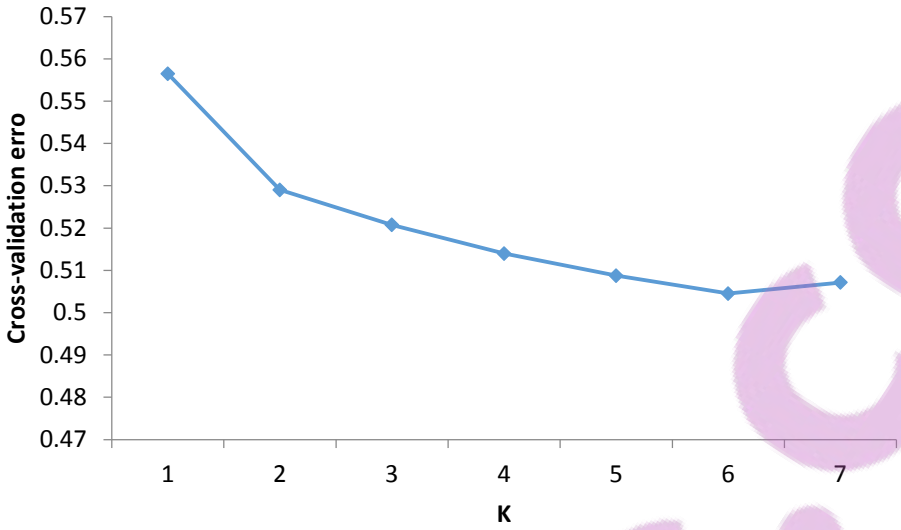


Figure 4.2 Cross validation plot for six cattle breeds in South Africa. Based on cross validation error the plot indicated that k=6 is optimal for data set

Results

SNP polymorphism and within breed genetic diversity

Parameters for SNP validation that included the level of polymorphism, minor allele frequency (MAF) and deviation from Hardy Weinberg equilibrium (HWE) for all six cattle breeds in this study were previously reported (Makina *et al.* – submitted). In summary, examination across breeds revealed that about 56% of SNPs were polymorphic in all breeds and the distribution of MAF showed that nearly half of the SNPs (41%) showed a higher degree of polymorphism (MAF≥0.05) across the breeds. With regard to deviation from HWE only between 5 and 6% of SNP were shown to deviate from HWE (P≤0.05) across the six breeds.

Table 4.1 presents three measures of within breed diversity across the breeds: Afrikaner cattle had the highest number of alleles per locus (A_R = 1.88) while the Nguni cattle had the lowest number of alleles per locus (A_R = 1.73). However, the Afrikaner cattle was observed to have the lowest level of expected heterozygosity (H_e=0.24) in this study. Among indigenous and locally-developed breeds the Drakensberg cattle (H_e=0.30) had the highest level of genetic diversity. Looking across all six breeds

Angus and Holstein cattle had the highest level of gene diversity (H_e=0.31). The level of inbreeding was low across the breeds in this study ranging from 0.005 (Nguni) to - 0.002 (Drakensberger).

Table 4.1 Sample size and genetic diversity within six cattle breeds in South Africa

Breed	**Code**	**n**	**A_R (SD)**	**H_e (SD)**	**F_{is}**
Afrikaner	AFR	42	1.88 (0.12)	0.24 (0.18)	0.004
Nguni	NGU	54	1.73 (0.11)	0.28 (0.17)	0.005
Drakensberger	DRA	47	1.85 (0.12)	0.30 (0.17)	-0.002
Bonsmara	BON	44	1.84 (0.11)	0.29 (016)	-0.017
Angus	ANG	31	1.80 (0.13)	0.31 (0.16)	-0.012
Holstein	HOL	29	1.81 (0.13)	0.31 (0.18)	-0.026

Analyses of molecular variance and population differentiation

Analysis of Molecular Variance illustrated that within breed genetic variation accounted for 91% among indigenous and locally-developed breeds. On the other hand when indigenous and locally-developed breeds were grouped together with *Bos taurus* cattle 92% of genetic diversity occurred within breeds while only 8% occurred between the breeds (Table 4.2).

Table 4.2 Analysis of Molecular Variance among six cattle breeds in South Africa

Data set	**Variance component (%)**		
	Among groups	**Among populations within group**	**Within populations**
All six cattle breeds	7.80	0.70	91.45
Indigenous and local developed breeds	7.80	1.40	90.80

Populations differentiation estimates showed that F_{ST} varied from 0.043 (Bonsmara-Drakensberger) to 0.081 (Afrikaner-Drakensberger) among indigenous and locally-developed breeds and from 0.078 (Drakensberger-Angus) to 0.159 (Afrikaner-Holstein) across all six breeds (Table 4.3).

Table 4.3 Wright fixation index (F_{ST}) pair-wise among six cattle breeds in South Africa

	Afrikaner	Nguni	Drakensberger	Bonsmara	Angus	Holstein
Afrikaner	***					
Nguni	0.064	***				
Drakensberger	0.080	0.044	***			
Bonsmara	0.071	0.044	0.043	***		
Angus	0.151	0.108	0.078	0.083	***	
Holstein	0.159	0.114	0.084	0.099	0.098	***

Genetic distance within and between cattle breeds

The average genetic distance between individuals drawn from the same breeds was 0.20±0.01 within the Afrikaner cattle, 0.23±0.01 within the Nguni, 0.25±0.01 with the Drakensberger, 0.24±0.01 within the Bonsmara, 0.25±0.02 within the Angus and Holstein 0.25±0.01. The average genetic distance between individuals drawn from different breeds ranged from 0.23±0.005 (Afrikaner-Nguni) to 0.29±0.004 (Angus and Holstein).

Topological relationships between breeds, from Neighbour-Joining tree clearly separated *Bos taurus* breeds (Angus and Holstein) from indigenous and locally-developed cattle breeds (Afrikaner, Nguni, Drakensberger and Bonsmara) (Figure 4.3). Three main groups were separated: the group formed by Nguni, Drakensberger and Bonsmara, the group formed by Afrikaner cattle and the group formed by the *Bos taurus* breeds (Angus and Holstein).

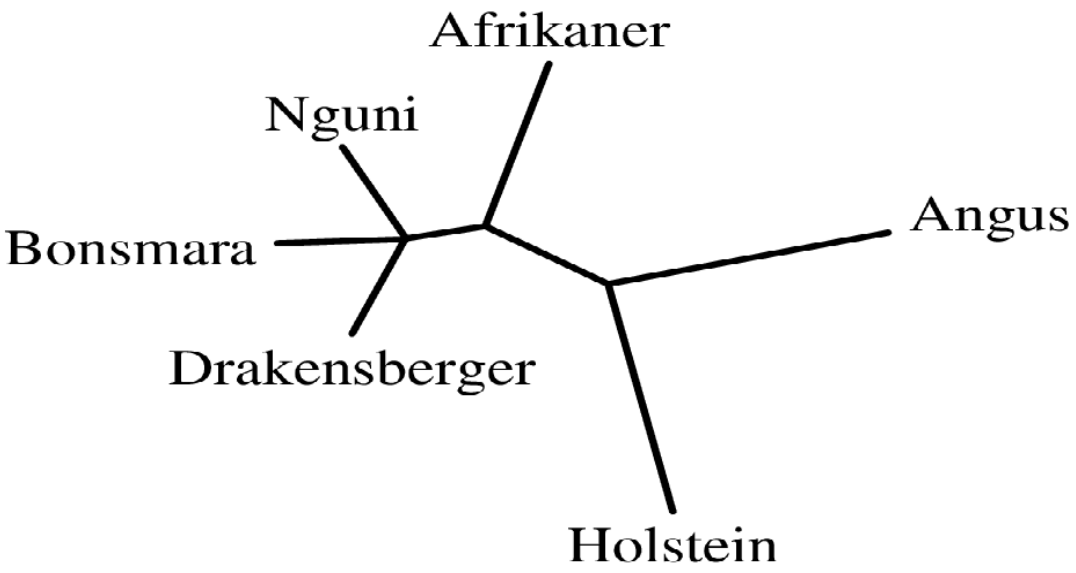


Figure 4.3 Genetic distances between six cattle breeds in South Africa: Neighbor-joining relationship tree of tested cattle breeds

Population structure analysis between six cattle breeds in South Africa

The proportions of individuals in each of the breeds in the six most likely clusters inferred by the ADMIXTURE are presented in Table 4.4 and this corresponded to the six different breeds included in the study. This revealed that 94 % of Afrikaner breed were assigned to cluster one, 84 % of Nguni were assigned to cluster two with 8 % of its genome assigned to cluster one, 81 % of Drakensberger were assigned to cluster three with 5 % of its genome assigned to clusters two, four and five, 89 % of Bonsmara were assigned to cluster four with 3 % of its genome assigned to cluster two, 93 % of Angus were assigned to cluster five and 97 % of Holstein were assigned to cluster six. The results presented in Figure 4.4 (k = 6) demonstrated that among the SA indigenous and locally-developed breeds (Afrikaner, Nguni, Drakensberger and Bonsmara), the Afrikaner population had the least level of admixture while the Drakensberger had the highest level of admixture. The Nguni cattle showed some signals of admixture with Afrikaner breed while the Drakensberger cattle revealed some signals of admixture with Nguni, Bonsmara and Angus. Bonsmara cattle shared more genetic links with the

Nguni cattle than with other indigenous breeds. When comparing all six breeds Afrikaner, Angus and Holstein populations showed the lowest level of admixture in the current study.

Table 4.4 Proportion of membership of the analysed South African cattle breeds in each of the six clusters inferred in the ADMIXTURE program

	Inferred clusters						
Predefined populations	1	2	3	4	5	6	n
Afrikaner	**0.938**	0.036	0.011	0.006	0.005	0.011	42
Nguni	0.083	**0.838**	0.032	0.032	0.007	0.009	54
Drakensberger	0.032	0.048	**0.806**	0.040	0.045	0.028	47
Bonsmara	0.005	0.030	0.013	**0.887**	0.017	0.006	44
Angus	0.003	0.012	0.005	0.034	**0.932**	0.015	31
Holstein	0.000	0.000	0.008	0.005	0.018	**0.969**	29

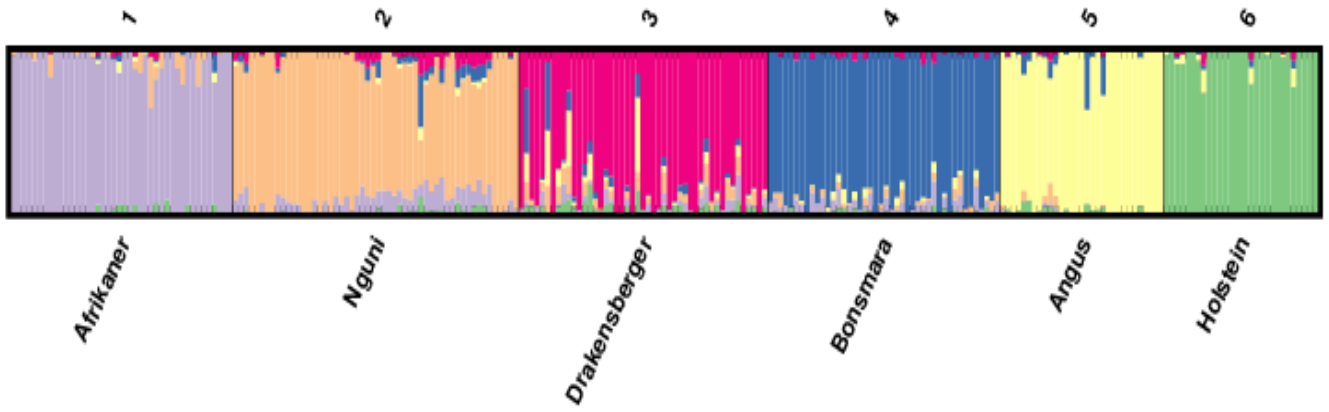


Figure 4.4 ADMIXTURE clustering of six cattle breeds in South Africa.

Discussion

Information about genetic diversity and population structure among cattle breeds is essential for genetic improvement, understanding of environmental adaptation as well as utilization and conservation of cattle breeds (Groeneveld *et al.,* 2010). This study investigated the genetic diversity and population structure among six cattle breeds in South Africa. Among indigenous and locally-developed breeds; Drakensberger cattle demonstrated the highest level of genetic variability (H_e=0.30) while the Afrikaner demonstrated the lowest level of genetic diversity. The lower level of genetic variability observed within the Afrikaner cattle could be due to the presence of strong selection and use of elite sires which is common among stud and commercial herds and small effective population size. This lower level should be noted in Afrikaner and step toward increasing diversity should be prioritized. This could include the exchange of bulls from the different genetic pools. The negative correlation observed between allelic richness and expected heterozygosity in the Afrikaner cattle could be attributed to the processes that differential affect these two measures of diversity, such as bottleneck, selection and increased gene flow between populations within the Afrikaner (Compos *et al*., 2001).

Angus and Holstein cattle (H_e=0.31) demonstrated the highest level of genetic variability compared to all other breeds. The highest genetic diversity observed in *Bos taurus* breeds were in agreement with the results of Lin *et al*. (2010) who reported higher genetic variability within *Bos taurus* compared to *Bos indicus* and also to Edea *et al*. (2013) who reported more genetic diversity in Hanwoo (H_e=0.41) breed than in Ethiopian cattle breeds (between H_e=0.37 to 0.38) based on SNP data. Heterozygosity values observed in this study were comparable to the previously reported heterozygosity among African (H_e=0.25) and European (H_e=0.30) cattle breeds using SNPs (Gautier *et al*., 2007). The levels of inbreeding observed in this study were lower across the breeds. However, it should be noted that this may not indicate the real status of inbreeding within these cattle breeds as allele frequencies may be poor estimate of inbreeding. Assessment of the inbreeding level should be done every five years to determine any unfavourable change in inbreeding levels, so that appropriate steps could be taken to prevent increases in inbreeding.

Analysis of molecular variance among indigenous and locally-developed breeds revealed that about 90 % of the genetic variation occurred within the populations. This was lower than the within-population genetic variation (99 %) observed among Ethiopia populations by Edea *et al*. (2013). Combining all six breeds showed that 92 % of total variation was within populations. This was higher than 81 % observed among Ethiopia and Hanwoo cattle populations.

As expected genetic differentiation (F_{ST}) among the indigenous and locally-developed breeds was lower than African-*Bos taurus* pairs, ranging from 4 to 8 %. This was lower than 12 % observed among West African cattle breeds by Gautier *et al*., (2007), but higher than 1% reported among Ethiopian cattle breeds (Edea *et al*., 2013). Among indigenous and locally-developed and *Bos taurus* cattle breeds genetic differentiation ranged between 8 and 15 %; this was comparable to 15 % reported between African and European breeds by Gautier *et al*., (2007) and 17 % reported by Edea *et al.* (2013) among Ethiopia and Hanwoo cattle populations.

The average genetic distance between pairs of animals drawn from the same breeds ranged from 0.20 (Afrikaner) to 0.25 (Angus and Holstein). Average genetic distance between pairs of animal (0.21) was previously reported within 19 cattle breeds (Bovine HapMap consortium, 2009). As expected average genetic distance between individuals drawn from different breed was higher than those drawn from within breeds, ranging from 0.23 (Nguni-Afrikaner) to 0.29 (Angus-Holstein).

Phylogenic analyses confirmed the closer relationship among indigenous and locally-developed breeds and clearly separated indigenous and locally-developed breeds from *Bos taurus* breeds; this was in agreement with the great divergence between African and European/British breeds observed by Gautier *et al*., (2007). It will be interesting to expand this breed level analysis in subsequent studies through the inclusion of all SA cattle breeds to better understand genetic relationship among SA cattle breeds.

Population structure analysis revealed some signals of admixture and genetic relationship between Afrikaner, Nguni, Drakensberger and Bonsmara. Nguni cattle shared some genetic links with the Afrikaner cattle, with about 8 % of its genome derived from the Afrikaner cattle. This may reflect co-ancestry regarding the origin of these breeds as both these came from the same migration route into Southern Africa (Scholtz, 2011). On the other hand, the Bonsmara cattle shared some genetic links with the Nguni cattle (3 %) but only limited genetic links with Afrikaner cattle (0.5 %); which was unexpected since the Bonsmara cattle was developed through crossbreeding of Afrikaner cattle with exotic breeds such as Hereford and Milk Shorthorn during the early sixties (Bonsma, 1980). However, it should be noted that when Afrikaner and Nguni cattle were brought to Southern Africa by the Khoi-Khoi people, Afrikaner cattle migrated along the western side of Southern Africa whilst the Nguni cattle migrated along the eastern side of Southern African (Scholtz *et al.*, 2011), and the Bonsmara cattle was developed in the eastern part of South Africa which predominantly consisted of the Nguni cattle. The observed low relationship between Bonsmara and Afrikaner may also be attributed to genetic drift or small sample size. The Drakensberger cattle was the most admixed breed in this study with about 5 % of its genome derived from the Nguni, Bonsmara and Angus and 3 % from Afrikaner

and Holstein; this was in agreement with the history of this breed which is believed to have unclear origin (Scholtz, 2010). Afrikaner cattle was the least admixed breed in this study, this was in agreement with the history of this breed as it was the first indigenous South African breed to form a breed society in 1912, thus this breeds may have been closed within the breeding society where only registered animals are allowed within the society. Limited genetic component was shared between indigenous *Bos taurus* breeds, this indicated distinct genetic resources in South Africa which should be utilized and conserved separately.

In general phylogenetic and population structure analysis revealed distinctiveness among South African (indigenous and locally-developed cattle breeds) and *Bos taurus* cattle breeds which is in agreement with their separate domestication and great time divergence (McKay *et al*., 2008). The presence of some admixture among South African cattle breeds was in accordance with previous results of genetic diversity studies among cattle breeds that are generally close together geographically (McKay *et al*., 2008; Edea *et al*., 2013). This indicated that the genetic diversity of breeds is directly linked to the areas of origin, suggesting that breeds which have diverged more recently have a generally closer relationship than breeds which diverged long time ago (Muadet *et al*., 2002).

Conclusion

This study revealed low to moderate genetic diversity within six cattle breeds in South Africa and showed a closer relationship among indigenous and locally-developed cattle breeds. Clear genetic divergence between South African (indigenous and locally-developed cattle breeds) and *Bos taurus* cattle breeds was observed which suggested distinct genetic resource in South Africa cattle breeds that should be proper utilization and conservation in order to cope with unpredictable future environments. Information generated from this study forms the basis for future management of these cattle breeds.

Conflict of interest

The authors declare that they have no competing interests

Acknowledgements

The authors would like to thank the breeders and research institutions for provision of animals for blood samples. Provision of semen on Holstein bulls by Taurus Co-operative is also acknowledged.

ARC-Biotechnology Platform is acknowledged for availing their laboratory resources for genotyping of samples. Financial support from the ARC is greatly appreciated.

References

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

Boettcher, P.J., Tixier-Boichard, M., Toro, M.A., Simianer, H., Eding, H., Gandini, G., *et al*. (2010). Objectives, criteria, and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics* 41 (1), 64-77.

Bonsma, J.C. (1980). Cross-breeding, breed creation and the genesis of the Bonsmara. In: JC. Bonsma (Ed.), Livestock production - A global approach (pg. 126–36). Tafelberg, Cape Town, Publishers Ltd.

Blott, S.C., Williams, J.L. & Haley, C.S. (1998). Genetic relationship among European cattle breeds. *Animal Genetics* 29 (4), 273-282.

Compos, B, Gömöry, D., Letouzey, J., Thiébaut, B. & Petit, R.J. (2001). Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics* 157(1), 389-97.

Edea, Z., Dadi, H., Kim, S., Dessie, T., Lee, T., Kim, H., *et al*. (2013). Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Frontiers in Genetics*, doi:10.3389/fgene.2013.00035.

Epstein, H. (1971). The Origin of the Domestic Animals of Africa, Vol 1 Dog, Cattle, Buffalo. Revised by I.L. Mason. Africana Publishing Corporation. New York, London, Munich.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin version 3.0: An integrated Software Package for Population Genetics Data Analysis. Computational and Molecular Population Genetic Laboratory (CMPG), Institute of Zoology, University of Berne, Switzerland.

FAO. (2010). Progress report of the FAO on selected activities related to agricultural biodiversity. http://www.fao.org/docrep/meeting/018/k5700e.pdf

Felsenstein, J. (1989). PHYLIP – Phylogeny Inference Package (version 3.6). *Cladistics* 5: 164-166.

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., *et al*. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetic Society of America* 177, 1059-1070.

Golden Helix, Inc. (2012). SNP & Variation Suite Manual, Version 8.1 Available at: http://www.goldenhelix.com

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals – a review. *Animal Genetics*, 41, (1) 6.31.

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, 336–339.

Hayes, B.J., Visscher, P. M., McPartlan, H.C., and Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* 13, 635–643.

Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V.H., Nicholas, F.W., and Raadsma, H. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* 4:e4668.

Lin, B.Z., Sasazaki, S. and Mannen, H. (2010). Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations analyzed by SNP markers. *Journal of Animal Science* 81281–289.

Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E. & Maiwashe, A. Genome wide linkage disequilibrium and effective population size in six cattle breeds in South Africa. Submitted to the *BMC genomic* Journal.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., *et al*. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE*, 4: e5350-63.

Maudet, C., Luikart, G. & Taberlet, P. (2002). Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. *Journal of Animal Science* 80(4), 942-950.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., *et al*. (2008). An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genetics* 9:37. doi: 10.1186/1471-2156-9-37.

Muchenje, V., Dzama, K., Chimonyo, M., Raats J.G. & Strydom, P.E. (2008).Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. *South African Journal of Animal Science* 2, 298-304.

Marufu, M.C., Qokweni, L., Chimonyo, M. & Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. *Ticks and Tick Borne Diseases* 2, 172-177.

Nei, M. (1987). Molecular evolutionary genetics. Columbia University Press, New York.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., *et al*. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Journal of Human Genetics* 81, 559–575.

Qwabe, S.O., van Marle-Köster, E. & Visser, C. (2012). Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. *Tropical Animal Health and Production* 45 (2), 511-6.

Rege, J.E.O. (1999). The state of African cattle genetic resources. I. Classification framework and identification of threatened and extinct breeds. FAO/UNEP Anim. Genomic Research. Info. Bull. 25: 1-25.

Scholtz, M.M. (2010). Beef breeding in South Africa. 2nd Edition, (Pretoria: Agricultural Research Council, South Africa), ISBN-13: 978-1-86849-391-3.

Scholtz, M.M., Bosman, D.J., Erasmus, G.J., & Maiwashe, A. (2010). "Selection as the base of improvement in beef cattle, in Beef Breeding in South Africa, 2nd edn., Ed MM Scholtz (Pretoria: Agricultural Research Council) 2-10.

Scholtz, M.M., Gertenbach, W. & Hallowell, G. (2011). The Nguni breed of cattle: past, present and future. (Bloemfontein: Thee Nguni Cattle Breeders' Society). ISBN: 978-620-51723-2.

Soma, P., Kotze, A., Grobler, J.P. & van Wyk, J.B. (2012). South African sheep breeds: Population genetic structure and conservation implications. *Small Ruminant Research* 103, 5926 pp. 528-532 DOI: 10.1126/science.1167936.

Szpiech, Z.A., Jackobson, N.A. and Rosenberg, N.A. (2008). A refraction approach for counting allele private to combinations of populations. Bioinformatics 24: 249802540.

Strydom, P.E., Naude, R.T. Smith, M.F. Scholtz, M.M. & van Wyk, J.B. (2000). Characterization of indigenous African cattle breeds in relation to carcass characteristics. *South African Journal of Anim. Science* 70 (2), 241-252.

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

The Bovine HapMap Consortium. (2009). Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. Science: Vol. 324 no. 5926 pp. 528-532 DOI:10.1126/science.1167936.

Visser, C., Hefer, C.A., van Marle-Köster, E. & Kotze, A. (2004). Genetic variation of three commercial and three indigenous goat populations in South Africa. *South African Journal of Animal Science* 34 (1), 24-27.

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

Chapter 5

Extent of genome wide linkage disequilibrium and effective population size within six cattle breeds in South Africa

S. O. Makina[1, 2#], F. C. Muchadeyi[2], E. van Marle-Köster[3], J. F Taylor[4], A. Maiwashe[1]

[1]Agricultural Research Council-Animal Production Institute, Private Bag X 2, Irene 0062, South Africa

[2]Agricultural Research Council-Biotechnology Platform, Private Bag X 5, Onderstepoort, 0110, South Africa

[3]Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

[4]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA

Prepared for publication

Extent of genome wide linkage disequilibrium and effective population size within six cattle breeds in South Africa

S. O. Makina[1, 2#], F. C. Muchadeyi[2], E. van Marle-Köster[3], J. F Taylor[4], A. Maiwashe[1]

[1]Agricultural Research Council-Animal Production Institute, Private Bag X 2, Irene 0062, South Africa

[2]Agricultural Research Council-Biotechnology Platform, Private Bag X 5, Onderstepoort, 0110, South Africa

[3]Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

[4]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA

Abstract

The extent of linkage disequilibrium (LD) is important for determining the minimum distance between markers for effective genome coverage for genome wide association studies. It can also provide insight into the evolutionary history of a population. This study evaluated the extent of LD and effective population size (N_e) within six cattle breeds in South Africa including Afrikaner (n=44), Nguni (n=56), Drakensberger (n=47), Bonsmara (n=46), Angus (n=31) and Holstein (n=29). LD was assessed using the squared correlation coefficient between SNP pairs (r^2) based on 30 484 (Afrikaner), 35 479 (Nguni), 40 789 (Drakensberger), 39 215 (Bonsmara), 39 801 (Angus) and 40 734 (Holstein) SNP markers. Effective population size by breed was estimated from the LD data based on the assumed relationship of recombination rates to physical distance of 1Mb = 1cM. Genomic distances between SNPs affected the extent of LD across the breeds. At an average marker distance of 40-60 kb, average r^2 values were 0.23, 0.15, 0.14, 0.16, 0.21 and 0.21 but decreased to 0.15, 0.08, 0.08, 0.10, 0.13 and 0.13 at average intermarker distances of 100-200 kb for Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein, respectively. Useful LD ($r^2 \geq 0.20$) extended up to 40-60 kb in Afrikaner, Angus and Holstein but only up to 10-20 kb in Nguni and Drakensberger and 20-40 kb in the Bonsmara. Results indicated that from 50 000 - 150 000 SNPs would be required for future genome wide association studies in South African cattle breeds. Effective population size for the Nguni (N_e=92), Drakensberger (N_e=83) and Bonsmara (N_e=58) exceed the FAO recommended level of 50 N_e per generation. However new breeding strategies may be required for Afrikaner (N_e=33) cattle to ensure the future fitness of this breed.

Keywords: Linkage disequilibrium, effective population size, SNP markers, South Africa, cattle breeds

[#]Corresponding author: qwabes@arc.agric.za

Background

South Africa (SA) has a rich variety of cattle breeds consisting of *Sanga types* (e.g., Afrikaner and Nguni), *Bos taurus* breeds (e.g., Angus, Hereford and Holstein), those of unclear origins such as Drakensberger and locally developed breeds (e.g., Bonsmara and Brangus). Nguni and Afrikaner cattle are indigenous breeds that have been farmed for many centuries in SA (Scholtz, 2010). Afrikaner cattle were crossbred with exotic breeds such as Hereford and Shorthorn in the development of the Bonsmara breed during the mid-20th century (Bonsma, 1980). Afrikaner, Drakensberger and Bonsmara cattle are used for beef production, while the Nguni is a dual purpose breed farmed for beef as well as for milk particularly in the traditional farming systems. These cattle breeds are farmed within a variety of biomes characterized by periodic drought, seasonal dry period, and nutritional shortages in the natural veld and are also subjected to a variety of external and internal parasites and stock diseases (Scholtz & van der Westhuizen, 2008). These breeds have adapted to extreme production environments. It has been suggested that a broad range of either natural or man-made factors (e.g. geography, environment, culture and directional artificial selection) has shaped the genetic constitution of these cattle and that these breeds were forced by natural selection to retain characteristics that ensure a better reproductive success compared to other animals in the same environment (Bonsma, 1980). The possibility of adapted genotypes in these populations has prompted several studies to investigate the adaptability of these breeds to ticks, parasites, drought and diseases (Muchenje *et al*., 2008a; Muchenje *et al*., 2009; Marufu *et al*., 2011), as well as their ability to produce quality beef (Strydom *et al*., 2001; Muchenje *et al*., 2008b). However, little is still known with regard to the genetic variation underlying these economically important traits.

Identification of genomic regions responsible for genetic variation in economically important traits in cattle breeds may contribute to more effective selection and breeding strategies (Gautier *et al*., 2007). Linkage disequilibrium (LD) which is the non-random association of alleles at different loci has been used to investigate genes underlying genetic variation in different cattle breeds and populations (Hayes *et al*., 2009). The successful application of LD information in different populations requires that there is a significant population-wide disequilibrium between the markers and QTL so that the marker alleles can predict the QTL alleles across the entire population (Hayes *et al*., 2009). Therefore, estimating the extent of LD is essential to association studies as it allows the characterization of the range at which SNPs present on marker panels can detect QTL (Prasad *et al*., 2008; Zhao *et al*., 2005). The extent of LD at a given inter-marker distance can also assist in determining the minimum distance between markers to effectively cover the entire genome to achieve reasonable power for the detection of selection signatures (Sargolzaei *et al*., 2008). In addition to these applications, the strength of LD in the genome is useful for inferring ancestral effective population sizes (Hayes *et al*., 2003; Tenesa *et al*.,

2007). Effective population size is an important population parameter that helps to explain how populations have evolved (Falconer & Mackay, 1996) and it can be used to improve the understanding and modelling of the genetic architecture underlying complex traits (Hayes *et al.*, 2003).

The objectives of this study were to quantify the extent of genome wide LD and estimate effective population sizes within six cattle breeds of South Africa including Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein using data generated from the BovineSNP50 BeadChip. The Holstein and Angus cattle breeds of SA were included in this study as reference groups since these breeds have been extensively characterized in other countries.

Materials and Methods

Animal resources

Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein cattle were included in the study. Breeders and research stations that maintain full-blood animals from these breeds were identified and requested to make available animals for blood sampling. About twenty animals were sampled per breeder or research station. All animal handling and sample collection procedures were conducted according to the regulations of the Animal Ethics Committee of the University of Pretoria (E087-12). To maximize the genetic diversity within each sampled population, pedigree data were used to avoid sampling full and half sib animals. Afrikaner cattle (n=44) were collected from nine commercial breeders located in three Provinces (Free State, Eastern Cape and Northern Cape). Nguni (n=56) animals were sampled from five research stations and five breeders located in six Provinces (Gauteng, Limpopo, North West, Northern Cape, Western Cape and Mpumalanga). Drakensberger (n=47) samples were collected from one research station and eight breeders located in two Provinces (KwaZulu Natal and Mpumalanga). Bonsmara (n=46) were sampled from four research stations and five breeders located in four Provinces (Gauteng, Limpopo, North-West and Northern Cape). Angus (n=31) cattle were sampled from three research stations and three breeders located in three Provinces (Northern Cape, Limpopo and Gauteng). The sampling of these animals included collection of 10 ml of whole blood using EDTA VACUETTE® tubes. Holstein (n=48) semen samples were obtained with permission from an artificial insemination company (Taurus, South Africa). However, to maximize the genetic diversity within the Holstein samples, identity by descent analysis was performed using data generated from the BovineSNP50 BeadChip to select the least related Holstein bulls. A total of 29 of the least related bulls were selected for the purpose of this study.

Genotyping and quality control

Genomic DNA was extracted at the ARC-Biotechnology Platform laboratory from the whole blood and semen samples using the Qiagen DNeasy extraction kit (Qiagen) according to the manufacturer's protocol. The protocol was modified for the semen samples where Dithiothreitol (DTT) was added with proteinase K in the first step. Genomic DNA for all samples was quantified using a Qubit® 2.0 Fluorometer and the Nanodrop Spectrophotometer (Nanodrop ND-1000). In addition, gel electrophoresis was performed to quantify DNA quality.

Genotyping was conducted at the ARC-Biotechnology Platform with the Illumina BovineSNP50 BeadChip v2 which features 54 609 SNP probes distributed across the whole bovine genome with an average spacing of 49.9 kb (Matukumalli *et al.*, 2009). Samples were processed according to the Illumina Infinium–II assay protocol (Illumina, Inc. San Diego, CA 92122 USA). Genotype data generated from the iScan reader were processed using Illumina Genome Studio version 1.9.0 software (Illumina, Inc. San Diego, CA 92122 USA), which uses algorithms to perform primary data analysis, including raw data normalization, clustering and genotype calling. A final custom report was created from the Genome Studio output using PLINK plug-ins (Purcell *et al*., 2007), which generated a ped (Pedigree file) and Map (SNP panel file) file for use in downstream analysis.

Quality control was performed within breed and included removing any SNPs with less than 95% call rate, SNPs with less than 0.05 minor allele frequency (MAF) and SNPs which deviated significantly from Hardy Weinberg Equilibrium ($P < 0.001$) (Purcell *et al*., 2007) (SM1). Samples with more than 10% missing genotypes were also excluded from further analysis. Markers were further filtered to exclude loci assigned to unmapped contigs as well as SNPs on the sex chromosomes (Purcell *et al*., 2007). SNP Variation Suite (SVS) version 8.1 (SVS 8.1; Golden Helix Inc., Bozeman, Montana) (Golden Helix, 2012) was used to determine the chromosomal length (Mb), number of SNPs per chromosome and average gap between SNPs in this study. Details of the physical position for the markers used in this study were obtained from Illumina map file (http://support.illumina.com/array/array_kits/bovinesnp50_v2_dna_analysis_kit/downloads.html)

Minor allele frequency

Minor allele frequency (MAF) for the studied breeds was calculated using PLINK genetic software under default settings (Purcell *et al*., 2007).

Linkage disequilibrium analysis

The extent of LD in this study was measured between syntenic SNP pairs for distances up to 10 000 kb and between all syntenic adjacent SNPs using the squared correlation coefficient between SNP pairs (r^2) (Hill & Roberson, 1968) implemented in the LD module of the SNP Variation Suite (SVS) version 8.1 (SVS 8.1; Golden Helix Inc., Bozeman, Montana) (Golden Helix, 2012) and computed as:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

Where D = *f*(AB) - *f*(A)*f*(B) and *f*(AB), *f*(B), *f*(a), *f*(B) and *f*(b) are estimated frequencies of haplotype AB, and alleles A, a, B and b respectively. The squared correlation coefficient (r^2) was chosen over the *D'* parameter to estimate the extent of LD to allow for comparisons of the extent of LD in this study with previous studies in cattle and other domestic animals and also because *D*' tend to be inflated with small sample sizes or at low allele frequencies (McRae *et al.*, 2002). Furthermore this measure is accepted as an ideal measure of LD in the context of QTL mapping because it can be used to estimate the number of loci required for association studies (Pritchard & Przeworski, 2001).

Effective population size

The relationship between LD and effective population size (N_e) in the presence of mutation was estimated by the following equation (Tenesa *et al*., 2007):

$$r^2 = \frac{1}{4cN_e + 2}$$

where c represents the linkage map distance between two SNPs in Morgans and r^2 represent the LD between two SNPs. Physical map distances were transformed into Morgan genetic map distances using the assumed conversion of 1 Mb = 1cM (Arias *et al*., 2009). Genome wide effective population size was estimated as suggested by Hayes *et al*. (2003) to reduce the variability of estimates of N_t (effective population size at t generations in the past) caused by finite population size. Average r^2 were binned according to SNP intervals ranging between 20 and 10 000 kb and used to estimate historical (2000 - 100 generations ago) effective population size and more recent effective population size (50 to five generations ago). The age of N_e for any distance was calculated as t =1/2c, t being generations ago (Hayes *et al*., 2003), assuming a generation interval of 5 years.

Results

Marker statistics

The BovineSNP50 BeadChip containing 54 609 SNPs (Illumina, Inc. San Diego, CA 92122 USA) was used to investigate the extent of LD and N_e within six cattle breeds in SA. The Afrikaner cattle had the lowest percentage of polymorphic SNPs after data filtering, while the Drakensberger and Holstein cattle had the highest percentage of polymorphic SNPs (Table 1). Qwabe *et al*. (2013) suggested that the lower proportions of polymorphic loci found among South African cattle populations may be due to the fact that the majority of breeds used for the development of the Bovine SNP50K BeadChip were European *Bos taurus* (Holstein and Angus).

Table 1 summarizes SNP statistics by breed after quality control. These SNPs covered approximately 2.49 Gb of the bovine autosomal genome. The distributions of SNPs varied amongst chromosomes and chromosome BTA 1 which is the largest bovine autosome had the largest number of variable SNPs (2040 - 2674) after filtering (Table 1). Chromosomes 25, 27 and 28 had the lowest number of variable SNPs in Afrikaner cattle, BTA 27 and 28 had the lowest number in Nguni and BTA 27 had the lowest number in the Drakensberger, Bonsmara, Angus and Holstein cattle breeds. Chromosomes 25, 27 and 28 are among the smaller chromosomes and thus it was expected that these would have fewer numbers of variable SNPs. The largest average physical distance between SNPs that passed quality control was observed within Afrikaner (81.65 kb) whilst the lowest average inter-marker interval was observed in Drakensberger (62.44 kb) and Holstein (61.08 kb).

Table 5.1: Summary of SNPs distribution per chromosome by breed

		Afrikaner		Nguni		Drakensberger		Bonsmara		Angus		Holstein	
Chromosome	Length (Mb)	Number of SNP	Average spacing (kb)	Number of SNP	Average spacing (kb)	Number of SNP	Average spacing (kb)	Number of SNP	Average spacing (kb)	Number of SNP	Average spacing (kb)	Number of SNP	Average spacing (kb)
1	158.03	2040	77.46	2287	69.1	2646	59.72	2562	61.71	2542	62.19	2674	59.1
2	136.66	1684	81.15	1929	70.83	2196	62.23	2155	63.42	2146	63.68	2118	64.52
3	121.14	1469	82.47	1762	68.75	2033	59.59	1880	64.44	1945	62.28	1987	60.97
4	120.36	1515	79.45	1687	71.35	1960	61.43	1892	63.62	1898	63.46	1932	62.44
5	121.08	1235	98.04	1467	82.53	1698	71.31	1657	73.07	1599	75.72	1738	69.67
6	119.05	1529	77.86	1749	67.98	1973	60.32	1937	61.44	1984	59.99	2043	58.27
7	112.27	1295	86.7	1508	74.53	1773	63.39	1733	64.85	1731	64.92	1753	64.06
8	112.91	1373	82.23	1622	69.61	1879	60.09	1818	62.11	1850	61.03	1873	60.28
9	105.46	1215	86.8	1345	78.41	1602	65.83	1511	69.8	1571	67.13	1596	66.08
10	104.17	1277	81.58	1518	68.63	1723	59.88	1695	61.46	1701	60.66	1693	61.53
11	107.14	1227	87.32	1546	69.3	1766	60.69	1715	62.47	1710	62.65	1771	60.52
12	90.82	922	98.5	1190	76.32	1341	67.82	1283	70.88	1304	69.74	1334	68.09
13	83.84	1004	83.5	1179	71.11	1433	58.52	1336	62.77	1377	60.9	1385	60.39
14	83.15	1100	75.59	1263	65.84	1431	58.11	1401	59.35	1372	60.61	1426	58.31
15	84.22	1034	81.45	1176	71.62	1349	62.43	1274	65.95	1279	65.69	1388	60.84
16	81.25	985	82.49	1156	70.28	1294	62.79	1261	64.36	1276	63.68	1275	63.73
17	74.78	905	82.63	1074	69.73	1254	59.72	1188	63.04	1228	60.96	1278	58.6
18	65.4	807	81.04	917	70.98	1053	61.88	990	65.82	1054	62.05	1068	61.24
19	63.51	819	77.55	939	67.64	1115	56.99	1031	61.63	1097	57.92	1123	56.58
20	71.59	954	75.05	1071	66.85	1252	57.18	1196	59.86	1222	58.44	1237	57.88
21	69.45	858	80.95	982	72.4	1130	62.92	1084	65.59	1086	65.47	1111	63.97
22	61.22	735	83.29	928	65.97	1031	59.38	993	61.65	992	61.71	1005	60.79
23	52.1	704	74	765	68.1	853	61.23	807	64.72	848	61.59	846	61.73
24	62.05	747	83.07	893	69.51	1019	60.94	993	62.54	1001	62.08	1011	61.47
25	42.62	567	75.17	688	62.13	752	56.92	767	55.81	768	55.73	810	52.84
26	51.58	673	76.65	770	66.29	865	59.63	842	60.62	855	59.59	857	59.46
27	45.25	579	78.15	665	68.05	750	60.44	708	64.03	741	61.18	770	58.87
28	46.18	578	79.9	654	70.62	773	59.75	723	63.88	771	59.9	780	59.21
29	50.83	654	77.72	749	67.89	845	60.48	783	64.97	853	59.91	852	59.98
All	2498.13	30484	81.65	35479	70.08	40789	61.09	39215	63.65	39801	62.44	40734	61.08

Minor allele frequency

The average MAF over all chromosomes after filtering for quality control was 0.25 ± 0.13 (Afrikaner), 0.26 ± 0.13 (Nguni), 0.27 ± 0.13 (Drakensberger), 0.26 ± 0.13 (Bonsmara), 0.28 ± 0.12 (Angus) and 0.28 ± 0.13 (Holstein). The MAF spectra for the six studied breeds are presented in Figure 1. Afrikaner and Nguni cattle had the highest percentage of SNPs with MAF in the range 0.05 – 0.1 while Holstein and Angus had the lowest percentage in this range.

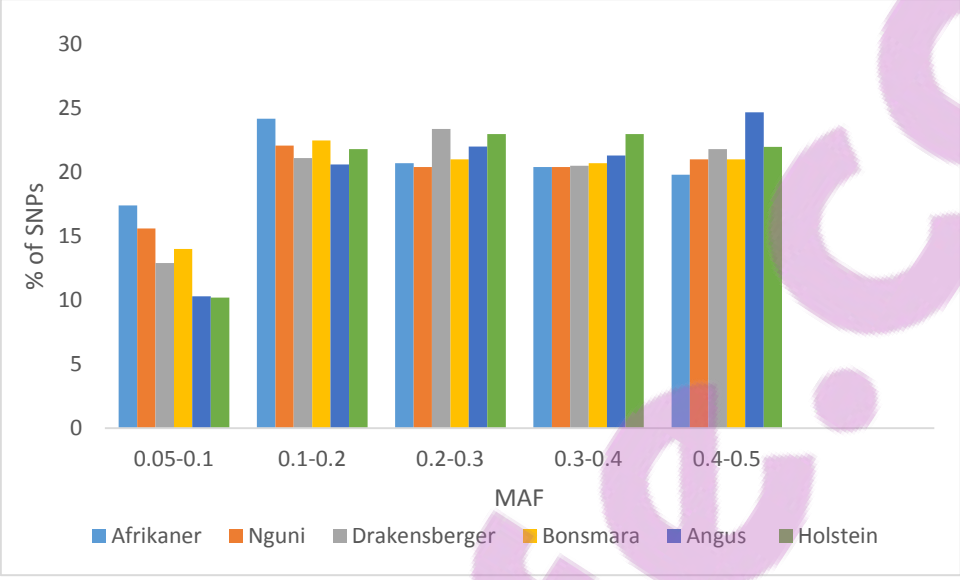


Figure 5.1: The minor allele frequency for SNPs that passed quality control by breed

Linkage disequilibrium and the effects of genetic distance, breeds and chromosomes

Genome-wide average LD was affected by genomic distance between the SNPs. Pairs of SNPs were binned according to inter-marker distances (0 – <10, 10 – <20, 20 – <40, 40 – <60, 60 – <100, 100 – <200, 200 – <500, 500 – <1000, 1000 – <2000, 2000 – <5000, 5000 - <10000 kb) and genome-wide average LD was plotted for each bin for each population. Table 5.2 present genome-wide average r^2 values and the percentage of SNPs that had greater than 0.20 average LD within each distance class, while Figure 5.2 graphically displays the decay of LD with increasing genomic distance for the six breeds. Average LD declined with increasing genomic distances between SNPs for all six breeds. Linkage disequilibrium was high for pairs of SNPs that were close to each other; for markers separated by 40 to 60 kb, the average r^2 was 0.23, 0.15, 0.14, 0.16, 0.21 and 0.21 and the percentage of markers with $r^2 > 0.2$ was 35.38, 22.18, 21.14, 23.33, 33.07 and 32.43 % in Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein, respectively. However, when marker interval increased to between 100 and 200 kb the average r^2 declined to 0.15, 0.08, 0.08, 0.10, 0.13 and 0.13 and the percentage of

SNP pairs with $r^2 > 0.20$ declined to 24.11, 11.03, 11.01, 13.91, 20.65 and 20.93 % for these breeds. Moreover, average r^2 values fell to below 0.10 when locus separation was between 100-200 kb (Nguni and Drakensberger), 200-500 kb (Bonsmara) and 500-1000 kb (Afrikaner, Angus and Holstein).

Table 5.2: Mean r^2 estimated for syntenic SNP pairs for inter-marker distances of up to 10 Mb across the genome and percentage of SNP with $r^2 \geq 0.20$ within the Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein cattle breeds

	Afrikaner			**Nguni**			**Drakensberger**		
Distance (kb)	**Average r^2(SD)**	**n**	**Frequency $r^2 \geq 0.20$ (%)**	**Average r^2 (SD)**	**n**	**Frequency $r^2 \geq 0.20$ (%)**	**Average r^2 (SD)**	**n**	**Frequency $r^2 \geq 0.20$ (%)**
0-10	0.47 (0.41)	141	59.57	0.36 (0.37)	173	48.55	0.36 (0.37)	216	48.15
10-20	0.30 (0.32)	353	44.76	0.20(0.25)	470	31.49	0.22 (0.28)	618	33.50
20-40	0.28 (0.31)	10365	40.90	0.19 (0.25	13728	27.99	0.18 (0.25)	17954	27.14
40-60	0.23 (0.28)	8253	35.38	0.15 (0.21)	10897	22.18	0.14 (0.21)	14403	21.14
60-100	0.19 (0.25)	16485	29.79	0.12 (0.17)	22125	16.17	0.11 (0.17)	28975	16.21
100-200	0.15 (0.21)	40320	24.11	0.08 (0.13)	54006	11.03	0.08 (0.13)	71055	11.01
200-500	0.11 (0.15)	117800	18.04	0.05 (0.08)	158441	5.49	0.06 (0.09)	208781	6.67
500-1000	0.09 (0.13)	192808	12.96	0.04 (0.06)	259135	2.64	0.05 (0.07)	342619	4.62
1000-2000	0.07 (0.10)	376483	9.45	0.03 (0.05)	507655	1.57	0.05 (0.06)	669820	3.33
2000-5000	0.06 (0.08)	690800	6.15	0.03 (0.05)	697629	1.13	0.04 (0.06)	643459	2.51
5000-10000	0.06 (0.08)	34843	6.13	0.03 (0.04)	14140	0.75	0.04 (0.05)	5995	2.07

n- Number of SNP pair

Table 5.2 (ctd)

Distance (kb)	Bonsmara Average r^2 (SD)	Bonsmara n	Bonsmara Frequency $r^2 \geq 0.20$ (%)	Angus Average r^2 (SD)	Angus n	Angus Frequency $r^2 \geq 0.20$ (%)	Holstein Average r^2 (SD)	Holstein n	Holstein Frequency $r^2 \geq 0.20$ (%)
0-10	0.35 (0.37)	200	45.50	0.44 (0.40)	206	54.85	0.43 (0.37)	223	54.71
10-20	0.23 (0.28)	603	34.83	0.32 (0.34)	617	46.35	0.33 (0.28)	608	45.23
20-40	0.20 (0.23)	16723	29.62	0.27 (0.31)	17264	40.11	0.27 (0.26)	18081	39.53
40-60	0.16 (0.26)	13347	23.33	0.21 (0.26)	13847	33.07	0.21 (0.23)	14377	32.43
60-100	0.13 (0.19)	26894	19.22	0.17 (0.22)	27846	27.20	0.17 (0.19)	29010	26.94
100-200	0.10 (0.15)	65992	13.91	0.13 (0.18)	67984	20.65	0.13 (0.15)	70904	20.93
200-500	0.07 (0.11)	193255	9.37	0.10 (0.14)	199428	15.49	0.10 (0.11)	341827	16.02
500-1000	0.06 (0.09)	316690	6.74	0.08 (0.11)	326344	11.73	0.09 (0.09)	666869	13.33
1000-2000	0.05 (0.08)	620518	5.04	0.07 (0.10)	638523	8.90	0.08 (0.08)	208456	11.15
2000-5000	0.05 (0.07)	662500	3.56	0.06 (0.08)	654439	6.63	0.07 (0.07)	645129	9.67
5000-10000	0.04 (0.06)	8475	3.07	0.07 (0.09)	7997	8.27	0.07 (0.06)	5661	7.42

n- Number of SNP pair

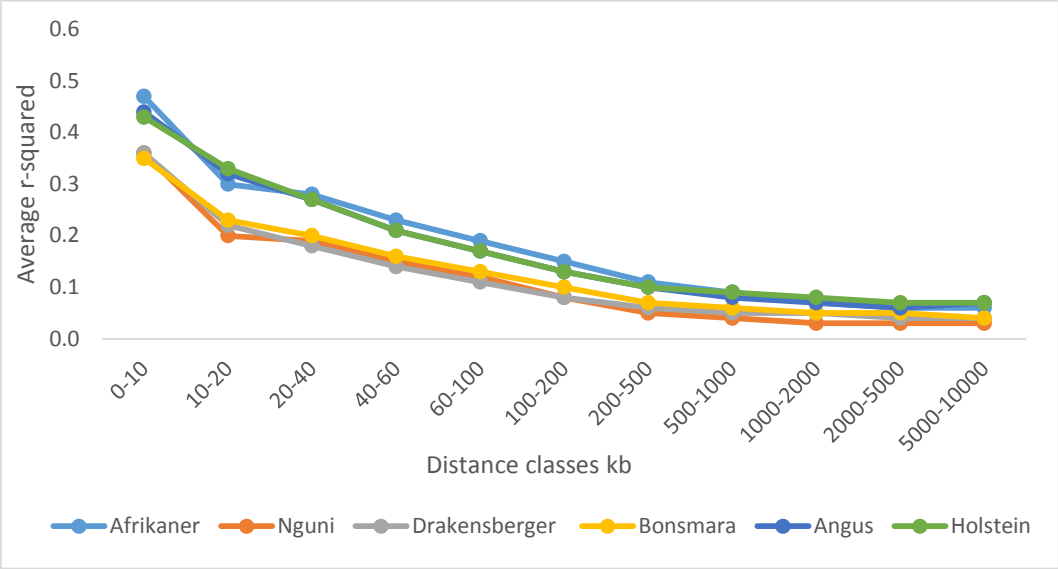


Figure 5.2: Average LD over genomic distance for Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein cattle breeds

Table 3 shows the average LD between adjacent syntenic SNPs across all chromosomes by breed in this study; the overall means for Afrikaner (0.25 ± 0.28), Angus (0.25 ± 0.27) and Holstein (0.25 ± 0.28) were higher than for the Nguni (0.17 ± 0.23), Drakensberger (0.18 ± 0.23) and Bonsmara (0.19 ± 0.24). Average LD was variable among chromosomes within breeds (Table 3). Chromosomes 12 (Afrikaner), 1 (Nguni), 4, 6, 7 & 18 (Drakensberger), 14 (Bonsmara), 7 &14 (Angus) and 14 (Holstein) had the highest average LD compared to the overall breed means, while chromosomes 23 (Afrikaner), 23 (Nguni), 19, 27, 28 (Drakensberger), 26 (Bonsmara), 29 (Angus), and 28 (Holstein) had the lowest average LD compared to overall breed means.

Table 5.3: Average LD between syntenic adjacent markers across all chromosomes by breed

BTA	Afrikaner r^2(SD)	Nguni r^2 (SD)	Drakensberger r^2 (SD)	Bonsmara r^2 (SD)	Angus r^2 (SD)	Holstein r^2 (SD)
1	0.26 (0.29)	**0.20** (0.25)	0.19 (0.24)	0.20 (0.25)	0.27 (0.29)	0.26 (0.30)
2	0.26 (0.29)	0.19 (0.24)	0.18 (0.24)	0.19 (0.25)	0.24 (0.27)	0.28 (0.30)
3	0.28 (0.30)	0.18 (0.24)	0.18 (0.23)	0.20 (0.25)	0.27 (0.29)	0.27 (0.29)
4	0.25 (0.28)	0.19 (0.24)	**0.20** (0.25)	0.20 (0.25)	0.27 (0.29)	0.25 (0.28)
5	0.27 (0.29)	0.17 (0.24)	0.17 (0.23)	0.19 (0.24)	0.27 (0.29)	0.25 (0.28)
6	0.28 (0.30)	0.19 (0.24)	**0.20** (0.25)	0.22 (0.27)	0.27 (0.29)	0.27 (0.30)
7	0.26 (0.29)	0.18 (0.24)	**0.20** (0.25)	0.20 (0.26)	**0.29** (0.30)	0.27 (0.29)
8	0.27 (0.29)	0.19 (0.24)	0.19 (0.24)	0.21 (0.25)	0.26 (0.28)	0.26 (0.29)
9	0.26 (0.29)	0.18 (0.24)	0.19 (0.24)	0.19 (0.25)	0.27 (0.29)	0.26 (0.29)
10	0.26 (0.29)	0.17 (0.22)	0.17 (0.23)	0.18 (0.24)	0.23 (0.27)	0.27 (0.29)
11	0.28 (0.30)	0.18 (0.24)	0.18 (0.24)	0.18 (0.24)	0.27 (0.30)	0.24 (0.28)
12	**0.29** (0.30)	0.16 (0.22)	0.17 (0.22)	0.18 (0.24)	0.24 (0.27)	0.24 (0.28)
13	0.27 (0.30)	0.17 (0.23)	0.16 (0.22)	0.21 (0.25)	0.28 (0.29)	0.26 (0.28)
14	0.25 (0.28)	0.18 (0.24)	0.19 (0.24)	**0.23** (0.26)	**0.29** (0.30)	**0.29** (0.30)
15	0.24 (0.27)	0.17 (0.23)	0.16 (0.21)	0.18 (0.23)	0.24 (0.27)	0.23 (0.27)
16	0.26 (0.28)	0.18 (0.23)	0.19 (0.24)	0.19 (0.25)	0.26 (0.29)	0.27 (0.30)
17	0.27 (0.29)	0.17 (0.23)	0.17 (0.23)	0.19 (0.24)	0.24 (0.27)	0.24 (0.28)
18	0.24 (0.28)	0.16 (0.22)	**0.20** (0.24)	0.18 (0.23)	0.24 (0.27)	0.22 (0.27)

BTA	Afrikaner r^2(SD)	Nguni r^2 (SD)	Drakensberger r^2 (SD)	Bonsmara r^2 (SD)	Angus r^2 (SD)	Holstein r^2 (SD)
19	0.24 (0.27)	0.16 (0.21)	***0.15*** (0.21)	0.18 (0.22)	0.24 (0.26)	0.23 (0.27)
20	0.27 (0.29)	0.18 (0.23)	0.17 (0.22)	0.18 (0.23)	0.23 (0.26)	0.26 (0.28)
21	0.28 (0.30)	0.17 (0.23)	0.17 (0.22)	0.19 (0.24)	0.24 (0.27)	0.26 (0.29)
22	0.25 (0.29)	0.16 (0.22)	0.16 (0.22)	0.18 (0.24)	0.23 (0.27)	0.25 (0.28)
23	***0.19*** (0.25)	***0.14*** (0.20)	0.16 (0.21)	0.17 (0.22)	0.21 (0.24)	0.24 (0.27)
25	0.25 (0.29)	0.17 (0.22)	0.18 (0.24)	0.18 (0.23)	0.24 (0.26)	0.23 (0.26)
26	0.23 (0.27)	0.15 (0.21)	0.17 (0.23)	***0.16*** (0.22)	0.22 (0.26)	0.24 (0.30)
27	0.24 (0.28)	0.16 (0.21)	***0.15*** (0.21)	0.18 (0.23)	0.22 (0.25)	0.21 (0.27)
28	0.22 (0.25)	0.15 (0.20)	***0.15*** (0.19)	0.17 (0.22)	0.23 (0.26)	***0.20*** (0.24)
29	0.22 (0.26)	0.16 (0.20)	0.16 (0.20)	0.18 (0.22)	***0.20*** (0.24)	0.22 (0.26)
Overall mean	**0.25 (0.28)**	**0.17 (0.23)**	**0.18 (0.23)**	**0.19 (0.24)**	**0.25 (0.27)**	**0.25 (0.28)**

Bold- higher than overall mean LD, bold and italic – lower than overall mean LD

Effective population size

In Figure 3, the estimated ancestral effective population size (N_e) (2 500 to 500 generations ago) is presented while Table 4 presents more recent estimates of N_e (100 – five generations ago). Looking at the N_e in the most distant past (between 2 500 and 500 generations ago), effective population sizes were 1 942, 2 970, 2 452, 2 560, 1 135, and 1 525 prior to domestication (2 500 generations ago) but declined to 636, 1 545, 1 539, 1 421, 956, and 938 following domestication (500 generations ago) in Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein, respectively. N_e has declined rapidly in all breeds in recent generations (100 to five generations ago) (Table 4). Five generations ago N_e was 33, 92, 83, 58, 27 and 32 for Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein, respectively. The Nguni, Drakensberger and Bonsmara breeds had higher effective population sizes than did the Afrikaner, Angus and Holstein in all generations.

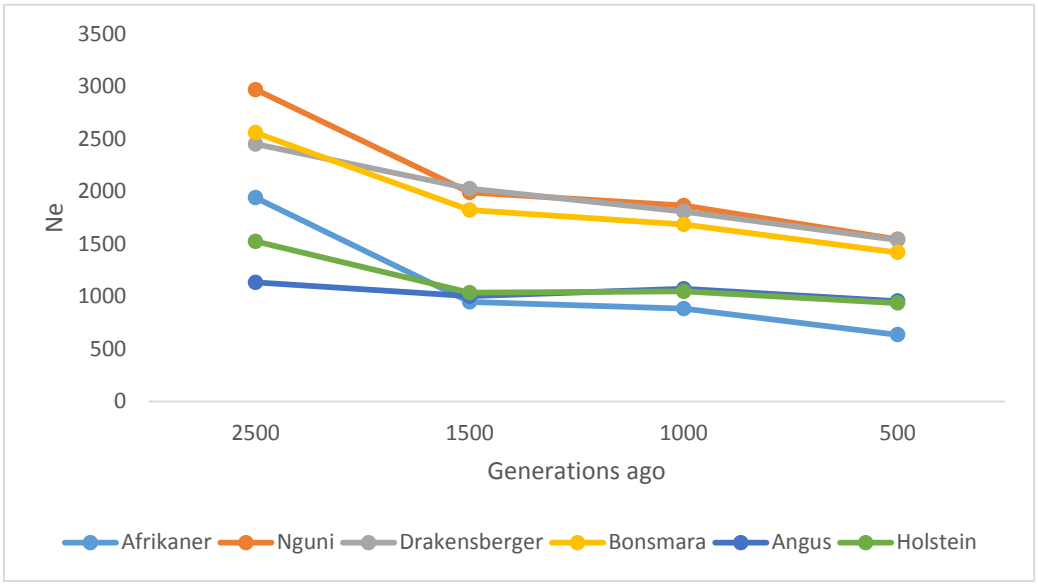


Figure 5.3: Genome wide effective population size (N_e) over the past generations based on linkage disequilibrium

Table 5.4: Effective population sizes (N_e) from 100 to five generations ago in the six cattle breeds

Distance (Mb)	Generations	Afrikaner	Nguni	Drakensberger	Bonsmara	Angus	Holstein
0.5	100	385.99	938.42	768.06	625.68	432.91	408.91
1	50	251.3	626.33	455.82	376.72	266.88	237.14
2	25	163.48	363.47	268.11	228.67	164.47	136.23
2.5	20	145.67	300.34	224	192.47	141.52	113.15
3.5	15	118.28	236.19	173.89	152.26	109.15	84.70
5	10	86.01	169.59	108.89	102.31	65.85	66.83
10	5	32.66	92.06	82.72	57.51	26.71	32.01

Discussion

The r^2 values decreased to less than 0.10 at a locus separation of between 100-200 kb (Nguni and Drakensberger), 200-500 kb (Bonsmara) and 500-1000 kb (Afrikaner, Angus and Holstein). Similar results were obtained in German Holstein (Qanbari *et al*., 2010) and Australian Holstein (Khatkar *et al*., 2008) and Angus (Lu *et al*., 2012) where r^2 values were 0.10 for SNPs with an inter-marker distance of 500 kb. Garcia-Gámez *et al*. (2012) investigated the extent of LD in Spanish Churra sheep and found similar trends in LD decay, where the average r^2 dropped from 0.20 for intermarker distances between 20-40 kb and to 0.06 for distances between 200-500 kb. Uimari & Tapio (2011) and Badke *et al*. (2012) assessed the extent of LD in pig breeds and found that pigs may have considerably higher LD compared to cattle as average r^2 values of 0.10 were observed even for SNPs that were separated by more than 1 Mb. In general, the results of the present study indicated that linkage disequilibrium persist over limited distances in cattle populations, which is consistent with previous reports (Khatkar *et al*., 2008; Bohmanova *et al*., 2010). The extent of LD between markers separated by 40 to 60 kb in Angus (r^2 = 0.21) and Holstein (r^2 = 0.21) was similar to results from previous studies that observed r^2 of 0.20 in Australian Holstein (Khatkar *et al*., 2008), 0.23 for North American Holstein (Bohmanova *et al*., 2010) and 0.23 for Angus (Lu *et al*., 2012).

Average LD varied between breeds. Afrikaner, Angus and Holstein cattle had higher average LD compared to the Nguni, Drakensberger and Bonsmara. Linkage disequilibrium is population specific and has some degree of heterogeneity between populations depending on the demographic histories of the animals under investigation. Variation in the extent of genome wide LD among the studied cattle breeds could therefore indicate that different evolutionary and molecular forces have acted on these cattle breeds (Lee *et al*., 2011). Selection is considered to be an important cause of

LD, however, its effect is likely to be localized around specific genes and thus has moderate effects on the average LD over the whole genome. Small effective population size is generally implicated as the key source of extensive LD in livestock populations (Hayes *et al*., 2003).

The average LD was variable among chromosomes within breeds. Chromosomes 12 (Afrikaner), 1 (Nguni), 4, 6, 7 & 18 (Drakensberger), 14 (Bonsmara), 7 &14 (Angus) and 14 (Holstein) had the highest average LD compared to overall breed means. Variation in LD observed among different chromosomes within breeds may be genetic drift or effects of selection within the studied breeds (Du *et al*., 2007). Selection reduces genetic variation in the next generation and produces LD among syntenic and non-syntenic loci (Bulmer, 1971).

The extent and patterning of LD within breeds in this study was used to assess the number of markers that would be required for genome wide association scans of six cattle breeds in South Africa. Useful LD can be defined as the r^2 between a marker and a QTL and this is the proportion of QTL variance that can be observed at the marker (Hayes *et al*., 2009). The threshold for useful LD was assumed to be $r^2 \geq 0.20$ in this study as established by Hayes *et al*. (2003) who recommended an LD higher than 0.20 for successful application of genomic selection and genome wide association studies. This will ensure a significant population-wide disequilibrium between the markers and the QTL so that the markers can predict the effects of the QTL across the entire population. Findings in this study indicated that for the Afrikaner, Holstein and Angus cattle breeds, SNPs spacing should be approximately between 40-60 kb for genome-wide scans. Assuming that any QTL will be at most in the middle of the interval, and therefore no more than 30 kb away from any marker, a minimum of 50 000 evenly spaced and informative markers would be sufficient to enable genome wide scans in these breeds. This agrees with McKay *et al*. (2007) who suggested that 50 000 SNPs would capture most of the LD information necessary for genome wide association studies in *Bos taurus* cattle populations. However, for the Nguni and Drakensberger cattle breeds, average r^2 of 0.20 and 0.22 only extended to 10-20 kb. This suggests that there should be an informative SNP every 20 kb to achieve the same power as in Afrikaner, Holstein and Angus, indicating that about 150 000 SNPs would be required in these breeds for genome wide association and selection studies, while 75 000 SNPs should be sufficient in Bonsmara. These results agrees with Gautier *et al*. (2007) and Khatkar *et al*. (2008) who suggested that 75 000–300 000 informative SNPs would capture most of the LD information within different cattle breeds.

Linkage disequilibrium structure can provide insight into the evolutionary history of a population (Hill, 1981), therefore in this study the strength of LD at different genetic distances between loci was used to estimate ancestral effective population sizes. We found a decline in N_e throughout time in all breeds. At about 2 500 generations ago Ne declined until 500 generations ago (1 942– 636), (2 970–1 545), (2 452-1 539), (2 560– 1 421), (1 135 – 956) and (1 525-938) in Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein, respectively. This decrease in N_e could be associated with the post-domestication events of human migration with cattle that ultimately led to

breed formation (Gautier *et al*., 2007). The most rapid decline in N_e occurred between 100 and five generations ago (approximately 500 and 25 years ago) in all breeds. This suggests a significant bottleneck occurred at breed formation and that population subdivision resulted in significantly reduced N_e (Daetwyler *et al*., 2010). De Roos *et al*. (2008) estimated effective population size in Australian Holstein-Friesian, Jersey and Angus cattle and found that the N_e for these breeds has decreased over the last 50 generations to approximately 100. Similarly, Villa-Angulo *et al*. (2009) estimated N_e for Angus and found a rapid decline in N_e in Angus over the last 100 generations. In addition, Decker *et al*. (2012) estimated N_e in North American Angus using molecular inbreeding coefficients and found N_e to be 94 over the last 63 generations. Results reported in the literature compare favorably with the results for N_e obtained at different generations for Angus and Holstein in this study (Lu *et al*., 2012; Villa-Angulo *et al*., 2009).

The lower effective population sizes for the Afrikaner, Angus and Holstein breeds compared to those of Nguni, Bonsmara and Drakensberger at more recent generations, could be due to intense selection, inbreeding and probably wide spread use of artificial insemination in South Africa and the use of relatively few elite sires after 1970 (Hayes *et al*., 2009). In order to maximize the net response in genetic gain, Food and Agricultural Organisation (FAO) (FAO, 1998) recommended a N_e of 50 per generation. In the current study, the most recent estimates of N_e (five generations ago) for the Nguni, Drakensberger and Bonsmara exceeded 50. However, care should be taken to maintain N_e in these breeds above the recommended threshold. The N_e for Afrikaner, Angus and Holstein were below the FAO recommended number. This suggested that these breeds are endangered and close to critical stage (FAO, 1998). Therefore pointing out the need for implementation of appropriate conservation programs as well as new selection and breeding strategies to ensure the long term fitness of these breeds. These could include increasing the number of animals contributing offspring to each generation by increasing the active cow populations (FAO, 1998). In addition, selection and the use of bulls through rotational breeding or the use of bulls from a number of unrelated sources. Crossbreeding is a well-known method of genetic improvement, however unsupervised practices of crossbreeding cold threaten the existence of pure breeds (FAO, 1998).

Conclusions

Small sample size may not be representative of the whole population and may only include the predominant animals within the breed. This could lead to erroneous conclusion that the entire population has higher or lower LD, when in fact it is the opposite. However, the results of this study for Holstein and Angus cattle were comparable to previous studies of LD with bigger sample size (>100) (Khatkar *et al*. (2008); Bohmanova *et al*. (2010); Lu *et al*. (2012).

The results of this study revealed significant differences in the extent of LD between South African breeds. Afrikaner cattle had the highest levels of LD compared to the other indigenous breeds. The higher LD suggests that Afrikaner cattle have experienced considerable bottlenecks restricting their effective population size in contrast to other indigenous breeds. This result also implies that this breed would require lower marker density panels relative to those required for the Nguni, Drakensberger and Bonsmara cattle to associate genetic variation with economically important traits. Effective population size for the Nguni, Drakensberger and Bonsmara were above the FAO recommended level. However care should be taken to ensure that these cattle breeds are viable for long term. On the other hand, new breeding strategies should be considered for the Afrikaner cattle to ensure the long-term fitness of this breed.

Acknowledgements

The authors would like to thank the breeders and research institutions for provision of animal for blood samples. Provision of semen on Holstein bulls by Taurus Co-operative is also acknowledged. ARC-Biotechnology Platform is acknowledged for availing their laboratory resources for genotyping of samples. Financial support from the ARC is greatly appreciated.

References

Arias, J.A., Keehan, M., Fisher, P., Coppieters, W. & Spelman, R. (2009). A high density linkage map of the bovine genome. *BMC Genetics*, 10 1471-2156.

Badke, Y.M., Bates, R.O., Ernst, C.W., Schwab, C. & Steibel, J.P. (2012). Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* 13, 24.

Bohmanova, J., Sargolzaei, M. & Schenkel, F.S. (2010). Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics* 11, 421-431.

Bonsma, J.C. (1980). Cross-breeding, breed creation and the genesis of the Bonsmara. In: JC. Bonsma (Ed.), Livestock production. A global approach, 126-136.

Bulmer, M.G. (1971). The effect of selection on genetic variability. *American Naturalist* 105, 201-211.

Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185 (3), 1021-1031.

Decker, J.E., Vasco, D.A., McKay, S.D., McClure, M.C., Rolf, M.M., Kim, J., Northcutt, S.L. Bauck, S., Woodward, B.W., Schnabel, R.D. & Taylor J.F. (2012). A novel analytical method, Birth Date Selection Mapping, detects response of the Angus (*Bos taurus*) genome to selection on complex traits. *BMC Genomics*. 13:606. doi: 10.1186/1471-2164-13-606.

De Roos, A.P., Hayes, B.J., Spelman, R.J. & Goddard, M.E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179(3), 1503-1512.

Du, F.X., Clutter, A.C. & Lohuis, M.M. (2007). Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences* 3, 166-178.

Falconer, D.S. & Mackay, T.F.C. (1996). Introduction to Quantitative Genetics, 4th edn. Pearson Education Limited, Malaysia, 1996.

FAO. (1998). Secondary guidelines for development of national farm animal genetic resources management plans: management of small populations at risk. Rome. (dad.fao.org/en/refer/library/guidelin/sml-popn.pdf).

FAO. (2010). Progress report of the FAO on selected activities related to agricultural biodiversity. http://www.fao.org/docrep/meeting/018/k5700e.pdf.

García-Gámez, E., Sahana, G., Gutiérrez-Gil1, B. & Arranz, J.J. (2012). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetic* 13, 43 http://www.biomecentral.com/1471-2156/13/43.

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

Golden Helix Inc. (2012). SNP & Variation Suite Manual, Version 8.1 Available at: http://www.goldenhelix.com

Hayes, B.J., Visscher, P.M., McPartlan, H.C. & Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* 13, 635–643.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92, 433–443.

Hill, W.G. & Roberson, A. (1968). Linkage disequilibrium in infinite populations. *Theoretical and Applied Genetics* 38, 226–231.

Hill, W.G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetic Research* 38, 209–216.

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

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 Australasian Journal of Animal Sciences* 24 (12), 1660-1665.

Lu, D., Sargolzaei, M., Kelly, M., Li, C., Voort, G., Wang, Z., Plastow, G., Moore, S. & Miller, S. (2012). Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Frontier in Genetics* 3, 152-160.

Marufu, M.C., Qokweni, L., Chimonyo, M. & Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. *Ticks Tick Borne Dis* 2, 172-177.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S. & Van Tassell, C.P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4, e5350-63.

McKay, S.D., Schnabel, D., Murdoch, M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Dias Neto, E., Gill, C.A., Gao, C., Mannen, H., Stothard, P., Wang, Z., Van Tassel, C.P., Williams, J.L., Taylor, J.F. & Moore, S.S. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* 8, 74-85.

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

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G. & Strydom, P.E. (2008a). Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. *South African Journal of Animal Science* 2, 298-304.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G., & Strydom, P.E. (2008b). Meat quality of Nguni, Bonsmara and Aberdeen Angus steers raised on natural pasture in the Eastern Cape, South Africa. *Meat Science* 79, 20-28.

Muchenje, V., Dzama, K., Chimonyo, M., Strydom, P.E. & Raats, J.G. (2009). Relationship between stress responsiveness and meat quality in three cattle breeds. *Meat Science* 81, 653-657.

Prasad, A. Schnabel, R.D., McKay, S.D., Murdoch, B., Stothard, P., Kolbehdari, D., Wang, Z., Taylor, J.F. & Moore, S.S. (2008). Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal Genetics* 39, 597–605.

Pritchard, J.K. & Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* 69, 1-14.

Purcell, S., Neale, B., Todd-Brown, K., Thomas Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. & Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81, 559-75.

Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. & Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics* 41, 346-356.

Qwabe, S.O., van Marle-Köster, E., Maiwashe, E. & Muchadeyi, F.C. (2013). Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. *South African Journal of Animal Science* 1, 43.

Sargolzaei, M., Schenkel, F.S., Jansen, G.B. & Schaeffer, L.R. (2008). Extent of linkage disequilibrium in Holstein cattle in North America. *Journal of Dairy Science* 91, 2106–2117. doi:10.3168/jds.2007-0553.

Scholtz, M.M. & van der Westhuizen, J. (2008). What is adaptation and how can it be measured? Report of the FAO/WAAP Workshop on Production Environment Descriptors for Animal Genetic Resources. Caprarola, Italy, 2 - 8 May 2008.

Scholtz, M.M. (2010). Beef Breeding in South Africa. 2nd Edition, ARC Pretoria, South Africa (ISBN-13: 978-1-86849-391-3).

Strydom, P., Naudé, R.T., Smith, M.F., Kotzé, A., Scholtz, M.M. & van Wyk, J.B. (2001). Relationships between production and product traits in subpopulations of Bonsmara and Nguni cattle. *South African Journal of Animal Science* 31(3), 181-194.

Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E. & Visscher, P.M. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Researc*h 17(4), 520-526.

Uimari, P. & Tapio, M. (2011). Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *Journal of Animal Science* 89, 609–614.

Villa-Angulo, R., Matukumalli, L.K., Gill, C.A., Choi, J., Van Tassell, C.P. & Grefenstette, J.J. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics* 10, 19. Doi:10.1186/1471-2156-10-19.

Zhao, H., Nettleton, D., Soller, M. & Dekkers, J.C. (2005). Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genetic Research* 86, 77–87. doi: 10.1017/S001667230500769X.

Chapter 6

Genome wide scan for selection signatures within and between six cattle breeds in South Africa

S. O. Makina[1, 3*], F. C. Muchadeyi[2], E. van Marle-Köster[3], J. F Taylor[4], M. L Makgahlela[1]& A. Maiwashe[1, 5]

[1]Agricultural Research Council-Animal Production Institute, Private Bag X 2, Irene 0062, South Africa

[2]Agricultural Research Council-Biotechnology Platform, Private Bag X 5, Onderstepoort, 0110, South Africa

[3]Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

[4]Division of Animal Science, University of Missouri, Columbia, MO 65211, USA

[5]Dept. Animal, Wildlife and Grassland Sciences, University of Free State, Bloemfontein 9300, South Africa

Prepared for publication

Genome wide scan for selection signatures within and between six cattle breeds in South Africa

S. O. Makina[1, 3#], F. C. Muchadeyi[2], E. van Marle-Köster[3], J. F Taylor[4], M. L Makgahlela[1]& A. Maiwashe[1, 5]

[1]Agricultural Research Council-Animal Production Institute, Private Bag X 2, Irene 0062, South Africa

[2]Agricultural Research Council-Biotechnology Platform, Private Bag X 5, Onderstepoort, 0110, South Africa

[3]Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

[4]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA

[5]Dept. Animal, Wildlife and Grassland Sciences, University of Free State, Bloemfontein 9300, South Africa

Abstract

Detection of selection signatures among cattle breeds may assist in locating regions of the genome that are, or have been, functionally important and targeted by selection. This study employed two approaches to detect signatures of selection within and between six cattle breeds in South Africa including Afrikaner (n=44), Nguni (n=54), Drakensberger (n=47), Bonsmara (n=44), Angus (n=31) and Holstein (n=29). The first approach was based on the detection of genomic regions for which haplotypes have been driven toward complete fixation within breeds. The second approach identified regions of the genome exhibiting elevated population differentiation (F_{st}). A total of 47 genomic regions were identified as harbouring potential signatures of selection using both methods. Eleven of the identified selected regions were shared among breeds and ten were also detected in previous studies. Thirty three of these regions were successfully annotated to identify candidate genes. Among these, were keratin genes (*KRT222*, *KRT24*, *KRT25*, *KRT26*, *KRT27*) and one heat shock protein (*HSPB9)* on chromosome 19 at 42,896,570 – 42,897,840 bp in the Nguni cattle. These genes have been associated with adaptation to tropical environments in Zebu cattle. Furthermore, a number of genes associated with the nervous system (*WNT5B*, *FMOD*, *PRELP*, *ATP2B*), immune response (*CYM CDC6*, *CDK10*), production (*MTPN*, *IGFBP4*, *TGFB1*, *AJAP1*) and reproductive performance (*ADIPOR2*, *OVOS2*, *RBBP8*) were detected to be under selection in this study. The results presented here provide a useful foundation for detection of mutations underlying genetic variation in traits of economic importance in cattle breeds of South Africa.

Key words: selection sweep, genetic variation, cattle breeds

[#]Corresponding author: qwabes@arc.agric.za

Introduction

Natural selection has shaped the pattern of genetic variation among cattle breeds worldwide over many centuries (The Bovine HapMap Consortium, 2009). Selection acts on a mutation and affects linked sites leaving its signature in the adjacent chromosomal regions. The signals observed on the selected genes may include a shift in allele frequency towards extreme (high or low) frequencies, an excess of homozygous genotypes, long haplotypes with high frequency and extreme population differentiation (Simianer & Qanbari, 2014). The availability of large scale SNP data has made it possible to scan the cattle genome for positions that may have been targeted by selection (Nielsen, 2005). The detection of selection signatures is of interest as it may assist in obtaining a comprehensive genomic understanding of how and where natural and artificial selection has shaped the patterns of variation in the genome. Such information provides important insights with regard to the mechanisms of evolution (Otto, 2000), selection of loci by breeding and selection programs (Vitalis *et al*., 2001) and is useful for the annotation of significant functional genomic regions (Nielsen *et al*., 2005).

Detection of selection signatures is challenging for several reasons. First, the effects of selection on the distribution of genetic variation can be confused with patterns of genetic variation caused by demographic events such as the size, structure and mating pattern of a population (Akey *et al*., 2002). Both adaptive hitchhiking and population expansion can result in an excess of rare alleles (Tajima, 1989). Secondly, most studies are conducted using the available SNP assays that were developed using an ascertainment process in which only common variation is incorporated in the assay. The variability, distribution of allele frequencies and levels of linkage disequilibrium are all strongly affected by this ascertainment bias (Nielsen *et al*., 2005). Despite these challenges detection of signatures of selection has been the focus of several theoretical and empirical studies (Sabeti *et al*., 2002, Nielsen, 2005, Ramey *et al*., 2013).

Several methods have been employed to detect selection signatures using molecular data, including those based on linkage disequilibrium, allele frequency spectrum and haplotype structure characteristics of selected populations (Helyar *et al*., 2011). These methods have been used to infer genomic regions that were affected by domestication, breed formation and selection for specific production traits of livestock. In chickens, Rubin *et al*. (2010) detected selective sweep regions potentially associated with domestication and the specialization of broiler and layer birds. They also found one putative region including *TSHR* that was associated with metabolic regulation and photoperiod control of reproduction in vertebrates. In pigs, putative selective sweeps were observed on chromosome 1 (Groenen *et al*., 2010). Genomic regions harbouring the genes *IGF2*, *PRLR* and *GHR* were shown to have undergone selection in pigs (Andersson and Georges, 2004). Furthermore, genomic regions associated with immune response and feed efficiency were detected based on F_{st} estimates of divergence in cattle (The Bovine HapMap Consortium, 2009). Using both population differentiation (F_{st}) and Integrated Haplotype Score (iHS) approaches, Qanbari *et al*.

(2011) identified a total of 236 genomic regions potentially under selection in Holstein cattle. Both approaches suggested selection in the vicinity of the Sialic acid binding Ig-lecitn 5 gene on chromosome (BTA18), a region that was shown to carry a major QTL with large effects on productive life and fertility traits in Holstein cattle.

South African cattle are known to possess adaptive traits that include tolerance to various diseases, tolerance to extreme temperatures and humidity, tolerance to fluctuations in the availability of feed, adaptation to low capacity management and the ability to survive, produce and reproduce for extended period of time (Muchenje *et al.*, 2008 & 2009; Marufu *et al.*, 2011). However, research to dissect the genomic basis of these traits of economic importance is limited. The aim of this study was to conduct a genome wide scan for signatures of selection within and between six cattle breeds in South Africa including Afrikaner, Nguni, Drakensberger, Bonsmara, Angus and Holstein using SNP data generated using Illumina Bovine SNP50 BeadChip (Illumina, San Diego, California, USA). This will be a first step towards understanding the nature of genetic variation underlying differences between these breeds.

Materials and methods

Animal samples and quality control

A total of 249 purebred animals from the Afrikaner (n=44), Nguni (n=54), Drakensberger (n=47), Bonsmara (n=44), Angus (n=31) and Holstein (n=29) breeds were genotyped using the Illumina BovineSNP50 BeadChip v2 which features 54,609 single nucleotide polymorphism (SNP) probes distributed across the whole bovine genome with an average spacing of 49.9 kb (Matukumalli *et al.*, 2009). The genotyped samples were derived from a previous study (Makina *et al.*, 2014) and were approved for this research by the University of Pretoria Ethical Committee (E087-12). Blood and semen were used as sources of genomic DNA. Only SNPs that were uniquely mapped to autosomes were included in the analyses. Samples with more than 10 % missing genotypes were excluded.

To detect signatures of selection, the data were quality controlled using two methods. The first approach detected selective sweeps within each breed by searching for local reduction in genetic variation using minor allele frequencies (MAF). Thus, the BovineSNP50 data were first filtered to retain loci with at least a 95% call rate per breed, leaving a total of 51,406 (Afrikaner), 50,870 (Nguni), 50,389 (Drakensberger), 51,242 (Bonsmara), 50,922 (Angus) and 52,294 (Holstein) SNPs. The second approach targeted the identification of signatures of divergent selection between breeds using population differentiation (F_{st}). Thus, SNPs with less than a 95% call rate and MAF less than 0.02 across all breeds (Purcell *et al.*, 2007) were removed leaving 45,657 SNPs across the breeds. Furthermore, SNPs that were in high Linkage Disequilibrium (LD) were pruned using *indep* 50 5 2 in PLINK (Purcell *et al.*, 2007). A total of 21,290 SNPs remained after pruning and were used for detection of signatures of selection. The pruning of SNPs that are in high LD has been shown to

reduce the mean SNP heterozygosity within the European cattle breeds that were used to discover common SNPs in the design of the BovineSNP50 assay and therefore partially counters the effects of SNP ascertainment bias (Kijas *et al*., 2009).

Identification of selection signatures

Two approaches were employed to detect signatures of selection in this study based on the fact that positive selected loci can result in exceptionally high F_{st} between populations while having an excess of low frequency allele within a population (Hayes *et al*., 2009). First, we searched for regions of the genome within each breed in which haplotypes had been driven to complete fixation. The theory behind this approach is based upon the fact that strong, on-going selection for variants ultimately leads to a complete loss of variation within the chromosomal region surrounding the selected variant and the complete fixation of the haplotype which harbours the selected variant (Ramey *et al*., 2013). To identify selective sweep regions, the BovineSNP50 data were analysed separately by breed noting that the total number of variable SNPs differed within each breed due to the ascertainment bias associated with SNP discovery in the design of the BovineSNP50 assay (Rubin *et al*., 2010). To identify selective sweeps within each breed; a breed-specific number of contiguous monomorphic loci (Table 6.1) of at least five adjacent SNPs spanning 100 kb (UMD3.1 coordinates) for which no SNP had a MAF greater than 0.01 was required. To allow for the possibility of new mutations, genotyping errors and assembly errors which may have incorrectly assigned a variable marker to a sweep region we allowed a minimum MAF of ≤ 0.01 (Ramey *et al*., 2013).

Table 6.1: Summary of animals genotyped for the six breeds

Breed	Breed type	Primary historical use	Number of animals genotyped with BovineSNP50	Contiguous BovineSNP50 loci*	Number of Monomorphic SNP50 loci
Afrikaner	Sanga	Beef	44	8	15,791
Nguni	Sanga	Beef	54	7	10,059
Drakensberger	Sanga	Beef	47	5	6,543
Bonsmara	Composite	Beef	44	6	8,278
Angus	Bos taurus	Beef	31	6	6,861
Holstein	Bos taurus	Milk	29	6	8,463

*Number of contiguous loci spanning at least 100 kb and with a minor allele frequency ≤ 0.01 required to declare a selective sweep in each breed.

To determine the appropriate number of contiguous SNPs within each breed with MAF ≤ 0.01 to declare a selective sweep a trade-off between Type 1 error and the size of the detected signature was required. According to Ramey *et al*. (2013) if 15 % of SNPs are monomorphic within a breed (Table 6.1), the probability that N contiguous SNPs are monomorphic is 0.15^N under the null hypothesis of no selective sweeps anywhere within the genome. For example, assuming independence, and testing of 51,406 (Afrikaner), 50,870 (Nguni), 50,389 (Drakensberger), 51,242 (Bonsmara), 50,922 (Angus) and 52,294 (Holstein) SNPs on 29 autosomes, we would expect to find 0.15^N x (52,294 – 29 x (N-1)) regions where N contiguous SNPs had fixed alleles. For N = 5 this corresponds to 4.0 false positives per breed but only 0.6 false positives when N = 6. While increasing the number of contiguous monomorphic SNPs decreases the Type 1 error, it also increases the size of the signature that can be detected to, on average, (N-1) x 47 kb (Ramey *et al*., 2013). Therefore, an intermediate balance of these conflicting constraints was chosen (Table 6.1) based on the idea that signatures identified in two or more breeds or any signature that was validated by the F_{st} analysis or any sweep that overlaps with previously reported sweep regions would be real and should share a common haplotype.

The second approach identified regions of the genome which showed elevated levels of population subdivision between the breeds (Akey *et al*., 2002; Kijas *et al*., 2009) using population-specific F_{st} (Weir & Cockerham, 1984). Unbiased estimates of F_{st} as described by Weir & Cockerham (1984) were calculated using SNP Variation Suite (SVS) version 8.1 (SVS 8.1; Golden Helix Inc., Bozeman, Montana) for each of the SNPs between all (fifteen) pairs of cattle breeds in this study. Values were interpreted using the qualitative guidelines proposed by Wright (1978) where F_{st} > 0.25 indicates very great differentiation, between 0.15 - 0.25 great differentiation, between 0.05 to 0.15 moderate differentiation and F_{st} < 0.05 indicates little differentiation among the populations.

Unbiased estimates of F_{st} can assume negative values, which do not have a biological interpretation, thus all negative values were set to 0.0 (Weir & Cockerham, 1984). To determine the inter-locus variation in allele frequency, an empirical genome distribution of F_{st} values for all autosomal SNPs was constructed across the breeds.

Based on the relationship between breed pairs the most differentiated breed pairs were selected as candidate pairs for the detection of signatures of selection. This resulted in the dairy breed (Holstein) being used as the control breed in comparisons to the other five beef breeds: Afrikaner *vs* Holstein, Nguni *vs* Holstein, Drakensberger *vs* Holstein, Bonsmara *vs* Holstein and Angus *vs* Holstein. In addition, the Angus beef breed (British origin and less adapted to tropical regions) was compared to all four South African beef breeds which are better adapted to tropical regions to search for signatures of selection that may be associated with adaptation to tropical environments.

A sliding window of 5 SNPs was used to compute averages for the F_{st} statistics and the resulting smoothed F_{st} values for each of the compared breeds pairs were plotted against chromosomal coordinates of the central SNP based on the UMD 3.1 assembly using SNP Variation Suite (SVS) version 8.1 (SVS 8.1; Golden Helix Inc., Bozeman, Montana) (Golden Helix, 2012). The most differentiated regions representing the largest 2 % of SNPs ($F_{st} \geq 0.25$) were identified and these were considered to be under selection.

Annotation and functional analysis of identified genomic regions

Genomic coordinates for all identified selected regions were used for the annotation of genes that were fully or partially contained within each selected region using the UCSC Genome Browser (Kent *et al*., 2002). The functions and pathways in which these genes are involved were assessed using Panther (Mi *et al*., 2013). In addition, the Bovine QTL database available online at http://www.animalgenome.org/cgi-bin/QTLdb/BT/search was explored to identify any overlapping of previously published cattle QTL with the candidate regions.

Results

Fixed haplotypes

Descriptive statistics including the minor allele frequency (MAF), percentage of polymorphic SNPs and Hardy-Weinberg Equilibrium (HWE) for the breeds in this study were previously reported (Makina *et al*., 2014). Table 6.2 shows putative selective sweep regions detected within each breed, identified by detecting haplotypes at complete fixation. A total of twenty genomic regions on 13 chromosomes were identified as harbouring selective sweeps. Signatures of selection were identified in all six breeds; ranging from one region (Nguni) to six regions (Holstein) per breed. Seventeen predicted signatures were breed specific and three were shared between breeds with one being shared between Drakensberger and Bonsmara (BTA5) and two between Angus and Holstein (BTA10 and 16) (Figure 6.1). The average size for the breed specific sweeps was 267.54 kb,

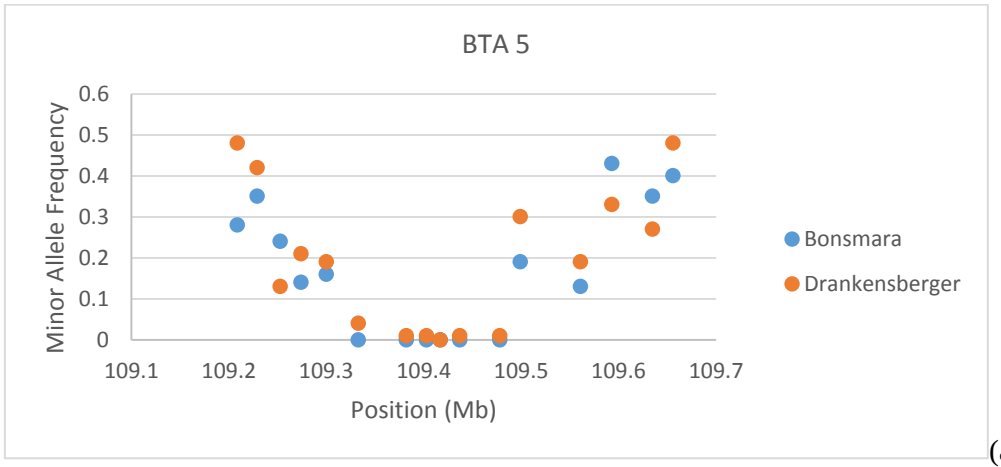
ranging from 162.16 to 530.46 kb while the average for the common signatures was 245.86 kb, ranging from 95.94 to 448.56 kb. No sweep regions were found to be common among Afrikaner, Nguni and Drakensberger.

Table 6.2: Potential candidate genes and previously detected QTL within detected selective sweep regions within breeds

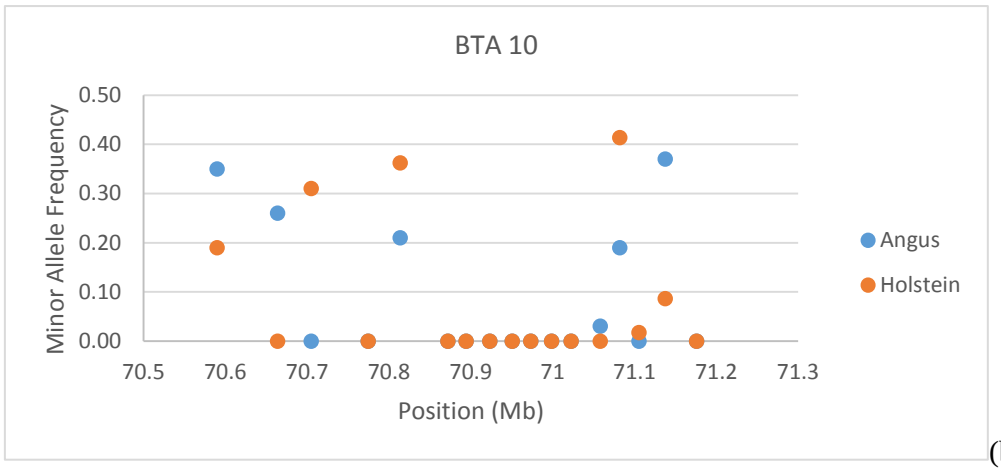
Breed[1]	**BTA**	**UMD3.1 coordinate (bp)**	**Number of SNPs**	**Size (kb)**	**Candidate genes**	**QTL**
ANG	1	89,563,554-89,734,339	6	170.79	*KCNMB3, PIK3CA, ZMAT3*	Body length, withers height, hip width
DRA	1	115,420,906-115,619,350	5	198.44	-	Non return rate, calving ease
BON	3	62,887,463-63,196,635	6	309.17	*GNG5, RPF1*	Milk protein percentage, marbling score, dystocia
AFR	4	102,570,116-103,100,577	6	530.46	*MTPN*	Parasites, marbling score, fat thickness,
DRA	5	28,859,701-29,043,711	5	184.01	*HOXC12, HOXC13*	Udder height, intramuscular fat, milk yield , longissimus muscle area
DRA & BON	5	109,333,059-109,478,057	6	145.00	*WC1, WC1.3*	Calving ease, milk fat, ovulation rate, milk yield, marbling score
BON	6	102,546,791-102,779,196	8	232.41	*ZNF280B, NUCB2, KBTBD1*	Interval to first oestrus after calving, marbling score
HOL	7	63,608,866-63,778,905	6	170.04	*ATOX1, G3BP1, GLRA1*	Somatic cell count, milking speed, tick resistance, heel depth, feed conversion ratio
	7	72,882,903-73,126,315	8	243.41	-	Somatic cell count, milking speed
BON	8	24,844,168-25,057,606	6	213.44	*KIAA1797*	First service conception rate, fat thickness, body weight, somatic cell and marbling score
AFR	10	40,135,969-40,460,414	5	324.45	-	Milk protein yield, milk fat, strength and body weight
ANG & HOL	10	70,871,943-71,022,679	7	150.74	*OTX2*	Milk protein yield, teat length, tick resistance, social separation walking and running
HOL	13	12,076,103-12,276,846	6	200.74	-	Body weight, somatic cell count, teat placement
	13	15,456,721-15,683,571	6	226.85	-	Body weight, somatic cell count, teat placement, udder depth
NGU	13	78,430,096-78,793,099	8	363.00	*KCNB1, PTGIS*	Residual feed intake, body weight (slaughter), weaning weight, teat length
ANG & HOL	16	45,425,579-45,874,144	7	448.57	*AJAP1*	Abomasum displacement, residual feed intake, carcass weight, bone percentage, calving ease
BON	16	51,195,450-51,357,613	6	162.16	*PDPN*	Abomasum displacement, residual feed intake, carcass weight, body weight (weaning and birth), calving

						ease
Breed[1]	**BTA**	**UMD3.1 coordinate (bp)**	**Number of SNPs**	**Size (kb)**	**Candidate genes**	**QTL**
	18	44,880,710-45,044,333	6	163.62	*DNAH2, TMEM88, GUCY2D*	Calf size, subcutaneous fat thickness, gastrointestinal nematode burden, residual feed intake, somatic cell
	19	27,734,700-28,060,683	9	325.98	*ALOX15B ALOX12B*	Calf size, residual feed intake, milk fat yield
BON	24	34,248,516-34,415,701	6	167.19	*RBBP8*	Stillbirth, udder depth, interval to first oestrus after calving, oleic acid content, weaning weight, somatic cell height

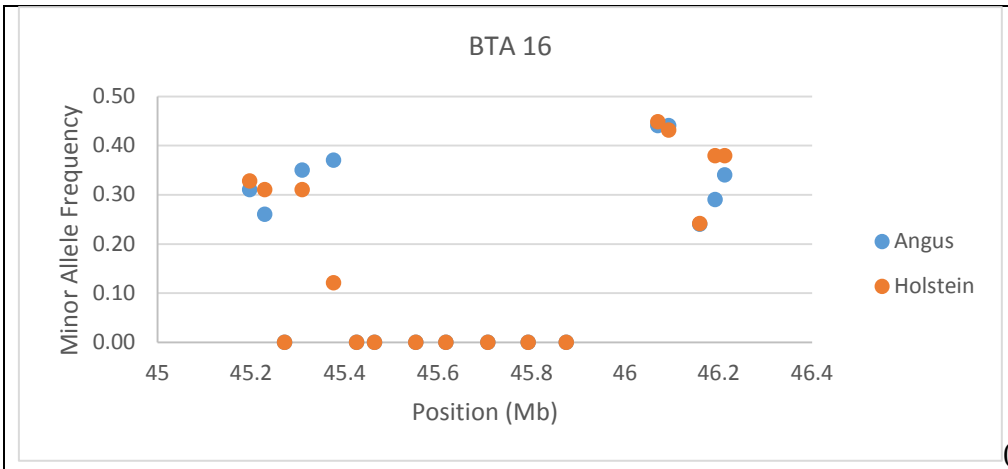
[1]AFR- Afrikaner, NGU-Nguni, DRA-Drakensberger, BON-Bonsmara. ANG-Angus and HOL-Holstein



(a)



(b)



(c)

Figure 6.1 Selective sweep regions shared between two breeds: (a) Bonsmara and Drakensberger, (b) Angus and Holstein and (c) Angus and Holstein

Highly differentiated genomic regions

The empirical genome-wide distribution of F_{st} values for all autosomal markers was constructed to examine inter-locus variation in allele frequency (Figure 6.2). The distribution was highly skewed towards small F_{st} values. About 31 % of SNPs possessed $F_{st} \leq 0.05$ while only 2 % had $F_{st} \geq 0.25$. This was consistent with other studies (Moradi *et al*., 2012; Kijas *et al*., 2012) which observed a skewed F_{st} distribution and is in accordance with the theory of natural selection (Akey *et al*., 2002).

Using the population differentiation approach, a total of 27 genomic regions were identified as potentially under positive divergent selection. These regions were distributed across 14 chromosomes (Table 6.3) indicating that about 8.5 Mb of sequence in these South African cattle breeds is under strong selection. The average size of the genomic regions under selection was 328.88 kb, with the largest region observed between Afrikaner - Holstein (860.14 kb) on BTA16 at 73,143-933,282 bp and the smallest observed between Bonsmara-Holstein pair (85.52 kb) on BTA20 at 11,932,262 – 12,017,779 bp.

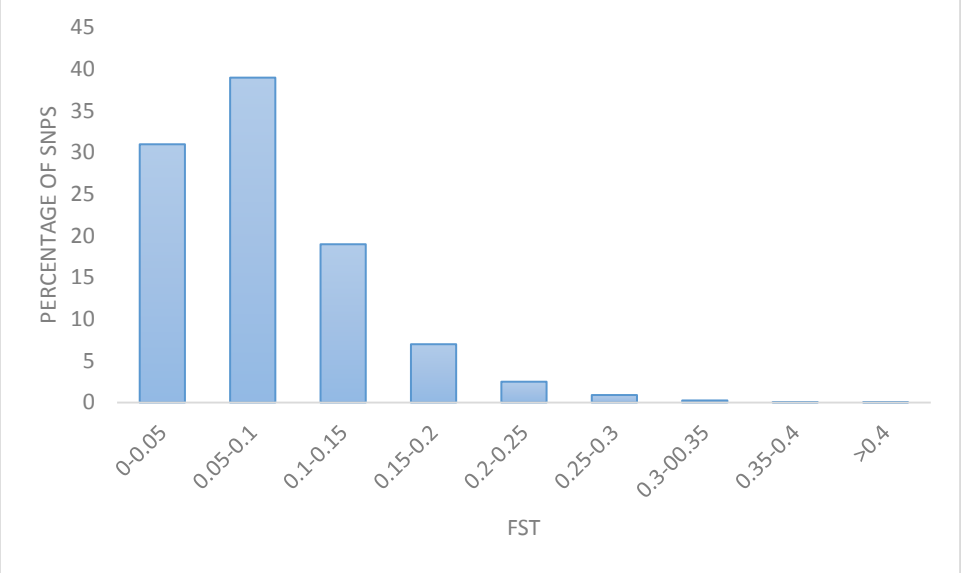


Figure 6.2 Genome wide distribution of F_{st} across all autosomes for all 15 breed comparisons

Figure 6.3 shows Manhattan plots of F_{st} values for the five breed comparisons that generated the greatest number of differentiated regions. The number of F_{st} peaks per chromosome varied from zero to two across these breed comparisons. Seven of these differentiated regions (BTA3, 5, 9 16, 18, 21 and 24) were shared among breeds pairs, with the Afrikaner *vs* Holstein and Nguni *vs* Holstein pairs sharing the most differentiated regions. Afrikaner *vs* Holstein pair had the highest number of differentiated regions (eight) while Angus *vs* Holstein pair had the lowest number (two). The most strongly differentiated region was observed between Afrikaner and Holstein on BTA9 at 105,263,583-105,587,941 bp. Comparisons of Angus *vs* Afrikaner, Nguni, Drakensberger and Bonsmara revealed a differentiated genomic region on BTA24 at 54,571,696-54,964,769 (Figure 4) which was shared by all of the South African cattle breeds.

Table 6.3 Genomic regions identified as being under divergent selection in six cattle breeds in South Africa and their associated QTL.

Breeds[1]	**BTA**	**UMD3.1 coordinate (bp)**	**Number of SNPs**	**Size (kb)**	**Smoothed F_{ST}**	**Candidate genes**	**QTL**
AFR *vs* HOL, NGU *vs* HOL	3	35,255,950-35,785,053	5	529.10	0.28	*KCNA2, CYM PROK1, PROK1, LAMTOR, SLC16A4, UBL4B*	Milk fat percentage, milk protein percentage, body weight, height, somatic cell count
AFR *vs* HOL	3	121,025,205-121,374,825	4	349.62	0.44	-	Shear force, fat thickness at the 12th rib
BON *vs* HOL	3	99,004,471-99,111,024	3	106.55	0.39	*SCP2, SLCA17, NDUFA12, SEMA4A*	Calf size, carcass weight, clinical tick-resistance mastitis, marbling score, calving index
DRA *vs* HOL	3	7,957,960-8,391,057	3	433.10	0.25	*NOS1AP*	Non return rate, body weight, longissimus muscle area, milk protein percentage, marbling score
AFR *vs* HOL, NGU *vs* HOL	5	4,472,786-4,598,476	4	125.69	0.42	-	Tenderness score, teat placement, shear force
AFR *vs* HOL, NGU *vs* HOL	5	114,085,555-114,594,935	3	509.38	0.48	*ERC1, FBXL14, WNT5B, ADIPOR2*	Hip height, rump length, calving ease, height, ovulation, type, rump angle
BON *vs* HOL	5	107,242,527-107,451,881	4	209.35	0.36	*OVOS2*	Ovulation rate, calving ease, marbling score, height, milk yield, milk fat
NGU *vs* HOL	7	71,038,040-71,240,079	4	202.04	0.25	*EBF1*	Somatic cell count, milking speed, tick resistance, heel depth, social separation--vocalisation
DRA vs HOL	7	46,109,256-46,700,828	4	591.57	0.34	*CXCL14, SLC25A48, FBXL21, LECT2, TGFBI*	Stillbirth, milking speed, body weight,, parasites, milk beta-casein percentage
AFR vs HOL, DRA *vs* HOL	9	105,263,583-105,587,941	3	324.36	0.50	*SFT2D1, BRP44L, RPS6KA2*	Chest depth, scrotal circumference, milk yield, milk alpha-casein percentage, milk protein yield
BON *vs* HOL,	9	15,767,136-15,991,964	3	224.83	0.37	*MYO6, IMPG1*	Clinical mastitis, weaning weight, longissimus muscle area, residual feed intake, milk fat yield

Breeds[1]	BTA	UMD3.1 coordinate (bp)	Number of SNPs	Size (kb)	Smoothed F_{ST}	Candidate genes	QTL
NGU *vs* HOL	9	46,632,366-47,246,008	4	613.64	0.43	*PREP*	Clinical mastitis, marbling score, milk protein yield
DRA *vs* HOL	14	81,125,493-81,269,892	3	144.40	0.40	-	Stature, body weight carcass weight, behaviour, height
AFR *vs* HOL	16	73,143-933,282	8	860.14	0.31	*FMOD, PRELP, OPTC, ATP2B4, LAX1, ZC3H11A, SNRPE, REN, TMEM51*	Milk protein yield, height, carcass weight, length of productive life
BON *vs* HOL, NGU *vs* HOL	16	49,386,191-49,867,758	3	481.57	0.41	*DNAJC16, CASP9, CELA2A, CTRC, EFHD2, TMEM51*	Abomasum displacement, milk, carcass weight, calving ease, bone percentage
AFR *vs* HOL	18	1,094,150-1,422,084	5	327.93	0.47	*DDX19A, DDX19B, AARS, EXOSC6, MRCL, PDPR, GLG1*	Weaning weight-maternal milk
DRA *vs* HOL, BON *vs* HOL	18	1,212,743-1,486,363	4	273.62	0.35	*PDPR, GLG1*	Weaning weight-maternal milk
NGU *vs* HOL	18	14,757,060 – 14,758,700	3	487.73	0.28	*CHMP1A, SPATA2, CDK10, FANCA, SPIRE2, TCF25, MC1R*	Dystocia, somatic cell score, longissimus muscle area, fat thickness at the 12th rib, carcass weight, stillbirth, skin pigmentation
NGU *vs* HOL	19	42,896,570 – 42,897,840	4	478.76	0.32	HSPB9, *WIPF2, CDC6, RARA, IGFBP4, TNS4, CCR7, SMARCE1*, K *KRT222, KRT24-27*	Intramuscular fat, average daily milk yield, milk capric acid percentage, lauric acid, myristic acid, milk c14 index, hair development
AFR *vs* HOL	20	11,932,262- 12,017,779	3	85.52	0.41	-	Body weight, average daily gain, longissimus muscle area, somatic cell score
AFR *vs* HOL, NGU *vs* HOL	21	43,246,618-43,399,424	4	152.81	0.30	-	Somatic cell score, calving ease, carcass weight

Breeds[1]	**BTA**	**UMD3.1 coordinate (bp)**	**Number of SNPs**	**Size (kb)**	**Smoothed F_{ST}**	**Candidate genes**	**QTL**
DRA *vs* HOL	21	59,640,020-59,787,612	3	147.59	0.26	*SERPINA3-8*	Calving ease, gastrointestinal nematode burden, weaning weight, body weight (birth), height (mature & yearling)
ANG *vs* HOL	22	32,930,704-33,076,318	3	145.61	0.28	*FRMD4B*	Non return rate, calf size, somatic cell
NGU *vs* HOL	23	2,019,985-2,247,046	3	227.06	0.34	-	Milk protein yield, height, carcass weight, percentage live sperm after thawing
ANG *vs* HOL	23	49,809,003-49,945,187	4	136.18	0.46	-	Body weight, dry matter intake
AFR vs ANG, NGU *vs* ANG, DRA vs ANG, BON *vs ANG*	24	54,588,817- 54,593,951		393.07	0.45 0.29 0.25 0.25	*DCC*	Gastrointestinal nematode burden, body weight, calving ease, udder attachment, feed conversion ratio, body weight
AFR *vs HOL*	27	35,734,689-36,117,365	4	382.68	0.26	-	Dystocia, marbling score, clinical mastitis

[1]AFR- Afrikaner, NGU-Nguni, DRA-Drakensberger, BON-Bonsmara, ANG-Angus and HOL-Holstein

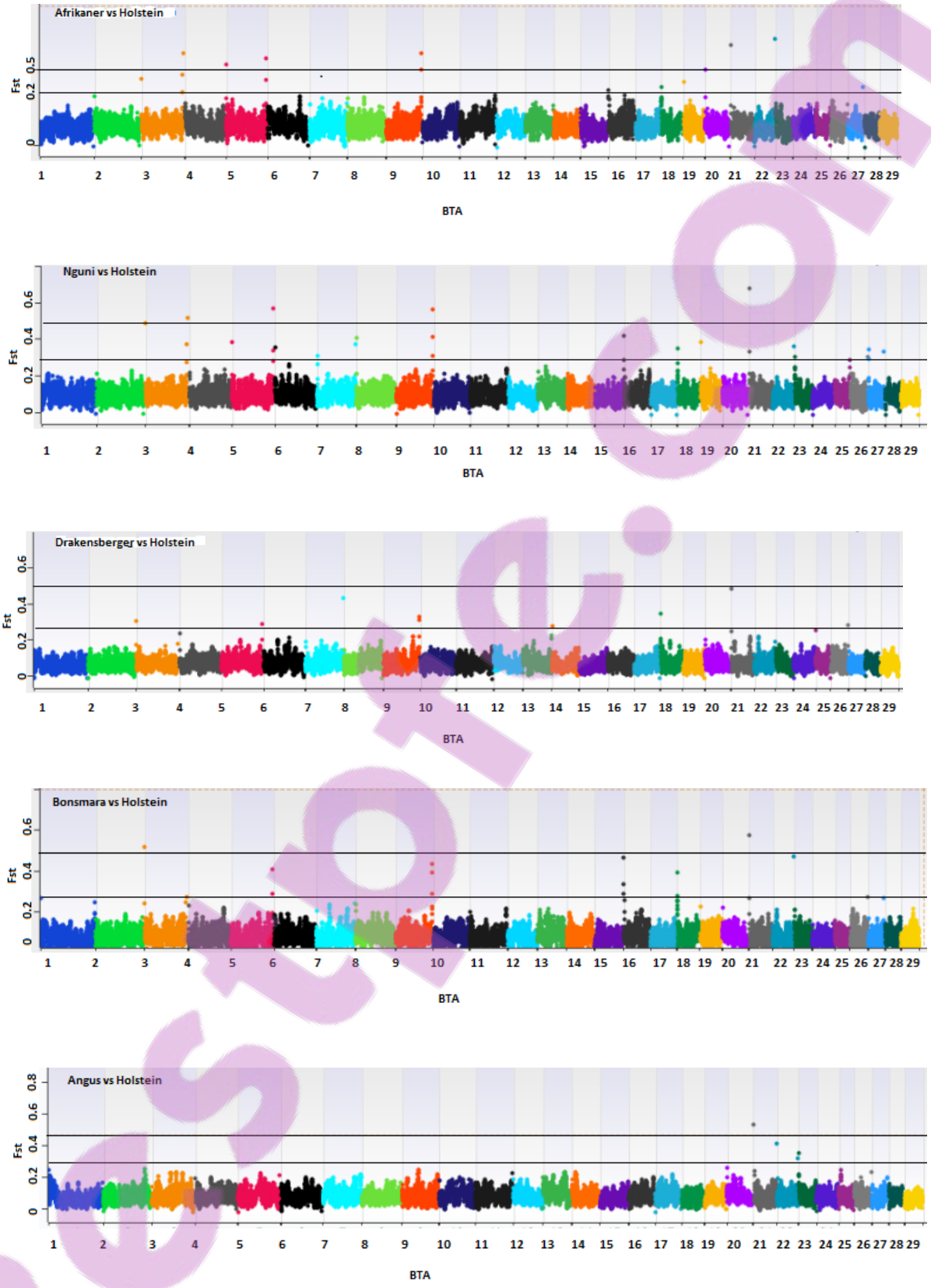


Figure 6.3 Smoothed F_{st} values for five breed comparisons across the autosomal genome.

Functional annotation of genomic regions showing evidence of selection

Using genomic regions obtained from both the within and between breed analyses, a total of 33 reference sequences (Refseqs) were annotated to identify potentially expressed genes. The number of genes obtained per Refseq varied from one to eight across the genomic regions. Using the PANTHER (Mi *et al*., 2013) website, a number of genes were linked to important biological functions and pathways in cattle. For example, the keratin family (*KRT222, KRT24, KRT25, KRT26, and KRT27*) and one heat shock protein (*HSPB9*) on BTA19 at 42,896,570 – 42,897,840 bp in the Nguni cattle were associated with tropical adaptation. *MTPN* (Afrikaner), *CYM* (Afrikaner and Nguni), *CDC6, CDK10, EBFI & TNS4* (Nguni), *NDUFA12, ALOX15B, ALOX12B* (Bonsmara) and SLC25A48, *SERPINA3-8* (Drakensberger) are related to immune response. *ADIPOR2* (Afrikaner) *PTGS* (Nguni*) HOXC12, HOXC13, WC13 & OVOS2* (Drakensberger *&* Bonsmara) are related to reproduction and *SLC6A17* and *PREP* are related to fatty acid biosynthesis.

Furthermore, genes related to nervous system development were also identified e.g. *WNT5B, FMOD, PRELP* (Afrikaner), *KRT25, CCR7* (Nguni) and *OVOS, SLC6A17* (Bonsmara). Genes involved in enzyme regulatory activities e.g. *MYO6, RBBP8* (Bonsmara), *CYM, LAX1* (Afrikaner), *ATP2B* (Nguni), and *SLC16A4* (Drakensberger) were also detected. Genes involved in growth and metabolic processes e.g. *DDX19A* (Afrikaner), *KCNB1, IGFBP* (Nguni), *TGFB1* (Drakensberger), *MYO6* (Bonsmara), AJAPI (Angus) and *ATOX1* (Holstein), genes involved in muscle organ development and skeleton development e.g. *KIAAI1797, EFHD2* (Bonsmara) and *MTPN, TMEM51* (Afrikaner) were also identified as being under selection. Finally, *MCIR* on BTA18 (13,486,389-13,974,114 bp) was detected to be under selection in Nguni cattle.

In addition, all of the genomic regions showing evidence of selection were further analysed to determine if some of these regions overlap with previously reported quantitative trait loci (QTL) in cattle. The online database of published Bovine QTL revealed that most of the genomic regions overlapped with previously reported regions harbouring QTL affecting milk, fat, carcass, body weight, stature, clinical mastitis, calving ease, tick resistance, gastrointestinal nematode burden and reproductive traits (Tables 6.2 and 6.3). For example, a region on BTA24 (figure 6.4) detected in Afrikaner, Nguni, Drakensberger and Bonsmara cattle overlapped with QTL previously found to be associated with gastrointestinal nematode burden (http://www.animalgenome.org/cgi-bin/QTLdb/BT/search).

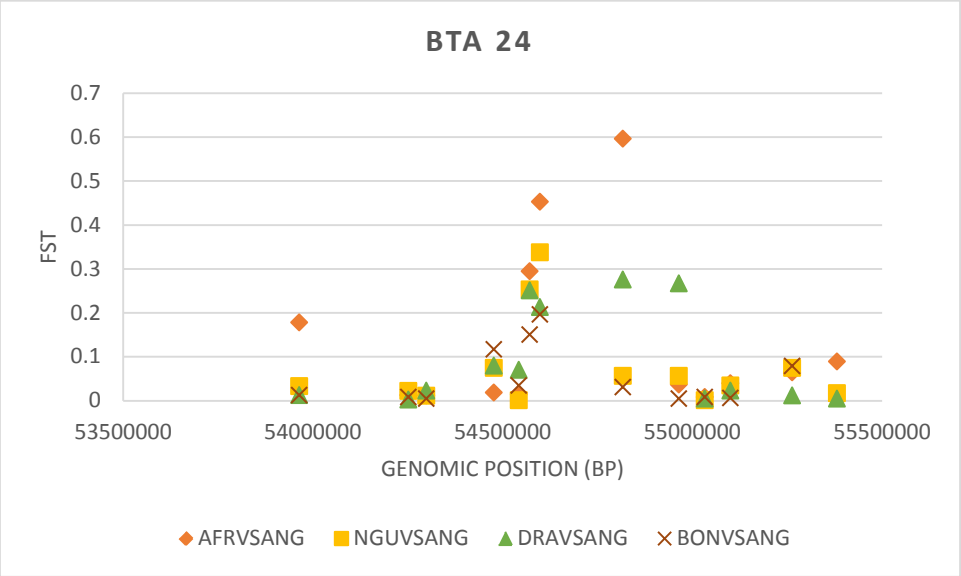


Figure 6.4 Distribution of *Fst* values for four breed pairs comparisons on BTA 24: AFR- Afrikaner, NGU-Nguni, DRA-Drakensberger, BON-Bonsmara and ANG-Angus

Signatures of selection observed in this study were also compared to previously detected sweep regions found in cattle and Table 6.4 presents regions identified in the current study that were also detected in previous studies of cattle. Ten genomic regions identified in this study were also detected in previous studies of signatures of selection. This highlights the importance of these genomic regions in the bovine.

Table 6.4 Overlapping regions possessing signatures of selection detected in previous studies in cattle

BTA	Position (bp)	Breed	Reference
1	89,563,554-89,734,339	Angus	(Chan *et al*., 2010)
3	99,004,471-99,111,024	Bonsmara	(Chan *et al*., 2010; Porto-Neto et al., 2013)
3	121,025,205-121,374,825	Afrikaner	(Gautier *et al*., 2009)
5	109,333,059-109,478,057	Bonsmara &Drakensberger	(Chan *et al*., 2010; Porto-Neto *et al*., 2013)
7	72,882,903-73,126,315	Holstein	(Ramey *et al*., 2013)
13	15,456,721-15,683,571	Holstein	(Ramey *et al*., 2013)
16	45,425,579-45,874,144	Angus and Holstein	(Ramey *et al*., 2013; Chan *et al*., 2010; Porto-Neto *et al*., 2013; Stella *et al*., 2010)
16	51,195,450-51,357,613	Bonsmara	(Porto-Neto *et al*., 2013; Stella *et al*., 2010)
22	32,930,704-33,076,318	Angus	(Gautier *et al*., 2009)

BTA	Position (bp)	Breed	Reference
24	54,588,817- 54,593,951	Afrikaner, Nguni, Drakensberger and Bonsmara	(Gautier *et al*., 2009)

Discussion

A total of 47 genomic regions potentially under selection were detected within and between six cattle breeds in South Africa. Twenty of these genomic regions were detected within breeds and 27 were detected as diverged between breeds. In addition, eleven of these genomic regions were shared between breeds and ten have been reported in previous studies (Ramey *et al*., 2013; Chan *et al*., 2010; Porto-Neto *et al*., 2013; Gautier *et al*., 2009). Natural selection is a process that is driven by environmental factors in which individuals with specific genotypes have a differential ability for contributing to the next generation. Thus, it affects allele frequencies in populations and result in a reduction in genetic variation as the frequency of alleles associated with higher fitness increases (Akey *et al*., 2002). This phenomenon is known as the hitchhiking effect or selective sweep (Akey *et al*., 2002).

The Melanocyte stimulating hormone receptor gene *(MC1R*) which influences the production of eumelanin and pheamelanin pigment and is responsible for pigmentation of the skin, eyes and hair (Seo *et al*., 2007), was found to be within the selected region in Nguni cattle on BTA 18 at 14,757,060 – 14,758,700 bp in this study. The presence of strong selection on this gene within Nguni cattle which are characterized by multi-coloured skin patterns that can be present in many different forms (white, brown, golden yellow, black, dappled, or spotty) is of interest. Identifying the mutation that underlies the strong selection detected on this gene would allow a better understanding of the role of *MC1R* on coat colour patterning in cattle. Domestication has caused considerable changes in the morphology and behaviour of livestock species, this was shortly followed by artificial selection in which specific traits were selected based on a goal (Hayes *et al*., 2009). During the process of domestication, coat colours were selected because of their immediate rewards to animal breeders and also because they were associated with improved individuals or because of cultural preferences (Helyar *et al*., 2011). The *MC1R* gene was also found to be under selection in domesticated cattle by Flori *et al*. (2009), Stella *et al*. (2010) and Ramey *et al*. (2013). This gene has two common alleles E^D and e and a less common allele, E^+, also called "wild type". In the presence of E^D which is the dominant allele in the series, an animal is typically black, while a recessive genotype (e/e) results in red coats. However, E^+ appears to act as a "neutral" allele in most breeds and it is thought that E^D /E^+ cattle are typically black and E^+/e cattle are typically red. However, $E^{+/}E^+$ cattle can be almost any colour since other genes, such as Agouti, can now be expressed to indicate the pigments produced (Chen *et al*., 2009).

Behavioural changes such as a reduction in fear and anti-predator responses as well as an increase in sociability are believed to have been selected during domestication (MacHugh *et al*., 1997). This study detected several selected genomic regions that are related to nervous system functions. For example, *WNT5B, FMOD, PRELP* (Afrikaner), *KRT25, CCR7* (Nguni), *OVOS, SLC6A17* (Bonsmara) are involved in nervous system development. The Bovine HapMap Consortium (2009) and Gautier *et al*. (2009) also reported selection signatures surrounding genes associated with the nervous system of cattle.

South African cattle breeds have a history of adaptation to the sub-tropical areas of Africa. A number of genes and families of genes that were previously associated with one or more performance attribute for tropical adaptation (Chan *et al*., 2010; Gautier *et al*., 2009) were found in this study. For example, a number of keratin genes (*KRT222, KRT24, KRT25, KRT26* and *KRT27*) and one heat shock protein (*HSPB9*) on chromosome 19 at 42,896,570 – 42,897,840 bp were found to be under selection in Nguni. Keratin (heteropolymeric structural protein) form the basis for structural constituent of epidermis during epidermis development. Epidermis development occurs in response to adaptation to different climatic conditions, including tick exposure (Wang *et al*., 2007). Heat shock proteins have been shown to be differentially expressed between indicine and taurine cattle in tropical environments of Africa and have been associated with tropical adaptation in Zebu cattle (Chan *et al*., 2010; Gautier *et al*., 2009).

In addition to the role that the keratins genes play during epidermis development, they also play a role in the formation of the hair shaft (Wu *et al*., 2008). Skin colour and the thickness of hair directly influence the thermo-resistance of cattle living in the tropics (Mattioli *et al*., 2000). Nguni cattle have smoother and shinier hair coats compared to European cattle breeds. These characteristics provide Nguni cattle with a greater ability to regulate their body temperature and to more efficiently maintain cellular function during heat (Muchenje *et al*., 2009) as well as the ability to resist tick infestation (Marufu *et al*., 2011).

Several other candidate genes related to antigen recognition, which is a key process underlying the development of immune response were identified in this study. For example, *MTPN* (Afrikaner), *CYM* (Afrikaner and Nguni), *CDC6, CDK10, KCNBI* and *TNS4* (Nguni), *NDUFA12, ALOX15B, ALOX12B* (Bonsmara) and SLC25A48, *SERPINA3-8* (Drakensberger) were detected. The CD immune response genes were described by Meissner *et al*. (2012) as being closely involved with MHC molecular functions and pathways. *TNFAIP8L2* has been recognized as a major player in individual immune homeostasis (Zhang *et al*., 2011) and *NDUFA12* has previously been reported as having diverged allele frequencies between taurine and Zebu cattle and was associated with tick resistance. These observations are consistent with the tolerance of Afrikaner, Nguni, Drakensberger and Bonsmara cattle to various tick and parasitic diseases (Muchenje *et al*., 2009; Marufu *et al*., 2011). Furthermore genomic regions harbouring *MTPN* and *PDPR* (Afrikaner), *DCC* (Afrikaner, Nguni, Drakensberger and Bonsmara), *OTX2* (Angus), *DNAH2*, *TMEM88* and *GUCY2D*

(Bonsmara), *EBF1* (Nguni), and *CXCL14, SLC25A48* (Drakensberger) overlap with previously identified QTLs underlying tick resistance and nematode tolerance in cattle (http://www.animalgenome.org/cgi-bin/QTLdb/BT/search).

Several candidate genes indirectly or directly involved in reproductive pathways including spermatogenesis, ovulation rate, oestrus process, testis development and prostaglandin development in cattle were found. These included *ADIPOR2* (Afrikaner), *OVOS2* (Bonsmara), *ADIPOR2* (Afrikaner and Nguni), *WC1* (Drakensberger and Bonsmara), *RBBP8* (Bonsmara), *SERPINA3-8, HOXC12, HOXC13* (Drakensberger), and *FBXL4* (Afrikaner and Nguni). The fact that Afrikaner, Nguni, Drakensberger and Bonsmara cattle have the ability to produce and reproduce under harsh environment conditions and are considered excellent dam lines for crossbreeding (Scholtz, 2010) supports the strong selection on reproductive loci that is likely to have occurred in their adaptation to South African conditions. In addition, these regions also overlapped previously reported QTL associated with reproduction in cattle (http://www.animalgenome.org/cgi-bin/QTLdb/BT/search).

Candidate genes related to growth and muscle development were also detected as being under selection, these included *DDX19A, TMEM51, MTPN* (Afrikaner), *IGFBP4,* (Nguni), *TGFB1, KCNB1,* (Drakensberger) *MYO6, KIAAI1797, EFHD2* (Bonsmara) *AJAP1* (Angus) and *ATOX1* (Holstein). In addition, some of these regions overlapped QTLs previously associated with stature, body weight and double muscling in cattle. Furthermore, selection signatures observed in this study also overlapped with previously reported QTLs that affect milk yield and quality (BTA3, 5, 10 16 and 23), feed efficiency (BTA13, 16 and18), fat thickness (BTA5, 18 and 19), marbling score and carcass weight (BTA3, 5, 16, 20 and 27) as well as somatic cell count (BTA3, 5, 7, 9, 18 and 22).

Conclusion

The overall goal of this study was to determine genomic regions that were targeted by selection within and between the major cattle breeds of South Africa. A significant proportion of regions observed in this study were validated by previous studies (Gautier *et al*., 2009; Chan *et al*., 2010; Ramey *et al*., 2013; Porto-Neto *et al*., 2013). This study provided insights into the genetic mechanisms of traits of economic importance among cattle breeds in South Africa in particular with regard to adaptation to tropical environments and to tick and parasite-borne diseases as well as reproduction and production potential.

The study represents the first attempt at locating genomic regions targeted by selection in South African breeds that should be prioritized for functional dissection. A number of genomic regions were identified that are directly or indirectly involved in tropical adaptation, immune response activation, tick and parasite resistance, production and reproductive performance. In addition, selected regions that overlap with QTL reported in the QTL database provide additional evidence of the significance of the detected selected regions.

References

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

Andersson, L. and M. Georges. (2004). Domestic-animal genomics: Deciphering the genetics of complex traits. *Nature Reviews Genetics* 5, 202-212.

Chan, E.K., Nagaraj, S.H. & Reverter, A. (2010). The evolution of tropical adaptation: comparing taurine and zebu cattle. *Animal Genetics* 41, 467-477.

Chen, S. Y., Huang, Y., Zhu, Q., Fontanesi, L., Yao, Y. G. and Liu, Y. P. (2009). Sequence characterization of the MC1R gene in Yak (*Poephagus grunniens*) breeds with different coat colors. *Journal of Biomedicine and Biotechnology*, 10.1155/2009/861046.

Flori, L., S. Fritz, F. Jaffrézic, M. Boussaha, I. Gu *et al*. (2009). The genome response to artificial selection: A case study in dairy cattle. *PLoS ONE* 4, e6595.

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloe, D., Gut, I., Moazami-Goudarzi, K. & Foulley, J.L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10, 550.

Golden Helix Inc, (2012). SNP & Variation Suite Manual, Version 8.1 Available 430 at: http://www.goldenhelix.com

Groenen, M.A., Amaral, H.J., Megens, G., Larson, A. & Archibald, L. *et al.* (2010). The Porcine HapMap Project: genome wide assessment of nucleotide diversity, haplotype diversity and footprints of selection in the pig. The International Plant & Animal Genome XVIII Conference. January 9-13, San Diego, California. W609.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., Maes, G.E., Diopere, E., Carvalho, G.R. & Nielsen, E.E. (2011). Application of SNPs for population genetics of non-model organisms: new opportunities and challenges. *Molecular Ecology Resources* 11, 1-14.

Hayes, B.J., Chamberlain, A.J., Maceachern, S., Savin, K., McPartlan, H., Macleod, I., Sethuraman, L. & Goddard, M.E. (2009). A genome map of divergent artificial selection between *Bos taurus* dairy cattle and Bos taurus beef cattle. *Animal Genetics* 40, 176-184.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research* 12, 996-1006.

Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., Mc William, S., Tang, D., McEwan, J,, Cockett, N., Oddy, V.H., Nicholas, F.W. & Raadsma, H. (2009). International Sheep Genomics Consortium: A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4, e4668.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto-Neto, L.R., San Critobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J. & Dalrymple, B. (2012). International Sheep Genomics Consortium Members: Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* 10, e1001258.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P., & Bradley, D.G. (1997). Microsatellite DNA variation and the evolution, domestications and phylogeography of taurine and zebu cattle. *Genetics* 146: 1071-1086.

Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E., MacNeil, M.D. & Maiwashe, A. (2014). Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Front. *Genetics* Doi:10.3389/fgene.2014.00333.

Marufu, M.C., Qokweni, L., Chimonyo, M. & Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. *Ticks and Tick Borne Disease* 2, 172-177.

Mattioli, R.C., Pandey, V.S., Murray, M. & Fitzpatrick, J.L. (2000). Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (*Bos taurus*) and trypanosusceptible Gobra zebu (*Bos indicus*) cattle. *Acta Tropica* 75, 263-277

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

Meissner, T.B., Liu, Y.J., Lee, K.H., Li, A., Biswas, A., van Eggermond, M.C., van den Elsen, P.J. & Kobayashi, K.S. (2012). NLRC5 cooperates with the RFX transcription factor complex to induce MHC Class 1 gene expression. *The Journal of Immunology* 188, 4951-4958.

Mi, H., Muruganujan, A. & Thomas, P.D. (2013). PANTHER in 2013: modelling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research* 41, D377-386.

Moradi, M.H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K.G. & McEwan, J.C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genetics* 13, 10.

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G. & Strydom, P.E. (2008). Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. *Animal* 2, 298-304.

Muchenje, V., Dzama, K., Chimonyo, M., Strydom, P.E. & Raats, J.G. (2009). Relationship between pre-slaughter stress responsiveness and beef quality in three cattle breeds. *Meat Science* 81, 653-657.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics* 39, 197-218.

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

Nielsen, R. & Yang, Z., 1998. Likelihood models for detecting positively selected amino acids sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929-936.

Otto, S.P. (2000). Detecting the form of selection from DNA sequence data. *Trends Genetics* 16, 526-529.

Porto-Neto, L.R., Lee, S.H., Sonstegard, T.S., Van Tassell, C.P., Lee, H.K., Gibson, J.P. & Gondro, C. (2013). Genome-wide detection of signatures of selection in Korean Hanwoo cattle. *Animal Genetics* 45, 180-190.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferriera, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 81, 559-575.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G. & Simianer, H. (2011). Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* 12, 318.

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D. & Taylor, J.F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14, 382.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, O., Bed'hom, B., Tixier-

Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K. & Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587-591.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P, Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platkon, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R. & Lander, E.S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832-837.

Scholtz, M.M. (2010). Beef Breeding in South Africa. 2nd Edition, ARC Pretoria, South Africa ISBN-13: 978-1-86849-391-3.

Seo, K., Mohanty, T.R., Choi, T. & Hwang, I. (2007). Biology of epidermal and hair pigmentation in cattle: a mini-review. *Veterinary Dermatology* 18, 392-400.

Simianer, H. Ma, Y. & Qanbari, S. (2014). Statistical Problems in Livestock Population Genomics. Proceedings, 10th World Congress of Genetics Applied to Livestock Production, Vancouver, BC, Canada, 17-22 August 2014.

Stella, A., Ajmone-Marsan, P., Lazzari, B. & Boettcher, P. (2010). Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* 185, 1451-1461.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585-595.

The Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528-532.

Vitalis, R., Dawson, K. & Boursot, P. (2001). Interpretation of variation across marker loci as evidence of selection. *Genetics* 158, 1811-1823.

Wang, Y.H., Reverter, A., Kemp, D., McWilliam, S.M., Ingham, A., Davis, C.A., Moore, R.J. & Lehnert, S.A. (2007). Gene expression profiling of Hereford Shorthorn cattle following challenge with *Boophilus microplus* tick larvae. *Australian Journal of Experimental Agriculture* 47, 1397-1407.

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

Wright, S. (1978). Evolution and the Genetics of populations. 'Variability within and among Natural Populations'. Volume 14. University of Chicago Press 1978, Chicago.

Wu, D.D., Irwin, D.M. & Zhang, Y.P. (2008). Molecular evolution of the keratin associated protein gene family in mammals, role in the evolution of mammalian hair. *BMC Evolution Biology* 8, 241.

Zhang, L, Shi, Y., Wang, Y., Zhu, F., Wang, Q., Ma, C., Chen, Y.H. & Zhang, L. (2011). The unique expression profile of human TIPE2 suggests new functions beyond its role in immune regulation. *Molecular Immunology* 48, 1209-1215.

Chapter 7

General discussion and recommendations

7.1 General discussion and recommendations

Genome wide characterization can assist to unravel the genetic history of livestock populations, shed light with regard to population structure and further reveal genomic regions targeted by selection (Boettcher *et al.*, 2010). In addition, genetic information provided by such studies is most relevant for managing the present and future genetic diversity and may contribute to the development of more effective utilisation, selection, breeding and conservation strategies (Gautier *et al*., 2007). In this study, the Bovine SNP50 BeadChip was used to characterize the genetic diversity and population structure of SA cattle breeds, determine the level of linkage disequilibrium and conduct a genome wide scan for signatures of selection among the Afrikaner, Nguni, Drakensberger, and Bonsmara using the Angus and Holstein cattle as reference groups since they have been characterized in other countries using similar tools.

The first experiment performed included the evaluation of the Bovine SNP50 BeadChip to determine its utility for genome wide based studies of South African cattle. The results revealed that over 50% of SNPs derived primarily from European cattle breeds sequences were polymorphic among South African cattle breeds, e.g. Nguni = 35 843, while over 75 % SNPs were polymorphic among European breeds (Holstein = 41 078 and Angus = 40 146). Matukumalli *et al*. (2009) observed a similar trend in studies with European and African breeds with 42 849 and 41 073 polymorphic SNPs in Holstein and Angus cattle, respectively, and only 28 869 and 35 084 SNPs in the African N'Dama and Sheko breeds. It was encouraging to observe that the BovineSNP50 array was equally informative for use in South African Sanga compared with the other African breeds that were included during the validation of the BovineSNP50 array.

Overall, this analysis demonstrated that the BovineSNP50 array will be useful for genome wide studies across cattle breeds that are widely used by South African farmers for dairy and beef production. However, it should be noted that some level of ascertainment bias associated with the design of the Bovine SNP50 arrays were observed in this study, for example just over 50 % of SNPs within the Bovine SNP50 array were polymorphic among Sanga cattle while over 75 % SNPs were polymorphic among *Bos taurus* breeds. As the cost of sequencing decreases it is envisaged that limitations associated with the current SNP arrays will diminish. Sequencing of South African indigenous and locally developed breeds will contribute additional SNPs to reduce the ascertainment bias associated with the currently available Bovine SNP arrays.

The second experiment contributed genetic information regarding the genetic diversity and population structure of South African cattle breeds. This revealed that Afrikaner cattle had the

lowest level of genetic diversity (H_e=0.24) while the Drakensberger cattle (H_e=0.30) had the highest level of genetic variation among indigenous and locally-developed cattle breeds. Animal genetic diversity is a prerequisite for genetic improvement and environmental adaptation (FAO, 2005). It is also critical for food security and rural development because it allows farmers to select stock or develop new breeds in response to changing conditions, including climate change, new or resurgent disease threats, new knowledge of human nutritional requirements, and changing market conditions or societal needs (FAO, 2010). Therefore, this lower level should be noted in Afrikaner and steps toward increasing diversity that include exchanging bulls from the different genetic pools and making use of natural breeding, random mating or pedigree-based mating (FAO, 2010) should be prioritized. The level of inbreeding was lower across the studied cattle breeds. High rate of inbreeding in livestock populations results in loss of genetic variation and the occurrence of inbreeding depression, which could increase the prevalence of rare lethal disorders (Szpiech *et al*., 2013). Therefore an assessment of the inbreeding level should be done every five years to determine any unfavourable change in inbreeding levels, so that appropriate steps could be taken to prevent increases in inbreeding.

As expected, the average genetic distance was the greatest between indigenous cattle breeds and *Bos taurus* cattle breeds but the lowest among indigenous and locally-developed breeds. This was in agreement with the great divergence between African and European/British breeds (Gautier *et al*., 2007) and suggested distinct genetic resource among South African cattle breeds that should be properly utilized and conserved in order to cope with unpredictable future environments. Population structure analysis revealed some signals of admixture and genetic relationship between Afrikaner, Nguni and Drakensberger and Bonsmara. Nguni cattle shared some genetic links with the Afrikaner cattle, with about 8 % of its genome derived from the Afrikaner cattle. This may reflect co-ancestry between these breeds as both breeds came from the same migration route into the Southern Africa (Scholtz & Ramey, 2011). On the other hand, the Bonsmara cattle shared some genetic links with the Nguni cattle (3 %) but only limited genetic links with Afrikaner cattle (0.5 %); which was unexpected since the Bonsmara cattle was developed through crossbreeding of Afrikaner cattle with exotic breeds such as Hereford and Milk Shorthorn during the 1950s (Bonsma, 1980). In general the results of this study gave insights on the genetic structure of the SA cattle breeds some of which is supported by the domestication and breed formation of South African cattle breeds.

The third experiment determined the extent of linkage disequilibrium and effective population size among the six studied cattle breeds in this study. Linkage disequilibrium analyses revealed significant differences on the extent of LD between the studied breeds. Afrikaner cattle had the highest extent of LD (with threshold of r^2 =0.2 extending up to 40-60 kb) compared to the other indigenous breeds. It is well documented that the extent of LD in the population depends on the

history of the population, especially its effective population size (Hayes *et al.*, 2003). For example, the estimated effective population size for Afrikaner cattle was 33 five generations ago, while Nguni and Drakensberger cattle (with threshold of r^2 =0.2 extending up to 10-20 kb) had a N_e of 92 and 83 respectively. Furthermore, Bulmer (1971) showed that selection reduces genetic variation in the next generation and produces LD among syntenic and non-syntenic loci. Indeed, in this thesis, the long range LD within the Afrikaner cattle was in agreement with its low level of genetic variation.

In general, the results of the population structure and effective population size suggested that the Afrikaner, Angus and Holstein breeds are endangered and close to critical stage (FAO, 2005). Therefore, pointing out the need for implementation of appropriate conservation programs as well as new selection and breeding strategies to ensure the long term fitness of these breeds. These could include increasing the number of animals contributing offspring to each generation by increasing the active cow populations (FAO, 2010). In addition, proper selection of bulls and rotational breeding strategies as well as the use of bulls from a number of unrelated sources could ensure fitness within this breeds. Crossbreeding is a well-known method of genetic improvement, however unsupervised crossbreeding could threaten the existence of pure breeds (FAO, 2007).

It is well known that natural and artificial selection has dramatically shaped the pattern of cattle genetic variation during domestication and breeding (The Bovine HapMap Consortium, 2009). The last experiment focused on a genome wide scan for signatures of selection to detect regions of the genome targeted by selection. This study detected a total of 47 genomic regions potentially under selection within and between six cattle breeds in South Africa. Twenty of these genomic regions were detected within breeds and 27 were detected as diverged between breeds. In addition, eleven of these genomic regions were shared between breeds and ten were previously detected in cattle (Ramey *et al*., 2013; Chan *et al*., 2010; Porto-Neto *et al*., 2013; Gautier *et al*., 2009). Among these, a number of genomic regions that are directly or indirectly involved in tropical adaptation (*KRT222*, *KRT24*, *KRT25*, *KRT26* and *KRT27*) and one heat shock protein (*HSPB9*) on chromosome 19 at 41,447,971-41,926,734 bp in the Nguni cattle were detected. This was in agreement with history of sub-tropical adaptation within the Nguni cattle.

The observation of the immune response activation genes *MTPN* (Afrikaner), *CYM* (Afrikaner and Nguni), *CDC6*, *CDK10*, *KCNBI* and *TNS4* (Nguni), *NDUFA12*, *ALOX15B*, *ALOX12B* (Bonsmara) and SLC25A48, *SERPINA3-8* (Drakensberger) as well as tick and parasite resistance genes - *MTPN* and *PDPR* (Afrikaner), *DCC* (Afrikaner, Nguni, Drakensberger and Bonsmara), *OTX2* (Angus), *DNAH2*, *TMEM88* and *GUCY2D* (Bonsmara), *EBF1* (Nguni), and *CXCL14*, *SLC25A48* (Drakensberger) were consistent with the tolerance of Afrikaner, Nguni, Drakensberger and Bonsmara cattle to various tick and parasitic diseases (Muchenje *et al*., 2008; Marufu *et al*., 2011).

Furthermore, the identification of genes related to nervous system development e.g. *WNT5B, FMOD, PRELP* (Afrikaner), *KRT25, CCR7* (Nguni), *OVOS, SLC6A17* (Bonsmara) were in agreement with the behavioural changes such as a reduction in fear and anti-predator responses as well as an increase in sociability that are believed to have been selected during domestication (MacHugh *et al*., 1997).

Coat colour is an important breed characteristics and have undergo strong selection in cattle (Ramey *et al*., 2013). Nguni cattle has unique multi-coloured skin, which can present many different patterns (white, brown, golden yellow, black, dappled, or spotty). This study identified the *MCIR* gene on BTA18 (13,486,389-13,974,114 bp) in the Nguni cattle which has previously been associated with coat colour in cattle (Seo *et al*., 2007). Further investigation of the *MCIR* gene would allow a better understanding of the role of *MC1R* on coat colour patterning in cattle.

In general, this experiment provided insights with regard to regions of the genome that are, or have been, functionally important and have thus been targeted by either natural or artificial selection among South African cattle breeds. It has provided a genomic understanding of how and where natural and artificial selection have shaped the patterns of variation in the genome of South African cattle breeds. In addition, it has shed important insight with regard to mechanisms of adaptation to tropical environments, tick and parasite tolerance, diseases resistance and reproduction and production potential. The results presented in this study provide a useful foundation for detection of mutations underlying genetic variation in traits of economic importance for South African cattle breeds.

7.2 Future studies

Future research studies should focus on expanding this breed level analysis through the inclusion of all major African cattle breeds (Gautier *et al*., 2009) together with the cattle breeds of the world (Dekkers *et al*., 2014). This could provide insights with regards to the genetic relationship shared among South African cattle breeds and cattle breeds of the world and hopefully reveal some history regarding African cattle domestication and shed more light with regard to genomic requirement for survival in African environments.

Analyses of distribution of other forms of genomic variants such as runs of homozygosity and their relationship with inbreeding among the studied breeds should be prioritized. Runs of homozygosity is a powerful tool for estimating inbreeding coefficients, evaluating its effects on traits of economic importance and controlling its level in selection and breeding programs (Bjelland *et al*., 2013).

7.3 Conclusions

Small sample size may not be representative of the whole population and may only include the predominant animals within the breed. This could lead to erroneous conclusion about the entire

population. However, the results of this thesis for all four experiments were in agreement with the previous studies of cattle populations with bigger sample size (>60) e.g. Matukumalli *et al*. (2009); Gautier *et al*., 2007 & 2009; Khatkar *et al*. (2008); Bohmanova *et al*. (2010); Lu *et al*. (2012); Ramey *et al*. (2013).

This study presented the first attempt to genomically characterize South African cattle breeds for neutral and selected variation using the recently developed Bovine SNP50 BeadChip. The study revealed low to moderate genetic diversity within six cattle breeds in South Africa and showed a closer relationship among indigenous and locally-developed cattle breeds. Clear genetic divergence between South African (indigenous and locally-developed cattle breeds) and *Bos taurus* cattle breeds was observed which suggested distinct genetic resource in South Africa cattle breeds that should be properly utilized and conserved in order to cope with unpredictable future environments.

In addition, this study revealed significant differences in the extent of LD between the studied breeds. Afrikaner cattle had the highest level of LD compared to the other indigenous breeds. This suggests that the Afrikaner cattle have experienced considerable bottlenecks restricting their effective population size in contrast to other indigenous breeds and would require less number of SNPs to perform genome wide studies compared to other indigenous breeds.

Furthermore, a genomic understanding of how and where natural selection has shaped the pattern of genetic variation among cattle breeds in South Africa was unveiled by identifying loci that are important to the development of South African cattle breeds.

The results of this thesis form the foundation for the development of more appropriate utilisation, selection, breeding and conservation strategies of South African cattle.

References

Bjelland, D.W., Weigel, K.A., Vukasinovic, N. & Nkrumah, J.D. (2013). Evaluation of inbreeding depression in Holstein cattle using whole genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science* 96, 4697-4706.

Boettcher, P.J., Tixier-Boichard, M., Toro, M.A., Simianer, H., Eding, H., Gandini, G., Joost, S., Garcia, D., Colli, L., Ajmone-Marsan, P. & the GLOBALDIV Consortium. (2010). Objectives, criteria, and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics* 41 (1) 64-77.

Bohmanova, J., Sargolzaei, M. & Schenkel, F.S. (2010). Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics* 11, 421-431.

Bonsma, J.C. (1980). Livestock Production – A global approach. Tafelberg, Cape Town.

Bulmer, M.G. (1971). The effect of selection on genetic variability. *American Naturalist* 105, 201-211.

Chan, E.K., Nagaraj, S.H. & Reverter, A. (2010). The evolution of tropical adaptation: comparing taurine and zebu cattle. *Animal Genetics* 41, 467-477.

Dekkers, 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., Regitano, L.C.A., 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 Genetics* 10(3), e1004254. http://dx.doi.org/10.1371/journal.pgen.1004254.

FAO. (2005). Genetic characterization of livestock populations and its use in conservation decision making, by O. Hannotte & H. Jianlin. In J. Ruane & A. Sonnino, eds. The role of biotechnology in exploring and protecting agricultural genetic resources, pg. 89–96. Rome 1.

FAO. (2007). The State of the World's Animal Genetic Resources for Food and Agriculture, Rome, Italy.

FAO. (2010). Progress report of the FAO on selected activities related to agricultural biodiversity. http://www.fao.org/docrep/meeting/018/k5700e.pdf.

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., *et al*. (2007). Genetic and Haplotypic Structure in 14 European and African cattle breeds. *Genetics Society of America* 177, 1059-1070.

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloe, D., Gut, I., Moazami-Goudarzi, K. & Foulley, J.L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10, 550.

Hayes, B.J., Visscher, P.M., McPartlan, H.C. & Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* 13, 635–643.

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

Lu, D., Sargolzaei, M., Kelly, M., Li, C., Voort, G., Wang, Z., Plastow, G., Moore, S. & Miller, S. (2012). Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Frontier in Genetics* 3, 152-160.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P. & Bradley, D.G. (1997). Microsatellite DNA variation and the evolution, domestications and phylogeography of taurine and zebu cattle. *Genetics* 146: 1071-1086.

Marufu, M.C., Qokweni, L., Chimonyo, M. & Dzama, K. (2011). Relationships between tick counts and coat characteristics in Nguni and Bonsmara cattle reared on semiarid rangelands in South Africa. *Ticks and Tick Borne Diseases* 2, 172-177.

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

Muchenje, V., Dzama, K., Chimonyo, M., Raats, J.G. & Strydom, P.E. (2008). Tick susceptibility and its effects on growth performance and carcass characteristics of Nguni, Bonsmara and Angus steers raised on natural pasture. *Animal* 2, 298-304.

Porto-Neto, L.R., Lee, S.H., Sonstegard, T.S., Van Tassell, C.P., Lee, H.K., Gibson, J.P. & Gondro, C. (2013). Genome-wide detection of signatures of selection in Korean Hanwoo cattle. *Animal Genetics* 45, 180-190.

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D. & Taylor, J.F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* 14, 382.

Scholtz, M.M. & Ramsey, K.A. (2011). The Nguni Breed of Cattle—Past, Present and Future, Nguni Cattle Breeders' Society, Bloemfontein, pp. 14-17 (Chapter 2).

Seo, K., Mohanty, T.R., Choi, T. & Hwang, I. (2007). Biology of epidermal and hair pigmentation in cattle: a mini-review. *Veterinary Dermatology* 18, 392-400.

Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zoellner, S., Rosenberg, N.A. & Li, L. (2013). Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics* 93, 90-102.

The Bovine HapMap Consortium, (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528-532.

Chapter 8

Addendum

Supplementary material - Chapter 5

Supplementary material 1 (SM1): Number of SNPs genotyped and left after quality control filters

Breeds	**Initial SNPs**	**SNPs removed due to**				**Final SNPs**		
		Call rate < 95%	**Deviate HWE (p < 0.001)**	**MAF <0.05**	**Sex & *unknown position**	**Polymorphic SNP (used)**	**Percentage of polymorphic SNPs (%)**	***Average Gap (kb)**
Afrikaner	54609	725	279	22623	911	30484	56	81.65
Nguni	54609	2191	237	16436	963	35479	65	70.08
Drakensberger	54609	925	327	11825	118	40789	74	61.09
Bonsmara	54609	1871	140	12862	1018	39215	72	63.65
Angus	54609	916	198	12787	1135	398301	73	62.44
Holstein	54609	646	121	12127	1161	40734	74	61.08

***SNP position from the UMD 3.1 bovine assembly was used in the study**