Whole genome investigation of the genetic structure of South African sheep breeds

By

Anel Retief

Submitted in fulfilment of the requirements for the degree

Magister Scientiae Agriculturae

MSc(Agric) Animal Science: Animal Breeding and Genetics

In the Faculty of Natural and Agricultural Sciences

Department of Animal Sciences

University of Pretoria

Pretoria

November 2020

Supervisor: Dr Carina Visser

Declaration

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

Signature: _____
Date: November 2020
Anel Retief
Pretoria

Acknowledgements

Abstract

In this study whole-genome SNP data from 1977 animals, originating from thirteen sheep breeds in South Africa was analysed. The animals were classified into three production types: dual-purpose (Afrino, Dohne Merino, Dormer, Merino, SA Mutton Merino), meat (Black Headed Persian, Damara, Dorper, Meatmaster, White Dorper) and indigenous (Fat-tailed, Namaqua Afrikaner, Pedi). All animals were genotyped using the Illumina Ovine 50K SNP BeadChip. The aim was to investigate the genetic diversity and inbreeding levels of both indigenous and commercial sheep breeds in South Africa. The populations were first investigated individually to analyse within population diversity, whereafter they were merged in order to perform between population diversity analysis. During sample-based quality control, a total of 207 animals were removed due to low call rates. The number of SNPs remaining after marker-based quality control ranged from 32 422 to 44 778. Average observed heterozygosity values of 0.360, 0.355 and 0.340 were observed for dual-purpose, meat and indigenous populations respectively. Average linkage disequilibrium (LD) estimates were highest for dual-purpose populations at 0.277, followed by meat type populations at 0.259, and the lowest LD was observed for indigenous populations at 0.255. Dual-purpose, meat and indigenous populations showed average minor allele frequencies (MAF) of 0.393, 0.427 and 0.444 respectively. Principal component analysis (PCA) and ADMIXTURE results showed a clear differentiation between the dual-purpose populations, and the meat and indigenous populations. Average inbreeding coefficient (Fis) estimates of 0.008, -0.029 and 0.003 was observed for dual-purpose, meat and indigenous populations respectively. All populations showed a decline in effective population size (Ne) across generations. For all populations, the highest percentage of ROH was found in the shortest length category (1-3.99 Mb), and the largest number of ROH were observed on the first ten chromosomes. The average F_{ROH} estimates was highest for dual-purpose populations at 0.015, followed by the meat type and indigenous populations at 0.011 and 0.005 respectively. An unrooted phylogenetic tree based on pairwise Fst estimates showed a divergence of the dual-purpose populations from the meat and indigenous populations. Results from this study indicated lower variation within dual-purpose populations, and higher variation within indigenous populations. High population differentiation was observed between the various production types.

Table of Contents

List of Tables

List of Figures

Abbreviations

AFR	Afrino
BHP	Black Headed Persian
DAM	Damara
DNA	Deoxyribonucleic Acid
DOM	Dormer
DOR	Dorper
EBV	Estimated Breeding Value
FG	Genomic Inbreeding
Fis	Inbreeding Coefficient
Fst	Fixation Index
FTT	Fat-tail
GADI	Grootfontein Agricultural Development Institute
GCTA	Genome-wide Complex Trait Analysis
Ha	Hectare
He	Expected Heterozygosity
Ho	Observed Heterozygosity
HWE	Hardy-Weinberg Equilibrium
IBD	Identical by decent
INTERGIS	Integrated registration and genetic information system
ISGC	International Sheep Genomics Consortium
K	Number of assumed populations
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
Mb	Mega basepairs
MER	Merino
MMR	Meatmaster
NAM	Namakwa Afrikaner
Ne	Effective population size
NSSIS	National Small Stock Improvement Scheme
OAR	Sheep chromosome
PCA	Principal Component Analysis

PED	Pedi
QC	Quality Control
RAPD	Random Amplified Polymorphic DNA
ROH	Runs of Homozygosity
SA	South Africa
SAMM	SA Mutton Merino
SNP	Single Nucleotide Polymorphism
USA	United States of America
WC	Western Cape Agricultural Trust
WDOR	White Dorper

Chapter 1: Introduction

1.1 Introduction

Sheep, together with goats, are widely recognised as one of the first animals to be domesticated (Arranz et al., 1998; Peter et al., 2007). The Domestic sheep originated in southwest Asia from the wild Asian mouflon, and was domesticated for various reasons, such as their relatively simple food requirements, their ease of handling, their importance as sacrificial animals in several religions and because of the versatility of products obtained from them (meat, milk and wool) (Peter et al., 2007). Therefore, the process of sheep domestication was important for the development of human civilization. Initially, sheep were solely reared for meat production, however around 4 000-5 000 years ago specialisation for milk and wool started occurring as a result of human selection practises (Chessa et al., 2009). During the spread of agriculture from the Middle East, through south-eastern Europe and to the European continent during the Neolithic period, various sheep breeds were developed (Ryder & Bridbury, 1984).

The development of these differentiated breeds was driven by two main factors; selection by man for production purposes (e.g. wool vs mutton) and adaptation to environmental challenges (Ryder & Bridbury, 1984). As of the 18th century, the divergence of sheep populations into breeds became more prominent through the increased use of systematic breeding combined with well-defined objectives (Ciani et al., 2015). Since the last century, the rate of genetic gain has increased drastically through the application of quantitative genetics as well as the use of artificial insemination, allowing the prioritization of genetically superior rams (Cesarani et al., 2019). It is estimated that around 200 pure breeds and 400 composite breeds of sheep are farmed globally (Rasali et al., 2006).

Sheep populations are distinguished by the presence of large phenotypic variation within breeds (Kijas et al., 2012; Ciani et al., 2015). As a result of the large variation in phenotypes and genotypes observed in sheep, combined with the wide range of environmental conditions and production systems used for sheep farming, different sub-populations often exist within the same breed (Cesarani et al., 2019). Investigating the patterns of genetic variation have proven to be an important tool in studies of domestication, estimating population structure and breed formation, and for determining the consequences of selection (Kijas et al., 2012).

Understanding the population structure and breed composition of populations can aid in designing optimal breeding strategies, for example in crossbred and composite animals, where it is important to optimise the exploitation of inter-breed nonadditive genetic effects (Kijas et al., 2009). Furthermore, investigating the genetic variation of breeds that have been specifically selected for adaptation under specific conditions, thus showing optimal survival and production under these

conditions, can be especially beneficial to breeders (Peters et al., 2010). Understanding the unique combinations of adaptive traits that allow indigenous and locally developed breeds to effectively respond to environmental pressures is therefore an important asset to breeders (Peters et al., 2010). Examples of these traits of adaptation include resistance and tolerance to several diseases and parasites, changes in feed quantity or quality, fluctuations in temperatures and extreme weather conditions, as well as the ability of these animals to survive and reproduce over extended periods of time (Hammond, 2000). Globally, several examples of sheep breeds that are adapted to local conditions have been noted. These include breeds such as the Chiapas from Mexico (Quiroz et al., 2008), the Muzzafarnagri from India (Arora & Bhatia, 2004), and the Latxa and Carranzana breeds from the Iberian Peninsula (Rendo et al., 2004). The Dorper and Meatmaster breeds are examples of well-known adapted composite breeds developed in South Africa (Peters et al., 2010).

Sheep breeds are considered to be a good representative species for the study of phenotypic variation for traits such as resilience, adaptation to environmental conditions and resistance to diseases (Cesarani et al., 2019). Small ruminants are also relatively more resilient to increased temperatures, and are more capable to adapt to an extensive range of climatic conditions than other livestock (Benhin, 2006; Rust & Rust, 2013). The range of sheep breeds that occur in South Africa presents a unique opportunity to investigate them in terms of their genetic diversity and differentiation.

Before the widespread availability of dense SNP data, pedigree-based inbreeding estimates were used to characterise and control the genetic diversity of populations. However, this approach had several limitations, such as the need for complete, high quality pedigree data and records, and the fact that pedigree information does not take Mendelian sample variation, or linkage disequilibrium caused by selection, into account (Oliehoek & Bijma, 2009; Hill & Weir, 2011). The use of molecular data has proven to be a useful tool in investigating genetic diversity as well as differentiation among species and breeds (Sanarana et al., 2015).

Molecular studies on sheep in SA are limited. Furthermore, earlier studies were performed using microsatellite markers, and in the more recent studies where SNPs were used, limited sample sizes and a small number of breeds were included. This study will be the first comprehensive investigation of diversity in SA sheep using SNP data.

1.2 Aim of the study

The aim of this study was to investigate the genetic diversity, population structure and inbreeding levels of both indigenous and commercial sheep breeds in South Africa, using SNP data from 13 South African sheep breeds, namely: Afrino, Black Headed Persian, Damara, Dohne Merino, Dormer, Dorper, Fat-tail sheep, Meatmaster, Merino, Namakwa Afrikaner, Pedi, SA Mutton Merino and White Dorper.

This aim was achieved by attaining three objectives:

1. To investigate the level of genetic diversity within and between the sheep breeds.
2. To estimate inbreeding levels using both Fis and ROH methodology and effective population (Ne) sizes for the sheep breeds.
3. To investigate breed relatedness between the various South African sheep breeds and production types by investigating phylogenetic relationships.

Chapter 2: Literature review

2.1 Introduction

Modern domesticated livestock species were formed through centuries of artificial selection by humans, along with natural selection. Different environmental conditions, as well as different human cultures and needs led to the selection of specific breeds in certain areas (Erhardt & Weimann, 2007).

Sheep and goats were among the first animal species to be domesticated. This was due to their small body size, meek behavioural characteristics and versatile products, such as milk, meat, wool as well as leather (Ryder & Bridbury, 1984). Evidence from archaeological and molecular genetic studies indicate that the wild ancestor of the domesticated sheep was the mouflon *(O. orientalis),* and that domestication took place around 11 000 years ago within the Fertile Crescent region of modern-day Iraq, Syria, Jordan, and Egypt (Zeder, 2008). The process of sheep domestication first influenced behavioural and morphological traits, resulting in a reduction in body size, decreased horn length, the growth of underwool, increased docility, as well as an increased breeding period (Cesarani et al., 2019).

Domestication can also have a profound effect on the genetic diversity of a species, for example a loss of founder alleles caused by genetic selection, or mating of related individuals to improve certain traits (Vostry et al., 2018). In domesticated animals, gene flow can be constrained by human actions such as reproductive isolation of different flocks, which may lead to decreased variation and increased genetic differentiation (Xuebin et al., 2005). Furthermore, the extinction of breeds that were adapted to local conditions, and their replacement by highly productive breeds constitute a substantial loss of genetic diversity (Figueredo et al., 2019).

Genetic diversity within a breed is crucial in the process of evolution as it provides the substrate needed for both natural and artificial selection (Qanbari & Simianer, 2014). The genetic diversity among domesticated breeds allows producers to alter certain characteristics through selection, in response to changes in the environment as well as consumer preferences (Erhardt & Weimann, 2007).

2.2 The SA sheep industry

South Africa (SA) can be classified as an arid country, as much of the land is situated in areas ranging from hyper-arid to semi-arid (Cloete & Olivier, 2012). Aridity is defined as the conditions produced by permanent disparities of the water availability in an area; such as low average annual precipitation combined with high spatial and temporal variability, which results in an overall low moisture content and low carrying capacity of the ecosystem (Maliva & Missimer, 2012). Nearly all

land in the western and north western parts of the country are designated as arid, while the south western, southern, central and northern parts are mostly classified as semi-arid (Mucina & Rutherford, 2006). Figure 2.1 shows the Köppen climate type as well as the average summer and winter temperatures in SA. The average annual rainfall for the country is approximately 450mm, while a large part of the western half of the country, with the exception of small areas in the South-Western Cape mountains and the Southern Cape seaboard, has an average annual rainfall of below 400mm (Benhin, 2006).

The aridity of the country, as well as the acidity of the soils and restrictions in soil fertility and texture, limit the potential for agriculture, especially with regards to intensive and semi-intensive crop production (Cloete et al., 2014), and only 16.5% of the approximate 86.2 million hectares of commercial agricultural land can be deemed arable (DAFF, 2019). Large areas of the country can only be used for extensive livestock production. In the more arid western and north-western parts of the country, where the grazing capacity can be well over 12 ha per large stock unit, extensive small stock production is the predominant livestock industry. The small stock industry is therefore of critical importance to the South African livestock environment.
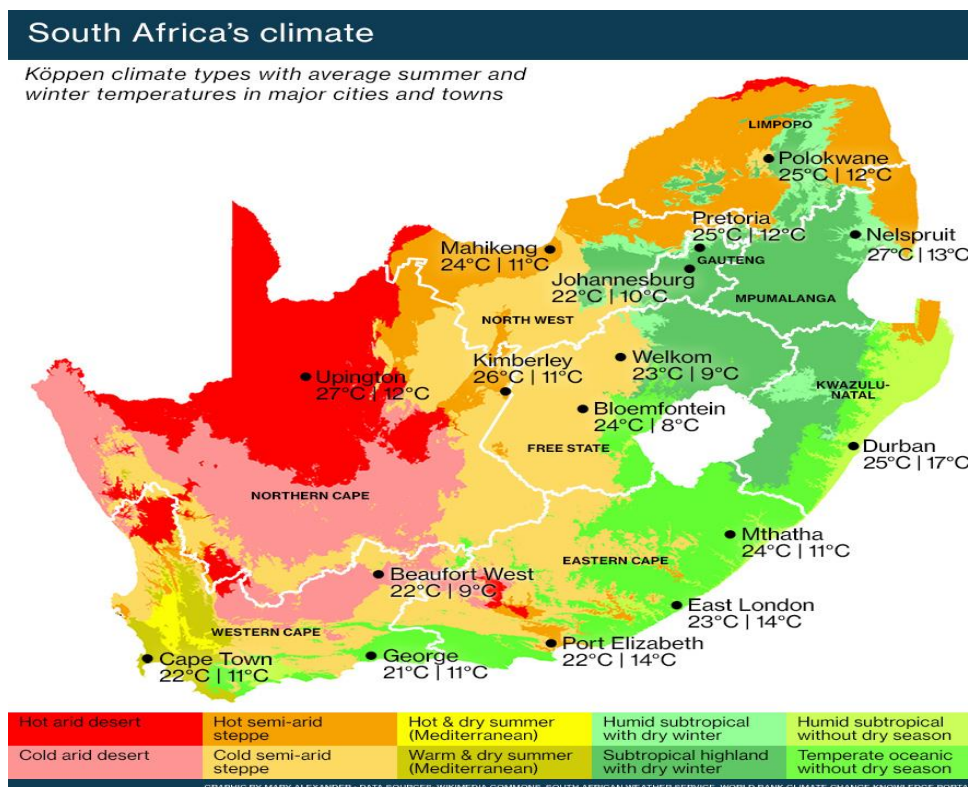
Figure 2.1 Map of climatic regions of South Africa (Alexander, 2018)

Sheep production in South Africa ranges from extensive production in arid areas with carrying capacities of less than one large stock unit per 40 ha, to intensive systems in pasture cropping areas and intensive horticultural regions (Cloete & Olivier, 2012). In November 2018, the estimated number of sheep in South Africa was approximately 22.3 million (DAFF, 2019). This includes both woolled and non-woolled sheep. The SA Animal Improvement Act listed a total of 23 registered sheep breeds in South Africa (Campher et al., 1998). Sheep farms are located throughout the country; however, the highest percentage can be found in the Eastern Cape, followed by the Northern Cape and the Free State. Sheep production in South Africa consist of approximately 8 000 commercial sheep farms, as well as roughly 5 800 communal farmers (DAFF, 2019). Sheep numbers in SA has declined steadily over the past 10 years, as a result of factors such as drought and stock theft, resulting in a decreased supply of mutton (DAFF, 2017). The average gross production value of mutton amounted to R4.57 billion per annum over the last 10 years (DAFF, 2017). As South Africa is a net importer of sheep meat, prices tend to be well connected to the global market, reacting to changes in supply and demand conditions that occur in major exporting countries such as Australia and New Zealand (ABSA, 2018).

In South Africa, income gained from animal products constitutes approximately 51% of the gross domestic income from agricultural production (DAFF, 2019). The largest proportion of livestock income is contributed by poultry production at 42-48%, while beef and veal contribute 16-28%, and milk 13-19%, respectively (DAFF, 2019). Small stock, consisting of sheep and goats, contribute approximately 8-10% of this animal product income. Mutton and lamb constitute the bulk of this value at 60.6%, followed by wool (31.4%), mohair (7.9%) and karakul pelts (0.2%) (DAFF, 2019). Despite the relatively moderate contribution of sheep production in financial terms, the industry plays an integral role in a regional context, and is strategically important in rural areas of the country (Schoeman et al., 2010).

In extensive pastoral areas where there are no viable alternative farming ventures, for example the vast Karoo regions, sheep farming permits sustainable production. The existence and maintenance of many rural communities are dependent on sheep production, and without the income derived from sheep products many of these rural towns would cease to exist (Cloete & Olivier, 2012). Sheep are also utilised in cropping systems, with sheep consuming crop residues. Many farmers depend on sheep farming to provide a livelihood for their employees and families during years of crop failure, drought and disease, and during periods of unstable grain prices (Cloete & Olivier, 2012).

2.3 SA sheep breeds

Several studies have been performed in an attempt to trace the migration patterns of domesticated species (Bruford et al., 2003; Chessa et al., 2009; Warmuth et al., 2012). Archaeological records show that the migration of sheep towards southern Africa occurred approximately 2000 years

ago. The truly indigenous breeds that moved south with the Khoi, Bapedi and Nguni people are the Damara, Namaqua Afrikaner, Ronderib Afrikaner, Pedi, Swazi and Zulu sheep. The South African Persian breeds are reportedly of Somalian or Saudi-Arabian origin, and have been found in South Africa for several centuries (Soma et al., 2012).

In South Africa, sheep breeds can be classified as hairy indigenous breeds, fat-tailed and fat-rumped breeds, and South African developed composite breeds, such as the SA Mutton Merino. Sheep can be used for meat and/or fibre production, and are sometimes used to produce dairy products (Soma et al., 2012). Production types include meat, wool and dual purpose, however there are very few that are used solely for wool production, as most wool sheep are also slaughtered for increased financial gain (Van der Merwe et al., 2019). Thirteen sheep breeds will be included in the current study, and are briefly described within their relevant production groups.

Indigenous breeds

Indigenous breeds are breeds that are native to a specific area. These sheep are best suited to extensive farming, as they are mainly animals that are well-adapted to harsh climates such as arid and hot areas (Hammond, 2000). They are generally of slender build with fat tails. Indigenous breeds in this study include Namakwa Afrikaner, Pedi and non-descript fat-tailed sheep. The importance of indigenous breeds is contained within their possible genetic diversity, as well as their adaptive abilities to harsh environments.

Namakwa Afrikaner: The Namakwa Afrikaner is an endangered, indigenous fat-tailed meat breed**.** It is one of the oldest sheep breeds found in South Africa (Cloete et al., 2014). Its normal production environment is the arid to semi-arid desert; most Namakwa Afrikaner sheep are kept in the Northern Cape Province. There are currently two flocks maintained by the Northern Cape Department of Agriculture. One flock is kept at the Carnarvon Experimental Station in the North-Western Karoo and the other at the Karakul Experimental Station near Upington (Snyman et al., 1993). This breed is also popular with communal farmers (Qwabe et al., 2012).

Pedi: The Pedi is an indigenous fat-tailed meat breed. It is a relatively small framed sheep, and is generally produced on sub-tropical bush veld to semi-arid savannah. It is a hardy breed which is well-adapted to the semi-arid bushveld. Most of the Pedi sheep in South Africa are found in the Limpopo province (Snyman, 2014b).

Meat type breeds

Mutton sheep are meat-type animals, which are solely kept for the production of meat, with no economic value added through fibre production. Meat type breeds included in this study are the Black Headed Persian, Damara, Dorper, White Dorper and Meat Master.

Black Headed Persian: The Black Headed Persian is a small framed, fat tailed sheep which originated in Somalia, and was imported to SA in 1870 (Soma et al., 2012). Thereafter the breed was crossbred with local breeds. It is a hair type sheep breed, and exists in three distinct varieties; Black head Persian, Red head Persian and Speckled Persian. The breed is adapted to hot and humid climates, and produced in dry semi-desert to savannah bushveld as well as coastal 'spekboom' veld. The bulk of the Blackhead Persians are farmed within the Northern Cape Province (Cloete & Olivier, 2012).

Damara: The Damara breed originated in Eastern Asia and Egypt, and currently the largest numbers of Damara sheep are found in Namibia and Angola (Soma et al., 2012) The breed was brought to SA between 200 and 400 AD. It is a fat-tailed and smooth haired sheep, which is adapted to the hot, dry savannah areas of SA, and most commercial farms are found in the Northern Cape. Smaller flocks are also found in the Free State and Gauteng provinces (Cloete & Olivier, 2012).

Dorper: The Dorper is a locally developed composite meat breed. It is the outcome of one of the most effective long-term livestock improvement programs in South Africa (Milne, 2000). It was developed by crossing and evaluating Persian and Dorset Horn sheep in a series of joint trials with the Department of Agriculture and sheep producers in the Northern Cape between 1933 and 1946. Normal production areas are arid to semi-arid grassveld and savannah grassveld. Dorpers are found across SA, but the majority occur in the more arid areas of the country such as the Karoo and Kalahari (Cloete et al., 2016).

Meatmaster: The Meatmaster is a composite non-fat-tailed meat breed that was developed locally (Peters et al., 2010)**.** The recommended composition of the Meatmaster is 50% Damara, with a Dorper component, and differing components of Ile de France, Van Rooy, SA Mutton Merino, Dormer, Wiltshire Horn and other breeds**.** This breed is predominantly found in the central and western extensive sheep breeding areas of South Africa (Peters et al., 2010).

Dual Purpose breeds

Dual-purpose sheep are bred for both wool and meat production. Studies have shown that production systems with dual purpose breeds may produce a larger income than farming solely with meat only breeds, but this is dependent on various factors, such as using a suitable breed to the climate and the farming system used (Louw, 2019). Dual purpose breeds included in this study include Afrino, Döhne merino, Dormer, Merino and SA Mutton Merino.

Afrino: The Afrino is a composite breed, made up of 25% Merino, 25% Ronderib Afrikaner and 50% SA Mutton Merino (Snyman & Herselman, 2005). The breed was developed at the Carnarvon Experimental Station in the North-western Karoo between 1969 and 1976, and was registered as a synthetic breed in 1980. It is a large-framed, white woolled breed that is mainly produced in the semi-

arid North Western Karoo veld, as well as the Southern Free State, Eastern Cape and Northern Cape. The Afrino breed society currently has 22 breeders (Schoeman et al., 2010).

Döhne merino: The Döhne is a locally developed white woolled composite breed, which has a medium to large frame. It was developed through crosses between local Merino ewes and SA Mutton Merino rams at the Döhne Agricultural Research Station near Stutterheim (Kruger, 2009). The breed was established in the 1950s as a dual-purpose fine-woolled Merino-type. The bulk of the Döhne Merino sheep in South Africa are produced in the Eastern Cape, Free State and Western Cape provinces (Schoeman et al., 2010).

Dormer: The Dormer is a locally developed white woolled composite breed, and was produced through crosses between Dorset Horn rams and German Merino ewes. The breed was developed at the Elsenburg Research Station of the Department of Agriculture through slaughter lamb experiments performed over a period of more than 10 years, starting in 1927 (Van Wyk et al., 2009). The breed is adapted to winter rainfall areas in temperate climates. It was developed to be suited to the climate and grazing environment of the Western Cape. It is also relatively widely distributed in the Free State and Gauteng Provinces. A Dormer Breed Society was established in 1965 and the breed was recognised as a developing breed in 1970 (Cloete et al., 2014).

Merino: The Merino is a medium framed white wool sheep. The breed was introduced to SA in 1789 after the donation of two Spanish Merino rams and four Spanish Merino ewes by the Dutch Government. From 1891 onwards, the American Vermont, the Australian Wanganella and Peppin Merinos were introduced to SA(Vink, 2009). The SA Merino was a result of selecting for adaptation and functional traits over a period of 200 years. The SA Merino is found across South Africa, including the drier Northern Cape Province, in the winter rainfall regions of the Western Cape, as well as in the Karooveld and Grassveld areas of the Eastern Cape and Free State. Merino breeders are also found in the East Griqualand of KwaZulu-Natal and parts of Mpumalanga (Cloete & Olivier, 2012).

SA Mutton Merino: The SA Mutton Merino is a locally developed dual purpose meat and white woolled breed. Originally, it was known as the German Mutton Merino, as the first sheep were imported from Germany in 1932 by the South African Department of Agriculture for use in a breeding program at Elsenburg (Cloete et al., 2007). The SA Mutton Merino is not a true land race; however, directional selection over time improved wool quality and functional efficiency. The breed was recognized as unique in 1917 and named the SA Mutton Merino (Cloete & Olivier, 2012). This breed is produced in areas ranging from semi-arid regions to high rainfall sour grass veld regions. Most of the SA Mutton Merino producers are found in the Free State, followed by the Western Cape Province, Mpumalanga and North-West province (Cloete & Olivier, 2012).

2.4 History of sheep recording and improvement in SA

Selection and pedigree recording of sheep in SA started in 1904 through the founding of the South African Studbook Association, after which the then Department of Agriculture and Technical Services instituted performance recording in 1956 (Schoeman et al., 2010). Although fleece testing was practised from 1934, the formal fleece testing facility was only established in 1965 at the Grootfontein college of Agriculture at Middelburg in the Northern Cape using funds from the wool industry (Schoeman et al., 2010).

Sheep breeding in SA went through several changes as a result of developments in technology and improved statistical procedures. Initially research was conducted by several institutions on designated genetic resource flocks. These resource flocks were managed as stud flocks according to stud breeding criteria implemented during this time (Cloete et al., 2014). Within these flocks, the three foremost production traits, namely growth, reproduction and wool traits, were continuously selected for and worthwhile genetic progress was reported. This illustrated that these traits exhibited genetic variation which can be transferred to future generations (Schoeman et al., 2010). The successful genetic improvement achieved in these flocks laid the foundations for the implementation of the National Small Stock Improvement Scheme (NSSIS). A summary of the most important resource flocks, the breeds they focused on, selection objectives, time lines and locations are provided in Table 2.1.

The NSSIS was first implemented in 1964. Several commercial farmers contributed to data collection, but research was unsynchronised, and thus uncoordinated (Cloete & Olivier, 2012). Improvements in data capturing technology, genetic evaluation, and digital communications systems enabled easier recording of data from across SA on central databases, allowing optimization of the NSSIS in the 1990's for both woolled and meat sheep (Schoeman et al., 2010). The aims of the NSSIS were described as improving income per animal by increasing reproduction, decreasing the length of the production cycle and by optimizing fibre traits in woolled sheep (Cloete & Olivier, 2012). Sheep farmers participate in the NSSIS to record economically important properties of their animals, in order to use the information obtained to improve the productivity of their flocks.

Table 2.1 Summary of South African resource flocks used in various selection experiments

Research flock	Breed	Location	Time span	Selection objective	Selection lines	References
Klerefontein Merino flock	Merino	Carnarvon	1962-1983	Increasing fleece weight and improving conformation	Control line Fleece weight line Visual appraisal line	Snyman et al. (1996)
Klerefontein Namaqua flock	Namaqua Afrikaner	Carnarvon	1982-present	Conservation of indigenous fat-tailed breed	Live weight and reproduction traits recorded	Snyman et al. (1993)
Klerefontein Dorper flock	Dorper	Carnarvon	1993-2000	Comparison of lines within the Dorper breed	Hairy type Woolly type	Snyman & Olivier (2002)
Koopmansfontein flock	Dorper	Jan Kempdorp	1966-present	Selection for growth under different scenarios	Weaning weight Weaning weight in ewes, post-weaning feedlot gain in rams Subjective selection	Neser et al. (1995)
Tygerhoek flock	Merino	Riviersonderend	1969-present	Increasing fleece weight without changing fibre diameter	Control line Clean fleece weight line	Cloete & Scholtz (1998)

Upington flock	Karakul	Upington	1970-present	Improving pelt quality	Control line Hair length line Pattern line Hair quality line Curl development line	Greeff et al. (1993); Schoeman (1998)
Tygerhoek finewool flock	Merino	Riversonderend	1998-present	A reduced fibre diameter	-	Cloete et al. (2001)
Elsenburg flock	Merino	Stellenbosch	1986-present	Divergent selection for reproduction (number of lambs weaned)	H line (selected for) L line (selected against)	Cloete et al. (2004)

All recorded information is kept on the INTERGIS (Integrated registration and genetic information system). This computer system was established in participation with the Department of Agriculture, Fisheries and Forestry. Its goal is to integrate the pedigree and performance data of the animals, and it serves as the national databank for animal recording and improvement (DAFF, 2007). It is available to farmers and other interest groups. Breed associations make use of service providers for the genetic analysis of their herds. Other objectives of the NSSIS can broadly be described as supplying breeders with objective performance information, to provide information for research and refinement of breeding objectives, and to provide a valuable instrument for increasing human capacity (Schoeman et al., 2010). Table 2.2 provides an overview of the breeds and animals registered with SA Studbook and for the period January to December 2018.

Table 2.2 Number of flocks and animals registered with SA Stud Book in 2018

Breed	Number of Flocks	Number of Animals
Damara	5	223
Dohne Merino	90	274924
Dormer	126	23779
Dorper	62	15366
Hampshire Down	9	511
Ile de France	59	9906
SA Merino	114	98510
Merino Landsheep	20	3631
Meatmaster	76	21654
SA Mutton Merino	135	158199
Suffolk	22	3376
Van Rooy Sheep	13	2699
White Dorper	28	7606

2.5 The development of molecular technology

DNA markers and genomics are an important tool in the sustainable genetic improvement of small stock. The widespread use of genetic markers has become available for farm animals in the early 2000s, allowing the evaluation and detailed analysis of genetic diversity, as well as the detection of genes that influence economically important traits (Blasco & Toro, 2014). The development of tools such as whole genome sequencing and improved statistical analysis of molecular data allow for in-depth investigations into the genetic diversity between and within populations (Cesarani et al., 2019).

Genomics can play an important role in selection for traits that are difficult to select for using traditional methods, such as sex-limited traits or traits only expressed late in life (Blasco & Toro, 2014). Therefore, genotyping and genomic analysis is important for the progression of small stock production in South Africa.

As a result of their potential for higher genotyping efficiency, high data quality and genome wide coverage, as well as their analytical simplicity, SNPs are currently the marker of choice for several studies, including population ecology and conservation, as well as livestock breeding strategies and diversity studies (Grasso et al., 2014). In several domestic species, SNPs replaced microsatellite markers in parentage testing, owing to their higher abundance and lower mutation rates, as well as the ease of standardization between different laboratories (Gudex et al., 2014). SNPs are bi-allelic markers that are abundantly spread throughout the genome, and can be found in both coding and noncoding regions, indicating a specific polymorphism in a population. SNPs that occur in coding regions of the genome can be directly associated with protein function, and their stable inheritance pattern make them suitable markers for selection over time (Beuzen et al., 2000; Stoneking, 2001). However, SNPs have lower informativeness compared to microsatellites, caused by their biallelic nature, therefore more SNPs are required to gain the same amount of information as with microsatellites (Lathrop et al., 2011).

The OvineSNP50 BeadChip was developed in 2008 by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC). The BeadChip contains 54 241 SNPs that uniformly span the entire ovine genome, with an average gap size of 50.9 kb, and a median gap size of 42.5 kb (Illumina, Inc. San Diego, CA 92122 USA). The array includes SNPs validated in many economically important breeds, including more than 75 *Ovis Aries* breeds, generating average call rates exceeding 99.9% with a mean MAF of 0.28.

Furthermore, Illumina and the ISGC also developed the Illumina OvineHD BeadChip (Illumina, Inc. San Diego, CA 92122 USA). This 600K SNP chip includes more than 600 000 genomic variants, and includes nearly all the content from the original OvineSNP50 array. It also includes 30 000 putative functional variants. The higher density of this array may increase the ability to identify key genes that are responsible for desired traits (ISGC, 2010).

An important factor to consider when using SNP data is ascertainment bias. This occurs when the selection of loci is from an unrepresentative sample of individuals, and the loci attained are not representative of the spectrum of allele frequencies in a population (Morin et al., 2004). In the context of SNP studies on South African sheep, the limited inclusion of loci representing indigenous breeds in international SNP chips may introduce systematic bias in estimates of variation within and between populations.

The information obtained from genotyping studies can be applied to several levels of management, such as investigations into population structure and genetic diversity for sustainable breeding programs, and as a management tool to control inbreeding and for parentage verification, ultimately leading to genetic progress.

2.6 Previous genetic diversity studies on sheep

Globally, numerous microsatellite studies have been conducted in order to investigate the genetic diversity of sheep breeds. However, these studies are difficult to compare due to the different numbers of microsatellites used, the different breeds investigated and differences in sample sizes between studies. A short summary of some of the more recent studies are given, in order to provide context on the global trend of sheep diversity. Arranz et al. (1998) investigated the genetic relationships among Spanish sheep using 19 microsatellites. Allele frequencies and mean heterozygosities reported in their study revealed that the greatest genetic variation was found in Merino sheep while the lowest was found in Awassi sheep. Peter et al. (2007) analysed the genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds using 31 microsatellite markers. Their study suggested that South-eastern European and Middle-Eastern sheep breeds had significantly more variation than north-western and western European breeds. Elbeltagy et al. (2015) conducted an investigation into the genetic diversity and structure of Egyptian indigenous sheep populations using 13 autosomal microsatellites. In this study they found that the 408 animals analysed separated into 9 genetic clusters, influenced by topography and domestication patterns. Naqvi et al. (2017) used microsatellite analysis to assess the genetic diversity and structure of major sheep breeds from Pakistan. Results from this study indicated that the genetic clustering of Pakistani sheep did not conform to geographical locations, however it was largely consistent with their purpose- only meat and coarse wool or meat, coarse wool and milk. Generally, high levels of genetic differentiation was observed within the breeds included in these studies, with substantial levels of genetic variation between breeds.

Recently, studies investigating the genetic structure of sheep populations using SNPs have become more common. Han et al. (2016) used SNP data to investigate the genetic structure and relationships among Chinese fine-wool sheep breeds. The study suggested a close genetic relationship between all fine-wool sheep populations, which is consistent with their breeding progress. Cesarani et al. (2019) investigated the genetic diversity and selection signatures between Sarda and Sardinian Ancestral black sheep using SNP data, reporting that the Sarda breed showed higher linkage disequilibrium levels and a smaller effective population size. O'Brien et al. (2020) utilized genome-wide SNP information in order to estimate the population structure and breed composition in a multi-breed sheep population. Results from this study suggested that the Suffolk and Border Leicester

breeds were the most genetically diverse from one another, while the Irish Blackface and Scottish Blackface could genetically be considered the same breed. Overall, these studies showed the same trend as studies performed with microsatellites, indicating substantial genetic variation within breeds.

Molecular studies on sheep in SA are limited. A study on the genetic and phenotypic diversity of Zulu sheep populations was performed by Kunene et al. (2009) using random amplified polymorphic DNA (RAPD) profiles. This study indicated that RAPD analysis could be used to demonstrate genetic diversity between and within Zulu sheep populations. A study was also done by Peters et al. (2010) on the genetic profile of locally developed Meatmasters based on 10 microsatellite loci. This study reported high levels of heterozygosity and allele numbers in the Meatmaster populations, indicating good retention of genetic diversity during the potential bottleneck created through intensive selection during the breed's early history. In 2011 Soma et al. (2012) conducted a study on the population genetic structure of 20 sheep breeds in SA using 12 microsatellite loci. The average unbiased heterozygosity found in this study was lowest in the fat-rumped breeds, while the composite and indigenous fat-tailed breeds had higher average heterozygosity values. Karakul and wool breeds showed even higher values of heterozygosity. Qwabe et al. (2012) investigated the genetic diversity and population structure of Namakwa Afrikaner sheep using 22 microsatellite markers. Results from this study showed limited hybridization between the Namaqua Afrikaner sheep, and Pedi and SA Mutton Merino breeds.

Sandenbergh et al. (2016) performed an investigation aimed at evaluating the use of the OvineSNP50 chip for use in four South African sheep breeds. The investigation showed that the Namaqua Afrikaner breed exhibited the lowest genetic diversity, as well as the highest inbreeding coefficient among the breeds investigated. Furthermore, it showed that the results obtained for the Dorper, SA Merino and SA Mutton Merino were comparable to international breeds, thus illustrating the utility of the OvineSNP50 chip for these breeds. Greyvenstein et al. (2016) conducted an investigation on Damara sheep in order to identify genomic regions for horn development. They performed a genome-wide association study for horn number with 43 Damara sheep, genotyped with 606 006 SNP markers. The analysis indicated a region with multiple significant SNPs on ovine chromosome 2, in a location different from the mutation for polledness in sheep on chromosome (OAR) 10. Molotsi et al. (2017) did a study on the genetic diversity and population structure of South African smallholder farmer sheep breeds using SNP data. Their study showed that for pure breeds, the average inbreeding coefficient was higher compared to the smallholder Dorper population. Dlamini et al. (2019) investigated the resistance of a Döhne Merino flock to *Haemonchus contortus* infection, using genome wide SNP data. The results gained from this study indicated that there is genetic

variation in host resistance against *H. contortus* present in the investigated Dohne Merino flock, and therefore breeding for resistance against nematodes in this population would be feasible.

2.7 Statistical parameters for the investigation of genetic diversity in sheep populations

Molecular data has proven to be a useful tool in investigating genetic diversity as well as differentiation among species and breeds (Sanarana et al., 2015). Investigation and documentation of diversity within and among breeds are an important tool for sustainable management practises, conservation programs as well as the development of breeding strategies (Bravo et al., 2019). Sheep can be seen as a good model species for investigating phenotypic variation in terms of resilience, adaptation to environment as well as resistance to diseases (Dominik et al., 2012). Loss of diversity can be caused by genetic drift, population bottlenecks, increased inbreeding rates as well as the loss of founder alleles through selection (Vozzi et al., 2007). Several genomic parameters are used to quantify levels of genetic diversity. A few of these will be discussed in more detail.

2.7.1 Heterozygosity

Population parameters such as the mean expected heterozygosity (He) and mean observed heterozygosity (Ho) can be used to investigate the genetic diversity within a population. He can be defined as the level of heterozygosity expected in a population under Hardy-Weinberg equilibrium (HWE), thus it is the heterozygosity when there are no evolutionary factors, such as mutation or migration, influencing the genetic variation in a population (Peter et al., 2007). Ho can be described as the level of heterozygosity actually observed within the population. If the observed heterozygosity is lower than the expected heterozygosity, it may be attributed to processes such as inbreeding or a population bottleneck. Furthermore, if the observed heterozygosity is higher than the expected heterozygosity, it might be attributed to admixture, which is the interbreeding of populations or breeds that were previously isolated (Bourdon, 2000)

2.7.2 Linkage disequilibrium

Genetic diversity can also be described by linkage disequilibrium (LD) estimates. LD is generated in a population through processes such as mutation, selection, drift, non-random mating and admixture (Espigolan et al., 2013). LD occurs when there is an absence of linkage equilibrium, resulting in a correlation between an allele at one locus and the allele at a different locus, thus it is the non-random association of alleles at different loci in a given population (Goddard & Hayes, 2009). The level of LD in a population can be measured using adjacent pairwise analysis, where LD analysis is performed on all adjacent pairs within a chromosome. The results are then measured as r^2 or $D^{'}$. The $D^{'}$ value indicates the occurrence of recombination between two loci, while the r^2 value indicates the correlation between two loci (Espigolan et al., 2013). Generally, the r^2 parameter is considered to be a more reliable estimation of LD, as the $D^{'}$ parameter tends to be inflated when used with small

populations, or low haplotype frequencies (Lee et al., 2011). The r^2 is the square of the correlation coefficient obtained between two indicator variables, with one indicator variable representing the presence or absence of a particular allele at the first locus, and the other representing the presence or absence of a particular allele at the second locus (García-Gámez et al., 2012).

2.8 Methods for the estimation of population structure in sheep populations

Investigating the population structure of sheep breeds is important, as it is necessary to know the genetic structure of a population in order to provide sound breeding advice to breeders and farmers (Molotsi et al., 2017). Knowledge of the population structure allows the design of optimal crossbreeding strategies in order to fully utilize the exploitation of inter-breed nonadditive genetic effects (O'brien et al., 2020), as well as limiting inbreeding and its adverse effects. Genomic data can be used to quantify and characterize population structure. A few parameters used to investigate the genetic structure of populations are discussed below.

2.8.1 Fixation index

Wrights fixation index (Fst) is a measure of population differentiation due to genetic structure(Weir & Hill, 2002). Fst can provide useful information on the evolutionary processes that influence the genetic variation within and between populations. Fst estimates can be used to identify genomic regions that have been the target of selection, and comparisons of Fst from different parts of the genome can provide insights into the demographic history of populations (Holsinger & Weir, 2015). Fst is directly related to allele frequency variation among populations and, thus to the degree of resemblance among individuals within populations. Fst ranges of 0 to 0.05 indicate low genetic variation, 0.05 to 0.15 indicate moderate genetic variation and 0.15 to 0.25 indicate large genetic variation (Frkonja et al., 2012).

2.8.2 Principal component analysis

The genetic structure of a population can be investigated by using principal component analysis (PCA). PCA is a multivariate technique of analysing a data table wherein observations are described by several inter-correlated quantitative dependent variables (Anderson et al., 2010). The goal of a PCA is to extract important information from the data table, in order to represent it as a set of new orthogonal variables called principal components, and to then display the pattern of similarity of the observations and of the variables as points on maps (Patterson et al., 2006). Thus, PCA assigns individuals to their population of origin using a common clustering algorithm.

2.8.3 Admixture

Population stratification is seen as a confounding factor in genetic association studies. It can be defined as the occurrence of a systematic difference in allele frequencies between subpopulations of a population, possibly caused by a difference in ancestry (Frkonja et al., 2012). Population

stratification is caused by non-random mating between populations, usually because of physical separation, and is followed by genetic drift of allele frequencies in each group. Therefore, determining the ancestry of individuals within a study is important to limit biased results. ADMIXTURE software, developed by Alexander et al. (2009) allows researchers to make a distinction between local and global ancestry. ADMIXTURE simultaneously estimates the allele frequencies of a population, along with ancestry proportions. A maximum likelihood approach is used, which can accommodate a large number of markers. The software assumes a dataset that consists of a large number of genotypes, with several SNPS from a large number of unrelated individuals, and that these individuals are drawn from an admixed population, contributed by 'K' number of postulated ancestral populations. The optimal K value is the value assumed based on the lowest cross-validation error estimate, and results are visualised through the generation of a bar plot, indicating the inferred ancestral population of each breed (Alexander et al., 2009).

2.9 Statistical parameters for the estimation of inbreeding in sheep populations

Inbreeding is a serious concern for animal breeders, because of its deleterious effects on the additive genetic variance and phenotypic values of populations. These negative effects of inbreeding are referred to as inbreeding depression (Gholizadeh & Ghafouri-Kesbi, 2016). Inbreeding may also increase the emergence of disorders caused by recessive gene action, leading to increased occurrences of defects such as spider-lamb syndrome (Selvaggi et al., 2010). The effect of inbreeding in sheep populations are often unfavourable, and has been shown to have a negative effect on several traits, such as milk production (Dario & Bufano, 2003), and reproductive efficiency (Selvaggi et al., 2010). The impact of inbreeding and inbreeding depression is related to a lack of genetic variation, resulting in the inability of populations to adapt to changes in environments, causing them to be more susceptible to challenges (Barros et al., 2017).

Evidence suggests that diverse breeds, populations as well as traits respond differently to inbreeding, and certain populations may show a pronounced change in a trait as a result of inbreeding, while other populations may not (Analla et al., 2002). Therefore, the degree to which populations are negatively affected by inbreeding depression can vary extensively, depending on the traits being examined, the lineage effect, population history as well as the environment (Selvaggi et al., 2010). The method of selection practiced in livestock breeding where animals are selected by truncation, based on EBVs across age classes often results in increased genetic gain, however this may also lead to increased rates of inbreeding. Inbreeding frequently leads to impaired response to selection, and decreased genetic diversity, and is therefore detrimental to selection strategies in the long term (Van Wyk et al., 2009). It is therefore important to estimate and control levels of inbreeding in livestock populations.

Before the widespread availability of dense SNP data, pedigree-based inbreeding estimates were used to characterise and control the genetic diversity of populations. However, this approach had several limitations, such as the need for complete, high quality pedigree data and records, and the fact that pedigree information does not take Mendelian sample variation, or linkage disequilibrium caused by selection, into account (Rodríguez-Ramilo et al., 2019). The increased availability of dense SNP data allows for more accurate estimations of inbreeding and information from genotypes allows more accurate genomic estimates of realized inbreeding coefficients (Vostry et al., 2018). Inbreeding levels within a population can be estimated through several methods, some of which will be discussed in more detail.

2.9.1 Inbreeding coefficient

One method of inbreeding estimation is through individual inbreeding coefficients (Fis). This can be defined as the probability that at any given locus, two alleles are identical by descent (IBD) and the probable proportion of an individual's loci containing genes that are IBD (Falconer & Mackay, 1996). The inbreeding coefficient describes deviations from Hardy-Weinberg. Therefore, genomic inbreeding (FG) for genotyped animal i, can be obtained through the formula $FG_i = G_{ii} - 1$, where G represents the genomic relationship matrix, and therefore values of FG < 0 can be interpreted as an individual that is more heterozygous than the average of the population. If base allele frequencies are used in the computation of G, FG is an estimate of inbreeding relative to the base population (VanRaden, 2008).

2.9.2 Runs of homozygosity

The introduction of next-generation sequencing and high throughput genotyping techniques has allowed the identification of continuous homozygous stretches of sequence, which are known as runs of homozygosity (ROH). ROH can be defined as genomic regions which display a series of consecutive homozygous genotypes (Broman & Weber, 1999). The characteristics of ROH, in terms of length, frequency, genomic distribution and abundance are varied depending on local recombination rate, guanine-cytosine content, positive selection and demography (Szpiech et al., 2013). ROH are generally classified into length classes; 1–3 Mb, 3–5 Mb, 5–10 Mb, 10–15 Mb, 15–20 Mb, 20–25 Mb, and > 30 Mb, where lengths of 0-5 Mb are seen as short runs, 5-15 Mb is seen as intermediate runs and more than 15 Mb is classified as long runs. The estimation of inbreeding from ROH (F_{ROH}) is considered to be the most efficient method of estimating inbreeding effects among a number of alternative methods (Keller et al., 2011), because of its ability to estimate the number of generations of inbreeding, as well as the history of recent selection events through the extent and frequency of ROH segments (Purfield et al., 2012).

The genomic inbreeding coefficient F_{ROH} can be estimated using the following formula:

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

where: L_{ROH} = the length of ROH in one individual

L_{AUTO} = the length of the genome covered by SNPs, excluding the centromeres

As recombination events interrupt chromosome segments, long ROH is often a result of recent inbreeding (up to five generations ago), which may be caused by population decline, unbalanced paternal contributions as well as selection, whereas a high frequency of short ROH (up to 50 generations ago) often reflect ancient inbreeding, due to ancestral family relatedness, and can be explained by the occurrence of an ancient founder effect or by a population bottleneck (Kirin et al., 2010; Mastrangelo et al., 2016)

2.9.3 Effective population size

The estimation of the effective population size (Ne) of a breed is an important indicator of genetic variation and inbreeding levels. Ne can be defined as the number of individuals in an idealized population, where no random mating and no selection takes place, that would cause the same rate of inbreeding that is observed in the real population (Gholizadeh & Ghafouri-Kesbi, 2016). Ne can therefore be much smaller than the actual size of the population, owing to the unequal contribution of genes from individuals to the subsequent generation (Goddard & Hayes, 2009). This population parameter aids in explaining the evolution of populations, and can also assist in the understanding and modelling of the underlying genetic architecture of complex traits (Hayes et al., 2003).

2.10 Methodology for the investigation of phylogenetic relationships between sheep populations

The differing populations within a species result from a complex demographic history, involving population splits, gene flow, and changes in population size (Pickrell & Pritchard, 2012). A phylogenetic tree can be defined as a graph used to illustrate possible evolutionary relationships between different individuals, organisms or entities (Waikagul et al., 2014). This can be a useful tool in researching the evolutionary progression of a species or breed. The construction of a phylogenetic tree in order to classify a domestic livestock population is an important base for determining the history and extent of certain breeds. It is also an effective method of estimating the potential distribution of special gene resources within a population (Sun et al., 2007).

There are several programs available that can construct phylogenetic trees using SNP data. One such program is the SPLITSTREE 4 software. This software uses either an alignment of sequences,

a distance matrix or a set of trees in order to compute a phylogenetic tree or network. It implements methods such as split decomposition, neighbour-net, consensus network, super networks methods or methods for computing hybridization or simple recombination networks (Huson & Bryant, 2006).

Another program that can be used is the TreeMix software (Pickrell & Pritchard, 2012). TreeMix utilizes large numbers of SNPs to estimate the historical relationships among populations, using a graph representation that allows both population splits and migration events to be visualized (Pickrell & Pritchard, 2012). Phylogenetic trees can also be computed based on the breed pair-wise Fst values using the APE package available in R software. This software allows the estimation of phylogenetic trees with distance-based DNA information in order to facilitate comparative and diversification analyses (Paradis et al., 2004).

2.11 Conclusion

Sheep production plays an important role in the livestock sector of South Africa, as well as forming an integral part of the sustainability of smallholder systems. Investigating the genetic diversity, population structure and inbreeding of the various sheep populations in South Africa is necessary in order to successfully manage and improve small stock genetic resources across the country.

Chapter 3: Materials and methods

3.1 Introduction

The aim of this study was to investigate the genetic structure of both indigenous and commercial sheep breeds in South Africa using genome-wide SNP data. Thirteen South African sheep breeds were included, consisting of a total of 1977 genotyped animals. The genotypes were analysed to quantify the genetic variation within and between breeds, as well as inbreeding levels and population structure. All genotypic data was provided by and used with approval from two institutes, the Grootfontein Agricultural Development Institute (GADI), and the Western Cape Agricultural Trust (WC). Ethics approval (NAS394/2019) for the use of all data in this study was obtained from the Ethics Committee of the Faculty of Natural and Agricultural Sciences at the University of Pretoria.

3.2 Materials

Genotypes from 13 South African sheep breeds, including dual purpose, meat and indigenous types, were included in this study. A total of 1977 animals were used, with 828 genotypes received from GADI and 1149 animals from WC. The animals used originated from several locations across South Africa, shown in Figure 3.1. Animals were previously genotyped as part of individual research projects, not as part of a national genotyping strategy. This explains the wide range of samples available per population. All animals were genotyped at the Agricultural Research Council, Biotechnology Platform (ARC-BTP), using the Illumina® Ovine 50K SNP BeadChip which contains over 54 000 SNPs distributed over the 27 autosomes and sex chromosomes. SNP-calling was done using the Illumina® Genome Studio software v2.0 (Illumina, San Diego, California 92122 U.S.A). The resulting genotype input files were converted into PLINK (ped and map files) input files using a plug-in in Genome Studio software v2.0. A summary of the number of animals used per breed and which institute provided the data, is given in Table 3.1.

3.3 Methods

Quality control was performed and within-breed diversity parameters were calculated for each population separately, based on all the individuals genotyped within that specific population. As the data was unbalanced in terms of numbers per population, each population was investigated individually for within-population analysis. For between population analyses, a representative sample consisting of a maximum of 60 animals from larger sample sizes was taken and merged to form one dataset.
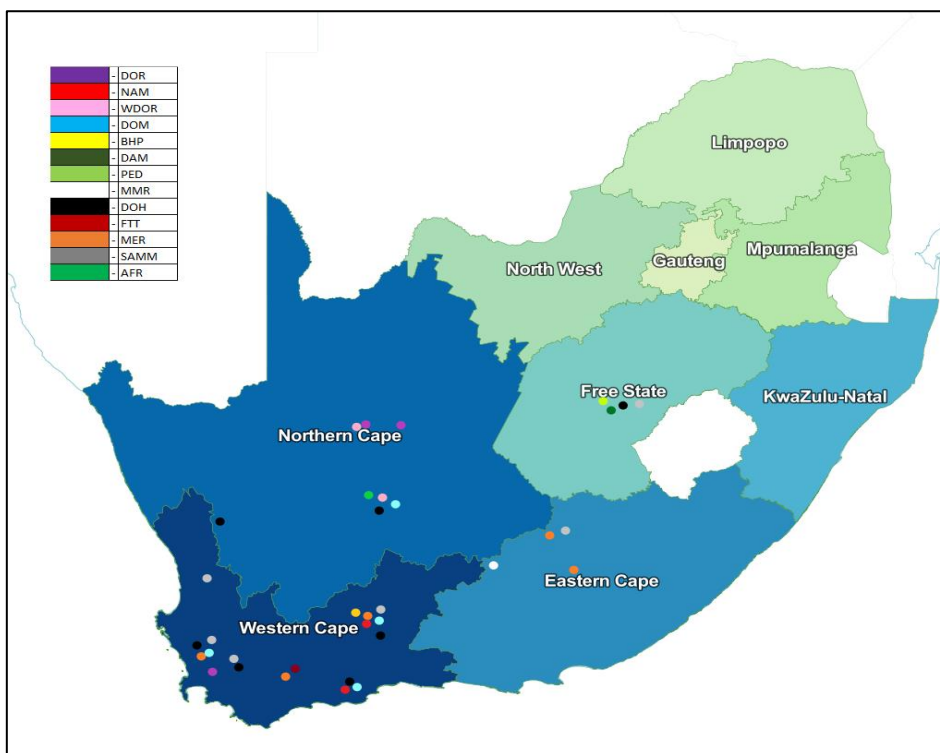
Figure 3.1 Map of SA illustrating the various locations the animals used in this study originated (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)

Table 3.1 Summary of genotyped animals used in this study

Type	Population	GADI	WC	Total
Dual purpose	AFR	179	0	179
	DMER	290	73	363
	DOM	0	44	44
	MER	304	672	981
	SAMM	0	75	75
Meat	DOR	7	60	67
	WDOR	0	28	28
	BHP	0	30	30
	DAM	0	30	30
	MMR	48	39	87
Indigenous	NAM	0	53	53
	PED	0	29	29
	FTT	0	16	16
Total		**828**	**1149**	**1977**

AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper, GADI= Grootfontein Agricultural Development Institute, WC= Western Cape Agricultural Trust

3.3.1 Quality control

Quality control (QC) was performed on the datasets per population using PLINK software (Purcell et al., 2007). Sample- and marker-based quality control were performed, in order to filter both non-informative SNPs and individuals from the dataset. Animals were removed according to missing genotype rates (using --mind), while SNPs were removed based on call rate (using --geno), MAF (using --maf) and HWE (using --hwe). The standard threshold levels used were mind 0.05, geno 0.05, maf 0.02 and hwe 0.001.

3.3.2 Within-population diversity

After QC procedures, marker-based summary statistics indicating genetic diversity were estimated per population using PLINK (Purcell et al., 2007). These included mean expected and observed heterozygosity (He and Ho), MAF and LD estimates.

He and Ho estimates were produced by using the --het command, importing the resulting .het file to Microsoft Excel (2013), and calculating the averages for expected as well as observed heterozygosity using the following formulas:

$$Ho = \frac{(N(NM) - O(Hom))}{N(NM)}$$

$$He = \frac{(N(NM) - E(Hom))}{N(NM)}$$

where: N(NM) = the number of non-missing genotypes

O(Hom) = the observed number of homozygous genotypes

E(Hom) = the expected number of homozygous genotypes

The MAF values were calculated using the –freq command, importing the resulting .frq file to excel and calculating the average MAF. The proportion of MAF per loci was then visualised as a bar graph using Microsoft Excel. LD, measured in r^2, was subsequently estimated for each population using PLINK's $–r^2$ command. The r^2 estimates were calculated using Microsoft Excel for all autosomal SNPs that passed quality control.

3.3.3 Between population comparison

A representative sample of 60 animals was selected from all the populations with larger sample sizes. This was done by selecting animals from each sample population, ensuring animals from each location, year and breeder was included. In order to maintain the highest possible number of animals, this was done after animals with high missing genotype rates (<98%), were removed. After sampling, the animal numbers per population ranged from 13 to 60. The final merged sample

consisted of 565 animals. The animal numbers remaining per population after selection for merged dataset is presented in Table 3.2. All 13 population datasets were then merged using the --merge command in PLINK, after which QC was performed again on the merged data set, at thresholds of: SNP call rate <98%, MAF <5%, and HWE <0.001, after which 50790 SNPS and 565 animals remained for downstream analysis.

Table 3.2 Animal numbers per population remaining after selection for merged dataset

Population	Animals
AFR	60
BHP	13
DAM	30
DMER	60
DOM	40
DOR	59
FTT	16
MMR	60
MER	60
NAM	51
PED	29
SAMM	60
WDOR	27
Total	**565**

AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper

3.3.4 Population structure

SNP-based genetic relatedness between individuals was calculated using GCTA version 1.24 (Genome-wide Complex Trait Analysis) (Yang et al., 2011). A genetic relationship matrix was created using the command –make-grm, followed by the estimation of eigenvalues and eigenvectors for the first three principal components using the command –pca 3. Microsoft excel was used to visualise PCA plots by importing the .eigenvectors file to excel and constructing scatterplots of the principal components (values for principal component one on x-axis and principal component two on y-axis). The package SNPrelate (Zheng et al., 2012), in R was used to estimate eigenvectors and eigenvalues for the first three principal components, after which R was used to plot the PCA plots.

ADMIXTURE 1.23 software (Alexander et al., 2009) was used to determine the genetic population structure of the animals through the maximum likelihood estimation of ancestry. The appropriate K-value for plot visualisation was determined based on the lowest cross-validation error estimate by adding the –cv command when running ADMIXTURE for 14 K-values (2-15), as the data set consisted of 13 breeds. When running ADMIXTURE for each K-value, a .Q as well as a .P file was generated as output. Thereafter bar plots were generated to indicate the inferred ancestral population of each breed, by using Genesis version 0.2.3 software (Buchmann & Hazelhurst, 2014).

3.3.5 Inbreeding estimation

Inbreeding in the various populations was investigated using PLINK (Purcell et al., 2007) to calculate average individual inbreeding coefficients (Fis), through the command –het. This produced a .het output file, which was imported to excel, where the average F value was calculated per population based on the heterozygosities.

ROH was also used to estimate inbreeding. The command –homozyg was used in PLINK to generate a –hom output file, which was then imported into excel. No pruning was performed based on LD, however ROH detection parameters were used to exclude short and common ROHs that derived from LD. These parameters are specified in Table 3.2. In excel the amount of ROHs per predefined length segments of 14 Mb, 4-8 Mb, 8-12 Mb, 12-16 Mb and >16 Mb was calculated. The number of ROH found per chromosome was also estimated for each population type and visualised using excel.

The genomic inbreeding coefficient FROH was estimated using the following formula:

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

where: L_{ROH} = the length of ROH in one individual

L_{AUTO} = the length of the genome covered by SNPs, excluding the centromeres

The effective population size (Ne) of each breed was calculated using the SNeP software tool (Barbato et al., 2015). Microsoft Excel was used to construct a line graph to visualise the N_e of each population.

Table 3.3 ROH detection parameters and code

Parameter	Code	Parameter used
Number of heterozygotes allowed	--homozyg-window-het	1
Number of missing calls allowed	--homozyg-window-missing	3
Window threshold to call a ROH	--homozyg-window-threshold	0.05
SNP threshold to call a ROH	--homozyg-snp	50
Sliding window size in SNPs	--homozyg-window-snp	1000
Allelic matching	--homozyg-match	0.98
Allowed distance between SNPs	--homozyg-gap	1000 kb
Overlapping ROH	--homozyg-group	-

3.3.6 Phylogenetic relationships

The pair-wise genetic differentiation among breeds (Fst) was calculated using ADMIXTURE 1.23 software (Alexander et al., 2009). Fst values can range from 0 to 1, where low Fst values among subpopulations indicate a low level of genetic divergence in the population, whereas a value of 0 indicates that there is no subdivision between the populations. The Fst was calculated using the following formula:

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

where: s^2 = the variance of allele frequency among populations

$\bar{p}$ = the mean allele frequency across population

The estimated pairwise Fst values were used to determine the genetic distance between populations in order to construct a phylogenetic tree. The phylogenetic tree was created and visualized using the APE package in R software (Paradis et al., 2004).

Chapter 4: Results

4.1 Introduction

The following chapter presents the results that were obtained from performing the analyses described in Chapter 3. Whole genome SNP data from 1977 animals representing 13 sheep populations in SA were analysed in order to evaluate their genetic and population structure, in terms of diversity as well as ancestral differentiation. The phylogenetic relationships between individuals were also investigated.

4.2 Quality control

Quality control was performed in order to remove uninformative animals and SNPS, as well as SNPS with low MAF to ensure the maximum number of informative markers for downstream analyses. Quality control was performed at varying thresholds per population, to maintain a high number of animals and SNPs for each population. In most populations, a sample call rate of <95% was applied. Furthermore, a minimum SNP call rate of 95%, a MAF threshold of <2% and a violation of HWE of P<0.001 were applicable to most populations. Table 4.1 provides a summary of animal-based quality control thresholds and numbers of animal removed, per population.

Table 4.1 Summary of animal-based quality control, indicating the number of animals before QC, animals removed as well as animals remaining

Population	N Before QC	Sample call rate	Removed	Remaining
AFR	179	90%	2	177
BHP	30	85%	17	13
DAM	30	95%	0	30
DMER	363	90%	6	357
DOM	44	95%	4	40
DOR	67	95%	8	59
FTT	16	95%	0	16
MMR	87	85%	3	84
MER	976	93%	163	813
NAM	53	95%	2	51
PED	29	95%	0	29
SAMM	75	95%	1	74
WDOR	28	95%	1	27
Total	**1977**	**-**	**207**	**1770**

AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper

Following animal-based quality control, marker-based quality control was performed per population on the remaining animals. Table 4.2 provides a summary of marker-based quality control thresholds as well as the number of markers removed.

Table 4.2 Summary of marker-based quality control results indicating the number of SNPs removed for each parameter

Population	Initial SNPs	SNP call rate (<95%)	MAF (<2%)	HWE P<0.001	Total removed	SNPS remaining
AFR	52047	5232	2813	17	8062	43985
BHP*	52047	8484	11103	38	19625	32422
DAM	52047	1885	9686	58	11629	40418
DMER	52047	5570	1701	480	7751	44296
DOM	52047	3429	5732	929	7896	44151
DOR	52047	2583	4811	108	7502	44545
FTT	52047	2548	7142	109	9799	42248
MMR	52047	6492	3122	449	10063	41984
MER	51728	4180	2030	4128	10338	41390
NAM	52047	1831	12104	309	8460	43587
PED	52047	1551	6822	87	8460	43587
SAMM	52047	2947	4180	142	7269	44778
WDOR	52047	3409	6501	50	9960	42087

*SNP call rate <80%. AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper

Within the indigenous populations, a higher number of SNPs were discarded as a result of low MAF, while within the Merino-type populations a relatively higher number of SNPs were lost due to HWE violation. The highest number of SNPS discarded due to MAF was seen in the BHP and NAM populations, while the least SNPs discarded due to MAF was seen in AFR, DMER and MER.

4.3 Within population diversity

The observed and expected heterozygosity levels per population were estimated, and is reported in Table 4.3, together with the average MAF and the average Fis per population, as well as the r^2 for each population.

Table 4.3 Summary of average observed and expected heterozygosities, MAF and linkage disequilibrium per population

Population	Ho	He	MAF	r^2
AFR	0.3603	0.3649	0.2772	0.4067
BHP	0.3522	0.3270	0.2431	0.4992
DAM	0.3371	0.3328	0.2471	0.4192
DMER	0.3710	0.3731	0.2860	0.3650
DOM	0.3537	0.3509	0.2652	0.4319
DOR	0.3525	0.3539	0.2676	0.4036
FTT	0.3260	0.3346	0.2502	0.4512
MMR	0.3678	0.3568	0.2702	0.3829
MER	0.3646	0.3720	0.2853	0.3624
NAM	0.3367	0.3442	0.2578	0.4828
PED	0.3558	0.3426	0.2560	0.3970
SAMM	0.3520	0.3553	0.2694	0.3973
WDOR	0.3631	0.3535	0.2667	0.4291

Ho= Observed homozygosity, He= Expected homozygosity, MAF= Minor Allele Frequency, AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper

The observed heterozygosity levels for all populations were relatively high, with levels ranging from 0.3260 (FTT) to 0.3710 (DMER). The average MAF was lower for indigenous and meat type populations, ranging from 0.2431 for BHP, to 0.2667 for WDOR; while the dual-purpose merino-type populations had higher MAF values, ranging from 0.2652 for SAMM, to 0.2860 for DMER. Generally, lower r^2 values were obtained for Merino-type populations while indigenous and meat type populations had higher r^2 values. The lowest r^2 calculated was for MER at 0.3624, while the highest r^2 was observed for BHP at 0.4992.

The distribution of Minor allele frequencies (MAF) was calculated for all 13 sheep populations respectively. This was done in order to observe the distribution of the SNPs within differing MAF intervals for each population. The MAF frequency for the various populations included in the study are illustrated in Figure 4.1. This represents the proportion of loci partitioned into MAF intervals of 10% ranging from 0% to 50% for each individual population.

Figure 4.1 illustrates that for Merino-type populations, the highest percentage of SNPs have MAF above 20%, while for indigenous and meat type populations the highest percentage of SNPs have MAF lower than 20%. The highest percentage of SNPs with a MAF of 0.5 was observed in the BHP population, although it was a relatively small proportion of loci (less than 5%).
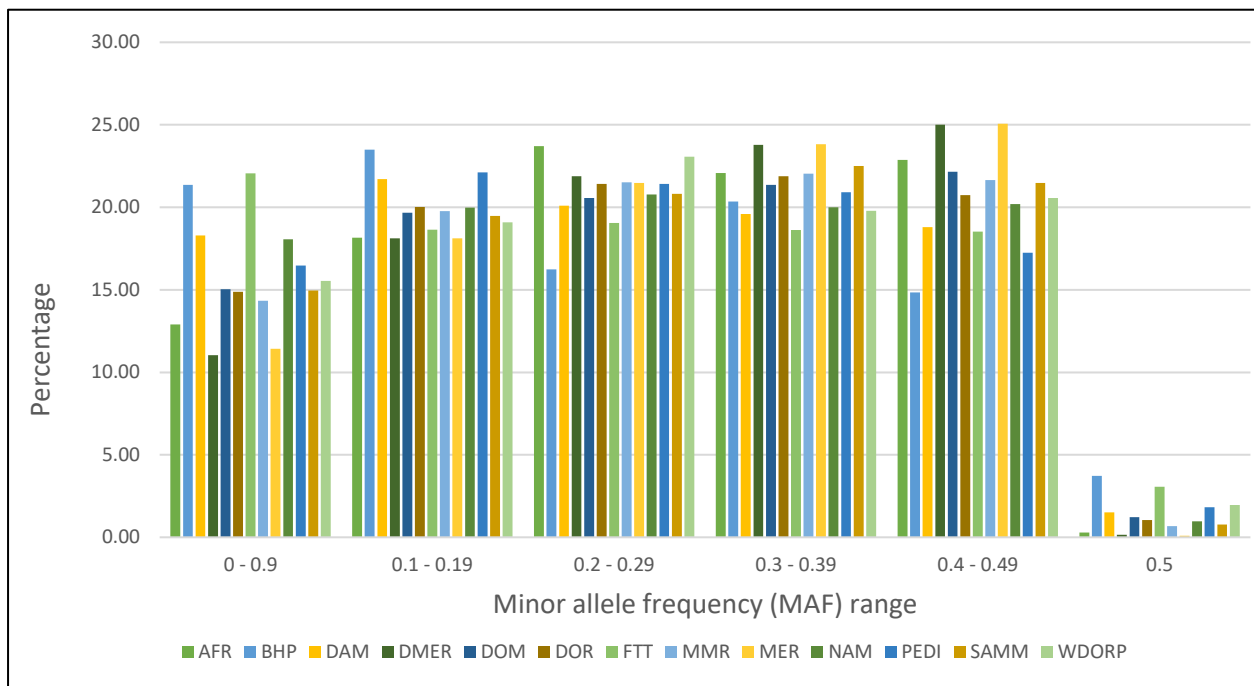
Figure 4.1 Proportion of loci for different MAF ranges between different populations (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)

4.4 Population structure

4.4.1 Principal component analysis

The genetic relatedness between the individual animals of the 13 different populations were investigated by principal component analysis (PCA). The first and second principal components were plotted against each other in Figure 4.2a, and the first and third principal components in Figure 4.2b.
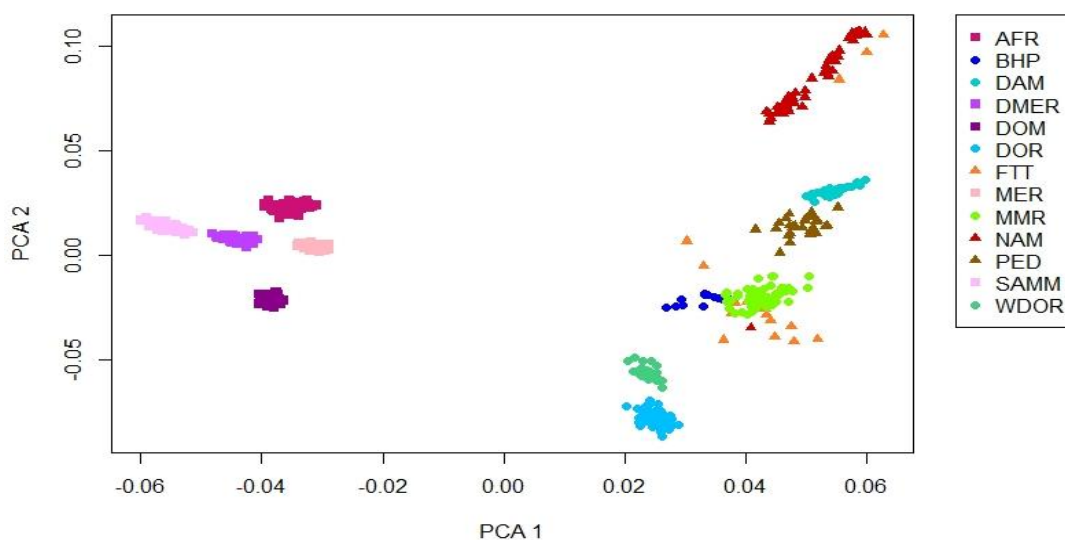
Figure 4.2a The genetic relationships among the 13 sheep populations as seen when plotting the first two principal components (PCA1 and PCA2) against one another (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)

Figure 4.2b The genetic relationships among the 13 sheep populations as seen when plotting the first and third principal components (PCA1 and PCA3) against one another. (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)
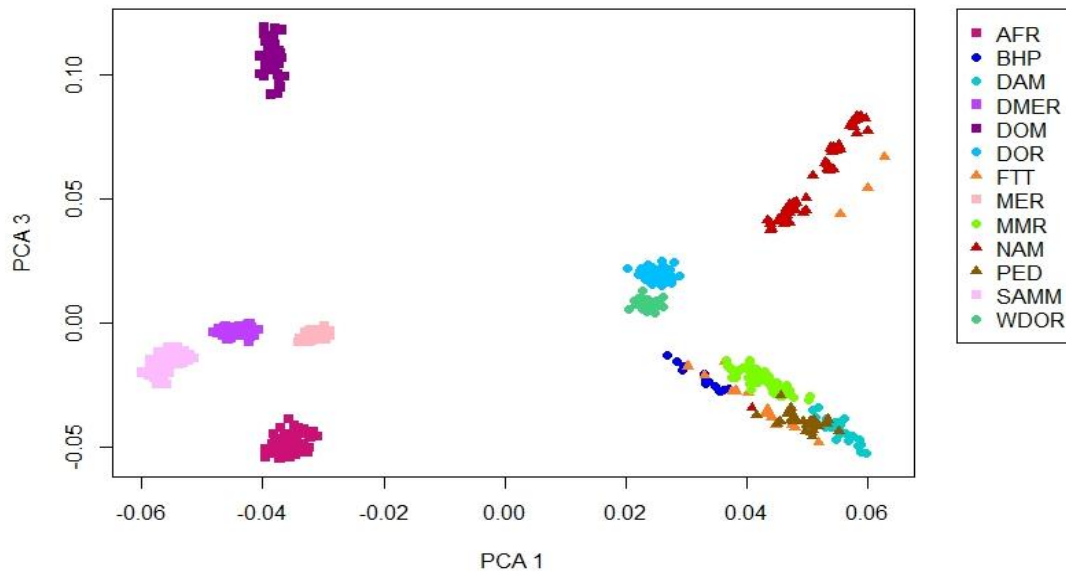
A clear distinction between the dual-purpose populations and the rest of the populations can be seen in Figure 4.2a, as all Merino-types clustered together when the two largest principal components were plotted against each other. The meat type and indigenous populations were clustered relatively close together, with the MMR, BHP and FTT forming overlapping clusters. The non-descript fat tailed population (FTT) formed a loose, dispersed cluster. The WDOR and DOR populations clustered further away, together. The NAM population clustered the furthest away from the other meat and indigenous populations, forming a spread-out line. This trend was maintained in Figure 4.2b, where it can be seen that Merino-types remain clustered together while indigenous and meat type populations cluster closer together, with the MMR, BHP, FTT, PED and DAM populations forming an overlapping cluster. The WDOR and DOR populations remain slightly separated from the other meat type populations, while remaining close together. The NAM population maintained its separate spread-out clustering. From the PCA plots, it could also be seen that three individuals of the FTT population cluster closely together with the NAM population.

4.4.2 Admixture

An ADMIXTURE plot of all thirteen populations was generated to investigate population structure. Cross-validation error estimates were obtained and used to identify the most probable number of genetic groups (K value with the lowest cross-validation score). The cross-validation scores for each K value from 2 to 16 were plotted in order to determine the correct K value, and is shown in Figure 4.3. From this graph, it could be seen that the lowest cross-validation error and inflection point

was found at K=15 (0.57943), thus the most probable number of inferred populations was chosen as 15.

From the bar plot shown in Figure 4.4, a clear differentiation between the dual-purpose populations, and the meat and indigenous populations was seen at low K values, such as K=2, and K=3. At the chosen K-value (K=15), the DOR, NAM and SAMM populations each differentiate into two distinct subpopulations, while the BHP and FTT populations can be seen as one population. At K=15, the MMR population shows the highest level of admixture. At each K value, three individuals from the FTT population can be seen to be similar to the NAM population. The ADMIXTURE results at K=15 is attached as an addendum (A1).
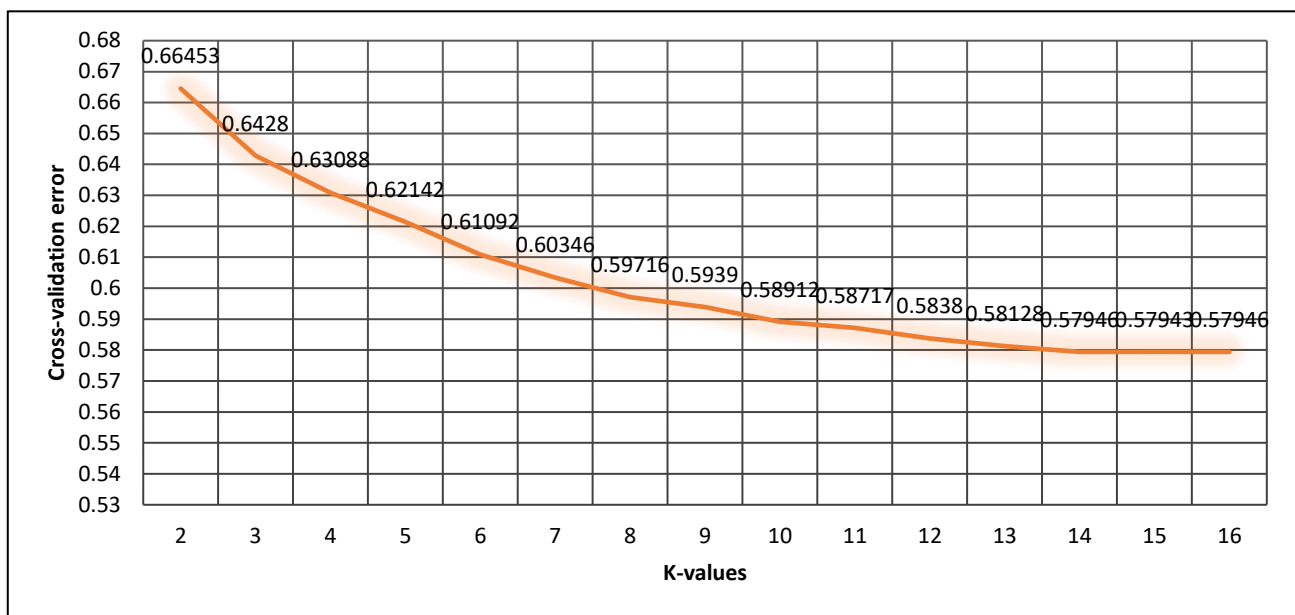
Figure 4.3 A cross-validation plot, indicating the cross-validation error rate for different K values

4.5 Inbreeding end effective population sizes

4.5.1 Wright's inbreeding coefficient

The inbreeding coefficient (Fis) for each individual as well as the average per population was calculated using the expected and observed heterozygosities. Table 4.4 presents the lowest and highest Fis obtained for an individual per population, as well as the average Fis per population.

The average Fis per population was generally negative or low positive. The lowest average Fis obtained was for the BHP population at -0.076, while the highest was estimated for FTT at 0.026. The lowest Fis for an individual was a BHP animal, at -0.288, while the highest Fis for an individual was for a MER animal at 0.354. The most variation in Fis within a population was observed for NAM, at 0.645.
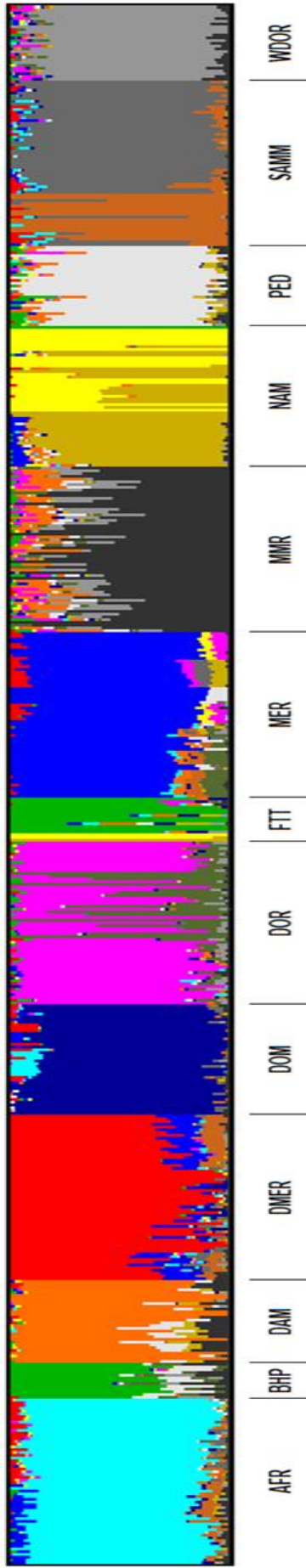
Figure 4.4 ADMIXTURE graph showing the proportions of ancestral populations for each individual at K=15 (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)

Table 4.4 Average inbreeding coefficients per population as well as lowest and highest inbreeding coefficient for individuals within each population

Population	Average Fis	Lowest	Highest
AFR	0.013	-0.050	0.193
BHP	-0.076	-0.288	0.176
DAM	-0.012	-0.099	0.239
DMER	0.005	-0.051	0.136
DOM	-0.008	-0.076	0.073
DOR	0.004	-0.054	0.116
FTT	0.026	-0.153	0.292
MMR	-0.031	-0.074	0.034
MER	0.021	-0.069	0.354
NAM	0.022	-0.128	0.336
PED	-0.034	-0.116	0.038
SAMM	0.009	-0.052	0.127
WDOR	-0.026	-0.087	0.022

AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper

4.5.2 Effective population size

The change in effective population size across generations for each population was calculated using SNeP. The change in Ne was plotted for each population from 985 to 13 generations ago. The change in Ne is illustrated in Figure 4.5.

Figure 4.5 Trends in historic effective population size (Ne) for each population type (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)
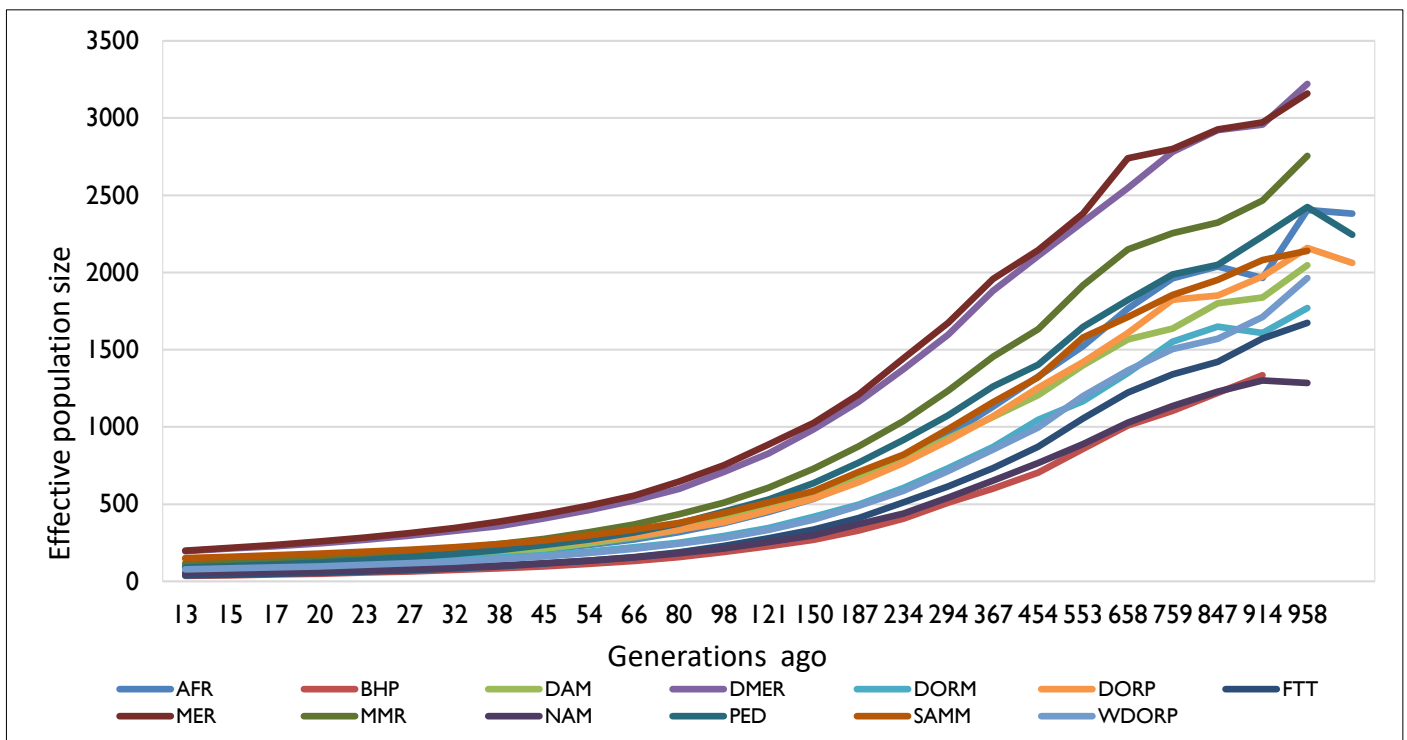
All populations showed a decline in Ne across generations. The populations showed a large variation in initial population sizes, with DMER having the largest population size at 958 generations ago, and NAM having the smallest. Therefore, the steepest decline was seen for the DMER population, while the most gradual decrease was seen for the NAM population.

4.5.3 Runs of homozygosity

The ROH for each population was estimated across four different length categories (1-3.99, 4-7.99, 8-11.99, 12-15.99 and ≥16 Mb) in order to distinguish between the degree of recent versus past inbreeding. Longer ROH segments indicate more recent inbreeding, whereas shorter segments indicate past inbreeding. Figure 4.6 indicates the number of ROH present for each population per length category.
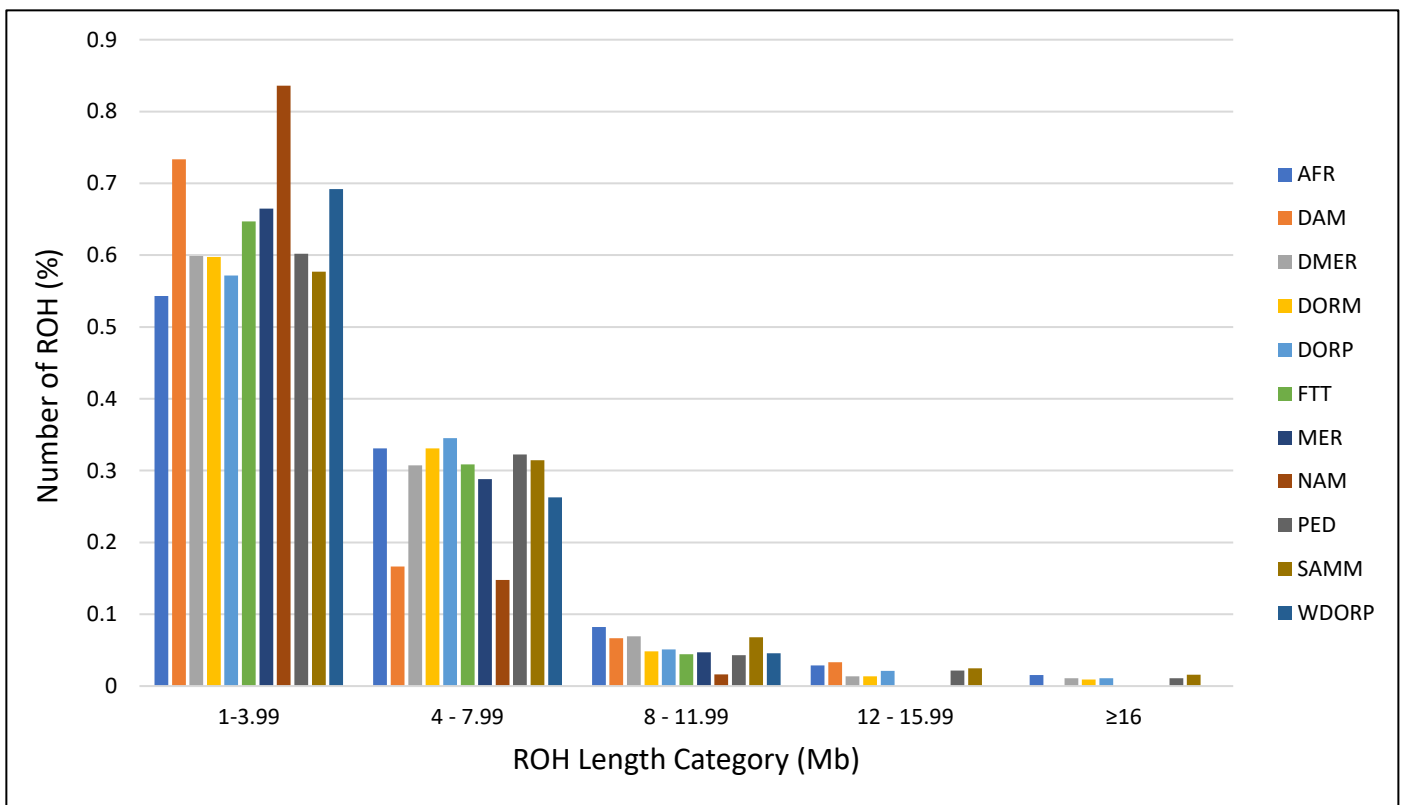
Figure 4.6 The number of runs of homozygosity per population within the defined length categories. (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)

It could clearly be seen that for all populations, the largest percentage of ROH fell within the shortest length category, less than 3.99 Mb. The highest percentage of short ROH was observed in the NAM population, while the lowest percentage of short ROH could be seen in the AFR population. The second largest percentage of ROH was observed within the category of 4-7.99, with a large decrease in ROH from the second (4-7.99), to the third (8-11.99) length category. A very small percentage of ROH was seen in the longest length category of ≥16 Mb.

The number of ROH present on each chromosome, except for the sex chromosomes, was calculated within each population type; dual purpose, meat type and indigenous populations respectively. Figure 4.7 indicates the distribution of ROH across 26 autosomes.
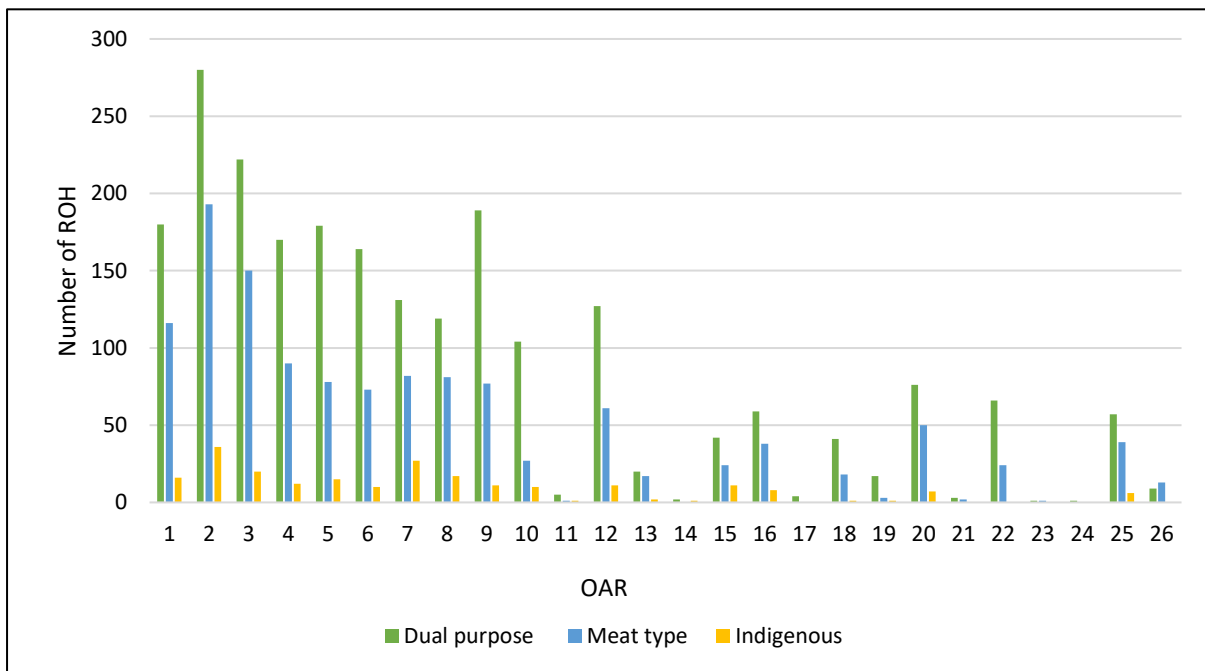
Figure 4.7 The number of ROH per chromosome for each population type

Figure 4.7 indicates that for all populations, the largest number of ROH was observed on the first ten chromosomes, while very few ROH was identified on chromosomes 11, 14, 17, 21, 23 and 24. It is noticeable that across all chromosomes, a significantly larger amount of ROH was observed for dual-purpose and meat type populations than for indigenous populations.

The overall level of inbreeding, measured in F_{ROH}, was estimated for each population in order to determine and compare the extent of inbreeding in all populations. Table 4.5 contains the F_{ROH} estimated for each population.

Table 4.5 F_{ROH} estimation for each population

Population	F_{ROH}
AFR	0.0174
BHP*	-
DAM	0.0017
DMER	0.0113
DOM	0.0193
DOR	0.0284
FTT	0.0068
MMR	0.0036
MER	0.0044
NAM	0.0016
PED	0.0060
SAMM	0.0244
WDOR	0.0116

AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper. *Could not be calculated as a result of small sample size.

The F_{ROH} estimates ranged from 0.0016 (NAM) to 0.0284 (DOR). The meat type and indigenous populations generally showed a lower F_{ROH} than the dual-purpose populations, with the exceptions of the MER population with a F_{ROH} of 0.0044, and the DOR population with a F_{ROH} of 0.0284. The F_{ROH} generally followed the same trend as the Fis estimates, with the dual-purpose populations overall showing higher F_{ROH} estimates than the meat type and indigenous populations.

4.6 Phylogenetic relationships

Previously in this study, ADMIXTURE results indicated that the SAMM, NAM and DOR populations each divide to form subpopulations. These subpopulations will be referred to as SAMM2, NAM2 and DOR2.

The mean pairwise Fst values were calculated with ADMIXTURE software and is reported in Table 4.6. The Fst values were generally low, with the lowest pairwise Fst value found between the MER and DMER populations at 0.173, while the highest Fst value was estimated between the DOM and NAM2 populations at 0.388.

These Fst values were then used to construct an unrooted tree based on genetic distances. The relationships between the thirteen populations was illustrated in Figure 4.8, which showed that the thirteen populations form three distinct branches. The dual-purpose type populations formed a separate branch on the tree, while the meat type and indigenous populations branched together, with the WDOR and DOR populations splitting off to form a third branch. It is notable that the SAMM, NAM and DOR populations each branched to form subpopulations. Furthermore, the BHP and FTT populations formed a single branch on the tree.

Table 4.6 Mean Fst values among populations

	DMER	BHP/FTT	MER	NAM2	DOR	AFR	DAM	PED	DOM	SAMM2	SAMM1	NAM1	DOR2	WDOR	MMR
DMER	0	0.282	0.173	0.344	0.219	0.187	0.277	0.243	0.222	0.175	0.204	0.29	0.259	0.237	0.237
BHP/FTT		0	0.299	0.379	0.246	0.295	0.297	0.253	0.324	0.3	0.323	0.316	0.292	0.268	0.265
MER			0	0.367	0.241	0.212	0.3	0.267	0.244	0.209	0.238	0.311	0.283	0.252	0.257
NAM2				0	0.348	0.355	0.347	0.324	0.388	0.361	0.387	0.288	0.387	0.365	0.342
DOR					0	0.242	0.276	0.235	0.254	0.237	0.266	0.292	0.189	0.196	0.225
AFR						0	0.289	0.258	0.251	0.204	0.218	0.298	0.282	0.257	0.255
DAM							0	0.226	0.324	0.297	0.323	0.287	0.314	0.292	0.261
PED								0	0.292	0.263	0.288	0.264	0.276	0.251	0.225
DOM									0	0.234	0.257	0.327	0.292	0.266	0.283
SAMM2										0	0.178	0.308	0.277	0.255	0.263
SAMM1											0	0.329	0.304	0.279	0.285
NAM1												0	0.331	0.305	0.28
DOR2													0	0.245	0.272
WDOR														0	0.229
MMR															0

AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper
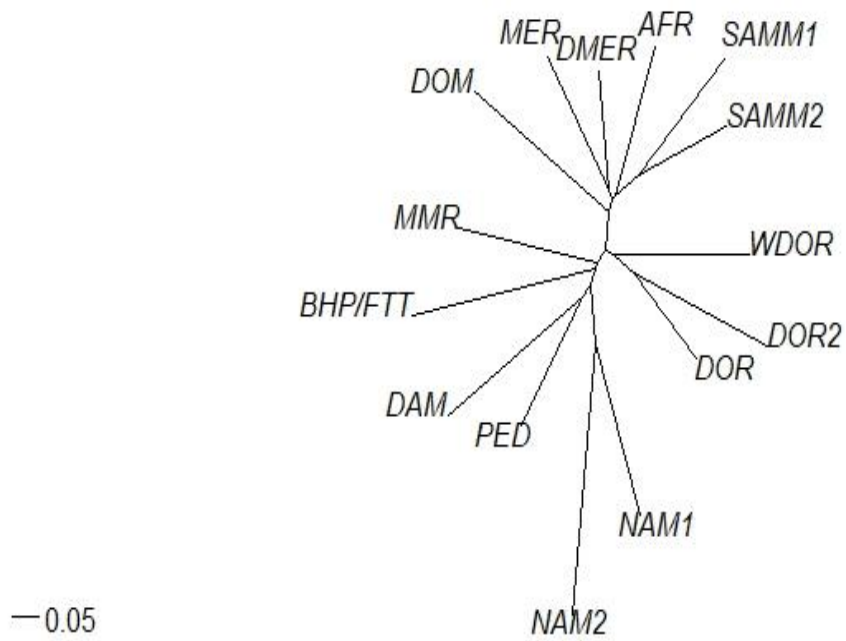
Figure 4.8 Genetic distance between populations based on pair-wise Fst estimates. (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED)

Chapter 5: Discussion

5.1 Introduction

The aim of this study was to use genome-wide SNP data to investigate the genetic structure of both indigenous and commercial sheep breeds in South Africa. Thirteen South African sheep breeds, including animals from different production types, namely dual-purpose, meat and indigenous populations. The genotypes were analysed to quantify the genetic variation within and between populations, as well as to estimate inbreeding levels and define the population structure of these animals.

The thirteen populations investigated consisted of varying sample sizes, complicating the investigation of the genetic variation between populations. In order to overcome this, the populations were first analysed individually to obtain as much information per population as possible, whereafter a representative sample of each larger population was taken, and all populations were merged and analysed to allow comparative investigation.

Limited similar studies have been done on South African sheep investigating genetic diversity, therefore limiting the amount of local studies available to compare with the results of this study. These include studies using both microsatellites (Peters et al., 2010; Soma et al., 2012; Qwabe et al., 2012), and SNP data (Sandenbergh et al., 2016; Molotsi et al., 2017; Dlamini et al., 2019). The current study is the first comprehensive analysis including a large number of sheep breeds in South Africa.

5.2 Quality control

It is important to perform quality control (QC), including sample-based QC and marker-based QC, before the analysis of SNP data, in order to avoid potentially biased results caused by sampling errors and incorrect genotype calling (Anderson et al., 2010). In the current study, the number of markers remaining per population after QC was performed, ranged from 40 418 (DAM), to 44 778 (SAMM), with the exception of the BHP population which had only 32 422 markers remaining. These numbers are lower than the 49 034 markers used by Kijas et al. (2012), but are comparable to the 43 500 markers reported by Molotsi et al. (2017).

Ajmone-Marsan et al. (2014) stated that because of the methodology of SNP discovery, ascertainment bias is the main drawback for the application of SNP genotyping arrays. Ascertainment bias is introduced when the sample of animals used for SNP discovery is small, or not wholly representative of all breeds within the species (Nielsen, 2004). As the identification of a single SNP depends largely on allele frequency, this causes a deficiency of SNP loci with rare alleles, as rare SNPS are more likely not to be discovered (Clark et al., 2005). The composition and size of the SNP discovery

population can therefore result in overcalling or under-calling of globally distributed SNPs (Clark et al., 2005).

In this study, lower levels of polymorphic loci were observed for indigenous and African fat-tailed breeds, which was probably caused by underrepresentation of these breeds during the development of the Ovine SNP50 beadchip (Kijas et al., 2012). The exceptionally low number of SNPs (32422) that remained for downstream analysis within the BHP population was further caused by poor quality samples with low DNA levels. The degree of polymorphism existing within different breeds of each species plays an integral role in the successful application of SNP arrays (Fan et al., 2010). The varying number of polymorphic loci between the populations used in this study may possibly be attributed to ascertainment bias introduced during the SNP discovery phase of the development of the ovine 50K SNP chip.

5.3 Within population diversity

Heterozygosity rate is an important parameter indicating population diversity, as genotypes showing low heterozygosity may indicate inbreeding, as well as low genetic diversity. Furthermore heterozygosity can also be used to identify human errors that may occur during genotyping, as excessively heterozygous genotypes could indicate contamination of a DNA sample (Anderson et al., 2010). Low genetic diversity, especially within small populations, is detrimental, as this limits the ability of the population to adapt to environmental changes, thus limiting the response of the population to selection pressure (Erhardt & Weimann, 2007).

The overall average observed heterozygosity (Ho), for all populations found in this study was 0.353, while the average expected heterozygosity (He) was 0.351. These are slightly higher than the overall average Ho of 0.341 and He of 0.312 reported by Molotsi et al. (2017) when investigating the genetic diversity of four South African sheep breeds, namely Dorper, SA Mutton Merino, Namaqua Afrikaner and smallholder Dorper animals, as well as the average Ho of 0.325 reported by Sandenbergh et al. (2016) when investigating four sheep breeds, Namaqua Afrikaner, SA Mutton Merino, Dorper and SA Merino animals.

The Ho found in this study for the Namaqua Afrikaner population at 0.337, was higher than that of both Sandenbergh et al. (2016), and Molotsi et al. (2017), at 0.28 and 0.295 respectively. However, the Ho estimate for the Dorper population of 0.353 was similar to the respective estimates of 0.34 and 0.348 found by Sandenbergh et al. (2016) and Molotsi et al. (2017). In terms of heterozygosity rates, the highest genetic diversity in this study was found for the dual-purpose populations (average Ho= 0.360; average He= 0.363), followed by the meat type populations (average Ho= 0.355; average He= 0.345), with the indigenous populations showing the lowest average diversity (average Ho= 0.339; average He= 0.340). This trend is supported by a study performed by Soma et al.

(2012), where microsatellites were used to investigate the genetic diversity of 20 sheep breeds. This study reported higher heterozygosity levels in wool breeds (average 0.659) compared to indigenous breeds (average 0.598). The lower heterozygosity rate and genetic diversity found in the indigenous populations might be caused by small population sizes, as well as underrepresentation of these breeds in the design of the Ovine SNP chip. This is supported by the findings of Sandenbergh et al. (2016). Kijas et al. (2012) indicated high levels of genetic diversity in Merino breeds worldwide, therefore supporting the high heterozygosity rates found for the dual-purpose populations in this study.

The average MAF values found in this study ranged from 0.243 (BHP), to 0.286 (DMER). The overall average MAF across all populations was 0.265, which is similar to the average MAF of 0.251 found by Molotsi et al. (2017), as well as the average MAF of 0.225 reported by Sandenbergh et al. (2016). The average MAF for indigenous populations (0.255) and meat type populations (0.259) was lower than for dual-purpose populations (0.277). This is supported by Sandenbergh et al. (2016) who reported a lower MAF value for a Namaqua Afrikaner population (0.19) and a Dorper population (0.23) than for a Merino population (0.26). The higher MAF seen in dual-purpose populations indicates a higher level of polymorphic markers, which corresponds to the higher levels of He found for these populations, indicating greater genetic diversity within these populations. The low MAF seen in indigenous populations possibly indicates lower genetic diversity and a larger number of fixed alleles (Engelsma et al., 2014).

The SNP density necessary for accurate genomic selection is determined by the extent of linkage disequilibrium (LD) across the genome. LD is influenced by factors such as breeding systems, population history, as well as the pattern of geographic subdivision (Goddard & Hayes, 2007), and will therefore vary between different breeds. The average linkage disequilibrium (r^2) values found in this study ranged from 0.362 (MER) to 0.499 (BHP). Indigenous populations had the highest average r^2 of 0.444, while dual-purpose populations had the lowest at 0.392. The high r^2 seen in die indigenous populations could be caused by small population sizes or population bottlenecks (Espigolan et al., 2013). The r^2 estimates were similar to a study done by Mastrangelo et al. (2017) on sheep of the Barbaresca breed.

5.4 Between population comparison

Principal Component analysis (PCA) is the most commonly used large-scale method of differentiating between different ancestries for the purpose of identifying genetically related samples (Anderson et al., 2010). Principal components represent uncorrelated components which are produced from potentially correlated SNPs, where the first principal component accounts for the largest proportion of variation, and subsequent components accounting for less variation (Anderson

et al., 2010). PCA and ADMIXTURE results from this study were in agreement and clustered the populations into production types.

The dual-purpose populations formed the tightest clusters in PCA analyses. The five dual-purpose populations also clustered close to each other, separating the dual-purpose populations from the meat and indigenous types. This is supported by the findings of Soma et al. (2012), who reported close groupings of wool populations in SA, as well as Kijas et al. (2009) who reported that Western breeds share high levels of genetic similarity. The major dual-purpose breeds in SA are descendant from European countries, such as the Netherlands, Spain and Germany (Vink, 2009). The WDOR and DOR populations formed two separate clusters close to each other, indicating that these populations form two distinct subpopulations of the Dorper breed (Milne, 2000).

The FTT and BHP showed overlapping clusters of animals, indicating similar ancestral genotypes between these populations, which was supported by the ADMIXTURE plots. As the BHP populations are classified as fat-tailed animals, this could indicate that the non-descript FTT animals are closely related to the BHP population. Animals of the MMR population also formed an overlapping cluster with the FTT and BHP populations, indicating shared genotypes. The similarity found between the composite MMR and these populations could indicate the use of these populations in the development of the composite MMR population (Peters et al., 2010).

The animals of the NAM population formed a loose cluster, indicating more variation in the genotypes of individuals of this population. As this is an indigenous population, less artificial selection for specific production traits are performed on these animals (Kijas et al., 2009), which could result in the loose clustering. Furthermore, indigenous and locally developed breeds have developed unique combinations of adaptive traits in order to respond to local environmental pressures (Peters et al., 2010), which could also explain why the animals of the FTT population was the most spread out, indicating that these animals have a higher variation on a genetic level. The production systems wherein these indigenous populations are kept may also cause the higher within breed variation seen compared to dual-purpose populations. Indigenous populations in SA are generally kept in small-holder systems, with limited management and breeding systems in place, thus forcing them to become more adaptable (Molotsi et al., 2017).

All dual-purpose populations in this study are Merino types, therefore the close clustering of these populations was expected. The tight clusters observed for the dual-purpose populations indicate that higher levels of selection have been carried out on these populations for traits of economic importance (Cloete et al., 2001). According to a review done by Schoeman et al. (2010), substantial genetic gains resulting from selection for a range of economically important traits were demonstrated

for breeds participating in the NSSIS. The level of participation of indigenous breeds in the NSSIS is low, possibly resulting in the loose clusters observed for these breeds (Cloete & Olivier, 2012).

The indigenous and meat type populations, with the exception of the NAM, WDOR and DOR populations, formed overlapping clusters when plotting PCA1 to PCA3 values, indicating admixture and shared genotypes between these populations. This could be a result of crossbreeding occurring between these populations, resulting from breeding systems aimed at improving the adaptative traits of meat type populations (Schoeman et al., 2010). Furthermore, the WDOR and DOR populations continued to cluster separately but close together, further indicating similar but distinct genotypes. The NAM population maintained a separate clustering from the other meat type and indigenous population, indicating a level of isolation of this population from other populations. This was supported by the findings of Qwabe et al. (2012), where microsatellite markers was used to perform genetic characterization of 144 animals from three NAM flocks in SA, and population structure results were compared to PED and SAMM sheep. This study reported limited hybridization between the NAM animals and animals from both of these populations. The low level of crossbreeding found in NAM populations may be explained by efforts to keep pure bred NAM populations, as it is estimated that there are only 2000 Namaqua Afrikaner sheep in SA, with two conservation flocks consisting out of 100 ewes each (Snyman, 2014a); indicating that the NAM breed may be at risk of becoming endangered.

ADMIXTURE separated the populations into wool and meat types when two ancestral populations were assumed (K=2), indicating that these two production types have distinct ancestral backgrounds. When a third ancestral population was introduced (K=3), the populations could further be distinguished, as indigenous populations were separated from the wool and meat populations. When four ancestral populations were assumed (K=4), it became noticeable that the NAM population separated into two distinct populations, this became increasingly clear as more ancestral populations were assumed. At K=12, the SAMM population also differentiated into two distinct populations. Across the various K values, it was noticeable that the BHP and FTT followed similar trends, indicating similar genotypes, supporting the PCA results. At the estimated optimal K value of K=15, the NAM population was clearly separated into two distinct populations, as was the SAMM population and the DOR population.

Two distinct NAM subpopulations were observed, and it was clear that the genotypes of these two populations originated from either the resource flock at Nortier, genotyped in 2013, or from Swellendam Drosdy Museum and Worcester Kleinplasie Museum, genotyped in 2016. Furthermore, the NAM animals originating from the Nortier research flock showed less admixture relative to the other NAM individuals, based on colour distribution seen in the ADMIXTURE plot. The two

subpopulations identified for the SAMM population was inspected, and it was found that the two subpopulations originated from the Nortier research flock genotyped in 2013, and from industry animals and show animals, genotyped in 2015 and 2016, respectively. Similar to the NAM population, the animals originating from the Nortier research flock showed lower levels of admixture and crossbreeding than the commercial animals. For the DOR population, the two subpopulations identified could again be assigned to animals genotyped at the Nortier research flock in 2013, and industry animals genotyped in 2015. Similar to the NAM and SAMM populations, the animals genotyped at the Nortier research flock showed a lower level of admixture when compared to industry animals. This indicates that animals kept at the Nortier research flocks maintain a higher level of breeding genotypes than animals involved in production flocks, causing genetic variation between these populations resulting in the formation of subpopulations. These varying levels of genetic diversity holds great importance for the creation of reference populations, as it is important that a reference population accurately represents the national flock.

5.5 Inbreeding

In the current study, the average Fis values were generally low positive or negative, indicating that none of these populations showed exceptionally high inbreeding levels. The average Fis for the dual-purpose populations was found to be the highest, indicating a lower level of heterozygosity in these populations (Gasca-Pineda et al., 2013), compared to the other investigated populations. The positive Fis values obtained for the indigenous populations, indicating a reduction in heterozygosity, could be a result of small population sizes (Molotsi et al., 2017). Furthermore, the negative Fis values observed for the meat type populations may indicate the use of crossbreeding systems (Kijas et al., 2009). The Fis estimates were comparable to estimates obtained by Sandenbergh et al. (2016).

The effective population size (Ne) is a parameter that can be used to assess inbreeding rates and thus genetic diversity within populations (Prieur et al., 2017). The Ne estimates for all populations included in this study show a decline from 958 generations ago. The steepest decline was seen for the DMER population, decreasing from 3221 animals to approximately 196 animals 13 generations ago. The BHP population had the lowest recent Ne, with only 36 animals at 13 generations ago. The decrease in Ne for all populations indicate a loss of genetic diversity over time (Gasca-Pineda et al., 2013). The general decrease in Ne seen for sheep breeds could possibly be attributed to increased selection for production traits, as well as inbreeding occurring within populations (Prieur et al., 2017). The trend observed with regards to Ne in this study was similar to other studies investigating Ne in sheep (García-Gámez et al., 2012; Gasca-Pineda et al., 2013; Prieur et al., 2017; Rodríguez-Ramilo et al., 2019)

Runs of homozygosity (ROH) is an important estimate of inbreeding rates within populations, as it provides a good measure of individual genome-wide autozygosity, and allows the distinction between recent and ancient inbreeding (Mastrangelo et al., 2016b). Selection for specific traits increase the homozygosity around the target region, thus it is expected to result in long fragments or high numbers of ROHs in regions that are under selection (Peripolli et al., 2017). It is important to set appropriate parameters to detect ROH, such as the minimum number of SNPs used to define a ROH, as it is necessary to determine the minimum number of SNPs needed to define an ROH based on the available SNP density (Signer-Hasler et al., 2017).

In the current study, all populations presented a relatively high number of short ROHs, with smaller numbers of long ROHs. The highest number of short ROHs was found within the NAM population, followed by the DAM and WDORP populations, possibly indicating a high amount of ancient inbreeding. None of the populations showed a significant amount of extremely long ROH, which would indicate high levels of recent inbreeding (Rodríguez-Ramilo et al., 2019).

The dual-purpose populations showed the highest number of ROHs across all lengths. This could be a result of the differences in populations sizes between dual-purpose populations and meat and indigenous populations. The relatively larger number of ROHs detected for dual-purpose populations than for meat type and indigenous populations could also be explained by increased selection pressure for production traits, which may result in increased levels of inbreeding (Van Wyk et al., 2009). All populations showed the highest number of ROHs distributed across the first 10 chromosomes, with the exception of the OAR 11 with very low number of ROHs for all populations. On chromosome 20, 22 and 25 a proportionally higher amount of ROHs was seen for meat and dual-purpose populations, with almost no ROHs found on the chromosomes for indigenous populations.

The F_{ROH} estimates calculated in this study ranged from 0.017 (AFR) to 0.028 (DOR). The F_{ROH} generally followed the same trend as the Fis estimates, with the dual-purpose populations overall showing higher F_{ROH} estimates than the meat type and indigenous populations.

5.6 Phylogenetic relationships

The fixation index (Fst) is a measure of genetic differentiation, which is calculated as the reduction in heterozygosity of subpopulations relative to the total population (Weir & Hill, 2002). High Fst values imply that there is greater genetic differentiation between subpopulations. This implies that individuals within subpopulations are more related to each other than to individuals between subpopulations. Therefore, a Fst of 1 would indicate complete isolation of the subpopulation from the population (Kelleher et al., 2017).

In this study, Fst values ranged from 0.173 to 0.388. These low Fst values indicated a level of relatedness between all populations examined in this study. The lowest Fst value was found between

the MER and DMER population, indicating that these populations are closely related, while the highest Fst value was observed between the DOM and NAM populations. A study done by Soma et al. (2012) indicated similar relationships between SA breeds, showing high identity between dual-purpose breeds.

The pairwise Fst values were used to determine the genetic distance between populations in order to construct a phylogenetic tree. The tree obtained in this study illustrated the divergence of the dual-purpose populations from the meat type and indigenous populations. It also supported the results of the PCA plot, as the DOR and WDOR formed a separate branch from the other meat type populations. The tree also illustrated the divergence of the DOR population into two separate populations, as well as the NAM and SAMM populations, supporting the results of the ADMIXTURE plots which indicated the formation of subpopulations within these populations.

Chapter 6: Conclusion and recommendations

6.1 Conclusion

It is important to understand the population structure and breed composition of populations, as this can assist in the design of optimal breeding strategies in populations where it is important to maximise the exploitation of inter-breed nonadditive genetic effects, for example in crossbred and composite animals. In this study, whole genome SNP data from 1977 animals representing thirteen populations of South African sheep were investigated in term of genetic diversity and inbreeding. During QC, a total of 207 animals and 126814 genotypes were removed. The number of polymorphic SNPs remaining for downstream analysis were the highest for dual-purpose populations. Higher heterozygosity levels were found in the dual-purpose and meat type populations compared to the indigenous populations. Linkage disequilibrium estimates were highest in indigenous populations, followed by meat type populations, while the lowest linkage disequilibrium was found in dual-purpose populations.

The dual-purpose populations formed the tightest clusters in PCA analyses, indicating lower variation between these breeds. ADMIXTURE results were consistent with the history of each population, and indicated high proportions of co-ancestry between animals of the same population or production type. These results also indicated the divergence of the NAM, SAMM and DOR populations into subpopulations, based on the location and year of genotyping. Inbreeding estimations indicated the highest levels of inbreeding within dual-purpose populations, however none of the populations showed significantly high levels of inbreeding. Fst estimates indicated a level of relatedness between all populations examined in this study. A phylogenetic tree based on Fst estimates supported the results of PCA and ADMIXTURE plots, showing lower variation within dual-purpose populations, and high variation between dual-purpose populations and meat and indigenous populations. The phylogenetic tree also indicated higher variation within meat type and indigenous populations, while supporting the results of the ADMIXTURE analysis, indicating the formation of subpopulations within the NAM, SAMM and DOR populations.

6.2 Recommendations

In this study, the sample sizes of animals investigated varied widely per population. Improved accuracy can be achieved through analysis of larger sample sizes from various geographic locations, in order to be truly representative of population diversity. In order to improve the accuracy of analysis, it is necessary to increase the sample sizes per population, especially in the case of the BHP (13), FTT (16), PED (29) and WDOR (27) and to ensure relatively equal sample sizes for unbiased comparison. Furthermore, to evaluate the status of genetic diversity in South African sheep populations, a

comparative approach by means of international collaboration would be beneficial. Due to its benefits compared to other molecular markers, the utilisation of SNP markers should be encouraged. SNP markers can be used to select for increased heterozygosity in genes for adaptation to climate change.

The development of accurate reference populations is important to ensure successful genomic selection, and to increase the accuracy of imputation between lower and high-density SNP chips. The variation in genetic diversity between resource flocks and commercial flocks observed in this study clearly indicate that reference populations for SA sheep breeds cannot be composed of resource flocks alone. These flocks contained a lower level of genetic variation, and showed different genetic compositions than industry flocks. The variation in genetic diversity between resource flocks and industry flocks also highlight the limitations of fragmented genotyping opportunities across SA, and highlights the importance of collaboration between institutes.

References

ABSA. 2018. Agricultural outlook Spring edition. *Agricultural outlook 2018-2027*. https://www.daff.gov.za/Daffweb3/Portals/0/Statistics and Economic Analysis/Statistical Information/Trends in the Agricultural Sector 2018.pdf 9 February 2020.

Ajmone-Marsan, P., Colli, L., Han, J.L., Achilli, A., Lancioni, H., Joost, S., Crepaldi, P., Pilla, F., Stella, A., Taberlet, P., Boettcher, P., Negrini, R. & Lenstra, J.A. 2014. The characterization of goat genetic diversity: Towards a genomic approach. *Small Ruminant Research*, 121(1): 58–72.

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

Alexander, M. 2018. South Africa's weather and climate. *South Africa Gateway*. https://southafrica-info.com/land/south-africa-weather-climate/ 12 September 2020.

Analla, M., Montilla, J.. & Serradilla, J.. 2002. Analyses of lamb weight and ewe litter size in various lines of Spanish Merino sheep. *Small Ruminant Research*, 29(3): 255–259.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T. 2010. Data quality control in genetic case-control association studies. *Nature Protocols*, 5: 1564–1573.

Arora, R. & Bhatia, S. 2004. Genetic structure of Muzzafarnagri sheep based on microsatellite analysis. *Small Ruminant Research*, 54: 227–230.

Arranz, J.J., Bayo, Y. & Primitivo, F.S. 1998. Genetic relationships among Spanish sheep using microsatellites. *Animal Genetics*, 29: 435–440.

Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M.W. 2015. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, 6(MAR): 1–6.

Barros, E.A., Brasil, L.H. d. A., Tejero, J.P., Delgado-Bermejo, J. V. & Ribeiro, M.N. 2017. Population structure and genetic variability of the Segureña sheep breed through pedigree analysis and inbreeding effects on growth traits. *Small Ruminant Research*, 149: 128–133. http://dx.doi.org/10.1016/j.smallrumres.2017.02.009.

Benhin, J.K. 2006. Climate Change and South African Agriculture: Impacts and Adaptation Options. *Africa*, (21): 113.

Beuzen, N.D., Stear, M.J. & Chang, K.C. 2000. Molecular markers and their use in animal breeding. *Veterinary Journal*, 160(1): 42–52.

Blasco, A. & Toro, M.A. 2014. A short critical history of the application of genomics to animal breeding. *Livestock Science*, 166(1): 4–9. http://dx.doi.org/10.1016/j.livsci.2014.03.015.

Bourdon, R.M. 2000. Understanding Animal Breeding. In Colorado: 56–65.

Bravo, S., Larama, G., Quiñones, J., Paz, E., Rodero, E. & Sepúlveda, N. 2019. Genetic diversity and phylogenetic relationship among araucana creole sheep and Spanish sheep breeds. *Small Ruminant Research*, 172(January): 23–30. https://doi.org/10.1016/j.smallrumres.2019.01.007.

Broman, K.W. & Weber, J.L. 1999. Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain. *American Journal of Human Genetics*, 65(6): 1493–1500.

Bruford, M.W., Bradley, D.G. & Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4(11): 900–910.

Buchmann, R. & Hazelhurst, S. 2014. Genesis Manual. http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf.

Campher, J.., Hunlun, C. & Van Zyl, G.. 1998. *South African Livestock Breeding*. P.O. Box 270, Bloemfontein 9300, South Africa: South African Stud Book and Livestock Improvement Association.

Cesarani, A., Sechi, T., Gaspa, G., Usai, M.G., Sorbolini, S., Macciotta, N.P.P. & Carta, A. 2019. Investigation of genetic diversity and selection signatures between Sarda and Sardinian Ancestral black, two related sheep breeds with evident morphological differences. *Small Ruminant Research*, 177(April): 68–75. https://doi.org/10.1016/j.smallrumres.2019.06.014.

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R.R., Pemberton, J.M., Beraldi, D., Stear, M.J., Alberti, A., Pittau, M., Iannuzzi, L., Banabazi, M.H., Kazwala, R.R., Zhang, Y.P., Arranz, J.J., Ali, B.A., Wang, Z., Uzun, M., Dione, M.M., Olsaker, I., Holm, L.E., Saarma, U., Ahmad, S., Marzanov, N., Eythorsdottir, E., Holland, M.J., Paolo, A.M., Bruford, M.W., Kantanen, J., Spencer, T.E. & Palmarini, M. 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science*, 324: 532–536.

Ciani, E., Lasagna, E., D'Andrea, M., Alloggio, I., Marroni, F., Ceccobelli, S., Delgado Bermejo, J. V., Sarti, F.M., Kijas, J., Lenstra, J.A. & Pilla, F. 2015. Merino and Merino-derived sheep breeds: A genome-wide intercontinental study. *Genetics Selection Evolution*, 47(1): 64. http://dx.doi.org/10.1186/s12711-015-0139-z.

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

Cloete, J.J.E., Cloete, S.W.P., Olivier, J.J. & Hoffman, L.C. 2007. Terminal crossbreeding of Dorper ewes to Ile de France, Merino Landsheep and SA Mutton Merino sires: Ewe production and lamb performance. *Small Ruminant Research*, 69(1–3): 28–35.

Cloete, S. & Olivier, J. 2012. South African sheep and wool industry. In *International Sheep and Wool Handbook*. 95–112.

Cloete, S.W.P., Cloete, J.J.E. & Scholtz, A.J. 2016. Genetic parameters for tick count and udder health in commercial and indigenous ewes in South Africa. *Veterinary Parasitology*, 230: 33–42. http://dx.doi.org/10.1016/j.vetpar.2016.10.014.

Cloete, S.W.P., Gilmour, A.R., Olivier, J.J. & Van Wyk, J.B. 2004. Genetic and phenotypic trends and parameters in reproduction, greasy fleece weight and liveweight in Merino lines divergently selected for multiple rearing ability. *Australian Journal of Experimental Agriculture*, 44: 745–754.

Cloete, S.W.P., Olivier, J.J., Sandenbergh, L. & Snyman, M.A. 2014. The adaption of the South Africa sheep industry to new trends in animal breeding and genetics: A review. *South African Journal of Animal Sciences*, 44(4): 307–321.

Cloete, S.W.P., Schoeman, S.J., Coetzee, J. & Morris, J.D.V. 2001. Genetic variances for liveweight and fleece traits in Merino, Dohne Merino and South African Meat Merino sheep. *Australian Journal of Experimental Agriculture*, 41: 145–153.

Cloete, S.W.P. & Scholtz, A.J. 1998. Lamb survival in relation to lambing and neonatal behaviour in medium wool Merino lines divergently selected for multiple rearing ability. *Australian Journal of Experimental Agriculture*, 38: 801–811.

DAFF. 2019. *Abstract of Agricultural statistics 2019, Department of Agriculture Forestry and Fisheries, Africa.* http://www.daff.gov.za/Daffweb3/Portals/0/Statistics and Economic Analysis/Statistical Information/Abstract 2017.pdf 21 February 2020.

DAFF. 2007. GOVERNMENT NOTICES. , 20(29516): 3–16.

DAFF. 2017. Trends in the Agricultural sector. *Trends in the agricultural section*, 9(2): 22–39.

Dario, C. & Bufano, G. 2003. Efeito da endogamia sobre a produção de leite na raça ovina altamurana. *Archivos de zootecnia*, 52: 401–404. http://www.uco.es/organiza/servicios/publica/az/php/img/web/18_13_55_13NotaDario.pdf.

Dlamini, N.M., Visser, C., Snyman, M.A., Soma, P. & Muchadeyi, F.C. 2019. Genomic evaluation of resistance to Haemonchus contortus in a South African Dohne Merino flock. *Small Ruminant Research*, 175: 117–125. https://doi.org/10.1016/j.smallrumres.2019.04.020.

Dominik, S., Henshall, J.M. & Hayes, B.J. 2012. A single nucleotide polymorphism on chromosome 10 is highly predictive for the polled phenotype in Australian Merino sheep. *Animal Genetics*, 43(4): 468–470.

Elbeltagy, A.R., Aboul-Naga, A.M., Hassen, H., Rischkowsky, B. & Mwacharo, J.M. 2015. Genetic diversity and structure in Egyptian indigenous sheep populations mirror patterns of anthropological interactions. *Small Ruminant Research*, 132: 137–142. http://dx.doi.org/10.1016/j.smallrumres.2015.10.020.

Engelsma, K.A., Veerkamp, R.F., Calus, M.P.L. & Windig, J.J. 2014. Consequences for diversity when animals are prioritized for conservation of the whole genome or of one specific allele. *Journal of Animal Breeding and Genetics*, 131(1): 61–70.

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

Espigolan, R., Baldi, F., Boligon, A.A., Souza, F.R.P., Gordo, D.G.M., Tonussi, R.L., Cardoso, D.F., Oliveira, H.N., Tonhati, H., Sargolzaei, M., Schenkel, F.S., Carvalheiro, R., Ferro, J.A. & Albuquerque, L.G. 2013. Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics*, 14(305): 1–8.

Falconer, D.S. & Mackay, T.F.C. 1996. Introduction to Quantitative Genetics (Fourth Edition). In *Trends in Genetics*.

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*, 23(7): 833–847.

Figueredo, J.S., Cruz, J.F., Sousa, L.S., Teixeira Neto, M.R., Carneiro, P.L.S., Brito, N.D., Pinheiro, R.G.S., Lacerda, K.S.O. & Mottin, V.D. 2019. Genetic diversity and population structure estimation of Brazilian Somali sheep from pedigree data. *Small Ruminant Research*, 179(September): 64–69. https://doi.org/10.1016/j.smallrumres.2019.09.010.

Frkonja, A., Gredler, B., Schnyder, U., Curik, I. & Sölkner, J. 2012. Prediction of breed composition in an admixed cattle population. *Animal Genetics*, 43: 696–703.

García-Gámez, E., Sahana, G., Gutiérrez-Gil, B. & Arranz, J.J. 2012. Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics*, 13: 43.

Gasca-Pineda, J., Cassaigne, I., Alonso, R.A. & Eguiarte, L.E. 2013. Effective Population Size, Genetic Variation, and Their Relevance for Conservation: The Bighorn Sheep in Tiburon Island and Comparisons with Managed Artiodactyls. *PLoS ONE*, 8(10): 20–22.

Gholizadeh, M. & Ghafouri-Kesbi, F. 2016. Inbreeding depression in growth traits of Baluchi sheep. *Small Ruminant Research*, 144: 184–190. http://dx.doi.org/10.1016/j.smallrumres.2016.09.012.

Goddard, M.E. & Hayes, B.J. 2007. Genomic selection. *Journal of Animal Breeding and Genetics*, 124(6): 323–330.

Goddard, M.E. & Hayes, B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6): 381–391. http://dx.doi.org/10.1038/nrg2575.

Grasso, A.N., Goldberg, V., Navajas, E.A., Iriarte, W., Gimeno, D., Aguilar, I., Medrano, J.F., Rincón, G. & Ciappesoni, G. 2014. Genomic variation and population structure detected by single

nucleotide polymorphism arrays in Corriedale, Merino and Creole sheep. *Genetics and Molecular Biology*, 37(2): 389–395.

Greeff, J.., Faure, A.., Minnaar, G.. & Schoeman, S.. 1993. Genetic and phenotypic parameters of pelt traits in Karakul sheep. *South African Journal of Animal Science*, 33: 156–161.

Greyvenstein, O.F.C., Reich, C.M., Van Marle-Koster, E., Riley, D.G. & Hayes, B.J. 2016. Polyceraty (multi-horns) in Damara sheep maps to ovine chromosome 2. *Animal Genetics*, 47(2): 263–266.

Gudex, B., Walker, M., Fisher, P. & Spelman, R. 2014. Validation of a single-nucleotide polymorphism panel for parentage testing of farmed red deer. *Animal Genetics*, 45(1): 142–143.

Hammond, K. 2000. A Global Strategy for the Development of Animal Breeding Programmes in Lower-input Production Environments. In *Animal Genetic Resources*. Rome, Italy: Animal Production and Health Division, FAO. www.agriculture.de/acms1/conf6/ws7globstrat.htm.

Han, J.L., Yang, M., Guo, T.T., Liu, J. Bin, Niu, C.E., Yuan, C., Yue, Y.J. & Yang, B.H. 2016. High gene flows promote close genetic relationship among fine-wool sheep populations (Ovis aries) in China. *Journal of Integrative Agriculture*, 15(4): 862–871. http://dx.doi.org/10.1016/S2095-3119(15)61104-2.

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(4): 635–643.

Hill, W.G. & Weir, B.S. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics research*, 93: 47–64.

Holsinger, K.E. & Weir, B.S. 2015. Genetics in geographically structured populations: defining, estimating and interpreting F ST GenAlEx (integrated software for analysis of genetic data with an interface to Excel). *Nature reviews. Genetics*, 10(9): 639–650. http://www.ncbi.nlm.nih.gov/pubmed/19687804%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4687486%0Ahttp://cmpg.unibe.ch/software/arlequin3BayeScan.

Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2): 254–267.

ISGC. 2010. International Sheep Genomics Consortium. https://www.sheephapmap.org/ 8 February 2020.

Kelleher, M.M., Berry, D.P., Kearney, J.F., McParland, S., Buckley, F. & Purfield, D.C. 2017. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. *Animal*, 11(1): 15–23.

Keller, M.C., Visscher, P.M. & Goddard, M.E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, 189:

237–249.

Kijas, J.W., Lenstra, J.A., Hayes, B., Porto Neto, L.R., Paiva, S., San Cristobal, M., McCulloch, R., Whan, V., Gietzen, K., McEwan, J., Raadsma, H., Boitard, S., Barendse, W., Ciani, E., Dalrymple, B. & Servin, B. 2012. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology*, 10(2): e1001258.

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., Raadsma, H., Barris, W., Bishop, S.C., Coltman, D., Crawford, A., Eggen, A., Erhardt, G., Forage, R., Hanotte, O., Hunt, P., Jianlin, H., Li, K., Marsan, P.A., Miller, J.E., Pemberton, J. & Schibler, L. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE*, 4(3): 1–13.

Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., Mckeigue, P.M. & Wilson, J.F. 2010. Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE*, 5(11): e13996.

Kruger, L. 2009. Dohne Merino. *South African Indigenous breeds*. https://web.archive.org/web/20090829200641/http://www.indigenousbreeds.co.za/local/sheep/dohne 15 March 2020.

Kunene, N.W., Bezuidenhout, C.C. & Nsahlai, I. V. 2009. Genetic and phenotypic diversity in Zulu sheep populations: Implications for exploitation and conservation. *Small Ruminant Research*, 84(1–3): 100–107.

Lathrop, M., Gut, I., Gärke, C., Weigend, S., Ytournel, F., Bed'hom, B. & Simianer, H. 2011. Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Animal Genetics*, 43(4): 419–428.

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.

Louw, M. 2019. Dual Purpose Sheep Breeds in South Africa. *Sheep Farming in South Africa*. http://southafrica.co.za/dual-purpose-sheep-breeds-south-africa.html 9 February 2020.

Maliva, R. & Missimer, T. 2012. Aridity and drought. In *Environmental Science and Engineering (Subseries: Environmental Science)*. Springer Berlin Heidelberg: 21–39.

Mastrangelo, S., Portolano, B., Di Gerlando, R., Ciampolini, R., Tolone, M. & Sardina, M.T. 2017. Genome-wide analysis in endangered populations: A case study in Barbaresca sheep. *Animal*, 11(7): 1107–1116.

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M.T. & Portolano, B. 2016a. Genomic inbreeding estimation in small populations: Evaluation of runs of homozygosity in

three local dairy cattle breeds. *Animal*, 10(5): 746–754.

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

Van der Merwe, D.A., Brand, T.S. & Hoffman, L.C. 2019. Application of growth models to different sheep breed types in South Africa. *Small Ruminant Research*, 178(May): 70–78.

Milne, C. 2000. The history of the Dorper sheep. *Small Ruminant Research*, 36(2): 99–102.

Molotsi, A.H., Taylor, J.F., Cloete, S.W.P., Muchadeyi, F., Decker, J.E., Whitacre, L.K., Sandenbergh, L. & Dzama, K. 2017. Genetic diversity and population structure of South African smallholder farmer sheep breeds determined using the OvineSNP50 beadchip. *Tropical Animal Health and Production*, 49: 1771–1777.

Morin, P.A., Luikart, G. & Wayne, R.K. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, 19(4): 208–216.

Mucina, L. & Rutherford, M.C. 2006. The Biomes and Vegetation of South Africa. *The vegetation of South Africa, Lesotho and Swaziland.*

Naqvi, A.N., Mahmood, S., Vahidi, S.M.F., Abbas, S.M., Utsunomiya, Y.T., Garcia, J.F. & Periasamy, K. 2017. Assessment of genetic diversity and structure of major sheep breeds from Pakistan. *Small Ruminant Research*, 148: 72–79. http://dx.doi.org/10.1016/j.smallrumres.2016.12.032.

Neser, F.W.C., Konstantinov, K. V & Erasmus, G.J. 1995. Estimated genetic trends for weaning weight in three Dorper lines with different selection criteria. *S.Afr.J.Anim.Sci.*, 25(3): 65–69.

Nielsen, R. 2004. Population genetic analysis of ascertained SNP data. *Human genomics*, 1(3): 218–224.

O'brien, A.C., Purfield, D.C., Judge, M.M., Long, C., Fair, S. & Berry, D.P. 2020. Population structure and breed composition prediction in a multi-breed sheep population using genome-wide single nucleotide polymorphism genotypes. *Animal*, 14(3): 464–474.

Oliehoek, P.A. & Bijma, P. 2009. Effects of pedigree errors on the efficiency of conservation decisions. *Genetics Selection Evolution*, 41: 9.

Paradis, E., Claude, J. & Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2): 289–290.

Patterson, N., Price, A.L. & Reich, D. 2006. Population structure and eigenanalysis. *PLoS Genetics*, 2(12): e190.

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

Peter, C., Bruford, M., Perez, T., Dalamitra, S., Hewitt, G. & Erhardt, G. 2007. Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Animal Genetics*, 38(1): 37–44.

Peters, F.W., Kotze, A., van der Bank, F.H., Soma, P. & Grobler, J.P. 2010. Genetic profile of the locally developed Meatmaster sheep breed in South Africa based on microsatellite analysis. *Small Ruminant Research*, 90(1–3): 101–108. http://dx.doi.org/10.1016/j.smallrumres.2010.02.005.

Pickrell, J.K. & Pritchard, J.K. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, 8(11): e1002967.

Prieur, V., Clarke, S.M., Brito, L.F., McEwan, J.C., Lee, M.A., Brauning, R., Dodds, K.G. & Auvray, B. 2017. Estimation of linkage disequilibrium and effective population size in New Zealand sheep using three different methods to create genetic maps. *BMC Genetics*, 18(1): 1–19.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3): 559–575. https://linkinghub.elsevier.com/retrieve/pii/S0002929707613524.

Purfield, D.C., Berry, D.P., McParland, S. & Bradley, D.G. 2012. Runs of homozygosity and population history in cattle. *BMC Genetics*, 13: 70.

Qanbari, S. & Simianer, H. 2014. Mapping signatures of positive selection in the genome of livestock. *Livestock Science*, 166(1): 133–143. http://dx.doi.org/10.1016/j.livsci.2014.05.003.

Quiroz, J., Martinez, A.M., Zaragoza, L., Perezgrovas, R., Vega-Pla, J.L. & Delgado, J. V. 2008. Genetic characterization of the autochthonous sheep populations from Chiapas, Mexico. *Livestock Science*, 116: 156–161.

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): 511–516.

Rasali, D.P., Shrestha, J.N.B. & Crow, G.H. 2006. Development of composite sheep breeds in the world: A review. *Canadian Journal of Animal Science*, 86: 1–24.

Rendo, F., Iriondo, M., Jugo, B.M., Mazón, L.I., Aguirre, A., Vicario, A. & Estonba, A. 2004. Tracking diversity and differentiation in six sheep breeds from the North Iberian Peninsula through DNA variation. *Small Ruminant Research*, 52: 195–202.

Rodríguez-Ramilo, S.T., Elsen, J.M. & Legarra, A. 2019. Inbreeding and effective population size in French dairy sheep: Comparison between genomic and pedigree estimates. *Journal of Dairy Science*, 102(5): 4227–4237.

Rust, J.M. & Rust, T. 2013. Climate change and livestock production: A review with emphasis on

Africa. *South African Journal of Animal Sciences*, 43(3): 256–267.

Ryder, M.L. & Bridbury, A.R. 1984. *Sheep and Man.* London, UK: G. Duckworth & Co.

Sanarana, Y., Bosman, L., van Marle-Köster, E., Maiwashe, A., Visser, C. & Nephawe, K. 2015. Genetic diversity in South African Nguni cattle ecotypes based on microsatellite markers. *Tropical Animal Health and Production*, 48(2): 379–385.

Sandenbergh, L., Cloete, S.W.P., Roodt-Wilding, R., Snyman, M.A. & Bester-van der Merwe, A.E. 2016. Evaluation of the ovineSNP50 chip for use in four South African sheep breeds. *South African Journal of Animal Sciences*, 46(1): 90–93.

Sandenbergh, L., Cloete, S.W.P.W.P., Roodt-Wilding, R., Snyman, M.A.A., Bester-van der Merwe, A.E. & Van der Merwe, A.E. 2016. Genetic diversity and population structure of four South African sheep breeds. *South African Journal of Animal Sciences*, 46(1): 90–93.

Schoeman, S.J. 1998. Genetic and environmental factors influencing the quality of pelt traits in Karakul sheep. *South African Journal of Animal Sciences*, 28(3): 125–126.

Schoeman, S.J., Cloete, S.W.P. & Olivier, J.J. 2010. Returns on investment in sheep and goat breeding in South Africa. *Livestock Science*, 130(1–3): 70–82. http://dx.doi.org/10.1016/j.livsci.2010.02.012.

Selvaggi, M., Dario, C., Peretti, V., Ciotola, F., Carnicella, D. & Dario, M. 2010. Inbreeding depression in Leccese sheep. *Small Ruminant Research*, 89(1): 42–46. http://dx.doi.org/10.1016/j.smallrumres.2009.12.005.

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

Snyman, M.A. 2014a. South African sheep breeds : Namaqua Afrikaner. : 2. gadi.agric.za/InfoPacks/2014023 South African Sheep breeds - Namaqua Afrikaner.pdf.

Snyman, M.A. 2014b. South African sheep breeds : Pedi. *Info-pack ref. 2014/025. Grootfontein Agricultural Development Institute.*

Snyman, M.A. & Herselman, M.J. 2005. Comparison of productive and reproductive efficiency of Afrino , Dorper and Merino sheep in the False Upper Karoo. , 35(2): 98–108.

Snyman, M.A., Olivier, J.J. & Cloete, J.A.. 1993. Productive and reproductive performance of Namaqua Afrikaner sheep. *Karoo Agric*, 5(2): 21–24. http://gadi.agric.za/articles/Snyman_MA/productive.php 15 March 2020.

Snyman, M.A., Olivier, J.J. & Olivier, W.J. 1996. Variance components and genetic parameters for body weight and fleece traits of Merino sheep in an arid environment. *South African Journal of Animal Sciences*, 26(1): 11–14.

Snyman, M.A. & Olivier, W.J. 2002. Productive performance of hair and wool type Dorper sheep under extensive conditions. *Small Ruminant Research*, 45: 17–23.

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(2–3): 112–119. http://dx.doi.org/10.1016/j.smallrumres.2011.09.041.

Stoneking, M. 2001. Single nucleotide polymorphisms: From the evolutionary past. *Nature*, 409(6822): 821–822.

Sun, W., Chang, H., Yang, Z., Geng, R., Tsunoda, K., Ren, Z., Chen, H. & Hussein, M.H. 2007. Analysis on the Origin and Phylogenetic Status of Tong Sheep Using 12 Blood Protein and Nonprotein Markers. *Journal of Genetics and Genomics*, 34(12): 1097–1105.

Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A. & Li, J.Z. 2013. Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics*, 93(1): 90–102.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11): 4414–4423.

Vink, N. 2009. The global sheep industry. *Merino Focus*: 3–6.

Vostry, L., Milerski, M., Schmidova, J. & Vostra-Vydrova, H. 2018. Genetic diversity and effect of inbreeding on litter size of the Romanov sheep. *Small Ruminant Research*, 168(September): 25–31. https://doi.org/10.1016/j.smallrumres.2018.09.004.

Vozzi, P.A., Marcondes, C.R., Bezerra, L.A.F. & Lôbo, R.B. 2007. Pedigree analyses in the Breeding Program for Nellore Cattle. *Genetics and Molecular Research*, 6(4): 1044–1050.

Waikagul, J., Thaenkham, U., Waikagul, J. & Thaenkham, U. 2014. Methods of Molecular Study: DNA Sequence and Phylogenetic Analyses. *Approaches to Research on the Systematics of Fish-Borne Trematodes*, 6(6).

Warmuth, V., Eriksson, A., Bower, M.A., Barker, G., Barrett, E., Hanks, B.K., Li, S., Lomitashvili, D., Ochir-Goryaeva, M., Sizonovg, G. V., Soyonov, V. & Manica, A. 2012. Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21): 8202–8206.

Weir, B.S. & Hill, W.G. 2002. Estimating F-statistics. *Annual Review of Genetics*, 36: 721–750.

van Wyk, J.B., Fair, M.D. & Cloete, S.W.P. 2009. Case study: The effect of inbreeding on the production and reproduction traits in the Elsenburg Dormer sheep stud. *Livestock Science*, 120(3): 218–224. http://dx.doi.org/10.1016/j.livsci.2006.10.005.

Xuebin, Q., Jianlin, H., Lkhagva, B., Chekarova, I., Badamdorj, D., Rege, J.E.O. & Hanotte, O. 2005. Genetic diversity and differentiation of Mongolian and Russian yak populations. *Journal of*

Animal Breeding and Genetics, 122(2): 117–126.

Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. 2011. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1): 76–82. http://dx.doi.org/10.1016/j.ajhg.2010.11.011.

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

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. & Weir, B.S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24): 3326–3328.

Addendum
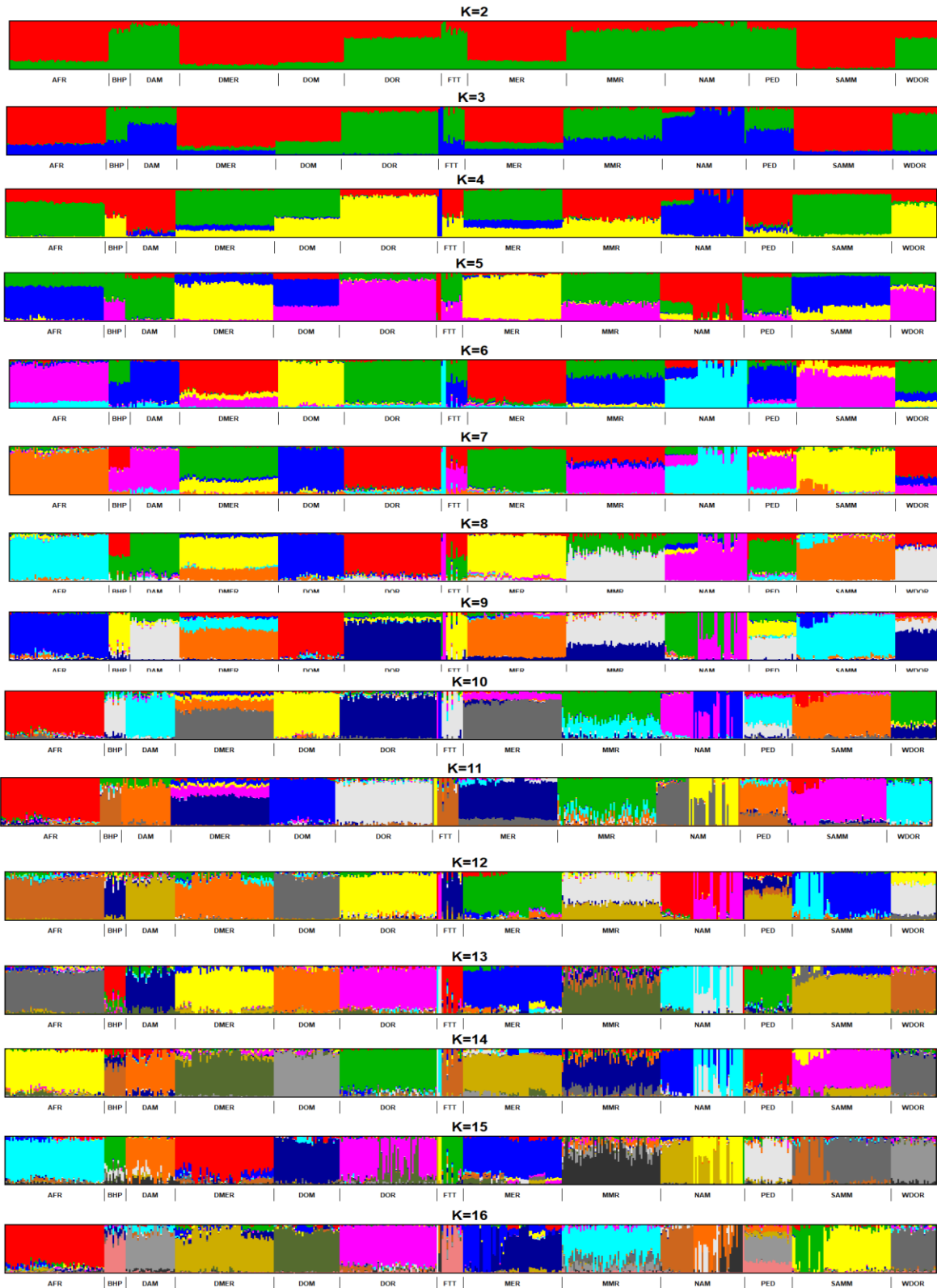
Figure 1A ADMIXTURE plots showing the proportions of ancestral populations for each individual from K=2 to K=16 (AFR= Afrino, BHP= Black Headed Persian, DAM= Damara, DMER= Dohne Merino, DOM= Dormer, DOR= Dorper, FTT= Fat-tail, MER= Merino, MMR= Meatmaster, NAM= Namakwa Afrikaner, PED= Pedi, SAMM= SA Mutton Merino, WDOR= White Dorper)