Genetic diversity and differentiation of Pelt, Mutton and Wool Sheep Breeds of South Africa Using Genome-Wide Single Nucleotide Polymorphisms

by

Edgar Farai Dzomba

Submitted in fulfilment of the academic requirements of

Doctor of Philosophy

in

Animal Science

School of Agricultural, Earth and Environmental Science
College of Agriculture, Engineering and Science
University of KwaZulu-Natal
Pietermaritzburg
South Africa

2021

PREFACE

The research contained in this thesis was completed by the candidate while based in the School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The research was financially supported by University of KwaZulu-Natal, National Research Foundation (NRF-DST).

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

Signed: Edgar Farai Dzomba

Date: 15/5/2021

Signed: Professor Michael Chimonyo

Date: 16 May 2021

DECLARATION 1: PLAGIARISM

Note that two declaration sections are required if there are papers emanating from the dissertation/thesis. The first (obligatory) declaration concerns plagiarism and the second declaration specifies your role in the published papers.

I, Edgar Farai Dzomba declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

_____

Signed: Edgar Farai Dzomba

Date: 15/5/2021

DECLARATION 2: PUBLICATIONS

1. **Dzomba E.F.**, Chimonyo M., Snyman M.A. and Muchadeyi, F.C. (2020) The genomic architecture of South African mutton, pelt, dual-purpose and nondescript sheep breeds relative to global sheep populations. Animal Genetics, 51: 910-923. https://doi.org/10.1111/age.12991

2. **Dzomba, E.F.**, Snyman, M., Chimonyo, M. and Muchadeyi F.C. (2018) Population genetic structure and selection footprints in sheep breeds from divergent production systems of South Africa. Proceedings of the World Congress on Genetics Applied to Livestock Production. Paper 11

3. **Dzomba, E.F.**, Snyman, M., Chimonyo, M. and Muchadeyi F.C. (2016) Assessing the genomic status of South African mutton, pelt and dual purpose sheep breeds using genome-wide single nucleotide genotypes Journal of Animal Science, Volume 94, Issue suppl_4, September 2016, Page 106, https://doi.org/10.2527/jas2016.94supplement4106x

4. **Dzomba, E. F.** and Muchadeyi, F. C. (2019) Medium density beadchip genotype data reveals genomic structure of South African merino-based breeds. Proceedings of the 23rd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Armidale, New South Wales, Australia, 27th October-1st November 2019 pp.452-455

5. **Dzomba E.F.**, Chimonyo M., Pierneef R. and Muchadeyi, F.C. (2021) Runs of homozygosity analysis of South African sheep breeds from various production systems investigated using OvineSNP50 data. BMC Genomics **22,** 7 https://doi.org/10.1186/s12864-020-07314-2

EXTENDED ABSTRACT

Sheep, *Ovis aries*, are a versatile species that has, over hundreds of years, been adapted to South African environmental conditions resulting in more than 40 breeds that are raised for various objectives and production systems and constituting a population of close to 30 million animals. The South African sheep genetic resource presents unique and distinct phenotypes and genotypes that, put together, contribute to the global biodiversity observed in sheep that ought to be conserved and used for improved human livelihoods and economies. South Africa shares its sheep genetics with the global world, through both exportation and importation of germplasm. The broad objective of the study was to profile the genomic architecture of South African sheep populations to provide information for optimal utilization, conservation and improvement. Four hundred South African sheep belonging to 13 breeds of mutton, wool, dual purpose (mutton and wool), pelt and uncharacterised non-descript indigenous sheep were sampled and genotyped. In addition, 623 genotypes from the International Sheep Genomics Consortium representing European, Asian, African sheep breeds were subsampled.

A series of statistical genomic analyses were pursued. In Chapter 3, genetic diversity, population genetic structure and divergence between South African sheep breeds was investigated using the OvineSNP50 Beadchip. A total of 400 sheep belonging to 13 breeds representing mutton, pelt and mutton and wool dual-purpose breeds and Nguni sheep as a representative of indigenous non-descript genotypes were genotyped. To gain a clearer understanding of the genetic diversity of South African breeds relative to other breeds, 623 genotypes from six African, two Asian and eight European breeds were included in the analyses. The study demonstrated low genetic diversity ($H_O \leq 0.27$) in small and geographically restricted populations of Namaqua Afrikaner; Nguni, and Blackhead Persian relative to moderate to high diversity ($H_O \geq 0.38$) in Merino and Merino-derived commercial breeds (*i.e.* Dohne Merino, Australian Merino and Chinese Merino). Overall, the African and Asian populations were the most inbred populations with F_{IS} ranging from 0.17 ± 0.05 in Grey Swakara and Ronderib Afrikaner sheep to 0.34 ± 0.07 in the Namaqua Afrikaner.

Principal component analysis separated the fat-tailed sheep (*i.e.* Swakaras, Nguni, Black Head Persian, Ethiopian Menzi, Meatmaster) from the rump-tailed sheep of Merino and Dorset Horn *etc.*, as well as according to breed history and production systems. Similarly, ADMIXTURE-based clustering revealed various sources of within- and amongst-breed genomic variation associated with production purpose, adaptation and history of the breeds. An analysis of F_{ST}-

based breed differentiating SNPs suggested selection and population divergence on genomic regions associated with growth, adaptation and reproduction. Overall, the analysis gave insight into the current status of the sheep genetic resources of South Africa relative to the global sheep population highlighting both genetic similarities as well as divergence associated with production system and geographical distribution and local adaptation.

The second set of analyses (Chapter 4) focused on assessing the genetic diversity, population structure and breed divergence in 279 animals including the three Merino-derived breeds and five presumed ancestral populations of Merinos and non-Merino founding breeds of Damara, Ronderib Afrikaner and Nguni. Highest genetic diversity values were observed in Dohne Merino with H_o = 0.39 ± 0.01 followed by Meatmaster and South African Merino with H_o = 0.37 ± 0.03. The level of inbreeding ranged from 0.0 ± 0.02 (Dohne Merino) to 0.27 ± 0.05 (Nguni). Analysis of Molecular Variance (AMOVA) showed high within population variance (>80 %) across all population categories. The first Principal Component (PC1) separated the Merino, South African Mutton Merino (SAMM), Dohne Merino and Afrino from the Meatmaster, Damara, Nguni and Ronderib Afrikaner. PC2 aligned each Merino derived breed with its presumed ancestors and separated the SAMM from the Merino and SA Merino. Within population selection based on |*iHS*| indices yielded selection sweeps across the AFR (12 sweeps), Meatmaster (4 sweeps) and Dohne Merino (29 sweeps). Genes associated with hair/wool traits such as *FGF12*, metabolic genes of *ICA1*, *NXPH1 and GPR171* and immune response genes of *IL22 IL26*, *IFNAR1* and *IL10RB* were reported. Other genes included *HMGA* which was observed as a selection signature in other populations, *WNT5A* important in the development of the skeleton and mammary glands, *ANTXR2* associated with adaptation to variation in climatic conditions and *BMP2* which has been reported as strongly selected in both fat-tailed and thin-tailed sheep. Using the Rsb analysis for selection sweeps, the Dohne Merino vs SAMM shared all six sweeps regions on chromosomes 1, 10 and 11 with the comparison for Afrino vs SAMM. Genes such as *FGF12* on OAR 1:191,3-194,7Mb and *MAP2K4* on OAR11:28,6-31,3Mb were observed. The selection sweep on chromosome 10 region 28,6-30,3 Mb, harbouring the *RXFP2* for polledness, was shared between Dohne Merino vs Merino, Meatmaster vs Merino and Meatmaster vs Nguni. The Dohne Merino vs Merino and the Meatmaster vs Merino also shared an Rsb-based selection sweep on chromosome 1 region 268,5 - 269,9 Mb associated with the Calpain gene, *CAPN7*. The study demonstrated some genetic similarities between the Merino and Merino-derived breeds emanating from common founding populations as well as some divergence driven by breed-specific selection goals.

Chapter 5 tested the hypothesis that production systems geared towards specific traits of importance or natural or artificial selection pressures influenced the occurrence and distribution of runs of homozygosity (ROH) in the South African sheep population. The ROH were screened and their distribution within chromosomes and between breeds were analysed to assess breed history and associated selected pressures. ROH were computed at cut-offs of 1-6 Mb, 6-12 Mb, 12-24 Mb, 24-48 Mb and >48 Mb. Analysis of the distribution of ROH according to their size showed that, for all breeds, the majority of the detected ROH were in the short (1-6 Mb) category (88 %). Most animals had no ROH >48 Mb. Of the South African breeds, the Nguni and the Blackhead Persian displayed high ROH based inbreeding (FROH) of 0.31 ± 0.05 and 0.31 ± 0.04, respectively. Highest incidence of common ROH per SNP across breeds was observed on chromosome 10 with over 250 incidences of common ROHs. Mean proportion of SNPs per breed per ROH islands ranged from 0.02 ± 0.15 (island ROH224 on chromosome 23) to 0.13 ± 0.29 (island ROH175 on chromosome 15). Seventeen of the islands had SNPs observed in single populations (unique ROH islands). The MacArthur Merino population had five unique ROH islands followed by Blackhead Persian and Nguni with three each whilst the South African Mutton Merino, SA Merino, White Vital Swakara, Karakul, Dorset Horn and Chinese Merino each had one unique ROH island. Genes within ROH islands were predominantly associated with metabolic and immune response traits and predomestic selection for traits such as presence or absence of horns. In line with observations in Chapter 3, the frequency and patterns of distribution of ROH observed in this study corresponded to the breed history and implied selection pressures exposed to the sheep populations under study.

Chapter 6 investigated (i) LD between adjacent SNPs, (ii) LD decay with increased marker distance, (iii) trends in effective population size over time and (iv) consistency of gametic phase in 13 South African sheep breeds South African Merino (n = 56), Merino (n =10); Mutton Merino (n = 10), Dohne Merino (n = 50), Meatmaster (n = 48), Blackhead Persian (n =14) and Namaqua Afrikaner (n = 12), the four pelt-colour based Swakara subpopulations of Grey (n = 22); Black (n = 16); White-vital (n = 41) and White-subvital (n =17) Dorper (n = 23); Afrino (n = 51) and unimproved Nguni sheep (n = 30). Linkage disequilibrium (r^2) averaged 0.16 ± 0.021and ranged from 0.09 ± 0.14 and 0.09 ± 0.13 observed in the SA Merino and Dohne Merino respectively to 0.28 ± 0.29 observed in the Blackhead Persian sheep. Chromosome 10 had the highest LD with r^2 values ranging from 0.10 ± 0.15 (SA Merino) and 0.12 ± 0.18 (Dohne Merino) to 0.28 ± 0.30 in Blackhead Persian and 0.29 ± 0.30 (SA Mutton Merino). Across the 14 breeds, LD decayed from 0.27 ± 0.30 at 0-10Kb window to 0.02 ± 0.03 at 1000-

2000 Kb window. A progressive decrease in N_e across generations across all populations was observed with effective population size of <500 for all the populations 66 generations ago decreasing to <250, 23 generations ago and well below 100, 13 generations ago. Highest correlations in gametic phase were observed within the 0-10kb window between pairs of Merino and Merino-derived breeds. The highest correlation observed with Nguni sheep was with Dorper sheep (0.33) within the 0-10kb window, which was similar to that observed with Blackhead Persian sheep and Dorper (0.32) again within the same window. The study reported considerable LD persistent over short distance in the South African sheep breeds. The implications of the observed LD, LD decay and consistency in gamete phase on applications such as GWAS, QTL mapping and GS were discussed.

It was concluded that the South African sheep population is highly diverse with that diversity found both within and between populations. Genetic differences between fat tailed sheep population, Merino type breeds and the English Dorset were demonstrated as well as low levels of genetic diversity in small and indigenous breeds such as the Nguni, Namaqua Afrikaner and Blackhead Persian. The frequency and patterns of distribution of ROH observed in this study corresponded to the breed history and implied selection pressures exposed to the sheep populations under study. The utility of the OvineSNP50 Beadchip as a genomic tool for the South African Sheep population was also demonstrated.

Keywords: *Ovis aries; SNP data; genomic structure; production system; selection signatures; ROH*

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Professor Michael Chimonyo, for his priceless tutelage during the entire research process from conceptualisation up until the preparation of this final report. I am eternally gratified and indebted to him, for his criticism and valuable contribution to the completion of this thesis.

Special thanks to Professors Gábor Mészáros and Johannes Sölkner at University of Natural Resources and Life Sciences, Vienna (BOKU) who were kind enough to offer me space in their lab during my sabbatical leave in 2017. I am particularly grateful to them for opening my mind to the vast opportunities to mine genomic data as well as for introducing me to a wide range of methods for data analysis. It was a very fruitful and enjoyable time.

Dr M.A. Snyman at Grootfontein Agriculture Development Institute (Middelburg) and Dr P. Soma at the Animal Production Institute (Irene) are acknowledged for providing some of the DNA samples of sheep that I used in this study. Additionally, I express gratitude to Drs Magriet van Der Nest and Rian Pierneef at the Biotechnology Platform (Ondersterpoort) for their treasured support which was really influential in shaping parts of my experimental methods as well as for critiquing my results.

My gratitude extends to Professors Samson Mukaratirwa, Carola Niesler and Ademola Olaniran at the School of Life Sciences at the University of KwaZulu-Natal for the encouragement and for the funding opportunity which was necessary to undertake parts my studies.

I would like to thank my friends in the animal breeding and genetics field dotted across the world, lab mates at the UKZN and BTP, my students and colleagues at the School of Life Sciences at the UKZN for providing an incredidle environment for our continued growth as researchers.

In undertaking this PhD, I received tremendous support from my family. My wife Farai has been a constant voice of reason and encouragement urging me to keep working, giving great suggestions and advice. I love you to the moon and back Farai. My kids were a source

of inspiration and gave me immesuarable support that made all this possible and worthwhile. Their love, friendship and encouragement has been comforting and amazing.

DEDICATION

This thesis is dedicated to my wife Farai and our children Kumba, Nzira, Zadziso and Chitsidzo, my parents Patrick and Beatrice and my mother-in-law Grace who have all taught me, in their simple ways, the virtues of life. I was surrounded by these people while writing this thesis during an unprecidented and unforgettable time in my life - often puntuated by faltering health, dismal economic outlook and the COVID-19 pandemic. You all made it worthwhile.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Rationale for the research

Domestic sheep (*Ovis* genus and *Ovis aries*) were first domesticated more than 10 000 years ago and have since been bred for a variety of uses including meat, milk and wool production (Taberlet *et al.,* 2011). According to Muigai and Hanotte (2013) African sheep were domesticated outside Africa and share a common ancestry with both European and Asian sheep. Africa has both fat tailed and thin tailed sheep, with the fat-tailed sheep widely distributed in North (Egypt to Algeria) and Eastern and Southern Africa (Eritrea to South Africa) (Bruford and Townsend, 2006; Meadows *et al.,* 2011) whilst the thin-tailed sheep are mainly found in Morocco, Sudan parts of West Africa. The first sheep entered Africa via the Isthmus of Suez and/or the southern Sinai Peninsula, between 7 500 and 7 000 BP (Muigai and Hanotte, 2013). According to mtDNA and Y chromosome analysis, African sheep share a common maternal ancestral origin with a different paternal ancestry for thin and fat-tailed sheep. There are more than 1 000 sheep breeds documented to be in current existence that were created or evolved to serve these diverse purposes and generally classified according to their products (wool, meat, milk, pelt, hides, or as dual/multi-purpose). Over 1 229 were listed on the Food and Agriculture Organization of the UN (FAO) as of 2006 (FAO, 2007).

South Africa has more than 40 sheep breeds distributed across the nine provinces of the country. The South African sheep sector farms with wool (*e.g.* Merino), mutton (*e.g.* Dorper, SAMM, Dormer, Blackhead Persian) and dual-purpose (Dohne Merino) breeds (NDA, 2006). Non-descript genotypes are farmed by smallholder farmers for multi-purpose goals (Qwabe *et al.,* 2015; Molotsi *et al.,* 2017). There is also a small sector farming with Karakul derived breeds for pelt production in Southern Africa (Malesa, 2016).

Sheep are a common small ruminant livestock and are a viable option for food security and improved livelihoods in Southern Africa where the bulk of the agricultural landscape is arid or semi-arid and not suitable for crop farming (Cloete and Schalk, 2010; Molotsi *et al.,* 2020). South Africa has close to 30 million sheep categorised into about 40 breeds and raised for various objectives and production systems. Sheep in South Africa are raised under commercial or smallholder agriculture and consists of breeds and populations selected and reared for mutton, wool, pelt and dual-purposes of either meat or wool. In addition, the country also hosts non-descript ecotypes reared by smallholder farmers under communal farming systems.

Altogether, this diverse genetic resource contributes to household food security, improved livelihood particularly of marginalised communities and national economic growth through their contribution to GDP.

The South African sheep genetic resource presents unique and distinct phenotypes and genotypes that, put together, contribute to the global biodiversity observed in sheep that ought to be conserved and used for improved human livelihoods and economies. South Africa shares its sheep genetics with the global world, through both exportation and importation of germplasm. Breeds such as South African Dohne Merino are farmed in other countries such as Australia, South America and Russia (https://dohnemerino.com/) whilst breeds such as Blackhead Persian are originally from Somalia/Saudi Arabia Region and have been established in South Africa (Synman, 2014). Similarly, the Damara sheep that are successfully farmed in South Africa and other southern African countries are reported to have originated from South East Asia and Egypt (Synman, 2014). Such interactions are however not comprehensively quantified and the extent of influence of the global sheep populations on the South African sheep gene pool is not well characterised.

The major sheep breeds of South Africa include the Merino, Dorper, Mutton Merino, Dohne Merino, Dormer and the Black-headed Persian (https://www.nda.agric.za/docs/Infopaks/Sheep_breeds.pdf). These breeds consist of indigenous breeds as well as exotic breeds that have been bred and reared in the country for an extended time to evolve and adapt to the local conditions. Whilst the Merino breeds were initially imported from Spain, Germany etc, the South African Merino breeds have undergone extensive natural and artificial selection to be considered as indigenous South African breeds. In addition, these Merino breeds have been used to crossbreed to local breeds in the development of composite merino-derived breeds that includes the Dohne Merino, Afrino, Meatmaster, Dormer etc. The distinction of these Merino derived breeds although transparent in the differences in phenotypic performance and morphological traits ought to be characterised and quantified particularly at the genomic level. Such information is valuable particularly for future breed development and improvement programs.

Recently, it has become apparent that the diversity of some populations is decreasing and that many of these breeds are at risk of extinction (Kunene *et al.*, 2007; Qwabe *et al.*, 2015; Molotsi *et al.*, 2017; Selepe *et al.*, 2018). South Africa faces ever greater threats as some of the sheep

breeds like the Namaqua Afrikaner, Blackhead Persian and indigenous ecotypes being at risk of extinction due to intensive production and increased commercial demands contributing to the threats facing improved commercially reared breeds (Soma *et al*., 2012; Synman, 2014; Molotsi *et al.,* 2020). Commercial production is predominantly focused on only a few breeds mostly the Merino and Merino-type breeds, to the detriment of rare or minority breeds of Namaqua Afrikaner, Blackhead Persian indigenous Nguni and Pedi sheep (Molotsi *et al.*, 2020), which are likely to be important genetic resources because of their local adaptation, disease resistance, high fertility and unique product qualities (Kunene *et al*., 2014; Selepe *et al.,* 2018. These minority breeds have also lost some of their genetic purity through population fragmentation leading to loss of genetic diversity through inbreeding and genetic drift as well as introgression into large commercial populations (Qwabe *et al.*, 2013; Molotsi *et al.,* 2020).

The loss of diversity in sheep has important economic, ecological and social implications (Hoffman and Scherf, 2006; Molotsi *et al*., 2017). In response to these threats, there is a need to document the state of the South African sheep population so as to understand the evolutionary history of its breeds. There is also a need to generate data on genetic variation within and among breeds which is essential for decision-making (Rege and Gibson, 2003). Information on within-breed diversity is important for management at the breed level whereas at the among-breed diversity information can be used to identify divergent breeds that may harbour distinct genotypes. This would be useful in assessing the need for conservation of certain breeds.

Until recently, genetic analyses of sheep were performed mainly using maternally inherited mitochondrial DNA and high-polymorphic autosomal microsatellite markers. However, the ascertainment bias caused by typically selecting the most polymorphic markers may lead to reduced sensitivity for judging genome-wide levels of genetic diversity (Vali *et al.,* 2008).

Single nucleotide polymorphisms (SNP) has become the alternative marker due to their wide distribution across the genome, low mutation rate and simplicity to reproduce among different laboratories. Availability of SNP markers of genomewide coverage have facilitated comprehensive characterization of genetic diversity of livestock populations including sheep with provision to make inferences on breed history and evolution, population genetic structure

and viability through parameters such as effective population size and inbreeding coefficient. The development of high throughput sequencing and genotyping technologies also makes it possible to investigate the selective pressures of sheep at the genomic level and identify candidate genes associated with economic traits in order to better understand the mechanisms of adaptive evolution.

South Africa is one of the few countries in Africa with well-established sheep breeds and organized breed improvement programs. These breeds include the Merino and Merino type breeds, Swakara breeds, Dorpers, Ile de France, Blackhead Persian *etc* (Synman, 2014). A number of the commercially farmed breeds are represented by breed societies and there is a coordinated effort to select for and improve on traits of economic and socio-economic importance. Recently resources have been mobilized and efforts made in making genetic improvement in the smallholder sector making use of indigenous and local breeds. Whilst traditional genetic evaluations are well established for a number of commercial breeds, the era of genomic selection and genomics assisted breeding will provide the South African sheep sector to adopt this new technology and attain faster rates of genetic gains even for difficult to measure and novel traits relevant to the smallholder and commercial sector. The challenge faced by the sector is however to optimize the currently available genomic tools to ensure that they cost effectively and accurately attain the required genetic gains considering the architecture of South African sheep sector that is characterized by many breeds or small population sizes. Such characteristics make it difficult to implement for example genomic selection that requires large reference populations to attain accurate predictions.

Rapid development of chip array-based genotyping techniques is providing SNP data which could be useful in uncovering genome-wide structural and functional differences in livestock genomes. The OvineSNP50 Beadchip array was developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC) and provides a comprehensive genomic tool to support diverse sheep genomics programs. This beadchip contains 54 241 SNPs that were chosen for being uniformly distributed across the ovine genome with an average gap size and distance of 50.9 kb and 46 kb, respectively and for their high levels of polymorphism in the more than 75 economically important sheep breeds (OvineSNP50 Datasheet - Illumina

2010). The information provided by the OvineSNP50 Beadchip has found many applications, including genome-wide association studies (GWAS), identification of quantitative trait loci, characterization of genetic variability among breeds, genomic selection and genetic comparison between breeds (Kijas *et al.*, 2009; Alam *et al.*, 2011). With proper optimisation and organisation of breeds and breeding programs, the chip could find utility as a tool for driving genomic selection and genomic assisted breeding of South African sheep. In New Zealand, who have multi-breed sheep breeds contributing to the Gross Domestic Product (GDP), strategies that used the OvineSNP50 Beadchip and multi-breed reference populations were tested.

It is, therefore, appropriate to use the OvineSNP50 Beadchip array to characterise the South African sheep population, in the context of global sheep populations and make inferences on breed history and evolution and investigate the selection pressures that prevailed on the populations during selection and breed formation and their implications on local adaptation. Furthermore, it is important to investigate the utility of the OvineSNP50 Beadchip for multiple genomic applications including feasibility of multi-breed genomics selection and assisted breeding through an investigation of linkage disequilibrium characteristics.

1.2 Objectives and hypotheses

The broad objective of the study was to profile the genomic architecture of South African sheep populations to provide information for optimal utilization, conservation and improvement. The specific objectives were to:

- Investigate genetic diversity, population genetic structure and divergence between breeds of the South African sheep population in the context of global sheep breeds;
- Investigate population genetic structure, breed similarities and divergence of Merino-derived sheep breeds of South Africa; and
- Determine the occurrence and distribution of ROH with the aim to draw insights into how the South African sheep populations were in the past, as well as how their structure and demography have evolved over time;
- Assess the extent of genome-wide linkage disequilibrium, trends in effective population size over time and consistency of gametic phase in the South African sheep population.

The hypotheses tested were:

- The South African sheep sector is genetically diverse, sub structured according to production systems and unique from the global sheep populations and breeds.
- Breed specific production goals and selection pressures, resulted in divergence of breeds from their founding populations producing selection signals as demonstrated in the case of South African merino and merino derived breeds.
- Production system and within-breed selection pressure had pronounced effects on the genome architecture of South African sheep, resulting in reduced genetic diversity and frequency of runs of homozygosity similar to that observed in global populations.
- The OvineSNP50 Beadchip genotypes of the South African sheep population demonstrates high and persistent linkage disequilibrium and correlated consistence in gamete phase between breeds implying utility of this genomic tool for applications such as multi-breed genomic predictions.

1.3 Outline of dissertation/thesis structure

The thesis starts with a general introduction and literature review followed by four experimental chapters written in a manuscript format. Two of the chapters (Chapter 3 and Chapter 5) have been published in international journals whilst Chapter 4 is under peer review in an international journal. The general discussion and conclusions are contained in Chapter 7. The thesis chapters are outlined as:

Chapter 1: Introduces the rationale of the study by highlighting research gaps on which the hypothesis to be tested were formed. It also states the main aim and objectives of the study.

Chapter 2: Reviews current knowledge on the South African sheep genetic resource relative to global sheep populations, and the available genomics and statistical genomic tools for their study.

Chapter 3: The genomic architecture of South African mutton, pelt, dual purpose and non-descript sheep breeds relative to global sheep populations.

Chapter 4: A genomewide divergence and selection signature analysis of South African Merino-derived breeds from their founders.

Chapter 5: Runs of homozygosity analysis of South African sheep breeds from various production systems investigated using OvineSNP50 Beadchip data.

Chapter 6: Linkage disequilibrium pattern, effective population size and persistence in LD phase in multi-breed South African sheep populations.

Chapter 7: General discussion, conclusions and recocommendations of the study findings.

2 REVIEW OF LITERATURE: GENOMIC ARCHITECTURE OF SOUTH AFRICAN SHEEP BREEDS IN THE CONTEXT OF GLOBAL POPULATIONS

Abstract

Sheep represent one of the earliest domestic animals, first domesticated more than 10 000 years ago and have since been bred for a variety of uses including meat, milk and wool production. African sheep were domesticated outside Africa and share a common ancestry with both European and Asian sheep. Sheep play an essential role in the livelihood of people around the world as they are a source of meat, milk, wool, hide and manure, especially in developing countries. South Africa has diverse breeds and populations, which is a reflection of the various production systems, diverse ecology and associated breeding objectives and the different ethnic groups. Much of the diversity in these sheep populations remain uncharacterised. Major developments in sheep genomics have provided tools for generating and analysing genomic data and thus allowing characterisation of breeds and populations as well as making inferences on productive advantage and local adaptation. The review provides an overview of sheep domestication and distribution of diversity globally and in South Africa followed by what is known about the diversity retained in sheep as a species. The review looked at the developments and availability of sheep genomic tools together with the required statistical genomics programs available for the different applications.

Key words: *Ovis aries; Diversity; Domestication; Selection; Genomics tools; Genetic Variants*

2.1 Introduction

Domestic sheep are classified under the *Ovis* genus and *Ovis aries* species and belongs to the ruminant *Bovidae* family. Sheep were first domesticated more than 10 000 years ago and have since been bred for a variety of uses including meat, milk and wool production (Taberlet *et al.*, 2011). Their remains have been found at numerous sites of early human habitation in the Middle East, Europe, and Central Asia. Sheep were one of the first animals domesticated and both archaeological and genetic data spot the domestication centre in eastern Anatolia and North-west Iran. There has been controversy in the systematic analyses of the genus *Ovis*. The wild ancestor of today's sheep are thought to be the Mouflon (*Ovis musimon*) and Urial (*Ovis*

orientalis) with suggestions that the Urial breeds contributed to the formation of all European sheep breeds while the Mouflon are the progenitors of the breeds found in the Caucasus, northern Iraq, north-western Iran and Anatolia regions (Scherf, 2000). The Asiatic Mouflon are still found in the mountains of Asia Minor and Southern Iran while the European Mouflon are existing on the islands of Sardinia and Corsica (Taberlet *et al.*, 2011). The difference between these two are in coat colouration and horn configuration (Yeomans *et al.*, 2017). It is hypothesised that the European mouflon originated from descendants of feral Asiatic mouflon (Rezaei *et al.*, 2010). Muigai and Hanotte (2013) provided evidence that African sheep were domesticated outside Africa and share common ancestry with both European and Asian sheep. Africa has both fat tailed and thin tailed sheep, with the fat-tailed sheep widely distributed in North (Egypt to Algeria) and Eastern and Southern Africa (Eritrea to South Africa) (Bruford and Townsend, 2006; Meadows *et al.*, 2011). Thin-tailed sheep are mainly found in Morocco, Sudan parts of West Africa. African sheep were domesticated outside Africa. Muigai and Hanotte (2013) indicated that the first sheep entered Africa via the Isthmus of Suez and/or the southern Sinai Peninsula, between 7 500 and 7 000 BP and mitochondrial DNA analyses supports a common maternal ancestral origin for all African sheep, while autosomal and Y chromosome DNA analyses indicates a distinct genetic history for African thin tailed and sub-Saharan fat-tailed sheep. African sheep breeds migrated with various nations from Asia, Arabia and the Middle East into North Africa (Budarum, 2006). Epstein (1971) highlighted that nomadic black and coloured tribes inhabited North Africa many years ago and were restricted in that region due to tsetse fly breakout stretching along the equator across the whole of Africa. It was only recently (~2000 years ago) that sheep migrated to the Cape (Plug and Badenhorst, 2001).

Domestic sheep are multi-purpose animals serving diverse purposes. There are more than 1 000 breeds documented to be in current existence that were created or evolved to serve these diverse purposes. The Food and Agriculture Organization of the UN (FAO) listed 1 229 breeds as of 2006 (FAO, 2007). Sheep are generally classified according to their products as either wool, meat, milk, pelt, hides, or as dual/multi-purpose breeds. Chessa *et al.* (2009) suggested that sheep were first reared for meat and later for wool and milk between 4 000 and 5 000 years ago. There are about 20 sheep breeds in South Africa distributed across the nine provinces of the country (National Development Agency [NDA], 2006). The South African sector farms with wool (*e.g.* Merino), mutton (*e.g.* Dorper, SAMM, Dormer, Blackhead Persian) and dual-purpose (Dohne Merino) breeds (NDA, 2006). Non-descript genotypes are farmed by

smallholder farmers for multi-purpose goals (Qwabe *et al.*, 2015; Molotsi *et al.*, 2017). There is also a small sector farming with Karakul derived breeds for pelt production in Southern Africa (Malesa, 2016).

Domesticated sheep are also categorised according to morphological traits such as tail length and shape, face colour and presence or absence of horns (Muigai and Hannotte, 2013). South Africa farms both fat and thin tailed sheep. Breeds and populations are also classified according to the geographical topography for which the breed has been developed or is reared (Brown *et al.*, 1996). The Bonga and Horro are examples of highland breeds raised in Ethiopia (Gebretsadik, 2014).

Within the broad array of phenotypes raised in diverse production systems and for the different goals is genetic diversity, the majority of which still remain to be characterised and tapped into. This review examined the sheep production systems in South Africa and in the global context and explore our current knowledge of the genetic diversity within and between breeds, the level of population divergence and associated selection signals. The review also describes the genomic tools available for studying and maintaining both neutral and functional diversity in these populations. The review ends with an overview of the statistical genomic tools currently applied to provide relevant inferences on these populations.

2.2 Global sheep production

Sheep play an essential role in the livelihood of households around the world, as they are a source of meat, milk, wool, hide and manure, especially in developing countries (Ramsay *et al.*, 2000; Pollot and Wilson, 2009). Currently, the global sheep population stands at more than 1 billion head with 19 per cent found in Asia followed by Africa (https://thesheepsite.com/focus/5m/99/global-sheep-meat-market-thesheepsite). In Africa, sheep numbers are highest in Sudan Ethiopia (60.9 million); (30.7 million), and Tanzania (26.4 million) (Hedge, 2019; FAOSTAT, 2019). In spite of other regions having larger sheep numbers, Australia and New Zealand continue to dominate sheep product exports to other countries. China continues to increase imports for fine and course wool, and the United States and China remain net importers of sheep meat (Takadi *et al.*, 2020). Asia, Africa and Latin America consume sheep products within the region of production, and Africa and Latin America do not currently receive any sheep product imports (Takadi *et al.*, 2020).

2.2.1 Sheep production in South Africa

Sheep farming is of considerable economic importance to South Africa and is represented throughout the country. The variable climates and landscapes have favoured the adoption of a wide diversity of sheep breeds of differing functions that are adapted and performed to different objectives under a range of production systems.

There are approximately 8 000 commercial sheep farms and about 5 800 communal sheep farmers in South Africa (DAFF, 2013). The country has an estimated 28.8 million sheep which contributes to household food security and GDP (DAFF, 2013). In South Africa, where the majority of the land is not suitable for crop production, sheep farming allows sustainable production in extensive pastoral areas where no alternative farming ventures can be practiced. Examples of such areas are the vast extensive Karoo regions of the central part of South Africa. Sheep production is dominant in the dry regions i.e. in 21 % of the total area of the Free State, 38 % of Western Cape, 51 % of Eastern Cape and 82 % of Northern Cape (AGIS, 2007; Cloete and Olivier, 2010).

The sheep industry in South Africa is divided into commercial, emerging commercial and smallholder subsistence farmers. The industry is dominated by the commercial sheep farmers who own more than two thirds of the sheep in the country and supply meat products locally and wool products for export. Many communal communities depend on sheep for their existence and the maintenance of a viable local society. Without the income from sheep products many of these communities would cease to exist (Cloete and Olivier, 2010).

There are a number of sheep breeds in South Africa that include wool sheep, dual-purpose breeds and mutton breeds (https://agribook.co.za/livestock/indigenous-and-locally-developed-breeds/). Most of the indigenous sheep breeds are neglected and at the risk of extinction including breeds such as the Namakwa Afrikaner and Ronderib Afrikaner with less than 1000 breeding animals left (Qwabe *et al.,* 2013; Molotsi *et al.,* 2020). South Africa has a considerable complement of locally developed composite breeds such as the Dorper, Dormer, Afrino, Meatmaster (Buduram, 2004; Peters *et al.,* 2010; Snyman, 2010a-e). The non-descript sheep such as Nguni and Pedi are in high demand due to their adaptive traits (Kunene *et al.,* 2014; Selepe *et al.,* 2018; Maqhashu *et al.,* 2020). The sheep breeds reared in South Africa are a mixture of hairy indigenous genotypes, fat-tailed and fat-rumped genotypes (http://southafrica.co.za/indigenous-sheep-breeds-south-africa.html).

For well-established breeds, sheep farmers are represented by organizations and the Dorper Sheep Breeders' Society of South Africa and Merino SA are the most prominent. South African farmers farm for different products ranging from wool, meat, pelt and dual purpose functions. The smallholder farming sector farm for substance needs which on its own is multifaceted *i.e.* food security, socio-economic security and cultural purposes. Although sheep farming is done all over South Africa, it is traditionally concentrated in the more arid regions and the industry is vital to the rural and arid regions (Cloete and Olivier, 2010).

Several indigenous fat-tailed and fat-rump breeds are reared in Southern Africa (Farm Animal Conservation Trust, 2003) are associated with specific indigenous tribes and production systems (Kunene *et al.,* 2014; Molotsi *et al.,* 2018; Selepe *et al.,* 2018). Such breeds include Ronderib Afrikaner, Damara, Namaqua Afrikaner, Nguni and Pedi in South Africa and the Damara.

2.3 Sheep genetic diversity

Sheep have, through domestication and subsequent selection, adapted to thrive in diverse environments (Kijas *et al.,* 2012). During domestication, animals were recruited from the wild into a captive environment, a process that changed their morphology, behaviour, and genetics resulting in a broad spectrum of breeds specialised for diverse production needs i.e. wool, milk, and meat were produced (Chessa *et al.,* 2009). Diversity was also created through intensive selection for improved productivity which saw the division of animals into breeds (Fariello *et al.,* 2014). Whilst geographical isolation has created divergence that ultimately lead to new breeds, the global word have also allowed gene flow between breeds leading to the creation of composite and crossbreeds (Kijas *et al.,* 2012). The diversity in sheep breeds is demonstrated by the wide range of phenotypes. Much of this genetic diversity has, however, not been fully described. There are still many sheep populations which have not yet been studied.

2.3.1 Global population diversity

Zelder *et al.* (2006) argues that sheep represent one of the earliest domestic animals. The Asian and European Mouflon have observed differences in the mtDNA haplogroups they carry with the European Mouflon (*Ovis musimon*) carrying the B haplogroup (Hiendleder *et al.,* 2002; Bruford *et al.,* 2003). Hiendleder *et al.,* (2002) pointed out that sheep descend from one or more Asiatic mouflon (*Ovis orientalis*) populations that are mostly aligned with the *C* haplogroup (Meadows et al., 2010). Kijas *et al.,* (2012) analysed 2819 DNA samples from sheep belonging

to 74 breeds sampled across the world, using slightly over 50 000 SNP markers and observed that sheep breeds have maintained high levels of genetic diversity, in contrast to other domestic animals such as dogs. Meadows *et al.* (2005) observed high levels of gene flow between domestic sheep breeds of Asia and Europe. The implementation of quantitative genetics methodology, and the use of artificial insemination to prioritise genetically superior rams and more recently genomic selection associated with breeding and improvement of animals post domestication, has seen an increase in the number of breeds and overall global diversity of sheep (Kijas *et al.,* 2012; Deniskova *et al.,* 2019). Such dynamics have however resulted in within breed diversity due to intense selection within breed defined standards (Kristensen *et al.,* 2015). Domestication and subsequent selection of sheep over thousands of years has produced a spectrum of breeds specialized for the production of wool, milk, meat, pelt and other dual-purpose uses (Kijas *et al.,* 2012; Gouveia *et al.,* 2014). Parallel to that has been changes in sheep morphology, behaviour, and genetics through domestication and subsequent selection. Al-Atiyat *et al.* (2018) concluded that the genetic diversity observed in sheep and other livestock is a mirror of the ecology or geographical region of existence of these livestock. Globally genetic diversity Genetic diversity has been described in many ecotypes and local breeds (Qwabe *et al.,* 2015; Selepe *et al.,* 2018; Deniskova *et al.,* 2019) whereby results considered local breeds as reservoirs of genetic diversity that they have evolved over time (Taberlet *et al.,* 2008). Local and indigenous breeds present strong adaptations to harsh conditions, nutritional fluctuations, and resistance to diseases and parasites (Kunene *et al.,* 2007; Mukhongo *et al.,* 2014; Wei *et al.,* 2016). Human socio-cultural and economic networks have also been observed to shape the genetic diversity of sheep and other livestock populations (Tisdell, 2003). Mwangi (2013) indicated that socio-economic and environmental factors such as population growth, urbanization and economic development, market demands, contributes to the changes in livestock numbers, which will directly affect diversity.

2.3.1.1 Genetic diversity of South African sheep populations

South Africa has diverse breeds and populations, which is a reflection of the various production systems, diverse ecology and associated breeding objectives and the different ethnic groups (Cloete and Olivier, 2010; Molotsi *et al.,* 2020). A summary of the South African breeds and their characteristics is found in Table 2.1. Wool production in South Africa is drawn mainly from Merino and other merino derived breeds and Karakul, but coarse and coloured types are also produced and marketed on a limited scale (DAFF, 2015). Different Sheep breeds are raised for either mutton, wool, pelt or as dual wool/mutton breeds (Cloete *et al.,* 2014; Soma *et al.,* 2012).

Table 2.1: List of South African Sheep Breeds and their attributes*

Breed	Category	Major Production System and Traits
SA Merino	Indigenous	Commercial, Dual purpose
SA Mutton Merino	Indigenous	Commercial, Dual purpose
Namaqua Afrikaner	Indigenous	Commercial, Hair
Blackhead Persian	Indigenous	Commercial, Meat breed
Nguni	Indigenous	Smallholder Meat
Dormer	Composite	Commercial, Dual purpose
Van Rooy	Composite	Commercial, Meat breed
Dohne Merino	Composite	Commercial, Dual purpose
Meatmaster	Composite	Commercial, Meat breed
Dorper	Composite	Commercial, Meat breed
Afrino	Composite	Commercial, Dual purpose
Pedi	Indigenous	Smallholder, Meat
Zulu	Indigenous	Smallholder, Meat
Damara	Indigenous	Commercial, Meat
White Vital Swakara	Exotic	Commercial, Pelt
White Sub-Vital Swakara	Exotic	Commercial, Pelt
Black Swakara	Exotic	Commercial, Pelt
Grey Swakara	Exotic	Commercial, Pelt
Karakul	Exotic	Commercial, Pelt
Ile De France	Exotic	Commercial, Meat breed
Suffolk	Exotic	Commercial, Dual purpose

- Primary information from Synman, 2014

South Africa has over 20 sheep breeds consisting of indigenous, exotic meat and crossbreds (Cloete and Olivier, 2010). Smallholder communal farmers keep non-descript uncharacterised breeds for multiple purposes ranging from being a source of meat, income generation through sales and as a source of security (Molotsi *et al*., 2017). Commercial sheep enterprises utilize exotic and synthetic breeds that have been purebred and selected for certain phenotypic traits, like colour and size, for example, the Dorper breed (Olivier and Cloete, 2006).

The South African Mutton Merino (SAMM) is a dual-purpose wool and meat sheep breed developed in South Africa from the German Merino and used primarily for wool but also for

meat production (Schoeman *et al.*, 2010). The Dohne Merino, Merino and Dorper breeds are the most prominent breeds of sheep in South Africa. Other sheep breeds found in the South African sheep industry include the Dormer, Ile de France, Meatmaster, Namaqua Afrikaner, Afrino, Merino Landsheep and South African Mutton Merino. The Namaqua Afrikaner is a hardy, indigenous fat-tailed breed that is indigenous to South Africa and primarily used in smallholder farming systems (Qwabe *et al.,* 2013). The Blackhead Persian were initially introduced as a hair breed that is able to tolerate heat better than wool breeds. The breed is raised primarily for meat in South Africa. Karakul sheep are known to have originated from Central Asia and are a multi-purpose breed, kept for milk, meat, pelts, and wool. Swakara sheep were derived from Karakul through intensive selection and are found in Namibia, South Africa and Botswana where they are raised predominantly for pelt production (Malesa, 2015; Muchadeyi *et al.,* 2015).

Using the Merinos and other indigenous breeds such as the Blackhead Persian, Namaqua Afrikaner, and Damara from east Africa, Merino derived breeds such as the Dohne Merino, Afrino, Meat Master were developed with the objective to improve the breed's robustness to harsh and challenging environments while maintaining good production performance in meat and wool traits. The Afrino for example is a synthetic dual-purpose breed consisting of Merino, Ronderib Afrikaner and SA Mutton Merino and thrives in the harshest conditions of South Africa. The Meatmaster was created through the crossing of South African Meat Merino and Damara sheep and is associated with improved meat characteristics. The Dorper, which is now numerically the second largest breed in South Africa and also found all over the world is a composite breed developed from interbreeding the Dorset Horn and the Persian breeds (Cloete and Olivier 2010). The Dohne Merino sheep were developed through intensive selection and interbreeding of Peppin-style Merino ewes and German Mutton Merino rams, whilst selecting for high fertility, fine Merino wool and good performance under commercial rangeland conditions and parasite infested environments.

A number of molecular studies have been conducted to characterise genetic diversity in South African sheep breeds. Due to intensive selection for specific production goals, the commercial breeds are genetically structured separate from the admixed indigenous sheep populations kept by smallholder farmers (Sandenbergh, *et al.,* 2015; Molotsi *et al.*, 2017; Molotsi *et al.,* 2020). Because of small effective population sizes and intensive selection, the purebreds generally accumulate homozygous loci for deleterious genes over time and resulting in reduced fitness, poor survival and low reproduction (Goddard, 2009). This decrease in genetic diversity through

linebreeding and artificial selection has been sustained in most purebreds because of the resultant increase in growth and production traits (Swanepoel *et al.,* 2007; Van Wyk *et al.,* 2009). In South African sheep populations, Molotsi *et al.* (2017) reported highest inbreeding coefficient in purebreeds of Namaqua Afrikaner, Dorper and South African Mutton Merino relative to smallholder populations of Dorper. Soma *et al.* (2012) observed higher levels of genetic diversity in other indigenous fat tailed sheep of Pedi, Swazi and Zulu.

Smallholder farmers make use of the non-descript indigenous breeds that are hardy or robust to survive in the local and compromised environments (Snyman *et al.,* 2005; Cloete *et al.,* 2013; Cloete *et al.,* 2016). Molotsi *et al.* (2017) suggested that the robustness of smallholder populations is attributed to their high genetic diversity that proves advantageous for their fitness. Kunene *et al.* (2009), for example, explained that the Nguni sheep have acquired high adaptation to harsh environmental conditions, resistance against diseases and parasites and the ability to walk long distances attributed to them carrying diverse and unique genotypes. Indigenous Nguni sheep breeds have a higher genetic diversity than Merinos using RAPD markers (Hlophe, 2011). Other smallholder breeds such as the Namaqua Afrikaner are endangered and at the risk of extinction (Makina *et al.,* 2015). Genetic diversity studies have reported low heterozygosity levels within the indigenous Namaqua Afrikaner breed (Sandenbergh *et al.,* 2016). The low levels of heterozygosity for this breed could be linked to its small effective population size (*Ne*). Low levels of heterozygosity have also been reported in other indigenous fat rumped sheep of Black headed Persian and Red head Speckled Persian (Soma *et al.,* 2012).

2.3.1.1.1 Population divergence and selection footprints

Breeding goals, management practises, structured mating programmes and natural selection pressures (geographic and climatic factors) have successfully imposed selection resulting in divergence of sheep into breeds and populations that are adapted to a diverse range of environments and to specialised production conditions (Kijas *et al.,* 2012; Gouveia *et al.,* 2014). Selection alters allele frequencies within the target population for both functional mutation(s) and their neighbouring SNP (Rochus *et al.,* 2018). Domesticated sheep have been exposed to artificial selection for the production of fibre, meat, and milk as well as to natural selection that are likely to have imposed distinctive selection signatures on the sheep genome (Fariello *et al.,* 2014; Paim *et al.,* 2018). The current large spectrum of sheep phenotypic diversity observed in sheep results from the combined effects of selection of sheep for different production traits such as wool, milk and meat, and its natural adaptation to new environments

(Otsanda *et al.*, 2018). A number of studies have been undertaken in different sheep populations investigating selection signals and inferring adaptation to local conditions.

A genome scan for selection signals in worldwide sheep populations by Kijas *et al.* (2012) revealed 31 regions containing genes for coat pigmentation, skeletal morphology, body size, growth, and reproduction. Strong selection signals on genes such as *RXFP2* have occurred in response to breeding for the absence of horns whilst selection for specialised coat pigmentation on genes such as *KIT, ASIP* and *MCIR* also represented breed-defining characteristics in sheep (Kijas *et al.*, 2012; Yang *et al.*, 2013) and other livestock (Randhawa *et al.*, 2016; Kim *et al.*, 2016; Bertolini *et al.*, 2018). Coat colour genes such as *KIT*, *ASIP*, and *MITF* have been observed to be under selection (Kijas *et al.*, 2012). Fariello *et al.* (2014) confirmed selection for morphology, colour and adaptation to new environments using Comparative genomic analysis between domestic sheep and Asian mouflon with regions that were enriched for genes involved in bone morphogenesis, growth regulation, and embryonic and neural development in domestic sheep (Wang *et al.*, 2019). Yurchenko *et al.* (2019) performed a high-density genotyping and comprehensive scans for signatures of selection in the genomes from 15 local sheep breeds reared across Russian populations and demonstrated that the genomes of Russian sheep breeds contain multiple regions under putative selection for traits related to morphology, adaptation, and domestication. The study also reported multiple genes that are putatively related to environmental adaptations were top-ranked in selected intervals (*e.g.*, *EGFR, HSPH1*, *NMUR1*, *EDNRB, PRL*, *TSHR*, and *ADAMTS5*). In China and many other tropical production systems, short tail is the result of both artificial and natural selection favouring specific set of genetic mutations. Zhi *et al.* (2018) analyzed the genetic differences between short-tail and normal-tail phenotypes and identified selection signals associated with vertebral development amongst other developmental mechanisms. Diseases and parasites present selection pressures on livestock production systems. McRae *et al.* (2014) undertook a study investigating selection and underlying genetics for resistance to gastrointestinal nematodes in Romney and Perendale sheep and reported candidate genes involved in chitinase activity and the cytokine response. Selection and adaptation to extreme environments associated with climate change was demonstrated in Chinese native sheep breeds using whole genome sequence data (Yang *et al*., 2016). Ostanda *et al.* (2018) scanned the genome of 25 Sasi Ardi and 75 Latxa sheep from the Western Pyrenees and reported selection in regions underlying local adaptation traditional dairy selection pressure and regions experiencing the specific effect of the modern genetic improvement program established for the Latxa breed during the last

three decades. Manzari *et al.* (2019) demonstrated selection in regions that play a role in skeletal system and tail, sugar and energy metabolisms, growth, reproduction, immune and nervous system traits going back to the domestication of Iranian sheep breeds reporting multiple candidate genes, such as *HOXB9, HOXB13, ACAN, NPR2, TRIL, AOX1, CSF2, GHR, TNS2, SPAG8, HINT2, ALS2, AAAS, RARG, SYCP2, CAV1, PPP1R3D, PLA2G7, TTLL7* and *C20orf10* were reported. Zhao *et al.* (2018) conducted detection of selection signatures in Sunite, German Mutton, and Dorper sheep and reported genomic regions subjected to positive selection and harbouring genes involved in muscle development, growth, and fat metabolism. Jiang *et al.* (2014) developed and analysed high quality reference genomes and transcriptomes of sheep and identified genes associated with rumen evolution to be under evolution. Overall, the studies that looked at selection signatures provided insights into evolutionary forces imposing selection pressures in sheep which is also common in other livestock species. The studies point to morphology *i.e.* horn presence and absences and shape, coat colour patterns etc. of sheep being under pressure as they are linked to adaptive and production traits. Production traits and adaptation traits are generally under selection as they influence survival and optimal productivity of sheep under different production conditions.

2.4 Genomic tools in sheep

The sheep genome was one of the first livestock genome to be sequenced in 2017 (https://www.ncbi.nlm.nih.gov/assembly/GCF_002742125.1/), after which a number of genomic tools evolved and have been improved to ensure integration of genomics in sheep research and development, characterization and production. These tools range from next generation sequencing (NGS) and SNP genotyping tools.

2.4.1 Next Generation Sequencing

2.4.1.1 *Ovis aries* (Sheep) reference genomes

The sheep (*Ovis aries*) genome was produced by the International Sheep Genome Consortium (ISGC) initially using a single Texel ewe and a single Texel ram (https://www.sheephapmap.org/). The current version of the sheep reference genome (*Ovis aries* v4.0) has seen improvements in the last decades from (i) Oar_v1.0 established in 2010, which was based on 454 data guided by the bovine genome; (ii) Oar_v2.0 based on Illumina data from the ewe from CSIRO, followed by gap-filling using Illumina data from the ewe and the ram by BGI; (iii) Oar_v3.1 available in 2012 from NCBI and CSIRO incorporated

additional Illumina data and MeDIP-seq, BAC data and 454 data to fill gaps in the Oar_v2.0 assembly and (iv) the current version of Oar_v4.0 available from NCBI incorporated Pacific Biosciences long read data to fill gaps in the Oar_v3.1 assembly (https://www.sheephapmap.org/).

Table 2.2 provides the global statistics of the *Ovis aries* v4.0 which when compared to other livestock species (i.e. goat) is fairly a good reference genome.

Table 2.2: Global statistics of *Ovis aries* v4.0

Total sequence length	2,615,499,683
Total ungapped length	2,587,499,057
Gaps between scaffolds	0
Number of scaffolds	5,465
Scaffold N50	100,009,711
Scaffold L50	8
Number of contigs	48,481
Contig N50	150,472
Contig L50	5,008
Total number of chromosomes and plasmids	27
Number of component sequences (WGS or clone)	48,481

The availability of a good and continuously improving reference genome for sheep is a major asset to the sheep community, allowing genomic applications and development of other more routine genomic tools like genome-wide SNP genotyping. Although initially built from a single breed (Texel), further versions have incorporated information from other breeds. The reference genomes have been used in a number resequencing projects undertaken in studying other breeds but overall adding to the genomic information available to the community. For example, Yang *et al.* (2016) conducted whole genome resequencing of Chinese native breeds to investigate adaptation to extreme environments. The reference genome Oar v3.1.75 was used for mapping and this study provided insights into rapid genomic adaptations to extreme environments in sheep and other animals, and presented a valuable resource for future research on livestock breeding in response to climate change.

Boloorma *et al.* (2019) investigate the feasibility of imputing genotype data to WGS, also using the Oar v3.1.75 as a reference for mapping. Wang *et al.* (2019) conducted deep genome resequencing of Chinese domestic sheep and revealed the role of artificial and natural selection in visual deterioration, plateau adaptability and high prolificacy.

Current initiatives to improve on the reference genomes include building a sheep genomes database comprising more than 100 million genomic variants from nearly 500 sheep breeds. In this project transcript sequences are being used to improve the Texel sheep assembly (Oar_v4.0).

2.4.1.2 Single nucleotide polymorphisms

The OvineSNP50 Beadchip is a comprehensive genome-wide genotyping array for sheep, providing superior power to interrogate genetic variation across many breeds. The BeadChip was developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC), comprising leading researchers from AgResearch, Baylor, UCSC, and Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO) (https://sheephapmap.org/news/datasheet_ovinesnp50.pdf). Featuring more than 54 000 SNPs that uniformly span the entire ovine genome, the OvineSNP50 Beadchip enables a broad range of applications such as GWAS (Zhang *et al.,* 2013; Al-Mamun *et al.,* 2015; Cao *et al.,* 2020) signature of selection (Zhu *et al.,* 2015), identification of quantitative trait loci (QTL) (Hazard *et al.,* 2014; Atlija *et al.,* 2016; Matika *et al.,* 2016; Usai *et al.,* 2019), genomic selection and breeding (Auvray *et al.,* 2014; Phua *et al.,* 2014; Mrode *et al.,* 2018), and breed characterization for evaluating biodiversity (Malesa, 2015; Molotsi *et al.,* 2017; Deniskova *et al.,* 2018). The OvineSNP50 Beadchip covers SNPs validated in many economically important breeds, including more than 75 *Ovis aries* breeds. The bead chip has been validated in a number of South African breeds (Sandenberg *et al.,* 2016; Malesa *et al.,* 2015; Molotsi *et al.,* 2017).

Ascertainment bias and low genome-wide linkage disequilibrium have been the major limitations of the OvineSNP50 Beadchip. Ascertainment bias has been reported particularly in local and indigenous breeds that were not represented in the SNP discovery using a limited genome assembly and validation (Sandenbergh *et al.,* 2016; Edea *et al.,* 2017; Prieur *et al.,* 2017). Generally, sheep have low genome-wide LD attributed to their derivation for a genetically diverse pool (Kijas *et al.,* 2012). Low LD has been reported based on the OvineSNP50 Beadchip genotypes in Sicilian dairy sheep breeds (Mastrangelo *et al.,* 2014);

highly diverse New Zealand sheep breeds (Prieur *et al.*, 2017) and Merino and merino derived breeds (Kijas *et al.*, 2014; Al-mamun *et al.*, 2015).

The amount of LD and extend of its decay has direct implications on use of SNPs chips for applications such as QTL mapping, GWAS and genomic selection (Qanbari *et al.,* 2010). Lower levels of LD which rapidly decays with increased marker distances implies that a denser SNP panel is required for genomic applications. The low LD levels that persist over very short SNP distances requires a higher density SNP panel for genomic applications i.e. diversity studies, GWAS, QTL mapping and genomic selection, who then recommended the use of the commercially available Ovine 600K SNP panel particularly for highly diverse breeds (Kijas *et al.*, 2014; Brito *et al.*, 2017). Very limited studies (Edea *et al.*, 2017; Rochus *et al.*, 2020) have, however, used this panel to provide a basis for comparison due mainly to the higher cost of using this panel over the SNP50K.

2.4.2 Statistical tools

Using SSRs, RAPDs and other low coverage genetic markers, statistical analysis for assessing genetic diversity were limited to heterozygosity estimates (H_o, H_e), inbreeding co-efficient, population fixation indices (*i.e.* Weir and Cockerham F_{ST} Statistics) and population clustering. In the advent of genomics, other tools that allow finer resolution of genomes have been availed. These include statistical tools such as Runs of homozygosity (RoH), Copy Number Variants (CNVs); Haplotypes and Haplotype Blocks and linkage disequilibrium (LD) based analysis.

2.4.2.1 Linkage Disequilibrium (LD) and LD decay

Linkage disequilibrium (LD) reflects the extent of non-random association between any two markers and underpins selection decisions in a wide range of livestock species that have adopted genetic technologies for selection purposes (Al Mamum *et al.*, 2015). Marker-assisted selection (MAS), genomic selection and GWAS all largely depend on the extent of LD within a population. It is the extent of LD that determines the minimum number of markers required for a successful genome-wide study; with LD remaining high over longer chromosomal segments, fewer markers are needed. Conversely, denser panels are required if LD decays rapidly. The pattern of LD decay also provides information on the evolutionary history of a population and can be used to estimate, *e.g.*, the ancestral N_e (Hayes *et al.*, 2003; Crispim *et al.*, 2013). With the advent of high density single nucleotide polymorphism (SNP) maps, it may be possible to fine map genes by exploiting linkage disequilibrium between genes of interest

and adjacent markers. However, the extent of linkage disequilibrium (LD) is generally unknown for livestock populations (McRae *et al.*, 2002). The levels of LD influences the power of QTL detection and accuracy of genomic predictions (Goddard, 2009). The LD levels indicate the minimum number of markers for successful genomic predictions. There is little knowledge about the degree of genome-wide LD in the sheep breeds included in this investigation. Comparison of the extent of LD between breeds is therefore informative about the overall diversity level in a species and can help us understand the patterns of selection that individual breeds have been subjected to. Due to its importance, various studies have reported LD estimates in various livestock species, *e.g*. cattle (McKay *et al.,* 2007; Porto-Neto *et al.,* 2014) pig (Uimari *et al.,* 2011), horse (Corbin *et al.,* 2010); chicken (Rao *et al.,* 2008) and sheep (Kijas *et al.,* 2012; Meadows *et al.,* 2008).

Brito *et al.* (2017) used LD and change in gametic phase to understand the genetic diversity and population history of New Zealand multi-breed sheep population using a collection of 74 sheep breeds and 49,034 SNP from the OvineSNP50 Beadchip. Kijas *et al.* (2012) observed a high variation in LD levels among breeds, with a Scottish breed (Soay) presenting the highest levels of LD and Qezel sheep (sampled in Iran) the lowest levels of LD. Using the HD SNP chip, Kijas *et al*. 2012 reported LD levels at 10 kb of 0.186, 0.191, 0.279, 0.221 and 0.339 for Merino ewes, Merino sires, Poll Dorset, Suffolk and Border Leicester, respectively. Prieur *et al.* (2017) estimated linkage disequilibrium and the associated effective population size in New Zealand sheep using three different methods to create genetic maps. Alvarenga *et al.* (2018) characterized LD structure in Brazilian Santa Inês sheep a commercial breed of importance for meat production, reproductive efficiency, and tropical adaptation.

Consistency of gametic phase influences the use of genomic tools across breeds. The more distant the relationship between individuals, the shorter the genomic distance over which the phase will be consistent. In a study on consistency of gametic phase, Brito *et al.* (2017) observed high consistency of gametic phase in closely related breeds of Lamb Supreme and Texel compared to distantly related breeds of Primera and Dual-Purpose breed-groups. Mdladla *et al.* (2016) observed a similar trend in South African goat species. In New Zealand, multiple closely related sheep populations are combined into multi-breed reference populations for genomic selection (Auvray *et al.,* 2014; Brito *et al.,* 2017). The accuracy of such multibreed based predictions is dependent on the consistency in gamete phase between populations (Hayes *et al.,* 2009; Cole *et al.,* 2016; Steyn *et al.,* 2019).

2.4.2.2 Haplotypes

The livestock genome has a haplotype block structure, such that it can be divided into discrete blocks of limited haplotype diversity with each block consisting of "tag SNPs," that are used to distinguish a large fraction of the haplotypes and in association studies. Studies indicated that haplotype-based analysis can be much more powerful than marker-by-marker analysis. Ghoreishi (2018) observed 1 217 haplotype blocks that covered 1.1 % of the genome (on average consisting of 2.24 ± 0.77 SNP and being 23.6 ± 46.4 kb in length) in Iranian sheep breeds. Ren *et al.* (2016) used linkage disequilibrium (LD) analyses and haplotype-based association tests, to map a genomic region comprising 132.0–133.1 Mb on chromosome 2 that contained the top 10 SNPs (including 4 significant SNPs) and 5 most significant haplotypes associated with the polycerate phenotype in sheep. Zhao *et al.* (2019) reported 14 960 haplotype blocks spanning 7 851 kb of the genome in a Chinese Meishan sheep population.

High LD and haplotype blocks have been used as proxies for selection footprints. It is based on this concept that LD based signature of selection analysis methods of |iHS| (Voght *et al.,* 2006), Rsb (Tang *et al.,* 2007) adaptation of sheep to the West African environment and XP-EHH (Sabeti *et al.,* 2002) have been derived and found application in livestock diversity studies. Using the OvineSNP50 Beadchip and *iHS* scores, Kim *et al.* (2016) reported selection signatures spanning several genes associated with traits for adaptation to hot arid environments, body size and development, energy and digestive metabolism and nervous and autoimmune response in tropical goats and sheep. Alvarez *et al.* (2020) used both his scores and XP-EHH on OvineSNP50 Beadchip genotypes and reported selection sweeps on genes associated with metabolic response to stress, including regulation of oxidative and metabolic stress and thermotolerance in line with adaptation of sheep to the West African environment. Zhao *et al.* (2016) screened the Chinesse Sunite, German Mutton and Dorper sheep using OvineSNP50 Beadchip genotypes and XP-EHHH scores and reported genomic regions harboured by genes, involved in muscle development, growth, and fat metabolism.

2.4.2.3 Runs of homozygosity

Domestication and the subsequent selection of animals for either economic or morphological features often leave imprints on the genome of a population. Genomic regions subjected to high selective pressures demonstrate reduced genetic diversity and frequent runs of homozygosity (ROH) (Purfield *et al.,* 2017), defined as long and continuous stretches of homozygous genotypes that result from parents transmitting identical haplotypes to their offspring (Ku *et*

al., 2011). ROH have been used as estimates of inbreeding (Zhang *et al.,* 2013). The availability of dense SNP assays allows for the determination of autozygous segments based on runs of consecutive homozygous genotypes (ROH) (Mastrangelo *et al.,* 2018). Purfield *et al.* (2017) used ROH based on OvineSNP50 Beadchip data to screen for genomic regions subjected to selection pressure in 3,191 sheep from six commercial meat breeds. Genomic regions associated with skin pigmentation, body size and muscle formation, characteristics that are often sought after in modern-day breeding programs were identified. Mastrangelo *et al.* (2018) screened for ROHs in Italian sheep breeds and used them to characterize autozygosity and identify genomic regions that frequently appeared in ROH within individuals (ROH islands). The effects of both ancient and recent inbreeding on the genome of the Italian sheep breeds were highlighted. Signer-Hasler *et al.* (2019) used a dataset consisting of both medium density SNP chip genotypes and high density genotypes to quantify genomic inbreeding based on runs of homozygosity and reported private signatures in proximity of genes/QTL influencing body size, coat colour and fatty acid composition in adapted breeds. Haile *et al*. (2019) suggested the use of ROH based genomic inbreeding estimate as a tool to control the level of inbreeding in smallholder sheep populations of Ethiopia. Meyermans *et al.* (2019) used the OvineSNP50 Beadchip to compare genetic diversity and relationship of the Belgian milk sheep in Flanders (Belgium) with the Flemish Sheep, the Friesian Milk Sheep, the French Lacaune dairy sheep and other Northern European. Recent studies have found associations between regions of extended homozygosity with both complex and simple gene disorders (Kijas *et al.,* 2012; Suarez-Vega *et al.,* 2013). Muchadeyi *et al.* (2015) used the OvineSNP50 Beadchip to screen for ROH and study the genetics of sub-vital performance in a pelt producing sheep breed of southern Africa.

2.4.2.4 Copy Number Variation

Copy number variation (CNV) caused by gene rearrangement is an important part of genomic structural variation (Jiang *et al.,* 2019) representing a major source of genomic variation (Yang *et al.,* 2017) and increasingly recognized as an important and abundant source of genetic variation and phenotypic diversity in livestock species (Liu *et al.,* 2013). Fontanesi *et al.* (2011) reported the first comparative map of CNVs of the sheep genome which was derived through cross-species array comparative genome hybridization (aCGH) with cattle. With advances in sheep genomic tools, follow up studies investigated genome-wide CNV in diverse sheep populations using the OvineSNP50 Beadchip data. Liu *et al.* (2013) constructed a sheep CNV map based on the OvineSNP50 Beadchip and screened CNVs in three sheep breeds of Chinesse

Sunite, German Mutton and Dorper. Yang *et al.*, (2017) identified 24,558 putative CNVs, merged into 619 CNV regions, and corresponding to ~ 6.9% of the genome of sheep from Africa, America, Asia, South-western Asia, Central Europe, Northern Europe and south-western Europe. Differences in CNV distribution between diverse breeds and populations linked to the population history of different sheep breeds were highlighted. Jiang *et al.* (2019) reported copy number variations in the region of the Src Homology 2 Domain Containing E (SHE) gene in a known QTL related to milk fat percentage and bone density in sheep.

2.5 Summary

Sheep play an essential role in the livelihood of people around the world as they are a source of meat, milk, wool, hide and manure, especially in developing countries. South Africa has an estimated 28.8 million sheep raised under approximately 8 000 commercial sheep farms and 5 800 communal farms through breeding and selection a number of breeds are reported in South Africa. Smallholder communal farmers on the other hand make use of the non-descript indigenous breeds that are hardy or robust to survive in the local and compromised environments. South Africa has diverse breeds and populations, which is a reflection of the various production systems, diverse ecology and associated breeding objectives and the different ethnic groups. The diverse gene pool, evidenced by multiple breeds raised under diverse production systems, needs to be characterised. Unveiling the influence the genetic history of the breed, production systems and environmental conditions on the genomic archicterure of the sheep populations will be useful in (i) understanding the genomic architecture of the breeds itself and (ii) aiding in informing future breed improvement programs. Analysis such as screening and dissecting the distribution of ROHs and profiling of signatures of selection are therefore crucial in understanding the complex genomic architecture of South African sheep populations. The South African gene pool is connected to the global sheep population through founder effects (some exotic breeds were used to establish SA sheep population) and constant gene flow through importation of exotic breeds and exportation of South African breeds to other countries. The genetic relationships between SA sheep breed and global populations as a result of the outlined factors need to be investigated. A number of molecular studies have been conducted to characterise genetic diversity in South African sheep breeds. These tools range from next generation sequencing (NGS) and SNP genotyping tools. The OvineSNP50 Beadchip has been validated in a number of South African breeds and presents a comprehensive genome-wide genotyping array for sheep, providing superior power

to interrogate genetic variation across many breeds. In the advent of genomics, other tools that allow finer resolution of genomes have been availed including LD based methods for QTL mapping and screening of selection footprints, haplotypes and haploblock analysis, ROHs and CNVs. Despite improvement made in sheep genomics and availability of both genomic and statistical genomic tools, a lot of breeds remain uncharacterized and there is still a gap in the information on the South African sheep gene pool, its genomic architectures and the forces of evolution that have shaped it.

3 THE GENOMIC ARCHITECTURE OF SOUTH AFRICAN MUTTON, PELT, DUAL PURPOSE AND NON-DESCRIPT SHEEP BREEDS RELATIVE TO GLOBAL SHEEP POPULATIONS[1]

[1]This manuscript is published in Animal Genetics: *Dzomba, E.F., Chimonyo, M., Snyman, M.A. and Muchadeyi, F.C. (2020), The genomic architecture of South African mutton, pelt, dual-purpose and nondescript sheep breeds relative to global sheep populations. Anim Genet, 51: 910-923.* https://doi.org/10.1111/age.12991

Abstract

South Africa has a diverse array of phenotypically distinct and locally adapted sheep breeds that have been developed for different production systems ranging from mutton, wool, pelt and some dual purpose and non-descript breeds kept by smallholder farmers. The current study investigated genetic diversity, population genetic structure and divergence between South African sheep breeds in order to gain an insight into breed history and genomic architecture aligned to breeding goals and production systems. The OvineSNP50 Beadchip was used to genotype 400 sheep belonging to 13 breeds representing mutton, pelt and mutton and wool dual-purpose breeds. Nguni sheep were included as a representative of indigenous non-descript breeds that are reared by smallholder farmers. Seeking a clearer understanding of the genetic diversity of South African breeds relative to global populations, 623 genotypes of sheep from worldwide populations were included in the analysis. These sheep breeds included six African, two Asian and eight European breeds. Across breeds, genetic diversity ranged from observed heterozygosity (H_O) = 0.26 ± 0.02 in Namaqua Afrikaner to observed heterozygosity (H_O) = 0.38 ± 0.01 in Dohne Merino, Australian Merino and Chinese Merino. The overall mean heterozygosity was 0.35 ± 0.04. The African and Asian populations were the most inbred populations with F_{IS} ranging from 0.17 ± 0.05 in Grey Swakara and Ronderib Afrikaner sheep to 0.34 ± 0.07 in the Namaqua Afrikaner. The South African Dohne Merino (F_{IS} = 0.03 ± 0.01), SA Merino (F_{IS} = 0.05 ± 0.04) and Afrino (F_{IS} = 0.09 ± 0.02) and other global Merino-derived breeds were the least inbred. The first principal component explained 28 % of the variation and separated the fat tailed sheep (*i.e.* Swakaras, Nguni, Black Head Persian, Ethiopian Menzi, Meatmaster) from the rump tailed sheep of Merino and Dorset horns. The second PC separated the Merino and derived breeds from the English breeds of Dorset Horns. Overall, South African indigenous breeds clustered together with indigenous breeds from other African and Asian countries. The optimal ADMIXTURE cluster (K = 20) revealed various sources of within and amongst breed genomic variation associated with production purpose, adaptation and history of the breeds. The Blackhead Persian, Nguni and Namaqua Afrikaner breeds differed significantly from other breeds particularly with the South African Mutton Merino and Dorset Horn. Breed differentiating SNPs were observed within genomic regions associated with growth, adaptation and reproduction. The current findings give insight into the current status of the sheep genetic resources of South Africa relative to the global sheep population highlighting both genetic similarities as well as divergence associated with production systems and geographical distributions.

Key words: *Sheep, SNP genotypes, breed diversity, production system, population structure*

3.1 Introduction

South Africa has a vibrant sheep industry contributing considerably to the total livestock gross domestic product through products such as lamb, wool and pelts (Cloete *et al.,* 2014). Several industrial breeds are reared mainly on commercial farms with extensively raised breeds found mainly in the smallholder areas (Molotsi *et al.,* 2017). South African sheep breeds are raised for either mutton, wool, pelt or as dual wool/mutton breeds (Soma *et al.,* 2012; Cloete *et al.,* 2014). Communal farmers keep non-descript uncharacterised breeds for multiple purposes ranging from being a source of meat, income generation through sales and as a source of security (Molotsi *et al.,* 2017). South Africa has over 20 sheep breeds consisting of indigenous, exotic and crossbreds.

The South African Mutton Merino (SAMM) is a dual-purpose wool and meat sheep breed developed in South Africa from the German Merino and used primarily for wool but also for meat production (Schoeman *et al.,* 2010). The Namaqua Afrikaner is a hardy, indigenous fat-tailed breed that is indigenous to South Africa and primarily used in smallholder farming systems (Qwabe *et al.,* 2013). The Blackhead Persian were initially introduced as a hair breed that is able to tolerate high ambient temperatures better than wool breeds. The breed is raised primarily for meat in South Africa. Karakul sheep are known to have originated from Central Asia and are a multi-purpose breed kept for milk, meat, pelts, and wool. Swakara sheep were derived from Karakul through intensive selection and are found in Namibia, South Africa and Botswana where they are raised predominantly for pelt production (Malesa 2015; Muchadeyi *et al.,* 2015).

Using the Merinos and other indigenous breeds such as the Blackhead Persian, Namaqua Afrikaner, and Damara from east Africa, Merino derived breeds such as the Dohne Merino, Afrino and Meat Master were developed with the objective to improve the breed's robustness to harsh and challenging environments while maintaining good production performance in meat and wool traits. The Afrino, for example, is a synthetic dual-purpose breed consisting of Merino, Ronderib Afrikaner and SA Mutton Merino and thrives in the harshest conditions of South Africa. The Meat Master was created through crossing of SA Mutton Merino and Damara sheep and is associated with improved meat characteristics. The Dorper, which is now

numerically the second largest breed in South Africa and also found all over the world, is a composite breed developed from interbreeding the Dorset Horn and the Persian breeds (Cloete and Olivier 2010). Dohne Merino sheep were developed through intensive selection and interbreeding of Peppin-style Merino ewes and German Mutton Merino rams, whilst selecting for high fertility, fine Merino wool and good performance under commercial rangeland conditions and parasite infested environments.

In addition to these well-defined indigenous, exotic and composite breeds, are the non-descript indigenous sheep kept by smallholder farmers in the marginal areas of South Africa. These uncharacterised sheep are often referred to as ecotypes and named after the ethnic groups under whom they are raised with examples such as Nguni and Pedi sheep.

The genetic diversity and population structure of South African sheep populations is thought to have been shaped by both natural and artificial selection processes when breeds were developed for both adaptive and functional traits that are now expressed in the diverse array of phenotypically-distinct breeds (Soma *et al.*, 2012). Whilst sheep genetic resources of South Africa have spread to other countries globally, they have also been movement of genotypes into South Africa. Simultaneous farmer- driven crossbreeding initiatives led to new sheep breeds flourishing that were better suited to produce optimally under the harsh production conditions of the country. Such processes, which are largely poorly documented, have shaped the genomic architecture of the current gene pool of South African sheep.

The OvineSNP50 Beadchip array was developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC) to support diverse sheep genomics programs. This beadchip contains 54 241 SNPs that were chosen for being uniformly distributed across the ovine genome with an average gap size and distance of 50.9 kb and 46 kb, respectively; and for their high levels of polymorphism in the more than 75 economically important sheep breeds (OvineSNP50 Beadchip Datasheet - Illumina 2010). The information provided by the OvineSNP50 Beadchip has found many applications, including genome-wide association studies (GWAS), identification of quantitative trait loci, characterization of genetic variability among breeds, genomic selection and genetic comparison between breeds (Kijas *et al.*, 2009; Alam *et al.*, 2011). In South Africa, the chip has found utility in characterising breeds from different production systems (Sandenbergh *et al.*, 2016; Molotsi *et al.*, 2017) with potential applications in genomic breeding as well (Cloete *et al.*, 2014). The objective of the current study was to investigate genetic diversity, population genetic structure and divergence

between breeds in order to gain an insight into breed history and genomic architecture of the various populations. Although the genetic diversity and population structure of South African sheep breeds have been explored previously using microsatellite markers (Soma *et al.,* 2012; Qwabe *et al.,* 2013), and using the OvineSNP50 Beadchip for a few isolated populations (Sandenbergh *et al.,* 2016; Molotsi *et al.,* 2017), a fine-scale investigation of a broad array of sheep breeds from the diverse spectrum of production systems (mutton, wool, pelt and dual-purpose and non-descript smallholder), was deemed necessary to confirm the overall genetic diversity, breed structure and uniqueness of the South African gene pool amongst itself and relative to other global populations.

3.2 Materials and methods

3.2.1 Animal Material

A total of 400 sheep belonging to 14 South African breeds/populations consisting of mutton, pelt wool and dual purpose breeds were used in this study. The list of breeds, their production systems are presented in Table 2.1. The mutton breeds were of South African Merino (n = 56), Mutton Merino (n = 10), Dohne Merino (n = 50), Meat Master (n = 48), Blackhead Persian (n = 14) and Namaqua Afrikaner (n = 12). The four pelt-colour based Swakara subpopulations of Grey (n = 22); Black (n = 16); White-Vital (n = 41) and White-Subvital (n =17) where used as local pelt producing breeds while the Karakul (n =10) was sampled from Halle (Germany) as an ancestral population of the Southern African Swakara breeds. Dual-purpose breeds used for both mutton and wool were the Dorper (n = 23) and Afrino (n = 51) and we also included the unimproved Nguni sheep (n = 30) in the experiment (Table 3.1). The commercial meat and wool breeds were sampled from Grootfontein Agricultural Development Institute (GADI) biobank and other commercial farms in the Eastern Cape and Northern Cape provinces of the country (Soma *et al.,* 2012). The Swakara sheep were sampled from Swakara pelt-producing farms in Namibia and from the Northern Cape province of South Africa. The Nguni is a non-descript indigenous sheep of South Africa reared by communal farmers in the KwaZulu-Natal region of South Africa from where it was sampled.

3.2.2 Genotyping and SNP quality control

The 400 sheep were genotyped using the Illumina OvineSNP50 Beadchip using the infinium assay platform at the Agricultural Research Council-Biotechnology Platform in South Africa. SNP genotypes were called using genotyping module integrated in GenomeStudio™ V2010.1

(Illumina Inc.). The SNPs were pruned for missing genotypes (GENO) > 0.95, genotype calling (MIND) > 0.95, minor allele frequency (MAF) > 0.01 and deviation from Hardy-Weinberg Equilibrium (HWE > 0.0001) resulting in 36 975 markers and 400 individuals remaining for further analysis using PLINK (Purcell *et al.*, 2007). SNPs were pruned per population using the same parameters as the whole population in addition to pruning of SNPs that were in LD >0.45 prior to linkage disequilibrium analysis. The dataset used of ADMIXTURE, PCA and Treemix analysis was pruned for identity by descent of > 0.45 to result in 29 465 SNPs. The resultant SNPs and individuals per population are given in Table 3.1.

3.2.3 Global sheep populations

An additional 623 genotypes from a global set of sheep breeds representing worldwide populations were included in the analysis. These sheep breeds included six African breeds of African Dorper (n = 21), African White Dorper (n = 6); Ethiopian Menz (n = 34), Namaqua Afrikaner (n = 12), Red Maasai (n = 45) and Ronderib Afrikaner (n = 17); two Asian population of Bangladesh Garole (n = 24) and Karakas (n = 18); and eight European breeds of Australian Poll Dorset (n = 108); Australian Industry Merino (n = 50), Australian Poll Merino (n = 98); Chinese Merino (n = 23), MacArthur Merino (n = 10); Dorset Horn (n = 21); Merino Landschaf (n = 24) and Black Headed Muttom (n = 24). The dataset was accessed with permission from the International Sheep Genome Consortium (ISGC) http://www.sheephapmap.org.

3.2.4 Data analyses

Minor allele frequency (MAF) distribution, heterozygosity and inbreeding estimates MAF and proportion of polymorphic SNPs were estimated per population using PLINK (Purcell *et al.*, 2007) under default settings. The mean MAF and standard deviation per population was calculated using the *PROC MEANS* procedure of the Statistical Analysis System (SAS) v 9.3 (SAS institute Inc., 2013). SNPs were categorized as either fixed (MAF = 0), rare (0<MAF<0.01), intermediate (0.01<MAF <0.05) or common (0.05< MAF <0.5). All SNPs with MAF > 0.05 were considered polymorphic.

Table 3.1: SNP and population quality control for genotyping rate, minor allele frequency and deviation from Hardy Weinberg equilibrium plus the resultant SNPs for downstream analysis

Breed	No. animals after QC	GENO 0.95	MAF 0.95	HWE 0.0001	Resulting SNPs
SA Merino	56	3115	7815	239	46193
SA Mutton Merino	10	3140	12965	0	41276
Dohne Merino	50	3119	7528	244	46477
Meatmaster	45	3143	8692	1403	44862
Dorper	22	3118	9086	331	44824
Afrino	51	3125	10641	89	43511
Namaqua Africana	10	3030	11722	0	40691
Blackhead Persian	14	3030	17305	8	35100
Nguni	30	3151	19471	763	34006
White Vital Swakara	40	3146	12189	866	41213
White Sub-Vital Swakara	17	3147	12215	100	41926
Black Swakara	16	3145	10933	97	43211
Grey Swakara	22	3145	12036	95	42110
Karakul	10	3103	15100	0	37313
African Dorper	21	0	7043	0	41991
African White Dorper	6	0	12038	0	36996
Merino*	138	0	2320	1229	45509
Australian Poll Dorset	108	5	4968	1202	42918
Australian Poll Merino	98	0	2359	5	45490
Chinese Merino	23	0	4326	1	44707
Karakas	18	0	5079	0	43955
Namaqua Afrikaner	12	0	17672	0	31362
Red Maasai	45	0	6811	6	42217
MacArthur Merino	10	0	18849	0	30185
Dorset Horn	21	0	11055	11	37968
Merino Landschaf	24	0	4346	148	44540
Ronderib Afrikaner	17	0	9543	480	39011
Ethiopian Menz	34	0	7396	113	41525
Black-Headed Mutton	24	0	5924	18	43092
Bangladeshi Garole	24	0	9954	7	39073
Overall	1023	0	1060	18298	29465

Per population proportion of observed and expected homozygous and inbreeding coefficients were estimated using Plink v.1.7 (Purcell *et al.*, 2007).

3.2.5 Population structure analysis

A principal component analysis (PCA) was used to assess the population structure using Golden Helix SNP Variation Suite (SVS) Version 8.1 (Golden Helix Inc., Bozeman, MT, 2012). In addition, ADMIXTURE 1.21 was used to infer the most probable number of ancestral populations based on the SNP genotype data (Alexander *et al.*, 2009). ADMIXTURE was run

from K =2 to K = 33 and the optimal number of clusters (K value) was determined as that which had the lowest cross validation error (CV-error).

The TreeMix software package (Pickrell and Pritchard, 2012) was employed for phylogenetic analyses to investigate interpopulation migration and gene flow among the complete set of 33 populations. A number of trees were plotted assuming n = 1-8 migration events and based on 29 465 SNPs and default parameter settings.

3.2.6 Per marker per population F_{ST}

To determine breed-differentiating SNPs, an outlier loci approach based on calculation of fixation index (F_{ST}) at different significance levels was used as a measure of genetic differentiation for each locus between highly differentiated breeds ($F_{ST} \geq 0.28$). The loci under selection were expected to show an allele frequency that deviates from that of neutral loci with an F_{ST} threshold ≥ 0.8.

3.2.7 Annotation of breed differentiating SNPs

A gene was considered breed-differentiating if it contained an unexpectedly high proportion of highly differentiated SNPs based on F_{ST} values ≥ 0.8. Annotated genes within ±1MB region of the significant SNPs were identified from National Centre for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov) databases. The sheep QTL database https://www.animalgenome.org/cgi-bin/QTLdb/OA/summary was also referenced for published QTLs within ±1MB region of the significant SNPs across breeds.

3.3 Results

3.3.1 Population parameters after quality control

Table 3.1 summarises the SNP and population statistics following standard quality control after genotyping. The global data set from the ISGC was already filtered standard quality criteria and genotyping rates using the similar parameters. South African indigenous breeds of Namaqua Afrikaner, Blackhead Persian, Nguni and the Swakara pelt producing breeds lost more (> 10 000) SNPs due to MAF (threshold of 0.05) relative to commercial breeds of Merino and its derivatives, and the Dorpers. The South African Nguni breed, which is non-descript breed, kept by smallholder farmers in South Africa had the largest (19471) SNPs below the

0.05 MAF thresholds. As a result, these indigenous breeds and the other African and Asian populations from the global dataset had less number of resultant SNPs for downstream analysis.

3.3.2 Minor allele frequency distribution

Using data that were quality controlled for only genotyping rate, the MAF distribution of all populations was further analysed and results are presented in Table 3.2. The proportion of fixed SNPs (MAF = 0) was high in South African indigenous breeds (*i.e.* Namaqua Afrikaner, Blackhead Persian, Nguni) and Karakul.

Table 3.2: Minor allele frequency distribution and average MAF per population

Breed	Fixed	Very rare	Rare	Moderate	Frequent	Total	Mean MAF	SD
SA Merino	1748	6648	8643	10364	23723	51126	0.2695	0.14470
SA Mutton Merino	9824	4155	8702	8618	19801	51100	0.2146	0.16210
Dohne Merino	1920	6059	8846	10338	23960	51123	0.2686	0.14480
Meatmaster	878	9520	9340	10174	21209	51121	0.2533	0.14620
Dorper	2867	6647	9809	9244	22556	51123	0.2582	0.14790
Afrino	3869	7579	8634	9668	21373	51123	0.2489	0.15250
Namaqua Africana	16863	0	9219	9769	15242	51093	0.1960	0.17420
Blackhead Persian	15993	6467	9401	4450	14783	51094	0.1714	0.16320
Nguni	13316	7065	7480	7614	15618	51093	0.1849	0.16550
White Vital Swakara	4271	9290	9059	9191	19287	51098	0.2324	0.15420
White Sub-Vital	5695	10436	8247	8979	17737	51094	0.2198	0.15630
Black Swakara	2535	11731	8978	8767	19085	51096	0.2268	0.15240
Grey Swakara	4451	8679	8449	10682	18837	51098	0.2352	0.15400
Karakul	11997	4436	4171					
African Dorper	3049	7959	8532	8986	20507	49033	0.2619	0.14600
African White Dorper	12037	5564	7229	6711	17492	49033	0.2111	0.16580
Merino	317	4978	8555	10380	24804	49034	0.2887	0.13470
Austraian Poll Dorset	1082	7738	8825	9688	21695	49028	0.2619	0.14600
Australian Poll Merino	395	4995	8598	10251	24789	49028	0.2880	0.13480
Chinese Merino	1527	6132	9478	8842	23055	49034	0.2693	0.14300
Karakas	3143	6145	9675	8152	21919	49034	0.2567	0.14870
Namaqua Afrikaner	14633	6655	5582	7813	14350	49033	0.1848	0.17000
Red Maasai	1937	9174	9206	9347	19369	49033	0.2401	0.15020
MacArthur Merino	18848	3216	6122	6505	14342	49033	0.1631	0.16870
Dorset Horn	4745	11271	8333	7726	16959	49034	0.2162	0.15640
Merino Landschaf	1762	5503	8945	10595	22229	49034	0.2162	0.15640
Ronderib Afrikaner	6599	8572	7465	9915	16482	49033	0.2195	0.15850
Ethiopian Menz	3375	7684	9115	9616	19243	49033	0.2421	0.15240
Black-Headed Mutton	2360	7129	9276	10038	20230	49033	0.2524	0.14840
Bangladeshi Garole	6553	6815	8637	9071	17957	49033	0.2266	0.15890
Overall	**40**	**3340**	**8011**	**11089**	**26223**	**48703**	**0.3034**	**0.12530**

The Nguni (13 316), Karakul (11 997) SA Muttom Merino (9 824) and Namaqua Afrikaner (8 692) had the highest of number of fixed SNPs. The average MAF was high (MAF = 0.29 ± 0.13) in Merino and Australian Poll Merino breeds and least (MAF =0.16 ± 0.16 -0.19 ± 0.17) in MacArthur Merino and South African breeds of Blackhead Persian, Namaqua Afrikaner and Nguni in the and other indigenous SA and African breeds.

3.3.3 Heterozygosity estimates and inbreeding coefficients

Across breeds, genetic diversity ranged from H_0 = 0.26 ± 0.02 (Namaquar Afrikaner) to H_0 = 0.38 ± 0.01 (Dohne Merino) with an overall mean of 0.35 ± 0.04 (Table 3.3). Inbreeding levels were high in the South African indigenous i.e. Namaqua Afrikaner (F_{IS} = 0.33 ± 0.07); Blackhead Persian (F_{IS} = 0.29 ± 0.04); Nguni (F_{IS} = 0.29 ± 0.04) and Pelt producing Swakara Sheep (F_{IS} = 0.17 in GVS – F_{IS} = 0.25 ± 0.05 in BVS). Contrary to other merinos, the MCM had high inbreeding of F_{IS} =0.32±0.14). Low inbreeding was observed in the South African Dohne Merino (F_{IS} = 0.03 ± 0.01), Afrino (F_{IS} = 0.09±0.02), Meat Master (F_{IS} = 0.09 ± 0.02) and Merino (F_{IS} = 0.05 ± 0.04) which however was elevated relative to the other global populations of Merino (F_{IS} =0.02 ± 0.03), Australian Poll Dorset (F_{IS} = 0.06 ± 0.04) and Chinese Merino (F_{IS} = 0.03 ± 0.05).

3.3.4 Population clustering based on PCA analysis

The first principal component, with 28 % of the total variation, separated animals into two main of (i) fat tailed sheep that included the Karakuls and Swakaras; South African indigenous breeds of Nguni and Namaqua Afrikaner; the Black Head Persian and Ethiopian and Kenyan breeds and (ii) rump tailed sheep of Merino and merino derived breeds and the Dorset horns (Figure 3.1). The second Principal Component explained approximately 20 % of the total variation and clustered the breeds according to their function and historical origin splitting the different Merino and Merino derived specialised breeds from the Dorset Horn. Lower PCs separated geographically restricted breeds such as the South African Nguni and the Black Headed Mountain breed of Central Europe from the other populations.

Table 3.3: Population sample size, homozygosities and inbreeding coefficient

Breed	No. animals	F_{IS}	H_O	H_E
SA Merino	56	0.05 ± 0.040	0.37 ± 0.010	0.39 ± 0.000
SA Mutton Merino	10	0.14 ± 0.040	0.33 ± 0.010	0.39 ± 0.000
Dohne Merino	50	0.03 ± 0.010	0.38 ± 0.010	0.39 ± 0.000
Meatmaster	45	0.09 ± 0.020	0.35 ± 0.010	0.39 ± 0.000
Dorper	22	0.19 ± 0.100	0.31 ± 0.030	0.39 ± 0.000
Afrino	51	0.09 ± 0.020	0.35 ± 0.010	0.39 ± 0.000
Namaqua Africana	10	0.34 ± 0.070	0.26 ± 0.020	0.39 ± 0.000
Blackhead Persian	14	0.29 ± 0.040	0.27 ± 0.010	0.39 ± 0.000
Nguni	30	0.29 ± 0.040	0.27 ± 0.020	0.39 ± 0.000
White Vital Swakara	40	0.23 ± 0.050	0.30 ± 0.020	0.39 ± 0.000
White Sub-Vital Swakara	17	0.24 ± 0.080	0.30 ± 0.030	0.39 ± 0.000
Black Swakara	16	0.25 ± 0.050	0.29 ± 0.020	0.39 ± 0.000
Grey Swakara	22	0.17 ± 0.050	0.32 ± 0.020	0.39 ± 0.000
Karakul	10	0.23 ± 0.030	0.30 ± 0.010	0.39 ± 0.000
African Dorper	21	0.10 ± 0.030	0.35 ± 0.010	0.39 ± 0.000
African White Dorper	6	0.14 ± 0.030	0.34 ± 0.010	0.39 ± 0.000
Merino	138	0.02 ± 0.030	0.38 ± 0.010	0.39 ± 0.000
Austraian Poll Dorset	108	0.06 ± 0.040	0.36 ± 0.010	0.39 ± 0.000
Australian Poll Merino	98	0.01 ± 0.020	0.38 ± 0.010	0.39 ± 0.000
Chinese Merino	23	0.03 ± 0.050	0.38 ± 0.010	0.39 ± 0.000
Karakas	18	0.06 ± 0.050	0.37 ± 0.020	0.39 ± 0.000
Namaqua Afrikaner	12	0.23 ± 0.030	0.30 ± 0.010	0.39 ± 0.000
Red Maasai	45	0.13 ± 0.030	0.34 ± 0.010	0.39 ± 0.000
MacArthur Merino	12	0.32 ± 0.140	0.27 ± 0.050	0.39 ± 0.000
Dorset Horn	21	0.14 ± 0.060	0.34 ± 0.020	0.39 ± 0.000
Merino Landschaf	22	0.04 ± 0.020	0.37 ± 0.010	0.39 ± 0.000
Ronderib Afrikaner	17	0.17 ± 0.060	0.32 ± 0.030	0.39 ± 0.000
Ethiopian Menz	34	0.14 ± 0.030	0.31 ± 0.010	0.39 ± 0.000
Black-Headed Mutton	24	0.12 ± 0.120	0.35 ± 0.040	0.39 ± 0.000
Bangladeshi Garole	24	0.24 ± 0.070	0.30 ± 0.020	0.39 ± 0.000
Overall	**1014**	**0.11 ± 0.100**	**0.35 ± 0.040**	**0.39 ± 0.000**

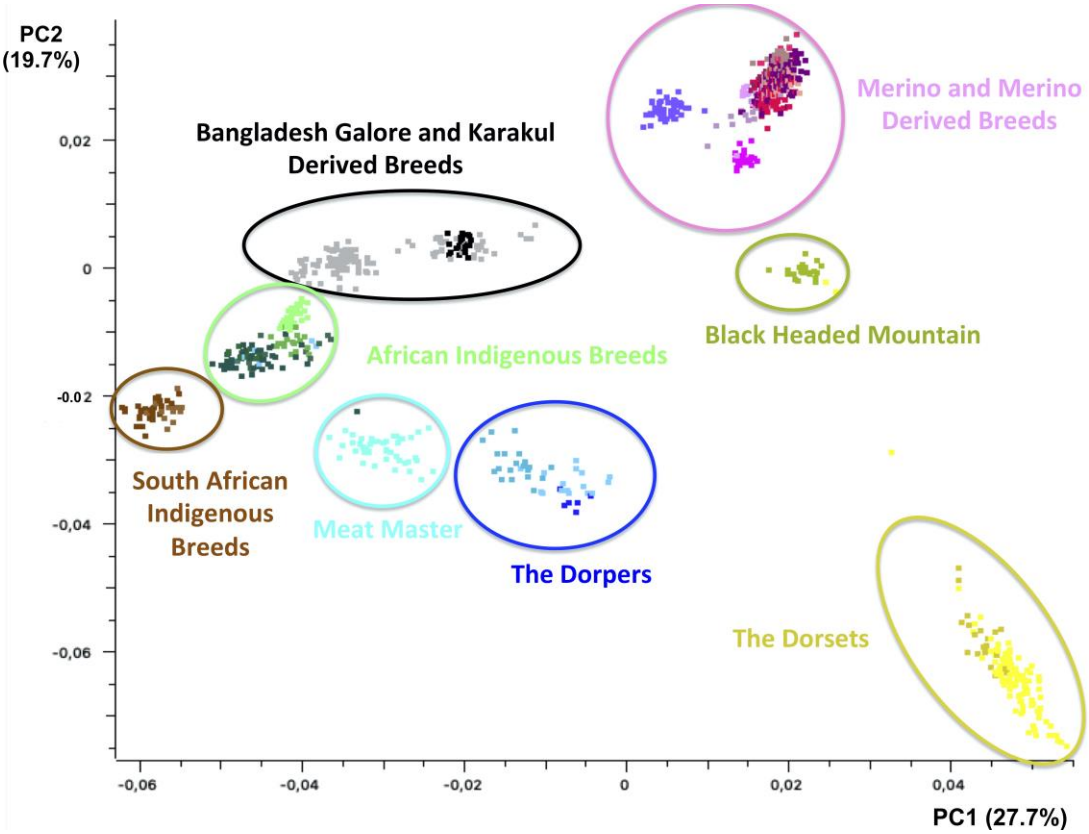


Figure 3.1: PCA based population structure 33 Sheep breeds/populations from South African commercial and indigenous breeds as well as global populations from the ISGC database.

3.3.5 Admixture analysis

Admixture based clustering uncovered the diversity and breed relations of the populations under study. At *K*= 7 (Figure 3.2), the (i) Nguni, (ii) Dorset Horn, (iii) Black Head Persian, (iv) Afrino and (v) Swakara, came out as unique and pure genomic signatures separated from the admixed clusters of (i) Merino/merino derived breeds, (ii) Meat Master and (iii) African breeds of Ethiopian Menz; Ronderib Afrikaner; Namaqua Afrikaner; and Red Maasai, (iv) Dorpers and Black Head Mountain sheep. At this *K*, the Meat Master showed admixed signatures from the Nguni, Black Head Persian and Dorset Horn while the Dorper was a mixture of Black Head Persian, Dorset Horns and Merinos. The clustering of breeds was maintained from K=7 upwards and increased *K* values only pulled out unique genetic signatures and diversity in the Swakara pelt colour based sub-populations, African sheep breeds and the Merino derived breeds. Overall, diversity was observed in all breeds except, Nguni, Dorset Horn, Bangladeshi Garole, Afrino, and Blackhead Persian.

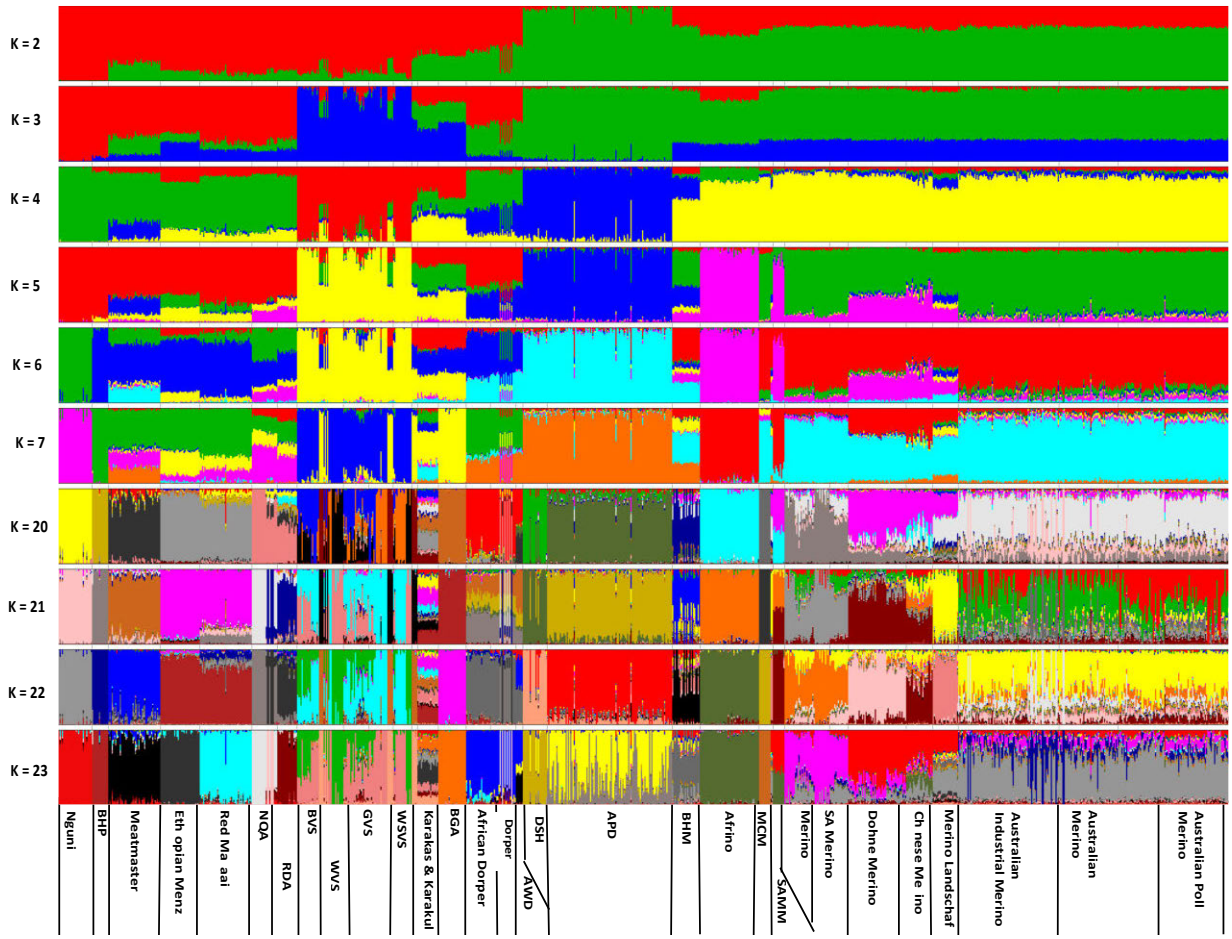


Figure 3.2: Admixture based population structuring of 33 Sheep breeds/populations from South African commercial and indigenous breeds as well as global populations from the ISGC database.

Treemix resulted in clustering that was analogous with the PCA clusters of (i) Dorpers; (ii) Swakara, Bangladesh Galore, and Karakas, (iii) Merino derived breeds (iv) Dorset Horns, (v) Black Head Mountain and (vi) cluster with the African indigenous breeds (Figure 3.3). Inputting the migration models revealed a picture observed from ADMIXTURE with arrows moving from the founder breeds of Black Head Persian, SA Mutton Merino and Bangladesh Galore to the composite breeds of Dorpers, Swakara, Meat Master and Dohne Merino. In addition to the revelation of the Afrino as a distinct cluster with minimum admixture, Treemix indicated gene flow from the Merinos and Ronderip Afrikaner into this composite breed.

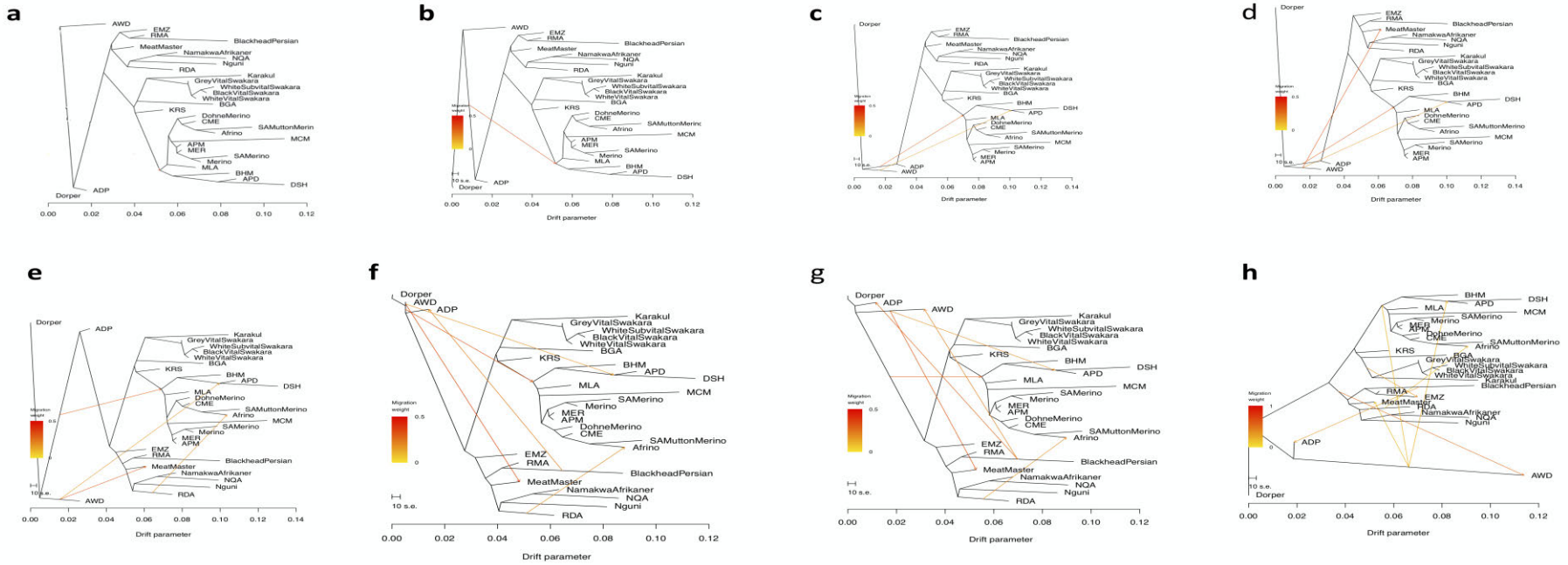


Figure 3.3: Treemix based phylogeny of 33 populations for migration model 3-6 in Figures 3a -d.

3.3.6 Per marker and population pairwise F_{ST} statistics

Population pairwise F_{ST} ranged from 0.01 (between the Swakara pelt producing sheep and Merino derived breeds of Merino versus Australian Polled Merino) to 0.35 between Blackhead Persian and MacArthur Merino and Blackhead Persian versus Dorset Horn. The Blackhead Persian was highly differentiated ($F_{ST} > 0.30$) from breeds such as MacArthur Merino, Dorset Horn, South African Mutton Merino, Namaqua Afrikaner and White Subvital Swakara. The South African smallholder non-descript Nguni breeds differed greatly ($F_{ST} > 0.35$) from MacArthur Merino, Dorset Horn, Karakul, Namaqua Afrikaner and South African Mutton Merino. Least diversity ($F_{ST} < 0.10$) was observed between merino/merino derived breeds *i.e.* Afrino versus South African Mutton Merino and Dohne Merino versus South African Mutton Merino as well as between Swakara pelt colour based subpopulations. Per marker per population pairwise comparison were plotted for population pairs that had an $F_{ST} \geq 0.30$ and included the (i) Nguni versus Karakul ($F_{ST} = 0.31$), SA Mutton Merino ($F_{ST} = 0.31$), MacArthur Merino ($F_{ST} = 0.33$)and Dorset Horn ($F_{ST} = 0.33$), (ii) Blackhead Persian versus MacArthur Merino ($F_{ST} = 0.35$), Dorset Horn ($F_{ST} = 0.35$), SA Mutton Merino ($F_{ST} = 0.32$) and Namaqua Afrikaner ($F_{ST} = 0.30$) and (iii) Namaqua Afrikaner versus SA Mutton Merino ($F_{ST} = 0.31$) and Karakul ($F_{ST} = 0.31$). Fixated and nearly fixated SNPs were observed to differentiate these populations as shown in Figure 3.4 (a- f) and Table 3.4 -3.6.

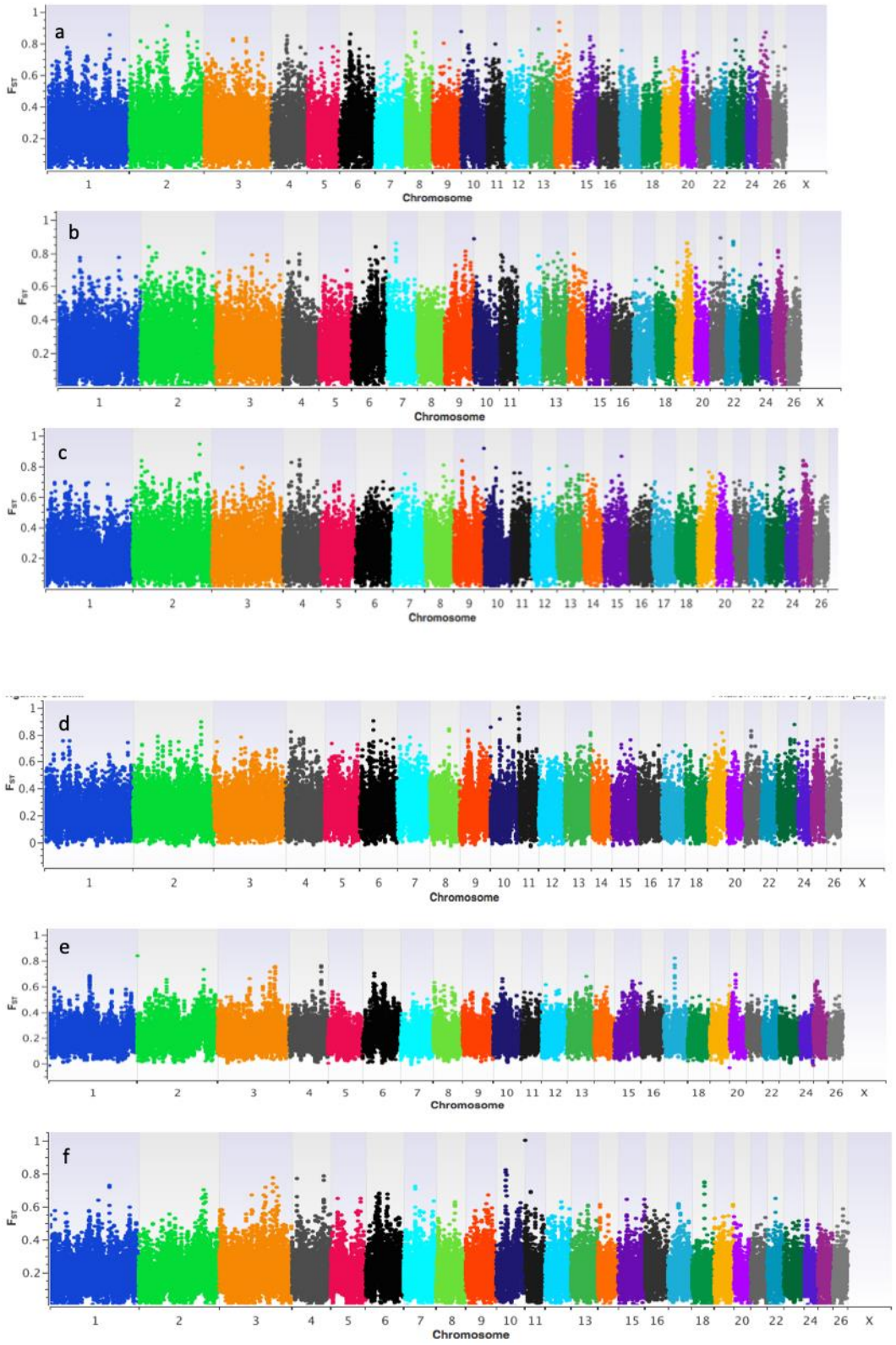


Figure 3.4: Per marker F_{ST} across all 26 autosomal chromosomes for (a) Blackhead Persian vs DSH (b) Blackhead Persian vs SA Mutton Merino (c) Nguni vs DSH (d)Nguni vs SAMM, (e) NQA vs DSH, and (f) NQA vs SAMM.

Table 3.4: Outlier SNPs (F_{ST} >0.79) and associated genes between Blackhead Persian versus DSH and SA Mutton Merino

Blackhead Persian *vs:*							
DSH				***SAMM***			
CHR	**Position**	**F_{ST}**	**Associated Genes**	**CHR**	**Position**	**F_{ST}**	**Associated Genes**
6	*OAR6_36682153.1*	0.91	*RAB44*	10	*OAR10_29538398.1*	0.87	*LYZL1*
6	*OAR6_36739717.1*	0.97	*GPR166P*	10	*OAR10_29546872.1*	1	*PTCHD3P1*
6	*OAR6_36763067.1*	0.97	*CPNE5*	10	*OAR10_29722772.1*	1	*SVIL*
1	*OAR1_210832168.1*	0.97	*HHAT*	10	*s18834.1*	0.95	*MIR604* *MIR938*
1	*OAR1_210858510.1*	0.94	*KCNH1*				
2	*OAR2_125283504.1*	0.96	*CNTNAP5*				
8	*OAR8_34529175.1*	1	*RPL21P80*				
8	*s20065.1*	1	*RPL21P80*				
10	*s52721.1*		-				
13	*OAR13_30278401.1*	0.91	*SLC7A1* *UBL3*				

Table 3.5: Outlier SNPs (F_{ST} >0.79) and associated genes between Nguni versus DSH and SA Mutton Merino

Nguni *vs:*							
DSH				***SAMM***			
CHR	**Position**	**F_{ST}**	**Genes**	**CHR**	**Position**	**F_{ST}**	**Genes**
2	*OAR2_210261780.1*	0.98	*SNAI1P1*	2	*OAR2_210232312.1*	0.88	*PKP4P1*
2	*OAR2_210295546.1*	0.98	*UNC80*	2	*OAR2_210261780.1*	0.91	*MEAF6P1*
2	*OAR2_210301097.1*	0.98	*RNA5SP118*	2	*OAR2_210295546.1*	0.91	*MAP2*
2	*OAR2_210232312.1*	0.96	*MAP2*	2	*OAR2_210301097.1*	0.91	
2	*OAR2_210373144.1*	0.83	*MEAF6P1*	6	*OAR6_41494878.1*	0.97	*FOXP4*
2	*OAR2_210597352.1*	0.83	*PKP4P1*	6	*OAR6_41558126.1*	0.97	*MIR4641*
2	*OAR2_210643482.1*	0.83	*CRYGFP*	6	*OAR6_41709987.1*	1	*MDFI; NPM1P51; TFEB*
4	*OAR4_18380514.1*	0.98	*LOC100286946*	10	*OAR10_29722772.1*	1	
4	*OAR4_26380336.1*	0.98	*ENPP7P10*	10	*OAR10_29737372.1*	1	
4	*OAR4_51489408.1*	0.98	*LOC101928803*	10	*OAR10_29722772.1*	1	
4	*OAR4_51441757.1*	0.96	*FAM86KP*	11	*OAR11_14785.1*	1	*PTDSS2*
4	*OAR4_39053460.1*	0.94	*ALG1L14P*	11	*OAR11_472905.1*	0.81	*RNH1*
4	*OAR4_41857328.1*	0.94	*FAM90A26*	11	*OAR11_486397.1*	0.81	*LOC101059906*
4	*OAR4_14572146.1*	0.92	*LOC101928910*	11	*OAR11_583043.1*	0.97	*HRAS*
4	*OAR4_29817460.1*	0.92	*USP17L10*	11	*OAR11_694267.1*	1	*LRRC56*
4	*OAR4_51346813.1*	0.92	*MIR548I2*	11	*OAR11_924831.1*	0.97	*C11orf35*
4	*s64770.1*	0.91	*RPS3AP19*	11	*OAR11_1085934.1*	1	*RASSF7*
4	*OAR4_52591717.1*	0.91	*ENPP7P11*				
4	*OAR4_26842281.1*	0.89	*FAM86MP*				
4	*OAR4_38425737.1*	0.89	*SLC2A9*				
15	*OAR15_41303143.1*	0.98	*PPP1R14D*				
15	*OAR15_55184101.1*	0.98	*SPINT1*				
15	*OAR15_48669726.1*	0.96	*RHOV*				
15	*OAR15_49442431.1*	0.96	*VPS18*				
15	*s53628.1*	0.94	*DLL4*				
15	*OAR15_54716840.1*	0.94	*OIP5*				
			NUSAP1				

Table 3.6: Outlier SNPs (F_{ST} >0.79) and associated genes between Namaqua Afrikaner versus DSH and SAMM

NQA *vs*:							
DSH				***SAMM***			
CHR	**Position**	**F_{ST}**	**Genes**	**CHR**	**Position**	**F_{ST}**	**Genes**
2	*s33265.1*	0.83		10	*OAR10_29538398.1*	0.87	*LYZL1*
4	*s39393.1*	0.97		10	*OAR10_29722772.1*	1	*MIR604*
							MIR938
4	*OAR4_98145910.1*	0.94	*STPG2-AS1*	10	*OAR10_29737372.1*	1	*PTCHD3P1*
4	*OAR4_98179556.1*	0.94	*STPG2*	10	*s18834.1*	0.95	*SVIL*
4	*OAR4_98242987.1*	0.92		11	*OAR11_14785.1*	1	*CICP23*
							LINC01001
							LOC101927
							0
17	*s40920.1*	0.82	*ASIC2*				
17	*OAR17_31805652.1*	0.85	*AA06*				
17	*OAR17_31839950.1*	0.85	*LOC147004*				
17	*OAR17_31867704.1*	0.85					
17	*OAR17_31905583.1*	0.94					

3.4 Discussion

The current study sought to investigate the within and between breed diversity of South African sheep populations and then compare that diversity to the global sheep populations. South Africa has a mixture of commercial and smallholder farming systems both of which rear sheep. Sheep are farmed for wool, mutton and pelt and as such the existing sheep population is categorised into different production systems each of which has well defined breeding goals. In addition to this, well-defined but multifaceted sector is the smallholder communal farming systems where non-descript sheep are kept for subsistence farming. The South African sheep genetic resources are a mixture of indigenous and exotic pure breeds as well as synthetic and composite breeds that were developed from crossing some exotic breeds to achieve higher yield in wool, mutton or pelts. Crossing indigenous with exotic breeds is used as a strategy to give the composite breeds robust performance under challenging production environments of diseases, extreme climate and poor nutrition. Based on such a complex system, it was hypothesised that the South African sheep genetic pool is diverse and complex demonstrating genomic signatures pooled from divergent founder populations and complex breeding system for the different production systems. To get a clear picture of the gene pool in South Africa, we therefore pooled in genotype data from the global sheep populations to investigate the relations and influence of exotic sheep populations on the South African genepool. From the International Sheep Genome Consortium data, breeds were selected from European, other African and Asian breeds that were assumed to have had some influence through crossbreeding or existence in similar production systems, on the South African gene pool.

Minor allele frequency pruning and distribution analysis gave an insight into the utility of the OvineSNP50 Beadchip in the South African sheep population. The OvineSNP50 Beadchip was developed in 2005 (ISGC 2009) and contains SNPs called from a broad range of populations from across the globe. Both Mutton and wool breeds were used in the development of this SNP panel and populations from Africa (*i.e.* African Dorper; African White Dorper; Ethiopian Menz; Namaqua Afrikaner; Red Maasai; Ronderip Afrikaner) were used to validate the chip. This study showed the South African indigenous breeds of the Blackhead Persian; Nguni and other African breeds had low MAF, which is in line with Kijas *et al.* (2009), who observed low MAF of geographically restricted and indigenous populations. However, with approximately 60 % polymorphic SNPs per population in most of these indigenous breeds, the OvineSNP50 Beadchip would still be considered of utility to even small local breeds of South Africa.

The heterozygosity estimates and inbreeding coefficients demonstrated low diversity and high inbreeding in South African indigenous breeds of Namaqua Afrikaner, Blackhead Persian, Nguni and Pelt producing Swakara. Low population sizes, small geographic spread and isolation from other commercial breeds, characterise majority of these indigenous breeds. Unlike the Merinos and other commercial breeds which have moderate to high population sizes and are well managed with structured breeding programs, the Namaqua Afrikaner, Nguni, Blackhead Persian and Swakara are kept by few farmers with small breeding flocks (Qwabe *et al.,* 2013). It is such production systems that would result in the observed low within population diversity. Heterozygosity deficiency and high inbreeding levels also reflect high selection intensities due to small effective population sizes and non-random mating strategies. Merino, Dohne Merino, Meat master and Afrino were the only breeds with inbreeding coefficients below the 10% safe thresholds. Breeds such as the Blackhead Persian, Nguni and Swakara are minor breeds consisting of smaller populations than the established industrial breeds of the Merinos.

The principal component analysis gave an insight into breed relations and demonstrated the (i) common founder effect and (ii) common breeding goals of Merino and Merino-derived breeds regardless of their geographic origin. The separation of fat tailed versus rump tailed sheep in the first cluster was evident and emanating from different founder populations as suggested by a number of authors (Marshal, 2000; Kijas *et al.*, 2012; Gifford-Gonzalez and Hanotte, 2011). According to Abulgasim *et al.* (2019) divergent selection was observed between fat and rump tailed sheep in Ethiopia. The Merino sheep are regarded as one of the oldest and most economically influential sheep breeds in the world (Al-Atiyat *et al.*, 2016) and known for their fine and soft wool. Merino and merino derived breeds have been widely distributed across the world as pure populations and admixed populations (Cian *et al.*, 2015). Merinos were introduced in South Africa, from Spain (Buduram 2004; DAD-IS, 2009) and have become adapted to South African climatic conditions (Snyman, 2014). In their study, Cian *et al.* (2015), suggested a more recent merino mediated gene flow to several merino derived breeds and populations. It was therefore hypothesized in this study that the South African Merino and other merino derived breeds would share common genes of founder effect that confer adaptation to the production environment of South Africa. The study also hypothesized that due to geographic isolation the South African merinos would carry unique alleles when compared to the global merino populations of Australia, New Zealand, Europe, Asia *etc*. Principle

component analysis however had all the merinos from South Africa and global populations clustering together as one cluster implying stronger influence of founder effects over geographic isolation and divergent selection for different production systems.

Historically, Swakara sheep have been shown to have developed through intensive selection of Karakul and other indigenous sheep of Namibia with the aim of producing sheep of distinct pelt quality (Malesa, 2015). Within the Swakara are four pelt colour based subpopulations of Black, White and Grey with the white subdivided into vital and non-vital white subpopulations. Regardless of the intensive selection for pelt colour, all the Karakas, Karakul and Swakaras clustered into one group separated from the wool, mutton and other dual purpose breeds. In terms of genetic distance, the Karakul/Karakas/Swakara cluster was close to that of the African indigenous breeds which confirms suggestions that these populations were crossed with the Karakul to improve pelt quality and robustness in the Swakara breed. African breeds of Ethiopian Menz, Red Maasai, Ronderib Afrikaner and Namaqua Afrikaner, clustered together and were close to the Nguni/Blackhead Persian cluster, as well as that of the Meat Master and Dorper. Such clustering could be a reflection of (i) low selection intensities in the African indigenous breeds which increases their within population diversity and as such relations to other breeds and (ii) the integration of indigenous genotypes when crossed with established breeds to produce breeds such as the Meat Master (cross between Merinos and Damara) and Swakara breeds (cross between Karakuls and Persians and Afrikaners). The Dorper sheep were developed by crossing Dorset Horn and the Blackhead Persian sheep (www.dorpersa.co.za) which explains why they clustered in between the Dorset and Nguni/BlackHead Persian cluster.

Admixture based clustering revealed the diversity within breeds and breed relations of the populations under investigation. The separation of Nguni, Dorset Horn, Blackhead Persian, Bangladeshi Garole, and Dohne Merino as distinct populations at $K = 7$ demonstrated their uniqueness as breeds regardless of the selection intensities exposed on breeds. Distinct and unique genomic signatures were identified in these breeds which showed that they are separated from the mainstream commercial sheep populations of Merinos and Swakara breeds. Although the Dohne Merino is a Merino-derived breed, it clustered separately in the ADMIXTURE analysis (from $K = 7$) which implied there have been intensive and divergent selection of this breed from its founder populations. The Dohne Merino is a synthetic breed developed from crossing German Mutton Merino commonly known as South African Mutton Merino and the South African Merino Ewe (Jordan, 2013). The breed is favoured for its robustness and

resistance to parasite infections in the Eastern Cape region of South Africa (Dlamini *et al.*, 2019). K = 20, which was the optimal K value with the least CV, demonstrated diversity and admixture in breeds such as Merinos, Merino derivatives excluding the Dohne Merino and the Swakara's. Overall, diversity was observed in all breeds except, Nguni, Dorset Horn, Bangladeshi Garole, Ethopian Menzi, Dohne Merino, Namaqua Afrikaner and Blackhead Persian. With Treemix, we set an assumption that that the South African genepool together with the global sheep, was a mixture of both founding populations and composite breeds. The results from Treemix support both the PCA, in clustering of breeds and ADMIXTURE, in the sharing of genomic signatures between founding populations and the composite breeds.

Population pairwise F_{ST} analysis supported population clustering by reporting high genetic distance between Nguni, Blackhead Persian and breeds such as Dorset Horn, South African Mutton Merino, Namaqua Afrikaner and the Swakaras. It therefore made sense that per marker F_{ST} analysis be conducted between Blackhead Persian, Nguni and Namaqua Afrikaner against the Dorset Horn and South African Mutton Merino that had high F_{ST} values above 0.30. Fixated and nearly fixated SNPs that were observed to differentiate these populations and generally showed the direction of selection and the forces behind breed formation in the compared breeds. SNPs at chromosome 2 for example differentiated the Nguni from the Dorset Horn and South African Mutton Merino and were found to be within QTLs associated with average daily gain, bone density (Campel *et al.*, 2003); body weight (Caranagh *et al.*, 2010), clinical mastitis (Banos *et al.*, 2017); change in haematocrit (Marshall *et al.*, 2013) and meat linoleic acid content (Karamichou *et al.*, 2006) as reported in the sheep QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/OA/summary).

The Nguni is a non-descript smallholder sheep population that has not been subjected to intensive selection towards any production trait. This is in contrast with breeds such as the South African Mutton Merino and the Dorset Horn which have been selected for growth and meat production. Genes such as *MAP2* that differentiated the Nguni from other commercial breeds are involved in microtubule cytoskeleton organization; central nervous system neuron development and regulation of cellular protein localization amongst other Biological processes (https://www.ensembl.org/Ovis_aries/Gene/Ontologies/biological_process), suggestive of the traits that differentiate commercial breeds of the SAMM and DSH from non-descript breeds like the Nguni.

The Blackhead Persian is a fat-tailed breed of domestic sheep originally from Somalia. The

breed is also a hair type sheep that is able to tolerate heat better than woolled breeds. SNPs that differentiated the Blackhead Persian with Dorset Horn and South African Mutton Merino were within QTLs associated with feacal egg count (*i.e FECGEN* on Chromosome 6), intramuscular fat (*TFAREA* on Chromosome), milk fat percentage, and response to parasite infection (*SFEC* on chromosome 1). The Namaqua Afrikaner which is also fat-tailed breed is a direct descendant of the sheep farmed with by the Namaqua Khoi-Khoi people and considered to be one of the oldest sheep breeds found in South Africa (Ramsay *et al.*, 2001). Development of the Namaqua Afrikaner occurred as a result of natural selection and since this breed fell into commercial obscurity during the 1930's, no selective breeding for the enhancement of commercially valuable attributes (i.e.. meat production yield and quality) transpired (Snyman *et al.,* 2013). The most prominent feature of the Namaqua Afrikaner is its fat-tail, an adaptive trait to offset nutritional needs in times of drought and forage scarcity. Instead of storing surplus energy in a thick subcutaneous fat (SCF) layer in times of forage abundance, as seen with the imported European and developed composite breeds, the Namaqua Afrikaner stores surplus energy in the fat-tail (Lawrie and Ledward, 2006). Breed differentiating SNPs between the Namaqua Afrikaner and Dorset Horn and South African Mutton Merino were within QTLs associated with faecal egg count (*FECGEN* on Chromosome 10, total lambs born (*LS* on Chromosome 10), milk protein percentage (*MY* on chromosome 2) (https://www.animalgenome.org/cgi-bin/QTLdb/OA/summary), also supportive of the differences in production challenges and breed development goals between the European breeds and the indigenous fat tailed sheep.

Although not highly differentiated as the Namaqua Afrikaner, Nguni and Blackhead Persian, the Dohne Merino showed a clear genomic signature and considerable F_{ST} from the Merino and other Merino derived breeds. Differentiating it from the other merino were SNPs found within QTLs for Feacal egg count, Response to parasitic infection particularly *Heamonchus corntortus.* This was quite interesting as Dohne Merino were established as a breed and predominantly raised in a Heamonchus corntortus challenged environment in the Eastern Cape province of South Africa (Dlamini *et al.*, 2019; Synman and Fisher, 2019). Other SNPs differentiating the Dohne Merino from the other Merinos were associated with body weight, wool crimp, ear size and carcass fat percentage.

3.5 Conclusions

The study overall provided a broad overview of the genomic architecture of South African breeds, with insights on within and between breed diversity as well as relations with global populations. The study used 14 sheep breeds representative of the broad range of production systems as well as South African sheep genetic pool. The study gave an initial assessment of how the purpose of breed and selection programs thereof could be shaping the population structure of the sheep breeds in South Africa and globally. There was evidence in this analysis that there is also close clustering of Merino and Merino-derived breeds separated from either the African breeds or the English Dorset breeds. Merino sheep are a predominant breed in South Africa that have been bred and selected over extended period of time to adapt to the local conditions of South Africa. In addition, Merino sheep have been crossbred with other local breeds to develop locally adapted and high performing breeds such as the Afrino, Meatmaster and Dohne Merino. An understanding of the evolutionary changes that have happened in the Merino and Merino-type breeds will elucidate the in-country breed developments and the genomic changes effected in the associated breeds.

Both PCA and ADMIXTURE analysis clustered populations according to breed founding history and production system while within population analysis showed the small, fragmented and localised breeds of the Nguni, Blackhead Persian, Namaqua Afrikaner and Swakara to exhibit high inbreeding values probably emanating from smaller founder effect. Further analysis that will shed more light on how the population history and production system have shaped the genomic structure of SA populations therefore becomes a prerequisite.

4 A GENOME-WIDE DIVERGENCE AND SELECTION SIGNATURE ANALYSIS OF SOUTH AFRICAN MERINO-DERIVED BREEDS FROM THEIR FOUNDERS[2]

[2] Preliminary results of this Chapter were published in the Proceedings of the AAABG conference: ***Dzomba, E. F.*** *and Muchadeyi, F. C. (2019) Medium density beadchip genotype data reveals genomic structure of South African merino-based breeds. Proceedings of the 23rd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Armidale, New South Wales, Australia, 27th October-1st November 2019 pp.452-455*

Abstract

Merino sheep are a breed of choice across the world popularly kept for their wool and mutton value. They are often reared as a pure breed or employed in crossbreeding and are a common component in synthetic breed development. The current study evaluated genetic diversity, population structure and breed divergence in 279 animals of Merino and Merino-based sheep breeds of South Africa using the Illumina Ovine50K BeadChip. The sheep breeds analysed included the three Merino-derived breeds and five presumed ancestral populations of Merinos and non-Merino founding breeds of Damara, Ronderib Afrikaner and Nguni. Highest genetic diversity values were observed in Dohne Merino (DM) with H_o = 0.39 ± 0.01 followed by Meatmaster and South African Merino (SAM) with H_o = 0.37 ± 0.03. The level of inbreeding ranged from 0.0 ± 0.02 (DM) to 0.27 ± 0.05 (Nguni). Analysis of Molecular Variance (AMOVA) showed high within population variance (>80 %) across all population categories. The first Principal Component (PC1) separated the Merino, South African Mutton Merino (SAMM), Dohne Merino and Afrino from the Meatmaster, Damara, Nguni and Ronderib Afrikaner. PC2 aligned each Merino derived breed with its presumed ancestors and separated the SAMM from the Merino and SAM. The |*iHS*| analysis yielded selection sweeps across the Afrino (12 sweeps), Meatmaster (4 sweeps) and Dohne Merino (29 sweeps). Genes associated with hair/wool traits such as *FGF12,* metabolic genes of *ICA1, NXPH1* and *GPR171* and immune response genes of *IL22 IL26, IFNAR1* and *IL10RB* were reported. Other genes included *HMGA* which was observed as selection signatures in other populations, *WNT5A* important in the development of the skeleton and mammary glands, *ANTXR2* associated with adaptation to variation in climatic conditions and *BMP2* which has been reported as strongly selected in both fat-tailed and thin-tailed sheep. The Dohne Merino vs SAMM shared all six sweeps regions on chromosomes 1, 10 and 11 with Afrino *vs* SAMM. Genes such as *FGF12* on OAR 1:191,3-194,7Mb and *MAP2K4* on OAR11:28,6-31,3Mb were observed. The selection sweep on chromosome 10 region 28,6-30,3 Mb harbouring the *RXFP2* for polledness was shared between Dohne Merino *vs* Merino, Meatmaster *vs* Merino and Meatmaster *vs* Nguni. The Dohne Merino *vs* Merino and the Meatmaster *vs* Merino also shared an Rsb-based selection sweep on chromosome 1 region 268,5 - 269,9 Mb associated with the Calpain gene, *CAPN7*. Overall, the study demonstrated some genetic similarities between the Merino and Merino-derived breeds emanating from common founding populations as well as some divergence driven by breed-specific selection goals.

Key words: *Merino-type sheep, population genetic structure, breed divergence, SNP genotypes, selection sweeps, EHH signatures*

4.1 Introduction

Globally, sheep and their products play a major role in the livelihoods of farmers (Ramsay *et al.,* 2001; Edea *et al.,* 2017). Sheep serve as an important source of mutton, manure and wool (DAFF, 2015). Other than this, sheep are considered as the most important farm animal in South Africa because of their role in the economy of the country (Groeneveld *et al.,* 2010). In other countries such as Ethiopia where crop production sometimes become none viable, sheep are regarded as an option for improved livelihoods (Edea *et al.,* 2017). Smallholder farmers in developing countries obtain cash from selling sheep and sheep have a key role in social security during periods of drought in crop farming areas (Mengesha and Tsega, 2012). Sheep become the ideal farm animal for smallholder farmers due to their body size (Hassan *et al.,* 2015). Furthermore, (Hassan *et al.,* 2015) affirms that sheep are sold to meet financial obligations and their ability to survive in harsh weather conditions allows resource-poor farmers to depend on sheep for food and human livelihood (Hassan *et al.,* 2015; Edea *et al.,* 2017). Sheep also fulfil different socio-cultural roles (Wilson *et al.,* 2011).

The Merino sheep breed is regarded as one of the oldest and most economically influential sheep breed in the world (Al-Atiyat *et al.,* 2016). The Merino are known for their fine and soft wool (DAFF, 2015). It is thought that the fine wool Spanish sheep of the 17-century contributed to the development of Merino sheep (Piper, 1998). In South Africa, the breed was introduced in the 1780's from Spain (Buduram, 2004; DAD-IS) and have become adapted to South African climatic and environmental conditions. The South African Merino is believed to be a composite breed between Spanish, Saxony, Rambouillet, American and Australian sheep breeds (Malson *et al.,* 1996). Through intensive natural and artificial selection, the Merino sheep could have evolved and are likely to carry different genes that confer adaptation to particular production environments (Peters *et al.,* 2010). Coupled to selection within breed, several Merino-based breeds have been developed (Hlophe, 2011) for either wool, mutton or as dual-purpose breeds. In South Africa, the Merino breed contributed to the development of composite breeds such as the Afrino, Meatmaster and Dohne Merino breeds.

The Dohne Merino is a composite breed (Buduram, 2004) developed from crossing the German Mutton Merino commonly known as SAMM and the SA Merino ewes (Kotze, 1951; Jordan,

2013). This breed was developed by the South African Department of Agriculture in the Eastern Cape province of South Africa by interbreeding Peppin-style Merino ewes and German Mutton Merino rams followed by interbreeding and selection for high fertility and fine wool under commercial rangeland conditions as well as for rapid lamb growth rate and robustness to the local production conditions. The objective of developing this breed in South Africa was to reduce fibre diameter and to maintain fleece weight (Naidoo *et al.,* 2005). Dohne Merino sheep are currently raised mainly as dual-purpose animals for both meat and wool production and are amongst the leading wool sheep breeds in South Africa. The breed is also farmed in Australia and New Zealand. These sheep are strong and hardy animals well adapted to their local environments and associated with resistance to parasites particularly Heamonchus Contortus Snyman and Fisher, 2019; Dlamini *et al.,* 2019).

The Afrino sheep originated from the crossbreeding of Ronderib Afrikaner, SA Merino and SAMM in a targeted ratio of 25:25:50 of the respective contributing breeds. Afrino sheep are a dual purpose breed that thrives in the harshest conditions. The breed originated during the depressed wool market of the late sixties when farmers began cross-breeding Merino ewes with mutton breeds. These crosses had coloured fibre (referred to as kemp), leading to contamination of the South African wool. The Afrino was then developed as a white-wooled breed for use as a terminal sire in crosses with Merino ewes (Bezuidenhout, 2020) and aimed to be robust and adapted to the harsh and dry conditions of Karoo sheep farms. Meatmaster is a breed that was developed to improve meat qualities of the fat hair tailed sheep breeds (Malson *et al.,* 1996). Literature indicates that Meatmaster are a composite of many breeds though was predominantly developed from SAMM, SA Merino, Damara and other indigenous breeds (Mason, 1996; Hlophe, 2011).

Outside South Africa, Merino and Merino-derived sheep breeds have been widely distributed across the world, both as purebred and admixed populations. They represent an economically and historically important genetic resource which over time has been used as the basis for the development of new breeds. Ciani *et al.* (2015) suggested that intensive gene flow, founder effects and geographic isolation are the main factors that determined the genetic makeup of current Merino and Merino-derived breeds. In South Africa, the Merino sheep were bred with local breeds in an effort to improve productivity and resilience of the breeds to the harsh local conditions whilst producing optimally (Cloete *et al.,* 2010; Synman and Fisher, 2019; Dlamini *et al.,* 2019). With climate change and other production challenges, the rationale of

crossbreeding and developing new and composite breeds may prevail, thus requiring an investigation and documentation of the genomic architecture of the current composite breeds and their evolution from the ancestral populations. It is, therefore, worthwhile for inventory purposes and to guide future breed development initiatives, to investigate the breed relationships and differentiations and the genomic regions targeted by selection through breed development of the SA Merino and Merino-derived breeds.

The OvineSNP50 Beadchip and other similarly designed beadchips provide unprecedented power to scan the genomes of livestock and investigate footprints of selection and their impact on the genetic potential of breeds to meet designed production goals. With the advent of genome-wide SNP genotyping, different statistical methods have been developed to interrogate genomes for signatures of selection and the associated effects on phenotypes. Signatures of selection in a genome are usually associated with either high-frequency derived alleles, highly differentiated allele frequencies between populations or long-range haplotypes with strong linkage disequilibrium (LD) (Grossman *et al.,* 2010). Statistical methods such as within population integrated haplotype homozygosity score (|*iHS*|) (Voight *et al.,* 2006) and between populations *Rsb* test (Tang *et al.,* 2007) have been developed to screen for high LD, long-range haplotype and infer on signatures of selection. A number of studies have used |*iHS*|, *Rsb* and other similar methods in investigation of selection sweeps in sheep (Paim *et al.,* 2018; Alvarez *et al.,* 2020) cattle (Chen *et al.,* 2016; Tijjani *et al.,* 2019) and other livestock species.

The objective of the current study was to investigate population genetic structure, breed similarities and divergence of Merino-derived sheep breeds of South Africa. The study first investigated population structure and admixture levels by referencing the composite breeds of Afrino, Dohne Merino and Meatmaster against their presumed ancestral breeds of Merino, South African Merino, South African Mutton Merino (SAMM), Ronderib Afrikaner, Damara and Nguni. Based on this analysis, the study went on to investigate regions habouring selection sweeps within the composite breeds as well as between each composite breed and its presumed ancestors. SNP genotypes for this analysis were limited to those generated in previous studies (Nxumalo *et al.,* 2018; Chapter 3) and the International Sheep Genome Consortium Hapmap data (http://www.sheephapmap.org). The study hypothesised that crossbreeding followed by intensive selection towards breed specific production goals resulted in genomic divergence between the South African Merino and Merino-derived breeds.

4.2 Materials and Methods

4.2.1 Animal genotypes

The study used OvineSNP50 Beadchip genotype data from a total of 277 animals obtained from five different sheep populations consisting of Merino (n = 8); South African Mutton Merino (n = 8), South African Merino (n = 45), Dohne Merino (n = 50), Afrino (n = 51) and Meatmaster (n = 48), Nguni = (n = 30), Ronderib Afrikaner (n = 17) and Damara (n = 20). The Afrino, DM, Meatmaster, SAM and SAMM genotypes were from samples kept in a biobank at Grootfontein College of Agriculture, South Africa and together with the Nguni sheep that were sampled from the KwaZulu-Natal region of South Africa, the SNP genotype data was generated and reported in Chapter 3. The Damara sheep genotypes were provided from a separate study (Nxumalo *et al.,* 2018), while the RDA and Merino genotypes were extracted from the ISGC (http://www.sheephapmap.org). The Afrino, Dohne Merino and Meatmaster are the Merino-derived composite breeds and their presumed ancestral breeds based on literature (Synman, 2014a; b and c) are presented in Table 4.1.

Table 4.1: South African Merino-derived breeds and their presumed ancestral breeds

Composite breed	SAMM (8)	SAM (45)	Merino (8)	RDA (17)	Damara (20)	Nguni (30)
Afrino (51)	X	X	X	X		
Dohne Merino (50)	X	X	X			
Meatmaster (48)	X	X			X	X

4.2.2 Genotype data quality control

In the current study, the Illumina OvineSNP 50K Bead genotypes from this study and from Nxumalo *et al.* (2018) and ISGC (http://www.sheephapmap.org) were subjected to quality control using PLINK v1.07 (Purcell *et al.,* 2007) and Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com) to ensure all SNPs had less than 5 % missing genotypes, a call rate more than 95 %, a minor allele frequency (MAF) less than 5 % and in Hardy Weinberg equilibrium ($P< 0.001$). Additional quality control measures ensured that individual animals had an IBD < 0.025. Further, in order to minimise bias, SNPs in strong linkage disequilibrium (LD > 0.2) with other SNPs were removed (Purcell *et al.*, 2007).

4.2.3 Determination of within breed genetic diversity

To determine the expected (H_E) and observed heterozygosity (*Ho*) values, PLINK V1.07 (Purcell *et al.,* 2007) was used by running the command *"--hardy"* on the data for each breed. Inbreeding coefficients were calculated as the difference between expected (H_E) and observed (Ho) heterozygosity values divided by the expected heterozygosity (H_E) values also in PLINK V1.07. Mean H_E and *Ho* values per breed were calculated using the PROC MEANS procedure in (SAS, 2013).

4.2.4 Analysis of Molecular Variation

Analysis of Molecular Variance (AMOVA) was used to determine the partitioning of total genetic variance within populations (F_{IS}), among populations within group (F_{SC}) and among groups (F_{CT}) using ARLEQUIN v3.5 (Excoffier and Lischer, 2009) The populations were categorised into (i) all 9 breeds; (ii) composite breeds of Afrino, Dohne Merino and Meatmaster; (iii) the presumed ancestral populations of Nguni, Damara, Ronderib Afrikaner, SAMM, SA Merino and Merino; (iv) composite and presumed ancestors of (a) Afrino and Ronderib Afrikaner, Merino and SAMM; (b) Meatmaster and Damara, Nguni, SAMM and SAM; and (c) the DM and the SAM and SAMM and Merino.

For each population category the following formula was implemented

$$\sigma_T^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2$$

where:

σ_T^2 = the total genetic variance *e.g*;

σ_a^2 = amongst breeds

σ_b^2 = amongst breeds within groups; and

σ_c^2 = within breeds.

The corresponding F statistics were then estimated as follows:

$$F_{CT} = \frac{\sigma_a^2}{\sigma_T^2}$$

$$F_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2 + \sigma_d^2}$$

$$F_{IS} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_d^2}$$

$$F_{IT} = \frac{\sigma_a^2 + \sigma_b^2 + \sigma_c^2}{\sigma_T^2}$$

4.2.5 Analysis of population structure

Principal Component Analysis (PCA) was carried out using Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). The Eigen values and Eigen vectors for the Principal components were estimated using Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com).

4.2.6 Inference of local genomic ancestry (PCAdmix)

PCAdmix v1.0 (Brisbin *et al.,* 2012) was used to infer local genomic ancestry in the composite breeds. The program utilises haplotypes from ancestral representatives to infer ancestry of focal individuals. In this study, the SAMM, SAM, Merino and Ronderib Afrikaner were treated as the ancestral representatives of the Afrino sheep, the SAMM, SA Merino and Merino were treated as the ancestral representatives for the Dohne Merino, while SAMM, SA Merino, Damara and Nguni were treated as the ancestral representatives for Meatmaster. The software algorithms perform the inference chromosome-wide through PCA, via short windows along each chromosome. Using a hidden Markov Model, PCAdmix then returns the posterior probability (PP) of ancestry from each reference population for each haploid individual for each window. PCAdmix requires phased genotypes, which were obtained using fastPHASE v1.2 (Scheet *et al.,* 2006) using default parameters.

4.2.7 Analysis of selective sweeps and differentiating genomic regions

Two complementary extended haplotype based statistics (EHH-based statistics) of the integrated Haplotype Score *(|iHS|)* (Voight *et al.,* 2006) and extended haplotype-based

homozygosity score test (*Rsb*) (Tang *et al.*, 2007) were used to assess genome-wide signatures of selection in the composite Merino-derived sheep of South Africa. The |*iHS*| tests detect partial selective sweeps within a breed, while the *Rsb* test detects selected alleles that have risen to near fixation in one and not the other breed/population (Tang *et al.*, 2007). These two approaches were chosen based on their robustness in minimising false positives that might arise due to the influence of the demographic history of the population on the results (Voight *et al.*, 2006; Tang *et al.*, 2007).

The ancestral alleles required for the computation of |*iHS*| were inferred as the most common alleles in the entire dataset as described by Bahbahani *et al.* (2015). Haplotypes were phased using Beagle (Browning and Browning, 2007), and used to calculate |*iHS*| scores for each SNP/haplotype within a breed/population and *Rsb* scores for each SNP/haplotype between breeds/populations. Haplotype frequencies were calculated using sliding windows of 20 SNPs that overlapped by 5 SNPs. For each locus, the |*iHS*| and *Rsb* scores were computed using the REHH package (Gautier and Vitalis, 2012) in R. For the analysis of within population |*iHS*| a score >3.0, and for the analysis of between population differences an *Rsb* score > 3, was used to infer the candidate genomic regions under selection.

4.2.8 Mapping region of differentiation to find genes

Significant SNPs (<0.05) were mapped for genes using ENSEMBLE genome browser and NCBI (NCBI; www.ncbi.nlm.nih.gov) ENSEMBL Ovine (Ovis Aries) genome build OAR3 implemented in Golden Helix SVS v8 (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Candidate genes were considered if their boundaries fell within 75 kb up or down-stream the selection sweep region defined. The associated genomic regions were also annotated using the Sheep QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/OA/summary). The Kyoto encyclopaedia of genes and genomes (KEGG) analysis was used to investigate pathways associated with each annotated gene within selection sweeps.

4.3 Results

4.3.1 Genetic diversity

All the 277 animals proceeded for further analysis following quality control. The number of SNPs retained for analysis ranged from 36 976 in Ronderib Afrikaner to 37 671 in Afrino (Table 4.1). Highest genetic diversity values were observed in Dohne Merino with H_O = 0.39±0.01 followed by Meatmaster and SA Merino with H_O= 0.37±0.03. Lowest diversity was

observed in the Nguni with H_O= 0.28±0.02. Inbreeding estimates ranged from 0.00±0.02 in Dohne Merino to 0.27±0.05 in the Nguni (Table 4.2).

Table 4.2: Expected and observed heterozygosity in five sheep breeds of South Africa

Breed	No. of animals	No. of SNPs	H_E±SD	H_O±SD	F_{IS}
Afrino	51	37671	0.39 ± 0.00	0.36 ± 0.01	0.06 ± 0.02
Meatmaster	48	36586	0.39 ± 0.00	0.37 ± 0.03	0.05 ± 0.07
Merino	8	37686	0.39 ± 0.00	0.35 ± 0.01	0.07 ± 0.03
SA Merino	45	37452	0.39 ± 0.00	0.37 ± 0.03	0.04 ± 0.08
SA Mutton Merino	8	37614	0.39 ± 0.00	0.34 ± 0.02	0.12 ± 0.04
Dohne Merino	50	37638	0.39 ± 0.00	0.39 ± 0.01	0.00 ± 0.02
Damara	20	37626	0.39 ± 0.00	0.31 ± 0.02	0.19 ± 0.04
Ronderib Afrikaner	17	36976	0.39 ± 0.00	0.33 ± 0.03	0.14 ± 0.07
Nguni	30	37634	0.39 ± 0.00	0.28 ± 0.02	0.27 ± 0.05
All breeds	269	37381	0.39 ± 0.00	0.35 ± 0.03	0.08 ± 0.09

4.3.2 Analyses of molecular variation in pure and developed breeds

Table 4.3 shows the partitioning of variation within breeds, among breeds and among breeds within categories of (i) ancestral breeds, (ii) composite breeds, (iii) each composite breed and its presumed ancestors. Within population variation was found to be 90 % in the composite breeds and 84 % within the presumed ancestral breeds whilst it was 83 % within all breeds (Table 4.3). High level of molecular variation was observed within populations in comparison to among populations and among individuals in population. In the category of a breed and its presumed ancestors, highest within population variation was observed in the Afrino and its presumed ancestors (92 %) followed by the Dohne Merino category (90 %) and least in the Meatmaster category (77 %). Among breeds within groups was highest in Meatmaster (21 %) and its presumed ancestors followed by the group of ancestral breeds (15 %) and least in the

Afrino and its presumed ancestors (6 %). Amongst breed variation was highest in the category consisting of all 8 breeds (17 %) and least in the category with ancestral breeds.

4.3.3 Analysis of population structure

In Figure 4.1, principal components PC1 and PC2 that explained 43 % of the total variation were plotted against each other. PC1 (28 % of the variation) separated the Merino, SA Merino, SAMM, Dohne Merino and Afrino from the Meatmaster, Damara, Nguni and Ronderib Afrikaner (Figure 4.1). PC2 (15 % of the variation) separated the Damara, Meatmaster, Merino, SA Merino and Dohne Merino from the SAMM, Ronderib Afrikaner and Afrino (Figure 4.1). In PC1, the Afrino were in the same axis as its Merino ancestors whilst separated from the Ronderib Afrikaner. The Dohne Merino clustered with the two Merinos and the Meatmaster on the other hand clustered in the same axis with its Nguni and Damara presumed ancestors separated from the Merinos

Table 4.3: Analysis of molecular variation

Dataset	Variance component (%)		
	Among breeds ***(F_{CT})***	**Among breeds within groups** ***(F_{SC})***	**Within breeds** ***(F_{IS})***
All 8 breeds	10.62 (17.23%)	-	91.52 (82.77%)
Ancestral breeds	0.011 (1.06%)	0.149 (14.77%)	0.158 (84.17%)
Composite breeds	0.029 (2.97%)	0.071(6.91%)	0.099 (90.11%)
Afrino & presumed ancestors	0.014 (1.42%)	0.065 (6.43%)	0.079 (92.15%)
Meatmaster & presumed ancestors	0.021(2.12%)	0.218 (21.38%)	0.235 (76.50%)
Dohne Merino & presumed ancestors	0.048 (90.21)	0.052 (4.95%)	0.098 (90.21%)

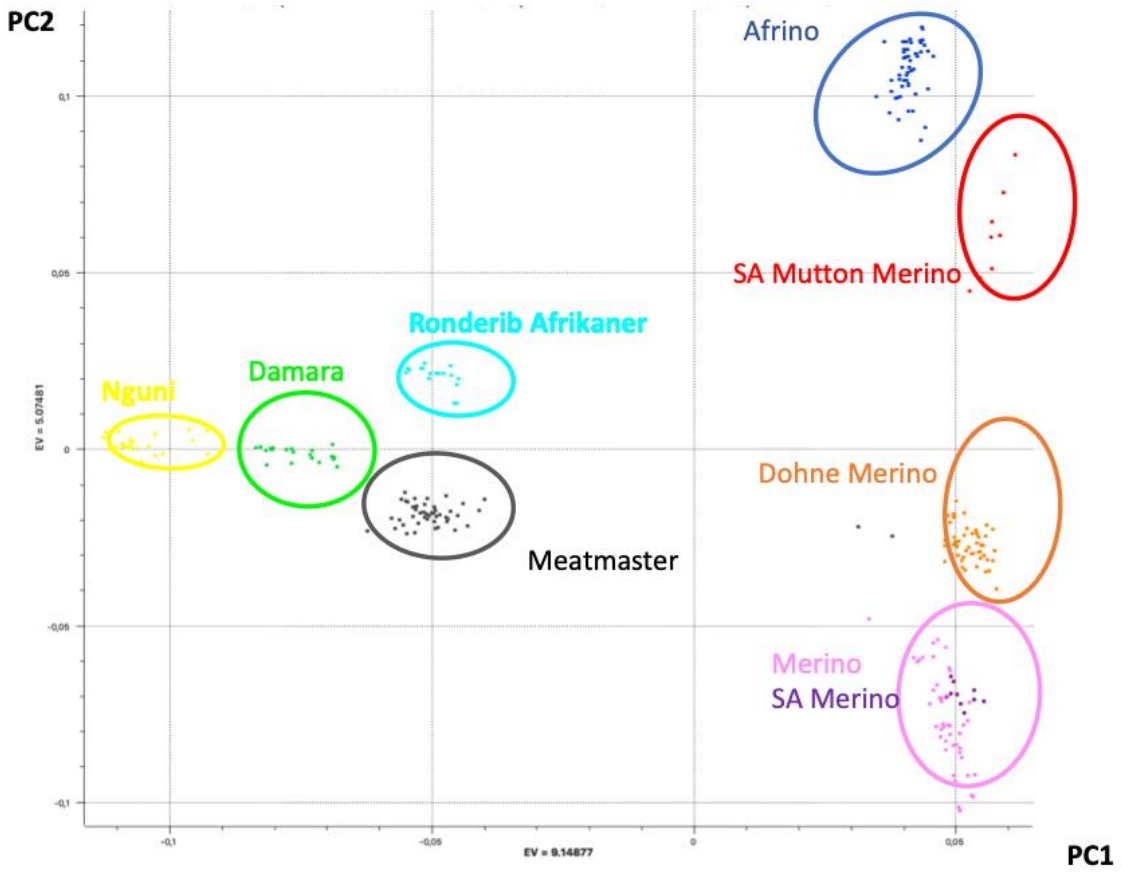


Figure 4.1: PCA based clustering of Merino, Merino-derived breeds and non-Merino presumed ancestors

4.3.4 PCAdmix based analysis of co-ancestry

The PCA admix results are illustrated in Figure 4.2. Using the PCAdmix algorithm, the genome of each composite Merino-derived breed was partitioned into segments of inferred ancestry at a resolution of chromosomal level. The PCAadmix of Afrino yielded tracts of ancestry consistent to predominantly SAMM (49.7%, 2.8 ± 0.03) followed Merino (28 %, 2.5 ± 0.03)) and Ronderib Afrikaner (22 %, 2.3 ± 0.02) (Figure 4.2a) consistent with a targeted ratio of 50:25:25 of the respective contributing breeds. The Dohne Merino that was developed from crossing SAMM and the SAM ewes, was predominantly Merino (37 %, 2.4 ± 0.03) and SAMM (40 %, 2.6 ± 0.03) and less of the SAM (23 %, 2.4 ± 0.02) (Figure 4.2b) The Meatmaster that are a composite of many breeds (Peters *et al.*, 2010), was largely Nguni (42 %, 2.4 ± 0.04) and Damara (32 %, 2.9 ± 0.03) and less of the Merino (27 %, 1.8 ± 0.02) breeds (Figure 4.2c).

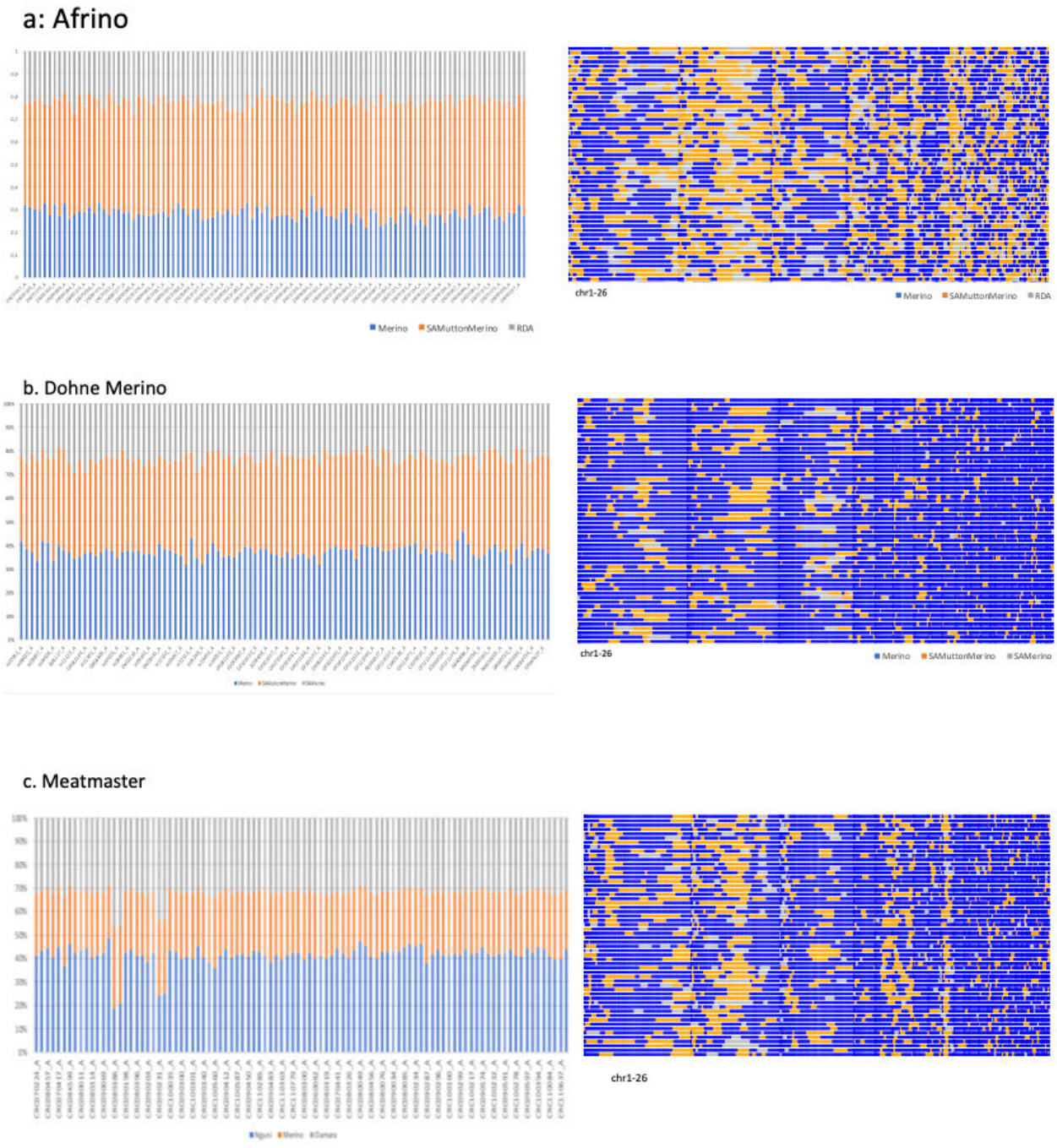


Figure 4.2: PCAadmix of Merino-derived (a) Afrino; (b) Dohne Merino and (c) Meatmaster breeds and their presumed ancestors

4.3.5 Signatures of selection

4.3.5.1 Signatures of selection - |*iHS*| approach

Twelve selection sweep regions (|*iHS*| >3.0) distributed across 9 chromosomes (OAR, 2, 3, 4, 6, 8, 10, 13 and 19) were identified in Afrino sheep (Figure 4.3) with a number of genes located within these twelve regions as illustrated in Table 4. We also identified two regions at the highest |*iHS*| on OAR1 (|*iHS*|=4.02; 192,7-194.0 Mb) and OAR 4 (|*iHS*|=4.02; 16.1–17.5 Mb). Gene *FGF12* was within the sweep on OAR1 region, whilst genes *ICA1* and *NXPH1* were observed within chromosome 4 sweep. Four selection sweep regions |*iHS*| >3.0) on four chromosomes (OAR 1, 2, 3 and 9) were identified in Meatmaster (Figure 4.3 and within these sweeps were genes *GPR171* on chromosome 1 and *HMGA* gene, *IL22* and *IL26* on chromosome 3. Twenty-nine selection sweep regions |*iHS*| >3.0) on fourteen chromosomes (OAR 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 15, 16 and 19 were identified in the DM sheep. Chromosome 1 region 119,0-121,0 Mbs harbours *IFNAR1; IL10RB; SLC5A3* and *CRYZL1* genes. The *WNT5A* gene is on 45,5Mb on chromosome 19. Selection sweeps on Chromosome 6 (93,5-95,2Mb) harbours the *FGF5* and *ANTXR2* genes while a region on 46,5 - 48,1Mb on chromosome 13 harbours the *BMP2* gene.

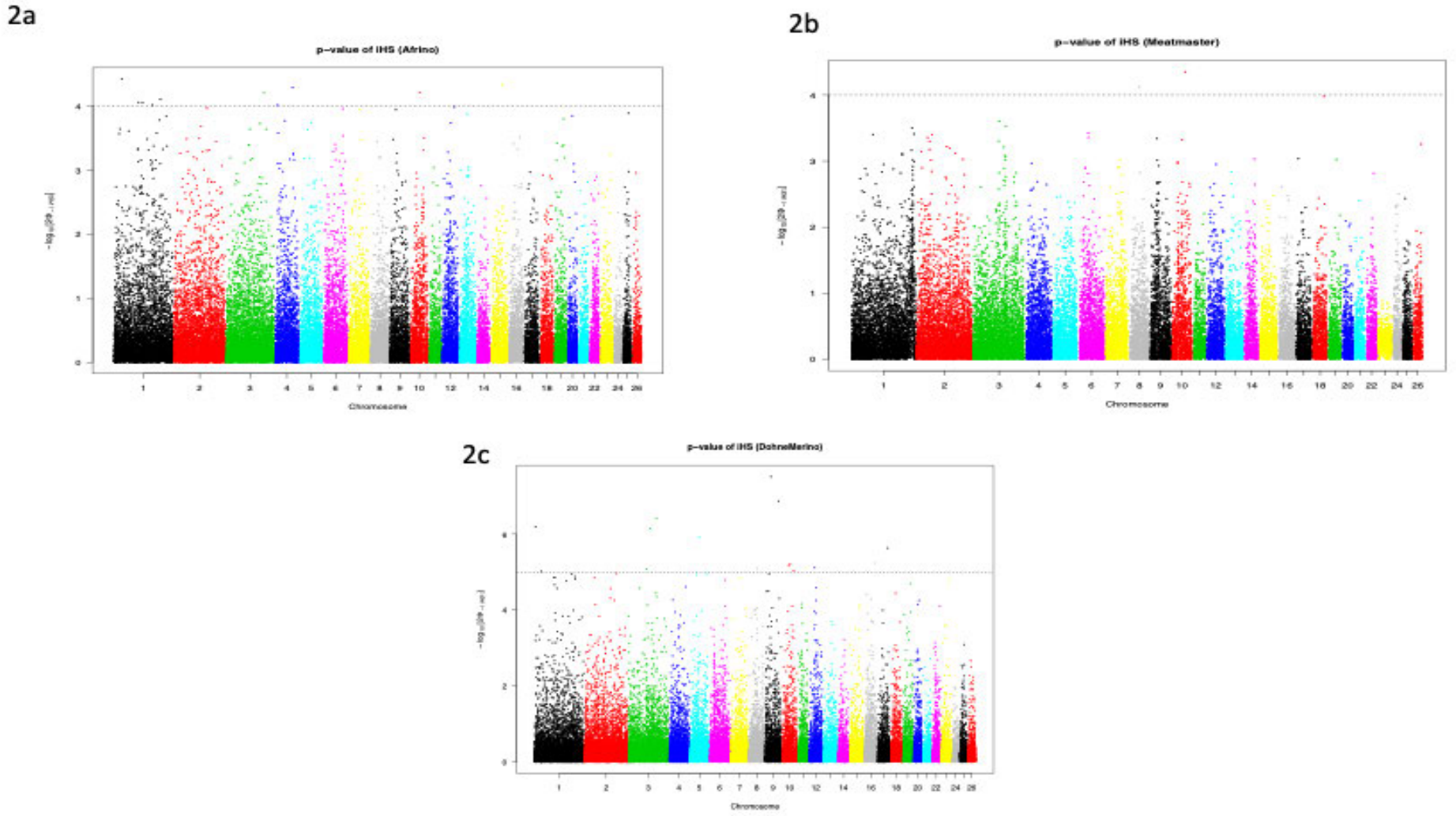


Figure 4.3: |*iHS*| scores of (a) Afrino, (b) Meatmaster and (c) Dohne Merino.

Table 4.4: |*iHS*| based Selection sweep regions *(|iHS|*>3.0) and associated genes in Afrino and Meatmaster and Dohne Merino sheep breeds.

OAR	SNP start (Mbp)	SNP end (Mbp)	Candidate genes	Cross Ref on pathways and associated functions
Afrino				
1	119,6	121,4	*SLC5A3; MRPS6; IFNGR2; IFNAR1; IL10RB*	Anti-inflammatory and immune response. (Dieckgraefe *et al.*, 2000)
1	185,8	187,2	*HEG1; ZNF148; LRCH3; SLC49A4; SLC51A*	QTL for muscle weight at 168.6-197.3 (Mbp) region
1	192,7	194,0	*FGF12; PLAAT1;ATP13A4*	*FGF12* - Hair/wool traits (Zhang *et al.*, 2013) reproductive traits (Woad *et al.*, 2009); (An *et al.*, 2018).
1	194,6	195,6	*IL1RAP; TPRG1;*	Reproductive traits
2	167,5	168,6	*LRP1B*	
3	114,5	116,4	*SYT1; PPP1R12A*	
4	16,1	17,5	*ICA1; NXPH1*	Glucose and fat metabolism; Kegg metabolic pathways
6	63,4	64,4	*GUF1; YIPF7; GNPDA2; GABRA4;*	*GNPDA2* body shape *KIT* gene (70.18-70.20 Mbp)
6	100,4	102,3	*SLC10A6; PTPN13; MAPK10*	Fiber traits
8	47,4	49,3	*LYRM2; BACH2; GJA10; GABRR1*	
13	46,5	48,0	*RASSF2; SLC23A2; BMP2*	(Fariello *et al.*, 2014); and (Kijas *et al.*, 2012); Tail Traits (Liu *et al.*, 2013).
19	46,7	48,1	*ACTR8; SELENOK; CACNA1D; IL17RB; NEK4; SPCS1*	
Meatmaster				
1	234,4	236,3	*IGSF10; P2RY12; P2RY14; WWTR1; RNF13; GPR171*	Feeding and metabolism (Ruiz-Larranaga *et al.*, 2018)
2	62,0	63,3	*TMC1; ALDH1A1*	
3	151,9	153,3	*IL22; IL26; C12orf66; HMGA*	(Fariello *et al.*, 2014) and 44; 56)
9	31,8	32,8	*SPIDR; PRKDC; MCM4*	
Dohne Merino				
1	5,3	6,9	*SH3BP4*	Sensitivity to environmental variation in Cattle (Uemoto *et al.*, 2020)
1	119,6	121,5	*IFNAR1; IL10RB; SLC5A3; CRYZL1*	Anti-inflammatory and immune response. (Dieckgraefe *et al.*, 2000) and Reproductive traits (implantation of conceptus to uterus (Zhang *et al.*, 2013)
1	123,5	125,3	*PAXBP1; SYNJ1; MIS1BA; HUNK; SCAF4*	
1	136,7	138,1	*TMPRSS15; CHODL; BTG3; CXADR*	
1	168,6	170,1	*ALCAM; CBLB*	
1	243,0	244,1	*DIPK2A; SLC9A9; CHST2*	
2	65,2	67,0	*FAM189A2*	

2	156,6	157,0	*RND3; LYPD6*	
3	138,7	140,5	*ZNF641; PFKM; SLC48A1; COL2A1; HDAC7; SLC38A4; SLS38A2; SLC38A1; ARID2*	
3	148,2	149,7	*YEATS4; LYZ; KCNMB4; SLC35E3; RAP1B; CCT2; MYRFL; IL22; IL26*	
3	169,8	171,5	*SLC5AB; SLC17AB; SCYL2; DRUM1; IGF1; ASCL1*	
3	217,4	218,6	*NFAM1; PACSIN2; A4GALT; PARVB; PARVG*	
4	62,5	63,5	*AVL9; NT5C3A; FKBP9; PDE1C*	
5	23,8	25,1	*SLC12A2; MEGF10; PHAX; TEX43*	
5	39,8	41,3	*PRSS57; PLPPR3; RNF126; STK11; REDX01*	
6	26,1	27,5	*STPG2*	
6	93,5	95,2	*ANTXR2; FGF5; GK2 FGFS*	Signature of selection identified by (Fariello *et al.,* 2014) and (Kijas *et al.*, 2012)
8	48,7	51,8	*RARS2; SLC35A; SPACA1; GABBRR1*	
9	34,2	36,1	*RGS20; SOX17; TGS1; TCEA1;*	
10	40,5	42,2	No genes	
10	61,4	63,3	*SLITRK5; SLC16A9*	Thyroid hormone signaling Protein Metabolism pathways Monocarboxylate transporter
10	70,1	71,8	*RUFY2; SLC25A16; COX20P*	Endocytosis mechanism; Thermogenesis
12	37,7	39,6	*VAMP4; FMO4; TNFSF18*	
12	43,8	45,5	*ACOT7; NOL9; HES3; RNF207*	
13	49,1	50,9	*BMP2*	(Fariello *et al.*, 2014) and (Kijas *et al.*, 2012)
15	54,4	56,0		
16	43,9	45,8	No genes	
19	25,8	27,3	No genes	
19	46,5	48,1	*WNT5A*	

4.3.5.2 Signatures of selection sweeps: *Rsb* approach

Figure 4.4 shows the *Rsb* based selection sweeps between composite Merino-derived breed and each of its presumed ancestors with details on the genes and pathways provided in Table 4.5 provides. The Dohne Merino *vs* Merino analysis yielded 8 selection sweeps within four chromosomes of OAR 1, 3, 10 and 11. Sweeps on OAR 1 yielded genes such as *GHSR*, *SPATA16* and *SLC7A14* and *SLC2A2*. The Dohne Merino *vs* Merino and the Meatmaster vs

Merino shared sweep on chromosome 1 region 268,5 - 269,9 Mb associated with the Calpain gene *CAPN7*. The selection sweep on chromosome 10 region 28,6-30,3 Mb is associated with the *RXFP2*. This region was shared between Dohne Merino *vs* Merino, Meatmaster *vs* Merino and Meatmaster *vs* Nguni. The region carried other genes such as *FRY* and *BSGLCT*. The Dohne Merino *vs* SAMM shared 3 regions with the Afrino *vs* SAMM which were on (i) OAR 1 with *ILIRAP; FGF12* and *GMNC* genes; (ii) OAR 10 with *SPRY2; SLITRK4* and *GPC* genes and (iii) OAR 11 with *SPTSSB, PPMIL* and *B3GALNT* genes. The Dohne Merino *vs* SAMM shared all six sweeps regions on chromosomes 1,10 and 11 with Afrino *vs* SAMM. Genes such as *FGF12* on OAR 1:191,3-194,7Mb and *MAP2K4* on OAR11:28,6-31,3 were observed.

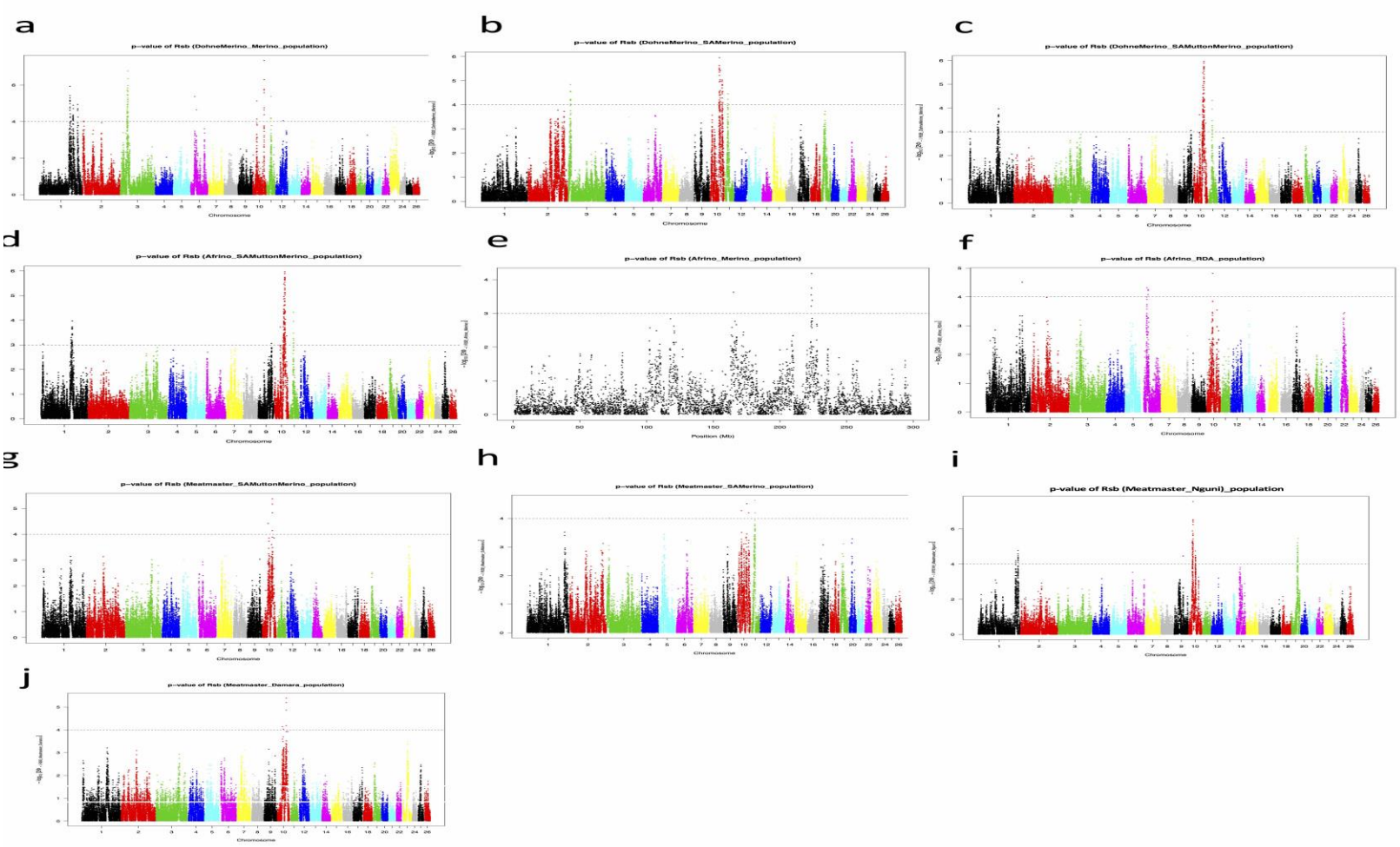


Figure 4.5: Genomewide *Rsb* scores of Merino-derived breeds against presumed ancestors.

Table 4.5: *Rsb* based Selection sweep regions (*Rsb*>3.0) and associated genes for Afrino, Dohne Merino and Meatmaster sheep breeds versus their presumed ancestors

OAR	SNP start (MBps)	SNP end (MBps)	Genes	Pathways
Dohne Merino *vs* Merino				
1	212,1	216,0	*SPATA16; ECT2; GHSR; SLC2A2; SLC7A14*	*GHSR*-growth and carcass traits in sheep (Bahrami *et al.,* 2012). *SPATA16*-Environmental variables in goats (Mdladla *et al.,* 2018) *SPATA16*-Male fertility in cattle *SLC7A14; SLC2A2*- nutrient transport and absorption
1	233,1	236,5	*C3orf33; PLCH1; DHX36; RAP25; P2RY1*	
1	265,6	267,5	*SETD4;*	
1	268,5	269,9	*CBR3* *CAPN7*	Upregulated in Lori-Bakhtiari sheep fat tailed breed (Bakhtiarizadeh *et al.,* 2019) CAPN7 muscle growth and maintenance (Bakhtiarizadeh *et al.,* 2019)
3	46,8	55,3	*LRRTM4; XPO1; FAM161A*	
10	28,6	30,3	*FRY; RXFP2; B3GCLT*	Highly expressed due to Heamonchus corntortus infection (Dieckgraefe *et al.,* 2000)
10	79,4	81,7		
11	32,5	34,4	*NCOR1; ZNF624; SLC47A2; SLC5A12; MAPK7*	Fleece traits; Muscle growth and development; Fiber diameter
Dohne Merino *vs* SA Merino				
3	15,4	17,2	*SOX11; RSAD2; RNF144A*	
11	24,6	26,5	*SLC13A5; TEKT1*	
Dohne Merino *vs* SA Mutton Merino				
1	191,3	194,7	*IL1RAP; FGF12; GMNC*	*FGF12* - Hair follicle and growth development
1	195,9	201,5	*TRPG1; LPP; RTP2; BCL6*	
10	57,6	59,4	*SPRY2; SLITRK4*	
10	59,6	64,2	*GPC5*	
10	64,9	72,5	*GPR180; GPC5; TGDS; DCT*	
11	28,6	31,3	*MAP2K4; DNAH9; MYH3; MYH4; MYH8*	MAP2K4-metabolic function in the liver

Continued

Afrino *vs* SA Mutton Merino				
1	191,3	194,7	*IL1RAP; FGF12; GMNC*	
1	195,9	201,5	*TRPG1; LPP; RTP2; BCL6*	
10	57,6	59,4	*SPRY2; SLITRK4*	
10	59,6	64,2	*GPC5*	
10	64,9	72,2	*GPR180; GPC5;TGDS;DCT*	
11	28,6	31,3	*MAP2K4; DNAH9; MYH3; MYH4; MYH8; GAS7*	
Afrino *vs* Merino				
1	222,5	224,6		
Afrino *vs* RDA				
1	227,4	228,6	*SHOX2; VEPH1; PTX3; CCNL1;*	
1	242,7	244,8	*DIPK2A; SLC9A9; CH5T2*	
2	107,9	109,5	*HPF1; CLCN3; NEK1*	
3	73,2	74,3	*NRXN1; U6*	
6	32,3	35,5	*CCSER1; MMRNA1; SNCA; FAM13A; ABCG2; SPP1;LCORl*	Milk traits Growth traits, Bone related and meat quality traits
6	41,9	43,9	*GBA3; PPARGC1A; DHX15; SOD3; LG12*	
10	38,6	42,2		
13	36,7	38,6	*BFSP1; RRBP1; DSTN; SNX5; KAT14; OVOL2*	
22	21,4	23,6	*PITX3; POLL; FBXW4; TLX1*	
Meatmaster *vs* SAMM				
10	67,3	69,4	*CTNNA3*	
Meatmaster *vs* Merino				
1	266,0	267,5	*CBR3* *CAPN7*	Upregulated in Lori-Bakhtiari sheep fat tailed breed (Bakhtiarizadeh *et al.*, 2019) *CAPN7* muscle growth and maintenance (Bakhtiarizadeh *et al.*, 2019)
5	36,6	38,6		
10	29,1	32,8	*RXFP2; FRY; BSGLCT*	Horns; Wool traits
10	62,0	63,4	*SLITRK5*	
10	67,3	69,2	*CTNNA3; GCD; TGDS; GPR180*	
10	79,6	81,5	*POGLUT2; SLC10A2*	
11	24,2	27,6	*SNORA62; SOX*	

11	28,9	33,6	*FSHB; SOX15; SAT2*	
Meatmaster *vs* Nguni				
1	265,2	266,7		
1	267,9	269,6		
10	27,4	29,3	*RXFP2; RXFP2; FRY; BSGLCT*	Horns
10	30,6	32,5		
10	47,8	49,3	*BORA; PIBF1; MZT1*	
14	27,0	28,1	*CDH8*	
19	41,2	43,6	*FAM3D; FAM107A; C3or167; ACOX2*	
19	44,2	48,8	*WNT5A; IL17RD; IL17RB; ACTR8; DCP1A*	
Meatmaster *vs* Damara				
10	67,3	69,4	*CTNNA3; GCD; TGDS; GPR180*	

4.4 Discussion

In conducting this study, we made use of available sheep genotypes from previous projects to make inferences on genetic diversity, breed relations and divergence amongst the Merino-derived breeds and their presumed ancestors. Our data was drawn from previous studies (Chapter 3) for the SAMM, SAM, Dohne Merino, Meatmaster and Afrino; (Nxumalo *et al.*, 2018) for the Damara sheep and (Kijas *et al.*, 2012) for the Ronderib Afrikaner and Merino sheep. These Merino and Merino-derived sheep breeds dominate the South African sheep industry contributing to mutton, wool and other sheep by-products. Merino sheep originate from Spain and are primarily useful and highly prized for their wool. In South Africa, their use in livestock farming can be traced to the late 18th century when few founder ewes and rams where donated by the Dutch government for experimental purposes to the Cape government (https://merinosa.co.za/history/).

The focus of the current study was to make inferences on genetic diversity. Observed and expected heterozygosity values together with the inbreeding coefficient were used to explain genetic diversity within the studied sheep breeds. Highest genetic diversity (H_E) was observed in the Dohne Merino followed by the Meatmaster, SA Merino and Afrino. The Dohne Merino, Afrino and Meatmaster are the three Merino-derived composite breeds and are of high-genetic diversity similar to that reported in Spanish and Australian Merino breeds by (Ciani *et al.*,

2015). The Nguni, Damara and Ronderib Afrikaner, which are the indigenous ancestral populations are raised by few fragmented communities (Nxumalo *et al.,* 2018; Selepe *et al.,* 2018; Qwabe, 2011), which would explain the low levels of within population genetic diversity. Although the founding population of Nguni, Damara and Ronderib Afrikaner had low genetic diversity, crossbreeding them to Merino breeds to develop the composite breeds resulted in increased diversity observed in the Afrino and Meatmaster breeds, which can be attributed to combination of two or more genetic pools.

To gain an insight into the genetic structure of breeds, analysis of molecular variation (AMOVA) was employed to determine the portioning of variance within and between populations and within populations amongst groups. In panmictic populations, the variance is expected to come from within samples (Excoffier and Lischer, 2009). If the variance occurs among samples within the population or comes from among populations, this would be regarded as evidence of the existence of population structure (Excoffier and Lischer, 2009). As expected, within population variation was high in the composite breeds in agreement to the high heterozygosities and low inbreed observed in Dohne Merino, Afrino in Chapter 3. The Afrino and Dohne Merino were established from predominantly well managed commercial breeds of SAMM, Merino and SA Merino and to a small extent, the Ronderib Afrikaner in the case of the Afrino. These breeds have moderate to high genetic diversity which explains the high within population genetic variation in this category. The Meatmaster on the other hand is based on small and less diverse breeds of the Damara and the Nguni which is reflected in the relatively lower (77 %) within population diversity in this group. Significant population substructure was therefore evident in the Meatmaster and its presumed ancestors' category, with an among breed diversity of 21 %. These Merino-derived sheep breeds that exist as widely distributed admixed populations represent economically and historically important genetic resources (Ciani *et al.*, 2015).

The first (28 %) and second (15 %) principal components together explained 43 % of the total variance, which is a significant contribution to the diversity in the population and similar to observations from other populations (Ciani *et al.,* 2015; Grasso *et al.,* 2014; Lancioni *et al.,* 2013). Similar to observations in Chapter 3, PC1 fat tailed sheep of Nguni, Damara, Afrikaner and Meatmaster from non-fat tailed sheep. In addition, PC1 separated the Merino breeds from non-Merino breeds with the exception of Meatmaster. Whilst the Meatmaster was bred from fat tailed sheep, there is intensive and directional selection against fat localisation and long tails

in the breed (www.meatmasters.co.za). Coupled to this, part of the breed standards for the Meatmaster are that it should be 50% Damara (Synman, 2014a; b). These selection criteria explains its clustering with the Damara, Nguni and Ronderib Afrikaner away from the Merinos and other Merino-derived breeds under PC1. In PC2 (15 % of variation), the Afrino clustered with the SAMM separated from the Dohne Merino, Merino and SA Merino. According to the breed standards (http://www.afrino.org.za), 80% of the income from Afrino is generated through meat production and 20% through wool production. This would be regarded as a biased selection objective towards growth and meat production traits which explains why the Afrino clustered with the SAMM. The Dohen Merino sheep were developed by interbreeding Peppin-style Merino ewes and German Mutton Merino rams followed by selection for high fertility and fine Merino wool under commercial rangeland conditions as well as for rapid lamb growth rate. Although currently raised as a dual purpose breed, it is one of the leading woollen sheep breeds in South Africa. Similarly, the Merino and SA Merino which cluster with the Dohne Merino, are all dual purpose breeds (https://Merinosa.co.za/). The Ronderib Afrikaner sheep are an improved form of the Namaqua Afrikaner Sheep (Epstein, 1960) and together with the Damara and Nguni sheep are fat tailed sheep (Peters *et al.,* 2010), which could have formed the basis of their clustering together in PC1.

PCAadmix confirmed the presumed ancestry of the Merino-derived breeds of Afrino, Dohne Merino and Meat Master. These Merino-derived breeds that represent admixed populations were developed to address specific production challenges so that the resultant synthetic breeds are suited to meet certain production goals (Ciani *et al.,* 2015). For example, the Afrino breed that was developed from the SAMM and SA Merino ancestral representatives, was developed to overcome the challenges of kemp in the populations while improving herd productivity for wool and mutton (Bezuidenhout, 2020). The Dohne Merino is a dual purpose breed developed from the SAMM and SA Merino representative to improve productivity and robustness to the production challenges particularly survival and enhanced production in parasite-infested regions of the Eastern Cape province of SA (Grossman *et al.,* 2010; Synman, 2014). The Meatmaster developed from the SAMM, SA Merino, Damara and Nguni breeds, is classified as a non-fat tailed hair-type sheep breed that is renowned for meat production ((Peters *et al.,* 2010).

|*iHS*| analysis is used to infer recent and generally segregating selection sweeps (Voight *et al.,* 2006) and has been used in humans (Liu *et al.,* 2013) and a number of livestock studies for dairy cattle (Cheruiyot *et al.,* 2018), pigs (Chen *et al.,* 2016), etc. For Afrino, *FGF12*, a

candidate genes for hair follicle development (Lv *et al.*, 2020) and reproductive traits (An *et al.*, 2018) and *ICA1* and *NXPH1*, associated with metabolic pathways (Akanno *et al.*, 2018) were some of the key candidate genes identified. Although the Afrino is predominantly selected or weighted for meat quality traits, the breed was established as a white-woollen breed for use as a terminal sire when crossing with Merino ewes (Synman, 2014a) in response to the presence of kemp (coloured fibre) in crosses of Merino ewes with mutton breeds. The signature for hair follicle development might be a reflection of this selection. The *ICA1* and *NXPH1* on the other hand are signatures of the intensive selection for weight put on the Afrino for meat traits. The four selection sweep regions identified in the Meatmaster breed corresponded to the *GPR171* on chromosome 1 which is associated with feed and metabolism (Ruiz-Larranaga *et al.*, 2018) the *HMGA* gene on chromosome 3 which was also observed as a selection signature in Sardinian Ancestral black sheep (Kijas *et al.*, 2012) and in Spanish breeds (Manunza *et al.*, 2016). *HMGA2* is involved in skeletal morphology and body size (Kijas *et al.*, 2012) and has been shown to be under selection in dogs with divergent stature (Jones *et al.*, 2008; Akey *et al.*, 2010). According to the breed society standards (https://www.meatmastersa.co.za/Breed-Standard.htm), Meatmaster sheep must be of average size, have a functional efficient body conformation and well placed legs with excellent walking ability. Such selection for body size and skeletal morphology could be the signature presented through the *HMGA2* gene and, together with *GPR171* gene associated with feed and metabolism, could ensure optimal performance for mutton production. *IL22* and *IL26* also on chromosome 3 are immune response genes that have been reported as under selection in some studies including (Fariello *et al.*, 2014). Detection of immune response genes is especially expected in breeds raised in arid environments with harsh and compromised production systems (Mdladla *et al.*, 2018).

Interestingly, more selection sweep regions were identified in the Dohne Merino sheep relative to the Afrino and Meatmaster. Dohne Merino shared two sweeps with Afrino, one on Chromosome 1 region 119,0-121,0 Mbs which harbours *IFNAR1; IL10RB; SLC5A3;CRYZL1* associated with anti-inflammatory and immune response (Uemoto *et al.*, 2020) and reproductive traits such as implantation of conceptus to uterus (Zhang *et al.*, 2013) and on chromosome 19 position 46,7-48,1Mb which harbours the gene *WNT5A* (45,5Mb), important in morphology particularly the development of limbs and skeleton (Fariello *et al.*, 2014) and reproductive traits (including mammary gland development) (Hao *et al.*, 2019). Other sweeps on Chromosome 6 (93,5-95,2 Mb) associated with the *FGF5* gene reported as a signature of selection in worldwide sheep breeds by (Kijas *et al.*, 2012) and (Fariello *et al.*, 2014) as well

as the *ANTXR2* gene associated with adaptation to variation in climatic conditions (Lv *et al.,* 2020). A selection sweep on chromosome 13 on region 46,5 - 48,1Mb in Afrino and 49,1-50,9 Mb in Dohne Merino was associated with the gene *BMP2* which has been reported as a signature of selection by (Kijas *et al.,* 2012) and (Fariello *et al.,* 2014) and strongly selected in both fat-tailed and thin-tailed sheep (Dong *et al.,* 2020).

Rsb results presented selection sweeps between a composite Merino-derived breed and each of its presumed ancestors. The Dohne Merino *vs* Merino analysis yielded 8 selection sweeps within 4 chromosomes of OAR 1, 3, 10 and 11. Sweeps on OAR 1 yielded genes such as the *GHSR* important for growth and carcass traits in sheep (Bahrami *et al.,* 2012), *SPATA16* associated with environmental variables in goats (Mdladla *et al.,* 2018) and male fertility in cattle (Wang *et al.,* 2014) and the *SLC7A14* and *SLC2A2* that are involved in nutrient transport and absorption (Wiedemar *et al.,* 2015). The Dohne Merino *vs* Merino and the Meatmaster *vs* Merino shared a sweep on chromosome 1 region 268,5 - 269,9 Mb associated with the Calpain gene *CAPN7*. The selection sweep on chromosome 10 region 28,6-30,3 Mb is associated with the *RXFP2* gene associated with polledness (Wang *et al.,* 2014; Wiedemar *et al.,* 2015). This region was shared between Dohne Merino *vs* Merino, Meatmaster *vs* Merino and Meatmaster *vs* Nguni. The region carried other genes such as the *FRY* gene which is associated with lambing percentage, ear size and coat phenotypes (Wei *et al.,* 2015) and the *BSGLCT* gene associated with wool traits. (Zhang *et al.,* 2013) suggested the role of *FRY* in sheep wool development. The Dohne Merino *vs* SAMM shared 3 regions with the Afrino and SAMM which were on (i) OAR 1 with *ILIRAP*; *FGF12* and *GMNC* genes; (ii) OAR 10 with *SPRY2*; *SLITRK4* and *GPC* genes and (iii) OAR 11 with *SPTSSB, PPMIL, B3GALNT* genes. The Dohne Merino *vs* SAMM shared all six sweeps regions on chromosomes 1,10 and 11 with Afrino vs SAMM. Genes such as *FGF12* on OAR 1:191,3-194,7Mb and *MAP2K4* on OAR11:28,6-31,3 were observed. Overall, the *Rsb* analysis revealed the direction of selection when these breeds were selected which focused on meat and wool production and robustness of breed through body confirmation, disease resistance and adaptability to the harsh production conditions of South Africa (Molotsi *et al.,* 2017; Kim *et al.,* 2016).

4.5 Conclusions

The variation within the breeds was greater compared to between breed diversity. Significant genetic variation is still maintained in South African Merino and Merino-derived sheep breeds. To develop these Merino-based breeds, it was important to include indigenous breeds in the

crossbreeding systems. There is limited information on the genomic influence of either Merino or other indigenous sheep on the different composite Merino-derived breeds. Results from this study demonstrated that whilst they maintained some genetic similarity with their founding breeds, the Merino derived breeds also diverged from their ancestral breeds at genomic regions aligned with the within breed selection pressures that included meat type and growth related trains in Meatmaster, wool traits in Afrino amongst other traits. The retained genomic similarities between Merino derived and ancestral populations could be of use in implementing across genomic selection programs. However, more appropriate detailed and focused analysis need to be included to investigate feasibility of such across breed genomic selection.

5 RUNS OF HOMOZYGOSITY AMONG SOUTH AFRICAN SHEEP BREEDS FROM VARIOUS PRODUCTION SYSTEMS[3]

[3] This manuscript was published in BMC Genomics: ***Dzomba, E.F***., *Chimonyo, M., Pierneef, R. et al. Runs of homozygosity analysis of South African sheep breeds from various production systems investigated using OvineSNP50k data. BMC Genomics* ***22,*** *7 (2021).* https://doi.org/10.1186/s12864-020-07314-2

Abstract

Population history, production system and within-breed selection pressure impacts the genome architecture resulting in reduced genetic diversity and frequency of runs of homozygosity islands. The current study tested the hypothesis that production systems geared towards specific traits of importance or natural or artificial selection pressures influenced the occurrence and distribution of runs of homozygosity (ROH) in the South African sheep population. The OvineSNP50 Beadchip was used to genotype 400 sheep belonging to 13 breeds from South Africa representing mutton, pelt and mutton and wool dual-purpose breeds, including indigenous non-descript breeds that are reared by South African smallholder farmers. To get more insight into the autozygosity and distribution of ROH islands of South African breeds relative to global populations, 623 genotypes of sheep from worldwide populations were included in the analysis. Runs of homozygosity were computed at cut-offs of 1-6 Mb, 6-12 Mb, 12-24 Mb, 24-48 Mb and >48 Mb, using the R package detectRUNS. The Golden Helix SVS program was used to investigate ROH islands. A total of 121,399 ROH with mean number of ROH per animal per breed ranging from 800 (African White Dorper) to 15,097 (Australian Poll Dorset). Analysis of the distribution of ROH according to their size showed that, for all breeds, the majority of the detected ROH were in the short (1-6 Mb) category (88.2%). Most animals had no ROH >48 Mb. Of the South African breeds, the Nguni and the Blackhead Persian displayed high Run of homozygosity based inbreeding (F_{ROH}) of 0.31 ± 0.05 and 0.31 ± 0.04, respectively. Highest incidence of common runs per SNP across breeds was observed on chromosome 10 with over 250 incidences of common ROHs. Mean proportion of SNPs per breed per ROH islands ranged from 0.02 ± 0.15 (island ROH224 on chromosome 23) to 0.13 ± 0.29 (island ROH175 on chromosome 15). Seventeen (17) of the islands had SNPs observed in single populations (unique ROH islands). The MacArthur Merino (MCM) population had five unique ROH islands followed by Blackhead Persian and Nguni with three each whilst the South African Mutton Merino, SA Merino, White Vital Swakara, Karakul, Dorset Horn and Chinese Merino each had one unique ROH island. Genes within ROH islands were associated with predominantly metabolic and immune response traits and predomestic selection for traits such as presence or absence of horns. Overall, the frequency and patterns of distribution of ROH observed in this study corresponds to the breed history and implied selection pressures exposed to the sheep populations under study.

Key words: *Sheep, production system, SNP genotypes, Runs of Homozygosity, autozygosity, ROH island*

5.1 Introduction

The genetic diversity of South African sheep populations is considered complex having been shaped by multifaceted production systems (Cloete *et al.,* 2014; Molotsi *et al.,* 2017) resulting from a combination of indigenous, commercial and synthetic/composite breeds raised to suit, various and often, extreme production conditions where natural selection forces are at play. Coupled to this have been farmer driven initiatives to crossbreed as an effort to develop breeds that are better suited to produce optimally under the harsh production conditions of the country. Whilst South African sheep genetic resources have been imported and introduced in other countries globally, there have also been movement of breeds into South Africa (Cloete *et al.,* 2016). The country has a combination of both large- and small-framed breeds where both inbreeding and outbreeding are considered dominant forces moulding their phenotypic appearance. Both natural and artificial selection of sheep, as well as regional variations due to drift have resulted in sheep breeds that differ extensively in phenotypes.

Production system and within-breed selection pressure have pronounced effects on the genome architecture and may cause reduced genetic diversity and frequency of runs of homozygosity islands (Szmatola *et al.,* 2019). Runs of homozygosity (ROH) are contiguous segments of homozygous genotypes that are present in an individual due to parents transmitting identical haplotypes to their offspring (Purfield *et al.,* 2012). The extent and frequency of ROHs are useful in providing information about the ancestry of an individual and its population (Purfield *et al.,* 2012; Joaqim *et al.,* 2019) with longer ROHs associated with more recent inbreeding within a pedigree while short ROHs are associated with ancient common ancestors (Szpiech *et al.,* 2019; Sams *et al.,* 2019). Shorter ROH can also be used to infer ancient relationships, information which in livestock is often missing due to limited recording. Long runs of homozygosity are persistent in inbred individuals, suggestive of unusually low mutation rates, high linkage disequilibrium (LD), and low recombination rates at certain genomic regions (Jemaa *et al.,* 2019). ROH accumulation in certain genomic positions has been used to analyze the demographic history in humans Halim *et al.,* 2015; Mooney *et al.,* 2018) and livestock populations (Upadhhyay *et al.,* 2016; Islam *et al.,* 2019). A study also used ROH to compare and characterize beef and dairy cattle breeds (Xu *et al.,* 2019). ROHs are also common in regions under positive selection and as such studies have associated accumulation of ROHs at specific loci to directional selection (Islam *et al.,* 2019; Chen *et al.,* 2018). In a number of studies, ROH have been used to estimate inbreeding levels and infer on signatures of selection

and genetic adaptation to production conditions (Sandenbergh, 2015; D`Ambrosio *et al.,* 2019; Aramburu *et al.,* 2020).

The OvineSNP50 BeadChip is a genome-wide genotyping array for sheep and was developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC). This BeadChip contains 54,241 SNPs that were chosen to be uniformly distributed across the ovine genome with an average gap size and distance of 50.9 Kb and 46 Kb, respectively, and were validated in more than 75 economically important sheep breeds (OvineSNP50 Datasheet, https://www.illumina.com/documents/products/datasheets/datasheet_ovinesnp50.pdf). The current study used the OvineSNP50 Beadchip to investigate the distribution of ROH in South African sheep breeds sampled from different breeding goals and production systems of mutton, wool, pelt and commercial versus smallholder sectors as well as various other sheep breeds obtained globally. The objectives of the study were to investigate the occurrence and distribution of ROH; characterize autozygosity and identify genomic regions with high ROH islands with the aim to draw insights into how the South African sheep populations were in the past, as well as how their structure and demography have evolved over time. The study presumed that the founder population establishing genetic processes and the extent of breeding control have differed greatly among the different sheep breeds of South Africa and globally. The study, therefore, hypothesised that production systems geared towards specific traits of importance such as mutton, wool, pelt or multiple traits (as with some dual-purpose breeds) or absence of selection programs *e.g*. in non-descript breeds kept by smallholder farmers influences the occurrence and distribution of ROH. In Chapter 3, the South African sheep breeds clustered according to breed and production system as illustrated in Figure 3.1. Using ROHs, the current study was used to infer the impact of breed history, inbreeding levels and selection generating the accumulation homozygous mutations in the diverse sheep populations. Global sheep populations accessed from the ISGC http://www.sheephapmap.org were used to further analyse the development and separation of populations from their presumed founder populations.

5.2 Methods

5.2.1 Animal populations

Four hundred animals belonging to 14 South African breeds/populations consisting of mutton (South African Mutton Merino (n = 10), Dohne Merino (n = 50), Meatmaster (n = 48),

Blackhead Persian (n = 14) and Namaqua Afrikaner (n = 12), pelt (Swakara subpopulations of Grey (n = 22); Black (n = 16); White-vital (n = 41) and White-subvital (n = 17) and Karakul (n = 10); wool (SA Merino (n = 56), dual purpose breeds (Dorper (n = 23); Afrino (n = 51) and non-descript Nguni sheep (n =30) were used in the study The details on the breeds used, the SNP genotyping is described in Chapter 3, Section 3.2.1 to 3.2.2. In addition 623 genotypes from a global set of sheep breeds representing worldwide populations were included in the analysis as described in Chapter 3, Section 3.2.3.

The two data sets were merged into a dataset that consisted of 1,019 animals from 31 sheep breeds/populations and 43,556 SNPs that were retained for analyses after global quality control of both the South African and ISGC sheep breeds (Table 5.1). Chromosomal coordinates for each SNP were obtained from ovine genome assembly 4.1 (OAR4.1). Markers were filtered to exclude loci assigned to unmapped contigs. Only SNPs located on autosomes were considered for further analyses. Moreover, the following filtering criteria were adopted to exclude certain loci and animals and to generate the pruned input file: (i) SNPs with a call rate <95 % and (ii) minor allele frequency <1 % and (iii) animals with more than 2 % of missing genotypes were removed. File editing was carried out using Plink (Purcell *et al.*, 2007).

5.2.3 Detection of common runs of homozygosity

Runs of homozygosity were computed using the R package detectRUNS and the consecutive runs method (Biscarini *et al.*, 2018). No pruning was performed based on LD, but the minimum length that constituted the ROH was set to 1 Mb to exclude short ROH deriving from LD. The following criteria were used to define the ROH: (i) one missing SNP and up to one possible heterozygous genotype was allowed in the ROH, (ii) the minimum number of SNPs that constituted the ROH was set to 30 (iii) the minimum SNP density per ROH was set to one SNP every 100 Kb and (iv) the maximum gap between consecutive homozygous SNPs was 250 Kb. The computed ROHs were then categorised into bins based on lengths of 1-6 Mb, 6-12 Mb, 12-24 Mb, 24-48 Mb and >48 Mb.

The mean number (MN_{ROH}) and average length (AL_{ROH}) of ROH per breed as well as the average sum of ROH segments per breed were estimated. The inbreeding coefficient (F_{ROH}) was estimated based on the ROH for each animal and averaged per breed. F_{ROH} was calculated within detectRUNS using the following formula:

$$F_{ROH} = L_{ROH} / L_{AUTO},$$

where:

L_{ROH} is the total length of ROH on autosomes and;
L_{AUTO} is the total length of the autosomes covered by SNPs, which was 2,453 Mb

For comparison, inbreeding coefficients were also estimated using variance between observed and expected heterozygosity (F_{HOM}). This was done using Golden Helix SVS software.

To identify the genomic regions most commonly associated with ROH for the meta-population and for groups on the basis of production purposes (mutton, wool and pelt and dual purpose breeds), Golden Helix SVS was used to analyse for the incidence of common runs per SNP, which was then plotted against the position of the SNP along the chromosome (OAR).

ROH islands were defined as clusters of runs that were >1000 Kb with a minimum of 30 SNPs and found in more than 20 samples and analysed for using Golden Helix SVS. For each sample, the proportion of SNPs in the ROH island was estimated. The mean proportion of SNPs per sample per ROH islands was determined using Proc MEANS procedure in SAS v9.4 (SAS, 2013). The variance in mean proportion of SNP in ROH islands amongst breeds was analysed using the Proc GLM in SAS v9.4 (SAS, 2013) using the following model:

$$Proportion\ of\ per\ ROH\ island = \mu + B_i + e$$

where:

μ = *overall mean;*

B_i = *Breed effect and;*

e = *random residual error*

5.2.3.1 Functional annotation of ROH islands

ROH islands that were constituted by SNPs from 1, 2 or 3 populations (considered as unique islands) and those common islands with SNPs from three quarters of the populations (>23

populations) were used for functional annotation. The genomic region associated with each of these island was annotated using the Sheep Quantitative Trait Loci (QTL) database (https://www.animalgenome.org/cgi-bin/QTLdb/OA/summary) and the Genome Browser (http://genome.ucsc.edu/). The genomic coordinates for these ROH islands were used for the annotation of genes that were fully or partially contained within each selected region using the Genome Browser (http://genome.ucsc.edu/) and submitted to the DAVID database (http://david.abcc.ncifcrf.gov/) for gene ontology (GO). Finally, the Kyoto encyclopedia of genes and genomes (KEGG) analysis was used to investigate pathways associated with each annotated gene within ROH islands. Significant enrichment in the candidate genes was indicated by a *p*-value of <0.05.

5.3 Results

5.3.1 Runs of homozygosity counts

The study identified 121 399 ROH in total with mean number of ROH per animal per breed ranging from 800 (African White Dorper) to 15 097 (Australian Poll Dorset) as illustrated in Figure 5.1. Analysis of the distribution of ROH according to their size showed that, for all breeds, the majority of the detected ROH were in the smallest 1-6 Mb in length category (88 %) ranging from 684 in African White Dorper (n = 684) to Australian Poll Dorset (n = 13 677). The longest ROHs (>48MB) were the least (n = 108) with most animals detecting no ROH in this category. The Black Head Mountain had largest number of long (>48 Mb) of 30 followed by Dorset Horn with 16 ROH >48 Mb as illustrated in Figure 5.1. The average length of ROH across breeds was 5.88 Mb and ranged from 2.60 Mb (Afrino) to 6.90 Mb (Nguni).

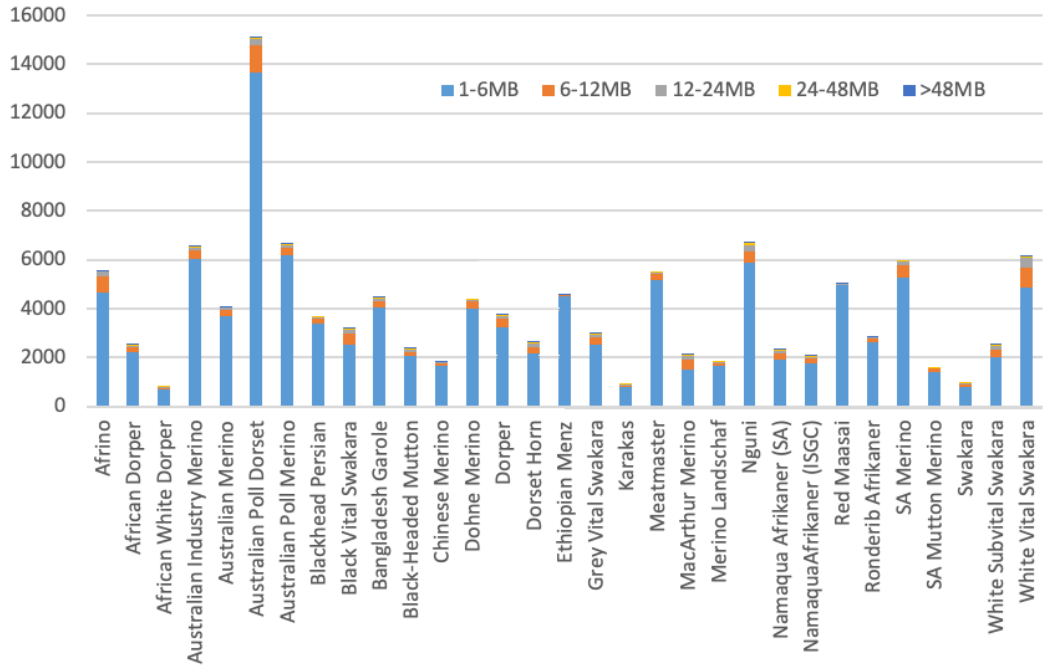


Figure 5.1: Runs of Homozygosity of different lengths per breed.

5.3.2 Inbreeding coefficient

McArthur Merino showed the highest value of inbreeding on the basis of ROH (F_{ROH} = 0.45 ± 0.03), whereas Australian Poll Merino (F_{ROH} = 0.08 ± 0.03) showed the lowest (Table 5.1). Of the South African breeds, the Nguni and the Blackhead Persian displayed high F_{ROH} of 0.31 ± 0.05 and 0.31 ± 0.04, respectively. Other breeds with high F_{ROH} included the Black Vital Swakara and the White Subvital and Vital Swakara with F_{ROH} >0.28 (Table 5.1). South African breeds with low F_{ROH} included Dohne Merino (F_{ROH} = 0.10 ± 0.02), the Meatmaster with F_{ROH} of 0.13 ± 0.02 and South African Merino with F_{ROH} = 0.14 ± 0.05. Inbreeding coefficient based on variance F_{HOM} are presented in Table 5.1. A correlation between F_{ROH} and F_{HOM} was observed, with breeds such as Blackhead Persian, Nguni displaying high F_{ROH} and F_{HOM}, respectively.

5.3.3 ROHs per chromosome per breed

The distribution of ROHs per chromosome per breed are illustrated in Figure 5.2. Runs were evenly distributed amongst chromosomes within breeds.

Table 5.1: Mean and Standard deviation of ROH based inbreeding (F_{ROH}) of South African and global sheep populations

Breed	No. animals	Mean F_{ROH}	SD	Mean F_{HOM}	SD
Afrino	51	0.162	0.0212	0.127	0.0210
African Dorper	21	0.155	0.0376	0.217	0.0996
African White Dorper	6	0.201	0.0333	0.163	0.0352
Australian Industry Merino	88	0.091	0.0272	0.104	0.0479
Australian Merino	50	0.103	0.0421	0.127	0.0381
Australian Poll Dorset	108	0.176	0.0402	0.117	0.0400
Australian Poll Merino	98	0.079	0.0275	0.049	0.0278
Blackhead Persian	14	0.309	0.0435	0.343	0.0434
Black Vital Swakara	20	0.292	0.0534	0.289	0.0515
Bangladesh Galore	24	0.233	0.0671	0.270	0.0628
Black-headed Mutton	24	0.184	0.1164	0.156	0.1202
Chinese Merino	23	0.096	0.0489	0.065	0.0492
Dohne Merino	50	0.103	0.0167	0.075	0.0180
Dorper	23	0.237	0.0890	0.217	0.0996
Dorset Horn	21	0.242	0.0604	0.193	0.0625
Ethiopian Menz	34	0.121	0.0387	0.179	0.0353
Grey Vital Swakara	22	0.205	0.0588	0.200	0.0562
Karakas	18	0.081	0.0553	0.095	0.0552
Meatmaster	46	0.126	0.0202	0.121	0.0209
MacArthur Merino	12	0.448	0.0332	0.351	0.1356
Merinolandschaf	22	0.101	0.0146	0.076	0.0153
Nguni	30	0.314	0.0521	0.348	0.0487
Namaqua Afrikaner (SA)	12	0.321	0.1318	0.313	0.0746
Namaqua Afrikaner (ISGC)	10	0.261	0.0272	0.222	0.0213
Red Massai	45	0.101	0.0234	0.169	0.0283
Ronderib Afrikaner	19	0.197	0.0665	0.194	0.0664
South African Merino	56	0.140	0.0467	0.141	0.0294
South African Mutton Merino	10	0.224	0.0366	0.191	0.0386
Swakara	6	0.262	0.0364	0.259	0.0354
White Sub-Vital Swakara	16	0.286	0.0795	0.280	0.0801
White Vital Swakara	40	0.282	0.0520	0.275	0.0505
Overall	**1019**	**0.162**	**0.0900**	**0.146**	**0.0992**

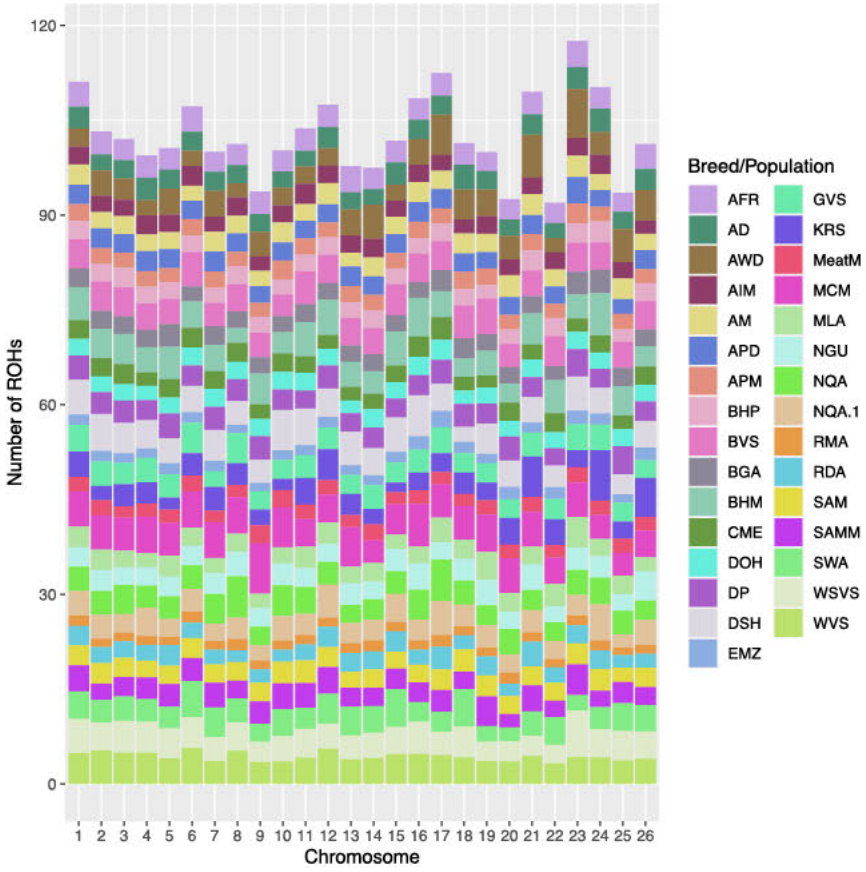


Figure 5.2: Number of ROHs per chromosome per breed

Legend:

AFR = Afrino

AWD = African White Dorper

AIM = Australian Industry Merino

AM = Australian Merino

APD = Australian Poll Dorset

APM = Australian Poll Merino

BHP = Blackhead Persian

BVS = Black Vital Swakara

BGM = Bangladesh Garole

BHM Blackheaded Mountain

CME = Chinese Merino

DOH = Dohne Merino

DP = Dorper

DSH = Dorset Horn

EMZ = Ethiopian Menz

GVS = Grey Vital Swakara

KRS = Karakas

MeatM = Meatmaster

MCM = MacArthur Merino

MLA = Merinolandschaf

NGU = Nguni

NQA = Namaqua Afrikaner

RMA = Red Maasai

RDA = Ronderib Afrikaner

SAM = SA Merino

SAMM = SA Mutton Merino

SWA = Swakara

WSVS = White Subvital Swakara

WVS = White Vital Swakara

5.3.4 Incidences of common runs per SNP

Using Golden Helix SVS, an analysis was conducted to investigate the incidence of common runs per SNP and results are illustrated in Figure 5.3. Highest incidence of common runs per SNP across breeds was observed on chromosome 10 with over 250 incidences of common ROHs at some of the SNPs (Figure 5.3). Other chromosomes such as 2, 6, 13, 15 and 19 were found to have moderate incidences of common SNPs averaging 150-160 (Figure 5.3). Across breeds, certain regions were observed to be absent of ROHs notable of which were chromosomes 10 (± 7 Mbs region); 21 (± 40 Mbs region); 22 (± 18 Mbs region) and 26 (± 8 Mbs region) as illustrated in Figure 5.4.

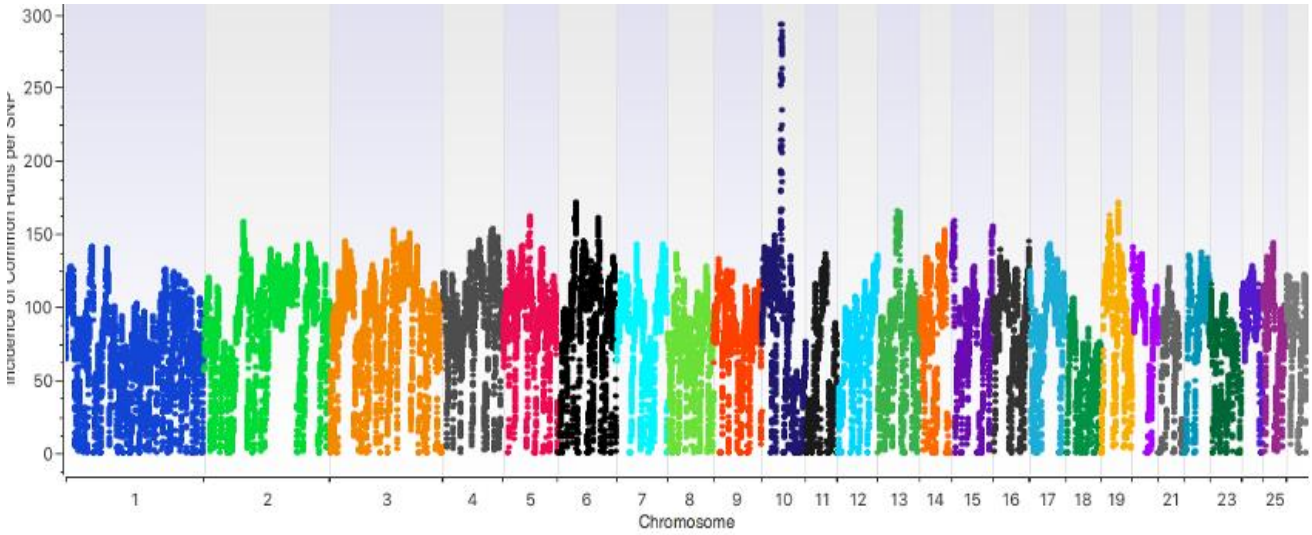


Figure 5.3: Incidences of common runs per SNP per chromosome.

5.3.5 ROH islands

A total of 244 ROH islands distributed across all 26 autosomes were observed. Mean proportion of SNPs in ROH island ranged from 0.02 ± 0.15 (island ROH224 on chromosome 23) to 0.13 ± 0.29 (island ROH175 on chromosome 15). Number of islands ranged from a minimum of two clusters per chromosome (on chromosome 22) to 32 clusters per chromosome (on chromosome 1). Seventeen of the islands were observed in single populations and considered unique ROH islands. Thirty-nine of the reported ROHs were each observed in three populations whilst the remaining 188 were each observed in more than 3 populations and considered common islands.

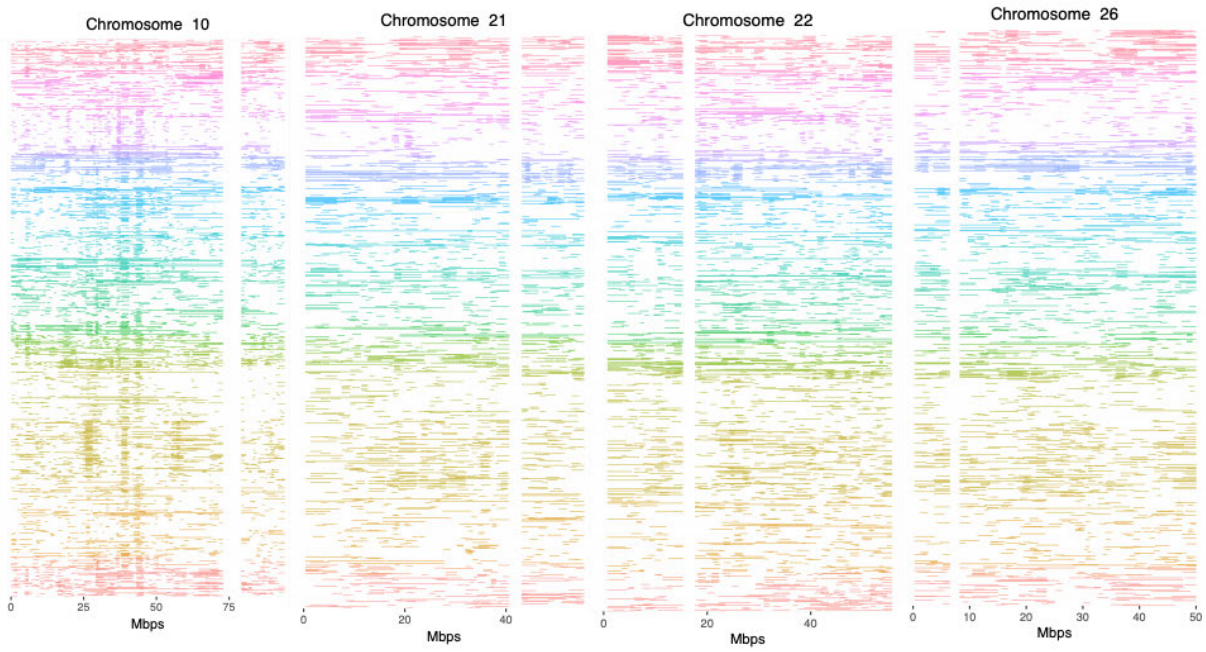


Figure 5.4: Plots for consectutiuve ROHs illustrating chromosomal regions with no ROHs on chromosome 10, 21, 22 and 26

The MacArthur Merino population had five unique ROH islands followed by Nguni and Blackhead Persian with three each whilst the South African Merino, South African Mutton Merino, White Vital Swakara, Karakul, Dorset Horn and Chinese Merino each had one unique ROH island. For those islands shared between two populations, the Blackhead Persian shared with South African Mutton Merino, SA Merino, Black Vital Swakara, Ronderib Afrikaner and MacArthur Merino; the MacArthur Merino shared with Nguni, Namaqua Afrikaner, Blackhead Persian and Bangladesh Galore; the Nguni shared with MacArthur Merino, White Vital Swakara and Karakul, the Dorper with NQA and RDA while the AWD shared with South African Mutton Merino. For ROH islands shared amongst three populations the Nguni shared one with White Vital Swakara and Karakul; the African White Dorper one with Namaqua Afrikaner and White Subvital Swakara; the Bangaldesh Galore with Blackhead Persian and South African Mutton Merino; the Dorset Horn with Nguni and White Vital Swakara while the BVS with RDA and South African Mutton Merino.

5.3.6 SNPs and gene annotation

Based on the Sheep QTL database, ROH7 on chromosome one, and unique to Blackhead Persian lies within a genomic region previously found to be harboring QTLs for selection for presence or absence of horns in Soay sheep and average daily gain in Awassi and Merino sheep. ROH island ROH15 on chromosome one and unique to MacArthur Merino lies in a genomic region harboring QTLs for faecal egg count and susceptibility to facial eczema. QTLs associated with carcass fat percentage (*FATP*) in Awassi and Merino sheep were also observed within island (ROH15).

A list of genes and associated KEGG pathways found with population restricted ROH islands are given in Tables 5.2 and 5.3. Using the KEGG pathway analysis, genes associated with metabolic pathways such as *ANPEP, HDDC3 and ST3GAL3* where observed with ROH islands unique to the Blackhead Persian (Table 5.2a). Two ROH islands (ROH100 and ROH101) unique to Blackhead Persian and Nguni, respectively, were observed in chemokine pathways (genes *GRK4*) and Natural killer cell mediated cytotoxicity pathways (genes: *SH3BP2*). Other pathways observed included thermogenesis (ROH199), interleukin signalling pathways (ROH11) and pathways associated with bacterial infections such as *Eschellichia coli, Salmonella,* and tuberculosis (TB) (ROH149) as shown in Table 5.2.

Table 5.2: ROH island observed in <3 populations and the associated genes and KEGG pathways

ROH island	CHR	Breeds	Genes	KEGG pathway
ROH2	1	BHP	*ST3GAL3.*	Metabolism (sphingolipid metabolism).
ROH199	18	BHP	*BTBD1*	Fanconi anaemia pathway
			WHAMM	Tight junction
			BLM	Fanconi anaemia pathway
			FES	Biosynthesis of secondary metabolites
			FANCI	Fanconi anaemia pathway
			PLIN1	Regulation of lipolysis in adipocytes; Thermogenesis
			PEX11A	Peroxisome
			ANPEP	Renin-angiotensin system; Hematopoietic cell lineage; Metabolic pathways
			HDDC3	Purine metabolism; Metabolic pathways
ROH100	6	BHP	*ADRA2C*	CGMP-PKG signalling pathway
			GRK4	Chemokine signalling pathway; Endocytosis
			SH3BP2	Natural killer cell mediated cytotoxicity
			PIGG	Prodigiosin biosynthesis; Metabolic pathways
ROH101	6	NGU	*ADRA2C*	CGMP-PKG signalling pathway
			GRK4	Chemokine signalling pathway; Endocytosis
			SH3BP2	Natural killer cell mediated cytotoxicity
			PIGG	Prodigios in biosynthesis; Metabolic Pathways
ROH108	7	SAM	*DAAM1*	Wnt signalling pathway
			L3HYPDH	Arginine and proline metabolism
ROH149	12	MCM	*IL10*	Cytokine-cytokine interaction; Toxoplasmosis; Tuberculosis; Various diseases
ROH11	1	NGU & MCM	*PRR4*	Circadian rhythm – plant
			S100A9	IL-17 signalling pathway
			CRTC2	Glucagon signalling pathway, Insulin resistance
			RAB13	Tight junction
			TPM3	Cardiac muscle contraction
			IL6R	Viral protein interaction with cytokine and cytokine receptor
ROH51	3	NQA & MCM	*TPRN*	Type I diabetes mellitus
			VAV2	Rap1 signalling pathway
			SARDH	Glycine, serine and threonine metabolism
			ABO	Metabolic Pathways
58	3	BHP & RDA	*FSHR*	Ovarian steroidogenesis
			LHCGR	Prolactin signalling pathway
75	4	BHP & MCM	*VPS41*	Salmonella infection; Autophagy – yeast
			AMPH	Inflammatory mediator regulation of TRP channels; Biosynthesis of secondary metabolites; Metabolic Pathways; Fc gamma R-mediated phagocytosis
			WNT16	Signalling pathways regulating pluripotency of stem cells; Melanogenesis

Legend: BHP= Blackhead Persian; Ngu = Nguni; MCM = MacArthur Merino; SAM = SA Merino; NQA Namaqua Afrikaner; RDA = Ronderib Afrikaner

Table 5.3: ROH island observed in <3 populations and the associated genes and KEGG pathways

ROH island	CHR	Breeds	Genes	KEGG pathway
			AASS	Lysine degradation; Metabolic pathways
ROH125	9	BHP & SAM	*MSC*	Biosynthesis of secondary metabolites; Sulfur metabolism; Degradation of aromatic compounds; Microbial metabolism in diverse environments; Metabolic pathways; Staphylococcus aureus infection
			TRPA1	Inflammatory mediator regulation of TRP channels
			RDH10	Retinol metabolism
			UBE2W	Ubiquitin mediated proteolysis
			LY96	NF-kappa B signalling pathway; Toll-like receptor signalling pathway; Salmonella infection; Pertussis; Toxoplasmosis
ROH200	18	BHP & SAMM	*CTSH*	Lysosome; Apoptosis
			RASGRF1	Ras signalling pathway; MAPK signalling pathway; Focal adhesion
			BCL2A1	NF-kappa B signalling pathway; Apoptosis
			FAH	Microbial metabolism in diverse environments; Metabolic pathways; Nitrogen metabolism
			IL16	Cytokine-cytokine receptor interaction
			MCEE	Carbon metabolism; Metabolic pathways; Microbial metabolism in diverse environments; Valine, leucine and isoleucine degradation
			FAN1	Fanconi anaemia pathway
			TJP1	Tight junction; Adherens junction; Gap junction; Vibrio cholerae infection; Epithelial cell signalling in Helicobacter pylori infection; Pathogenic Escherichia coli infection
ROH11	1	NGU & MCM	*S100A9*	IL-17 signalling pathway
			S100A8	IL-17 signalling pathway
			BGLAP	Parathyroid hormone synthesis, secretion and action
			ARHGEF2	Pathogenic Escherichia coli infection; Tight junction; Bacterial invasion of epithelial cells; Salmonella infection
			ASH1L	Metabolic pathways; Lysine degradation
			FDPS	Biosynthesis of secondary metabolites; Metabolic pathways
ROH232	24	BVS; RDA; SAMM	*STUB1*	Protein processing in endoplasmic reticulum; Ubiquitin mediated proteolysis
			SSTR5	Growth hormone synthesis, secretion and action; cAMP signalling pathway
			UBE2I	Ubiquitin mediated proteolysis; RNA transport
			GNPTG	Lysosome
			ORAI2	Calcium signalling pathway
			POR	RNA transport; ABC transporters; Protein digestion and absorption; Carbohydrate digestion and absorption; Glycolysis / Gluconeogenesis

Legend: BHP= Blackhead Persian; Ngu = Nguni; MCM = MacArthur Merino; SAM = SA Merino; NQA Namaqua Afrikaner; RDA = Ronderib Afrikaner; SAMM = SA Mutton Merino

5.4 Discussion

The domestication of sheep was a complex process that allowed both natural and artificial selection of breeds. Regional variations due to genetic drift in breeds became small and geographically restricted resulting in extensive and diverse phenotypes. Such processes, while well documented in other breeds, are unknown in most of the small and geographically restricted local populations. ROH have been extensively studied in humans and livestock populations and are an established method of inferring population history. ROH are continuous homozygous segments that are common in individuals and populations. The ability of these homozygous segments to give insight into a population's genetic events makes them a useful tool that can provide information about the demographic evolution of a population over time. Furthermore, ROH provide useful information about the genetic relatedness among individuals, helping to manage inbreeding rate, thereby exposing possible deleterious variants in the genome. ROH are widely used as predictors of inbreeding levels in populations. Calculating the inbreeding coefficient from ROH (F_{ROH}) is more accurate for estimating autozygosity and for detecting both past and more recent inbreeding effects than estimating inbreeding from pedigree data (Kim *et al.*, 2015).

South African sheep populations are a result of multifaceted production systems. Natural and artificial selection forces play vital roles through mixtures of indigenous, commercial and synthetic/composite breeds raised in extreme production conditions. ROH were used in the current study to investigate this population history with emphasis on breed relatedness. OvineSNP50 Beadchip genotypes of South African mutton, wool, pelt and dual purpose and non-descript breeds was analysed together with that of global populations of similar geographic background.

Breeds such as the Australian Poll Dorset followed by the South African Nguni, Australian Poll Merino and Australian Industry Merino and the White Vital Swakara sheep had the most ROHs ranging from 6 155 in White Vital Swakara to 15 097 in Australian Poll Dorset. Frequency of ROHs in different breeds reflect on the size of the breeds often positively associated with inbreeding levels (Kirin *et al.*, 2010). South African Nguni, Blackhead Persian, Namaqua Afrikaner and Swakara are small breeds restricted to specific production systems and geographic locations (Kunene *et al.*, 2014; Selepe *et al.*, 2018), which explains the high frequency of ROH in these populations. Similarly-raised worldwide populations include the

Ethiopian Menzi (Gizaw *et al.*, 2013; Edea *et al.*, 2017), Bangladesh Galore (Deb *et al.*, 2019), Black-headed Mutton (Schönherz *et al.*, 2020) that also had high numbers of ROH in the current study. ROH due to recent inbreeding tends to be longer, due to little opportunity for recombination to break up the segments that are identical-by-descent (Curik *et al.*, 2014). The Nguni (725) Afrino (845), White Vital Swakara (1 204) and Australian Poll Dorset (1 359) had the most moderately sized (6-24 Mb) ROH while the Australian Poll Dorset (61), DSH (75), BHM (90), Nguni (90) and White Vital Swakara (94) has the most of the large (>24 Mb) ROH presenting more ancient inbreeding. Ancient ROH are generally much shorter because the chromosomal segments have been broken down by repeated meiosis (Kirin *et al.*, 2010; Curik *et al.*, 2014). In the current study, the Nguni (5 855), Australian Industry Merino (6 040), Australian Poll Merino (6 188) and Australian Poll Dorset (13,677) had the highest number of short ROH implying more recent inbreeding events. The Nguni breed of South Africa is a non-descript breed kept by smallholder farmers under low-input communal farming systems (Qwabe *et al.*, 2013; Kunene *et al.*, 2014). Small flock sizes, sharing and retaining of bucks for multiple breeding cycles characterises the smallholder livestock production systems of South Africa inclusive of sheep. It has been suggested that such production factors lead to inbreeding in these populations (Hlophe, 2011).

Notable inbreeding has been observed in Zulu sheep (Kunene *et al.*, 2014). The spread of Zulu sheep into different areas of KwaZulu-Natal has fractured the sheep into isolated subpopulations occupying different ecological, social-cultural and management environments (Synman, 2014). The high frequency of both short and long ROH in the Nguni relative to other breeds is supported by the breed history and suggestive of a sustained high level of inbreeding due to both founder effects and the practice by smallholder farmers of raising the sheep as small fragmented populations. Contrary to the Nguni are the composite commercial breeds of South Africa, *e.g*. Afrino and Meatmaster that, although they experienced population bottlenecks during their formation which is now reflected by the high frequency of short ROH, the breeds are now well managed commercially and as such present minimum long ROH >24 Mb. The Blackhead Persian were initially introduced to South Africa by chance in 1869. A vessel damaged by a storm at sea carried a number of slaughter sheep. These sheep, one ram and three ewes, were taken to Wellington where the breed was further developed (Malesa, 2015). In 1948, the present Blackhead Persian Sheep Breeders' Society of South Africa was formed in De Aar and to date the Black-headed Persian is represented by a well-established breed society with breed standards and management practise. Such history of the Blackhead Persian explain

the predominating short ROHs reflective of ancient founder effects during its establishment in South Africa and the few long ROHs since it is now well managed with a representing breed society.

Overall, the patterns of distribution of ROH revealed in this study showed peculiar patterns of inbreeding of sheep breeds that corresponded with levels of selection pressure typical of trait of economic importance as well as the production system typical of their rearing. South African indigenous and local breeds of Nguni, Blackhead Persian, Namaqua Afrikaner and the pelt based subpopulations of the Swakara Sheep had high inbreeding estimates reflective of the small and fragmented populations. The Nguni and Namaqua Afrikaner for example are raised as small household flocks in geographically marginalised regions of the country (Kunene *et al.,* 2014) whereas the pelt based Swakara subpopulations are small populations raised for a unique production system of pelt (Mavule *et al.,* 2016). Challenges of small effective population size and inbreeding have been suggested in these populations (Hlophe, 2011; Kunene *et al.,* 2014; Muchadeyi *et al.,* 2015). Such high inbreeding levels imply that the breeds are of low genetic diversity and at risk of extinction. Conservation efforts are therefore required to minimise further loss of genetic diversity and extinction of some of these breeds. Phenotypic and genetic characterisation as well as sustainable utilisation of the genetic resources and implementation of structured and tailor-made breeding programs have been suggested as alternative approaches to conservation of threatened genetic diversity (Molotsi *et al.,* 2017). The global populations, the MCM from south west Europe (Kijas *et al.,* 2012) had a high F_{ROH} similar to that of the Swakara subpopulations and the Namaqua Afrikaner. The Dorper, Dorset Horn and South African Mutton Merino also shared a high F_{ROH} as the Swakara breeds, MCM and Namaqua Afrikaner.

ROH frequencies vary widely within and across chromosomes (Curik *et al.,* 2014). Chromosome 10 had the highest incidence of common runs per SNP across breeds with over 250 incidences of common ROHs at some of the SNPs. Chromosome 10 harbours a genomic region associated with horns in sheep (Kijas *et al.,* 2012). One of the regions on chromosome 10 harbours the *RXFP2* gene which has been reported as a main candidate for horns in Soay sheep (Johnston *et al.*, 2011). Other studies have identified *RXFP2* within a quantitative trait locus for horn size and highly heritable in sheep (Sim *et al.,* 2019). It was reported a 1.8 Kb insertion in the 3'-UTR of *RXFP2* to be associated with polledness in sheep (Wiedemar *et al.,* 2015). Presence and absence of horns and subsequently horn size have been the main

parameters under selection in sheep pre-and post-domestication. The *RXFP2* region has been observed to be under selection in some African sheep populations (Ahbara *et al.*, 2019). Other regions on chromosome 10 were SNPs within genes such as *Crystallin lambda 1(CRYL1)* which is associated with metabolic pathways particularly pentose and glucuronate interconversions and has been observed to be under trans-specific signatures of domestication in sheep and goats (Alberto *et al.*, 2018).

Chromosome 6 also harboured SNPs with high incidence of common ROH across breeds. Some of the associated genes included the *Secreted phosphoprotein 1* (*SPP1*) and *ligand dependent nuclear receptor corepressor-like* (*LCORL*) which are on a domain of 36.15-38.56 Mb and play an essential role in tissue and embryonic growth (Al-Mamun *et al.*, 2015; La *et al.*, 2019). *LCORL* was observed to be associated with height in cattle (Gudbjartsson *et al.*, 2008; Weedon *et al.*, 2008) and observed to be under selection in different sheep breeds (Al-Mamun *et al.*, 2015; Signer-Hasler, 2019) and in other species (Bolormaa *et al.*, 2016; Wood *et al.*, 2014; Saatchi *et al.*, 2014). La *et al.* (2019) observed significantly high expression of *SPP1* in the kidney of Hu sheep. The *Prostaglandin f2-alpha synthase (PGFS)* gene on chromosome 13 is associated with the Arachidonic acid metabolism and other metabolic pathways and were observed to be associated with wool growth regulation in Aohan fine wool sheep (Liu *et al.*, 2014).

The failure to observe ROHs in certain genomic regions (*i.e.* Chromosomes 10, 21, 22 and 26) could be attributed to gaps in marker coverage. Nandolo *et al.*, (2018) suggested that artefacts due to structural variants and gaps in marker coverage could influence the screening of ROHs. However, whilst this will be considered a SNP chip effect, affecting the screening of ROH similarly across breeds, the impact of such an artefacts might have minimal effects on the breed comparisons undertaken in this study.

In the current study, ROH islands were defined as clusters of runs that were >1,000 Kb with a minimum of 30 SNPs and found in more than 20 samples which represented 1.95% of the total population in this study. In other studies, a threshold of 1% was used (Perpolli *et al.*, 2018). Zhang *et al.* (2015) reported that ROH patterns are not randomly distributed across the genomes, and are seen to be distributed and shared among individuals as a result of selection events. The 244 ROH islands observed in this study varied from those with SNPs unique to one population (17 ROH islands), two-three populations (39 islands) and those distributed in more than 3 populations (188 islands). As expected, the highly inbred populations of the MCM,

Blackhead Persian and Nguni had the highest frequency of unique ROHs suggestive of small, fragmented populations with small effective population sizes and evolving independently from other populations. The Nguni and Blackhead Persian for example are small breeds kept by smallholder farmers in unique productionn systems (Kunene *et al.,* 2014). The MCM population is also a highly inbred population from west Europe (Zhang *et al.,* 2015). The other breeds such as White Vital Swakara and Chinese Merino are also equally small breeds highly selected for specific production purposes, *e.g.* pelt production in the case of White Vital Swakara sheep (Ciani *et al.,* 2015). Previous studies suggested that ROH islands are a result of intensive selection often found in populations of finite size (Perpolli *et al.,* 2018; Purfield *et al.,* 2018).

Based on the Sheep QTL database, both productive traits, *i.e.* average daily gain and carcass fat percentage, and adaptive traits, *e.g.* absence of horns, feacal egg count and susceptibility to facial eczema, were observed within reported ROH islands. The Blackhead Persian is a polled breed with both sexes lacking horns. The observed unique island, ROH7, associated with selection of absence of horns (Johnston *et al.,* 2011) will therefore be in line with ancient selection for horns in this breed. The identified regions under ROH islands associated with metabolic pathways (*e.g. ST3GAL3* on island ROH2*; FES* gene on island ROH199*)*, adaptive and innate immunity (*ADRA2C, GRK4* and *SH3BH2* genes on island ROH100), thermogenesis (*PLIN1* gene on ROH199*)* relate to key traits relevant to the Blackhead Persian's and other livestock's survival in harsh compromised environments smallholder populations of South Africa (Mdladla *et al.,* 2018). Similarly the Nguni sheep shared genomic regions (*i.e. ADRA2C, GRK4* and *SH3BH2* genes on island ROH101) as well as traits affected by different genomic regions such as island ROH11 on chromosome 1 that harboured the *S100A9* gene associated with the *IL-17* signalling pathway and *CRTC2* gene associated with the Glucagon signalling pathway. ROH islands that were shared between populations implied common selection pressures between/amongst affected breeds for example the Blackhead Persian and the South African Mutton Merino and South African Merino that shared ROH islands associated with metabolic and immune response pathways.

5.5 Conclusions

The study reported frequency and distribution of ROHs in South African sheep breeds, relative to global populations. The pattern of distribution of ROH corresponded to breed history and production system under which they are raised. Overall, the study showed peculiar patterns of

inbreeding of sheep breeds that corresponded with levels of selection pressure typical of traits of economic importance as well as the production system aligned to their rearing. Similarities in frequency and patterns of ROHs within South African breeds and between South African and other global breeds was observed especially when comparing the Merino-type breeds. Such similarities is suggestive of closely related breeds that share considerable genomic regions to allow between breed genomic improvement programs. Across breed genomic selection is particulary relavent in a sector characterised by breeds of small population sizes that could be joined together to constitute a reference population. Whilst ROH analysis did shed light into such breed similarities, there are analysis that can infer such possibilities with high accuracy and include analysis of linkage disequilibrium and consistence in gametic phase.

6 LINKAGE DISEQUILIBRIUM PATTERN, EFFECTIVE POPULATION SIZE AND PERSISTENCE IN LD PHASE IN MULTI-BREED SOUTH AFRICAN SHEEP POPULATIONS[4]

[4] This manuscript is in preparation for the Journal of Animal Breeding and Genetics

Abstract

The accuracy of genome-wide association studies, genomic selection and other applications is dependent on the level of LD across the genome. The current study investigated (i) LD between adjacent SNPs, (ii) LD decay with increased marker distance, (iii) trends in effective population size over time and (iv) consistency of gametic phase in 14 South African sheep breeds South African Merino (n=56), Merino (n =10); Mutton Merino (n = 10), Dohne Merino (n = 50), Meatmaster (n = 48), Blackhead Persian (n =14) and Namaqua Afrikaner (n = 12), the four pelt-colour based Swakara subpopulations of Grey (n = 22); Black (n = 16); White-vital (n = 41) and White-subvital (n =17) Dorper (n = 23); Afrino (n = 51) and unimproved Nguni sheep (n = 30). Linkage disequilibrium (r^2) averaged 0.16 ± 0.021and ranged from 0.09 ± 0.14 and 0.09 ± 0.13 observed in the SA Merino and Dohne Merino respectively to 0.28 ± 0.29 observed in the Blackhead Persian sheep. Chromosome 10 had the highest LD with r^2 values ranging from 0.10 ± 0.15 (SA Merino) and 0.12 ± 0.18 (Dohne Merino) to 0.28 ± 0.30 in Blackhead Persian and 0.29 ± 0.30 (SA Mutton Merino). Across 14 breeds LD decayed from 0.27±0.30 at 0-10Kb window to 0.02±0.03 at 1000-2000Kb window. A progressive decrease in N_e across generations across all populations was observed with effective population size of <500 for all the populations 66 generations ago decreasing to <250 23 generations ago and well below 100 13 generations ago. Highest correlations in gametic phase were observed within the 0-10kb window between pairs of Merino and merino derived breeds. The highest correlation observed with Nguni sheep was with Dorper sheep (0.33) within the 0-10kb window, which was similar to that observed with Blackhead Persian sheep and Dorper (0.32) again within the same window. Overall, the study reported, considerable LD persistent over short distance in the South African sheep breeds. The implications of the observed LD, LD decay and consistency in gamete phase on applications such as GWAS, QTL mapping and GS were discussed.

Keywords: *Multi-breed sheep; LD, gamete phase, OvineSNP50 Beadchip, genomic applications*

6.1 Introduction

The South African sheep population is estimated at 28.8 million and consists of many different breeds raised for different breeding objectives. Of these 28.8 million sheep, nearly tow thirds

are owned by commercial farmers (Directorate Statistics and Economic Analysis, 2013). There are approximately 8 000 commercial and 5 800 communal sheep farms throughout the country (Molotsi *et al.,* 2017). The commercial sheep farmers who own more than two thirds of the sheep supply meat and wool products for local consumption and for export. Meat and wool products from the emerging and smallholder sector are sold through the informal and formal markets and also used for household subsistence. Apart from meat and wool, sheep in South Africa also play other roles including provision of milk and manure and other religious and cultural roles. South African sheep industry contributes about 35 % of the meat produced in South Africa. The Dohne Merino, Merino and Dorper breeds are prominent breeds of sheep in South Africa as well as other sheep breeds of the Dormer, Ile de France, Meatmaster, Namaqua Afrikaner, Afrino, and South African Mutton Merino (Cloete *et al.,* 2014). The smallholder subsistence farmers keep the Nguni, Damara, Pedi, Namaqua Afrikaner and the Dorper that are adapted and have characteristics that make them produce optimally under local conditions (Amelda, 2011; Gwala *et al.,* 2015; Synman *et al.,* 2013).

To produce efficiently and to meet local and export demands, commercial farmers select for high production targeting traits such as improved growth, reproduction, meat and wool production (Schoeman *et al.,* 2010). Substantial genetic gains resulting from selection for a range of economically important traits have been demonstrated in the commercial sector (Cloete *et al.,* 2014) and South Africa is one of the few countries in Africa with formal and well-structured genetic evaluations for sheep and other small stock (van Marle Koster and Visser, 2018). There is minimal or no artificial selection that take place in smallholder sheep production systems, which is considered the cause of low production performance in this sector (Grobler, 2010; Cloete, 2013). It is hypothesised that indigenous breeds, through constant exposure to extreme climatic conditions and disease pathogens, are under constant natural selection for fitness traits such as survival and resistance to tick and diseases (Kunene *et al.,* 2014; Molotsi *et al.,* 2017; Selepe *et al.,* 2018).

Selection in commercial flocks has been achieved using traditional breeding methods which works very well with traits with high heritability, that are easily and routinely measured accurately and in farming sectors that are well organised and maintain animal performance records. The advent of genomics, accelerates genetic progress particularly for difficult to measure traits that are important in smallholder production (Van der Werf, 2009). There is much interest in genetically improving sheep in South Africa across production systems and

multiple breeding goals utilising current technologies such as genomic selection and genomics assisted breeding (Cloete *et al.,* 2014; van Marle Koster and Visser, 2018). The adoption of genomics in the sheep sector can play an essential role in satisfying the meat, wool and other multipurpose demands of the sector. Under genomic selection, candidates can be evaluated more accurately at very early ages, which increases the rate of genetic gain (Hayes *et al.,* 2013). Goddard and Hayes (2011) indicated that the accuracy of genomic selection is regularly higher than that of traditional pedigree-based selection such that there is potential to increase genetic gain in sheep by adopting genomic selection technologies. One major challenge in implementing genomics in breeding programs in South Africa is the small population sizes which makes it challenging to achieve accuracies particularly in highly diverse populations typical of South Africa (Chapter 3). In New Zealand and Australia, who face similar challenges, the sizes of reference populations have been increased through combining closely related populations into one population for genomics evaluations (Brito *et al.,* 2017).

Linkage disequilibrium is the non-random association between alleles of different variants mostly attributed to physical proximity (Qanbari *et al.,* 2010). The extent of genome-wide LD in a population determine the number of single-nucleotide polymorphism (SNP) markers required to have robust genomic evaluations and genome-wide association studies (Hayes and Goddard, 2001; Carlson *et al.,* 2004). The accuracy of genomic selection, is dependent on the genetic relationships within and between the training and validation populations and the level of LD across the genome (Wientjes *et al.,* 2013; Liu *et al.*, 2015), which is influenced by population history, the breeding systems used, *e.g.* natural mating or artificial insemination, and admixture among populations (Slatkin, 2008; Bohmanova *et al.,* 2010). Linkage disquilibrium analysis is therefore fundamental for the application of genetic selection to improve economically important traits in sheep and other livestock species. Linkage disequilibrium also influences the precision of QTL mapping in a genome wide association studies, for the investigation of the diversity among breeds, in tracing selective sweeps and in assessing the distribution of recombination events (Qanbari *et al.,* 2010). Genetic diversity within and amongst breeds, population demography such as changes in the effective population size through generations and tracing of selection and recombination events may be studied using LD information (Stephan *et al.,* 2006; Brito *et al.,* 2015; Khanyile *et al.,* 2015). Another application of LD is analysing the persistence of LD phase, which can be used to trace history of a species and relationships among individuals within that species. This is particularly for

genomic selection across breeds which depends on the relationship between breeds reflected in the persistence of LD phase.

Linkage diesequilibrium has been widely studied in various domestic species (Qanbari, 2020). The two most commonly used measures to evaluate LD, for bi-allelic markers, are r^2, which is the correlation between two loci and $|D'|$ that varies between 0 and 1. In sheep, the OvineSNP50 Beadchip have been used to investigate LD in diverse sheep breeds from different production systems (Kijas *et al.,* 2014; Prieur *et al.,* 2017; Ghoreishifar *et al.,* 2019). The objective of the current study was to assess the extent of genomewide linkage disequilibrium in the South African sheep population. The study investigated (i) LD between adjacent SNPs and (ii) LD decay in 13 South African sheep breeds and populations consisting of commercial and smallholder indigenous populations raised for multiple breeding objectives including wool, meat, pelt and dual purpose functions as well as adaptation to local environments. Using LD the study further estimated trends in effective population size over time. The consistency of gametic phase was also explored to determine the possible use of a multi-breed reference populations for genomic selection. Information generated in this study is considered a prerequisite for designing genetic improvement programs for the multi sheep breeds from complex production systems.

6.2 Materials and Methods

6.2.1 Animal material and SNP genotyping

Four hundred animals belonging to 14 South African breeds/populations consisting of mutton (South African Mutton Merino (n = 10), Dohne Merino (n = 50), Meatmaster (n = 48), Blackhead Persian (n = 14) and Namaqua Afrikaner (n = 12), pelt (Swakara subpopulations of Grey (n = 22); Black (n = 16); White-vital (n = 41) and White-subvital (n = 17) and Karakul (n = 10); wool (SA Merino (n = 56), dual purpose breeds (Dorper (n = 23); Afrino (n = 51) and non-descript Nguni sheep (n =30) were used in the study The details on the breeds used, the SNP genotyping is described in Chapter 3, Section 3.2.1 to 3.2.2.

6.2.2 SNP quality control

The SNPs derived from the OvineSNP50 Beadchip were pruned for missing genotypes (GENO) > 0.95, genotype calling (MIND) > 0.95, minor allele frequency (MAF) > 0.05 and

deviation from Hardy Weinberg Equilibrium (HWE > 0.0001) resulting in 36 975 markers and 400 individuals remaining for further analysis using PLINK (Purcell *et al.,* 2007).

The SNP marker map file was updated using Golden Helix SNP and variation suite (SVS) version 7 (Golden Helix, Inc., 2012), and information on the chromosome number and chromosomal position was extracted from the Ovies aries reference genome, which allowed further filtering of non-autosomal SNPs and those on contigs or not mapped to the latest reference assembly.

6.2.3 Linkage disequilibrium analyses

Plink 1.07 (Purcell *et al.,* 2007) was used to estimate the pairwise r^2 values per chromosome and population (Hill and Robertson, 1968). The r^2 value was then estimated using parameters defined by the plink option *'--r2 --ld-window-kb 2000 --ld-window-r2 0'* to calculate linkage disequilibrium (LD) among SNP pairs up to a distance of 2000 kb. Mean LD and standard deviation were calculated using the PROC MEANS procedure in SAS v 9.4 (SAS, 2013). The generalized linear model procedure implemented in SAS was used to determine the effects of breed, chromosome, distance between SNP markers and the interaction between breed and chromosome on r^2 values using the statistical model:

$$Y_{ijk} = \mu + A_i + B_j + (A \times B)_{ij} + bC_k + E_{ijk},$$

where Y_{ijk} is the observed r^2; μ is the overall mean; A_i is the fixed effect of breed); B_j is the effect of chromosome (j is chromosome 1–26); $(A \times B)_{ij}$ is the interaction effect of breed with chromosome; C_k is distance between SNP markers *(kb)*, which was treated as a covariate with a regression coefficient b; and E_{ijk} is the random residual error effect.

Linkage disequilibrium decay was analyzed by first categorizing SNP marker pairs into intermarker distance bins of 0–1, 1–10, 10–20, 20–40, 40–60, 60–100, 100–200, 200–500, 500–1000 and 1000–2000 kb followed by an analysis of mean LD within each bin.

6.2.4 Effective population size

Effective population size (N_e) was estimated based on the known relationship between r^2, N_e and the recombination rate (c) between two loci following Equation of Sved (1971) implemented in SNep (Barbato *et al.,* 2015). The following formula was used:

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

where N_t is the effective population size t generations ago calculated as $t = (2f(c_t))^{-1}$, c_t is the recombination rate defined for a specific physical distance between markers and optionally adjusted with the mapping functions mentioned above, r^2_{adj} is the LD value adjusted for sample size and α:= {1, 2, 2.2} is a correction for the occurrence of mutations according to Ohta and Kimura, (1971). The study corrected for sample size for each population using the second equation of Sved (1971) The different SNP marker distance bins used for r^2 analysis were used to obtain different estimates of N_e at different time points by calculating the number of generations (t) in the past as ½ c.

6.2.5 Consistency of gametic phase

Consistency of gametic phase, measured as the correlation between signed r-values (gametic phase), was assessed between all pairs of breeds (n =76) by examining the correlation of LD phase between all breed pairs. Of the SNPs used in LD analysis, a subset of 36 484 SNPs that were common across breeds was used for gametic phase analysis. For each marker pair, the gametic phase was estimated as signed r value using the *'--r2 --ld-window-kb 2000 – r'* command in plink (Purcell *et al.,* 2007). Gametic phase values were grouped into bins based on intermarker distances of 0–1, 1–10, 10–20, 20–40, 40–60, 60–100, 100–200, 200–500, 500–1000 and 1000–2000 kb. A correlation analysis of the signed r between breed pairs was estimated for each intermarker distance bin using PROC CORR in SAS 9.4 (SAS, 2013). Consistency of gametic phase between breeds was determined by plotting the correlation coefficient of the signed r against intermarker distance for each breed pair.

6.3 Results

6.3.1 Linkage disequilibrium variation

Linkage disequilibrium (r^2) in the South African population averaged 0.16±0.021. The mean r^2 for the different autosomal chromosomes per population is illustrated in Table 1a and 1b. The lowest LD was observed in the SA Merino and Dohne Merino with r^2 of 0.09 ± 0.14 and 0.09 ± 0.13, respectively. The highest r^2 was observed in the Blackhead Persian sheep with an average r^2 of 0.28 ± 0.29 and in SAMM with average r^2 of 0.27±0.28. Chromosome 10 had the highest LD with r^2 values ranging from 0.10±0.15 (SA Merino) and 0.12±0.18 (Dohne

Merino) to 0.28±0.30 in Blackhead Persian and 0.29±0.30 (SA Mutton Merino) as illustrated in Table 6.1 and 6.2**.**

Table 6.1: Linkage disequilibrium per chromosome per population and overall population for chromosome 1 to 13

Breed	CHR1	CHR2	CHR3	CHR4	CHR5	CHR6	CHR7	CHR8	CHR9	CHR10	CHR11	CHR12	CHR13	SD range
Afrino	0.15	0.15	0.14	0.15	0.16	0.17	0.14	0.14	0.13	0.21	0.15	0.17	0.17	0.17-0.24
BHP	0.28	0.29	0.30	0.26	0.27	0.26	0.25	0.31	0.25	0.28	0.26	0.27	0.31	0.27-0.31
BVS	0.20	0.20	0.22	0.23	0.21	0.23	0.20	0.21	0.19	0.20	0.20	0.19	0.20	0.22-0.25
Dohne Merino	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.12	0.10	0.09	0.09	0.13-0.18
Dorper	0.14	0.14	0.14	0.14	0.12	0.14	0.13	0.14	0.14	0.14	0.12	0.13	0.13	0.15-0.18
GVS	0.15	0.16	0.15	0.17	0.15	0.19	0.14	0.15	0.13	0.16	0.13	0.14	0.14	0.16-0.20
Meatmaster	0.11	0.12	0.12	0.12	0.10	0.12	0.11	0.11	0.13	0.14	0.10	0.11	0.11	0.14-0.16
NQA	0.22	0.24	0.26	0.25	0.25	0.23	0.24	0.21	0.27	0.23	0.23	0.26	0.23	0.25-0.28
Nguni	0.22	0.25	0.24	0.22	0.24	0.22	0.21	0.23	0.26	0.24	0.26	0.21	0.26	0.24-0.28
SAMM	0.27	0.24	0.26	0.26	0.28	0.26	0.28	0.25	0.27	0.29	0.26	0.27	0.27	0.26-0.28
SAMerino	0.09	0.10	0.10	0.10	0.09	0.11	0.09	0.09	0.10	0.10	0.10	0.10	0.09	0.13-0.15
WSVS	0.22	0.22	0.24	0.24	0.21	0.24	0.21	0.21	0.19	0.22	0.21	0.18	0.20	0.21-0.26
WVS	0.14	0.16	0.16	0.17	0.14	0.16	0.15	0.17	0.13	0.15	0.13	0.14	0.14	0.17-0.20
Overall	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.06	0.04	0.04	0.04	0.07-0.12

Table 6.2: Linkage disequilibrium per chromosome per population and overall population for chromosome 14 to 26

Breed	CHR14	CHR15	CHR16	CHR17	CHR18	CHR19	CHR20	CHR21	CHR22	CHR23	CHR24	CHR25	CHR26	Overall	Range SD
Afrino	0.14	0.16	0.15	0.14	0.15	0.14	0.13	0.15	0.15	0.14	0.14	0.14	0.14	0.15	0.16-0.19
BHP	0.27	0.26	0.26	0.26	0.30	0.25	0.22	0.22	0..28	0.24	0.26	0.26	0.29	0.28	0.25-0.29
BVS	0.19	0.20	0.20	0.21	0.22	0.21	0.19	0.21	0.22	0.21	0.16	0.19	0.21	0.21	0.20-0.25
Dohne Merino	0.09	0.09	0.10	0.09	0.10	0.10	0.10	0.08	0.10	0.09	0.08	0.09	0.08	0.09	0.10-0.14
Dorper	0.14	0.13	0.13	0.12	0.12	0.15	0.12	0.12	0.13	0.13	0.11	0.13	0.13	0.13	0.14-0.18
GVS	0.14	0.14	0.14	0.16	0.15	0.15	0.13	0.15	0.14	0.15	0.14	0.14	0.15	0.15	0.16-0.19
Meatmaster	0.11	0.10	0.10	0.10	0.10	0.11	0.10	0.10	0.11	0.10	0.10	0.11	0.10	0.11	0.13-0.16
NQA	0.22	0.25	0.23	0.24	0.21	0.25	0.23	0.26	0.26	0.23	0.23	0.27	0.23	0.24	0.24-0.29
Nguni	0.21	0.22	0.24	0.20	0.25	0.28	0.24	0.24	0.24	0.22	0.23	0.25	0.23	0.24	0.24-.29
SAMM	0.24	0.25	0.27	0.25	0.26	0.26	0.22	0.22	0.27	0.28	0.24	0.27	0.26	0.26	0.24-0.28
SAMerino	0.10	0.10	0.09	0.09	0.09	0.10	0.10	0.08	0.09	0.08	0.09	0.09	0.08	0.09	0.11-0.14
WSVS	0.20	0.24	0.22	0.22	0.21	0.20	0.20	0.20	0.21	0.22	0.17	0.18	0.22	0.22	0.20-0.25
WVS	0.13	0.15	0.13	0.14	0.14	0.14	0.13	0.14	0.15	0.15	0.14	0.11	0.16	0.15	0.14-0.18
Overall	**0.03**	**0.04**	**0.04**	**0.04**	**0.04**	**0.04**	**0.03**	**0.04**	**0.04**	**0.03**	**0.03**	**0.03**	**0.04**	**0.04**	**0.06-0.08**

6.3.2 Linkage disequilibrium decay

Linkage disequilibrium decay over marker distance is illustrated in Figure 6.1. Across 14 breeds LD decayed from 0.27 ± 0.30 at 0-10Kb window to 0.02 ± 0.03 at 1000-2000Kb window. The LD of the Nguni was highest at 0.52±0.39 at 0-10Kb window and decayed to 0.22 ± 0.24 at 1000-2000 Kb window. Overall, indigenous breeds such as Blackhead Persian, Namaqua Afrikaner and Nguni had high LD (>0.40) at short SNP intervals of 10-20Kb that persisted as moderate LD (>0.20) up to 2000 Kb. Similarly LD in SA Mutton Merino ranged from 0.20 at 2000 Kb to 0.45 ± 0.38 at 10Kb. Swakara pelt colour sub populations had moderated LD that ranged from 0.38 to 0.42 at 1-10Kb amd from 0.12 to 0.19 at 1000-2000 Kb. LD decayed from 0.33 ± 0.33 (1-10 Kb) to 0.06 ± 0.08 (1000-2000 Kb) in Dohne Merino sheep and from 0.33 ± 0.33 (1-10 Kb) to 0.07 ± 0.09 (1000-2000 Kb) in SA Merino and 0.33 ± 0.33 to 0.09 ± 0.12 in Meatmaster sheep.

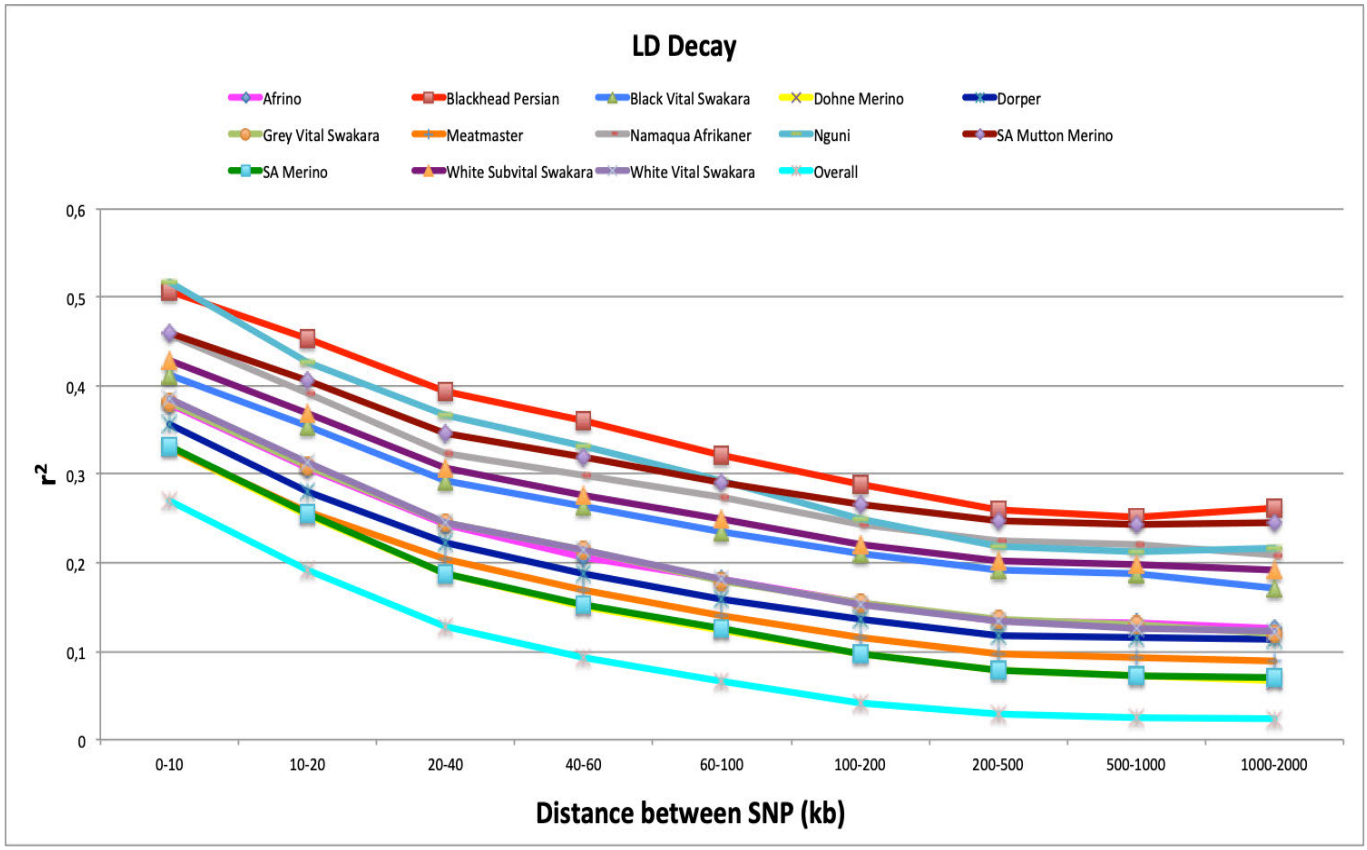


Figure 6.1: LD decay over marker distance in 14 SA sheep breeds

6.3.3 Estimates of N_e

Estimated N_e (t) at t generations in the past are shown in the plots of Figure 6.2 and in Table 6.3. Based on these results, a progressive decrease in N_e across generations across all populations was observed. The N_e indicated an effective population size of <500 for all the populations 66 generations ago and decreasing to <250 23 generations ago and well below 200 15 generations ago. Breeds such as the Dohne Merino, Dorper, SA Merino had relative higher N_e across generations relative to the indigenous and smallholder breeds of the Nguni, Blackhead Persian, Namaqua Afrikaner and Karakul.

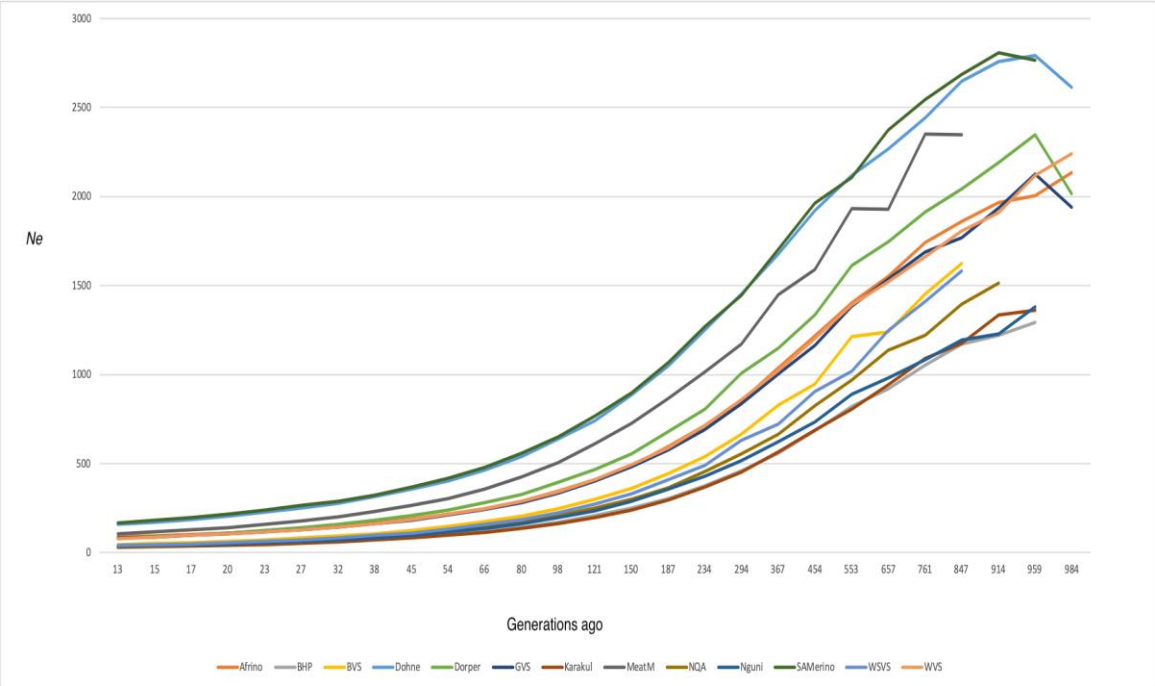


Figure 6.2: Effective population size over the past 984 generation

6.3.4 Correlation of gametic phase

Consistency of gametic phase was determined between pairs of population across all 14 populations as presented in Figures 6.3 a-m. Highest correlations in gametic phase were observed within the 0-10 Kb window between pairs of Merino and merino derived breeds. For example, correlations of 0.54 and 0.45 were observed between Dohne Merino and SA Mutton Merino and Afrino respectively, whilst correlation of 0.40 was observed between Afrino and SA Mutton Merino. Other Merino derived breed of Meatmaster had high correlations with Dorper (0.44), SA merino (0.40) (Fig. 6.3 a-e).

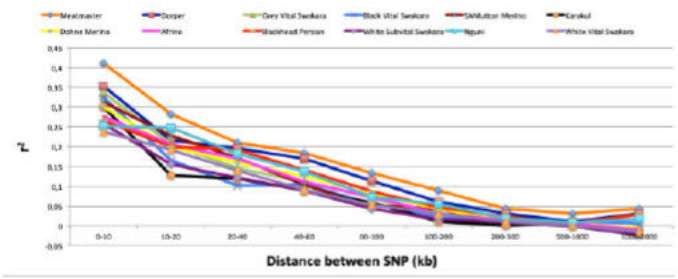
Table 6.3: Trends in effective population size of populations over 984 generations

GenAgo	Afrino	BHP	BVS	Dohne	Dorper	GVS	Karakul	MeatM	NQA	Nguni	SAMerino	WSVS
13	90	33	47	158	81	82	31	107	34	37	167	41
15	95	37	52	172	90	89	34	118	39	41	181	46
17	103	41	57	188	100	97	38	131	44	45	198	51
20	111	46	65	206	111	106	42	142	49	50	215	57
23	122	52	73	227	125	117	47	158	57	56	238	66
27	133	59	82	252	140	130	53	180	65	65	265	74
32	147	67	95	279	159	144	61	203	76	73	290	84
38	166	77	108	314	181	163	71	232	86	85	325	97
45	189	89	126	356	209	182	82	267	101	99	369	112
54	215	104	147	403	240	211	97	304	121	116	419	132
66	248	123	173	465	280	243	114	356	143	135	481	159
80	290	144	205	541	328	283	135	426	172	162	558	187
98	346	171	248	640	396	335	162	506	208	196	652	226
121	411	208	302	741	468	402	198	611	252	236	767	274
150	487	251	362	888	555	482	239	726	299	290	899	331
187	597	304	446	1050	681	578	296	869	365	357	1069	411
234	717	377	541	1253	808	694	369	1017	455	429	1269	492
294	853	460	666	1452	1007	836	451	1172	556	519	1447	633
367	1039	558	829	1679	1150	1004	566	1448	666	624	1704	722
454	1218	683	947	1920	1337	1165	687	1591	827	734	1963	907
553	1403	822	1212	2121	1613	1386	806	1933	972	889	2108	1020
657	1552	919	1239	2266	1745	1536	944	1928	1136	982	2375	1248
761	1741	1052	1454	2444	1914	1690	1091	2351	1221	1085	2547	1413
847	1861	1173	1623	2648	2045	1769	1176	2348	1396	1196	2686	1582
914	1968	1223	.	2757	2191	1935	1334		1513	1230	2809	
959	2004	1295	.	2793	2348	2125	1362			1380	2765	
984	2133	.	.	2615	2018	1940						

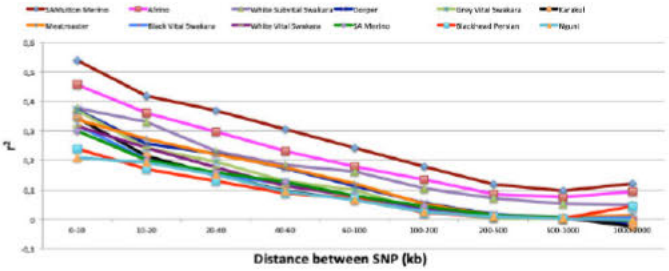
Legend: GenAgo= Generations ago; BHP = Blackhead Persian; BVS = Black Vital Swakara; GVS = Grey Vital Swakara; MeatM = Meatmaster; NQA = Namaqua Afrikaner; WSVS = White Subvital Swakara; WVS = White Vital Swakara.

The correlations in gametic phase Merino derived breeds of Dohne Merino and Afrino fell below 0.20 at the 60-100kb window High correlation of 0.52 was observed between Swakara pelt colour subpopulations of Black Vital Swakara vs White Vital Swakara and Grey Vital Swakara. The correlations in gametic phase persisted above 0.20 between White Vital Swakara and Black and Grey Vital Swakara up to window of 1000-2000 Kb (Fig. 6.5 i-m). The highest correlation observed with Nguni sheep was with Dorper sheep (0.33) within the 0-10kb window. Similarly the highest correlation observed with Blackhead Persian sheep was with Dorper (0.32) again within the 0-10kb window (Fig. 6.4 f-h). Even within the 0-10kb window, correlations with the Nguni and BHP and other breeds went as low as 0.18 (Nguni vs Black Vital Swakara sheep) and 0.15 (BHP vs Karakul).

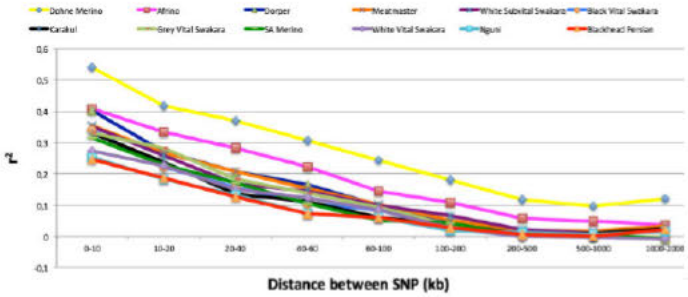
a: SA Merino



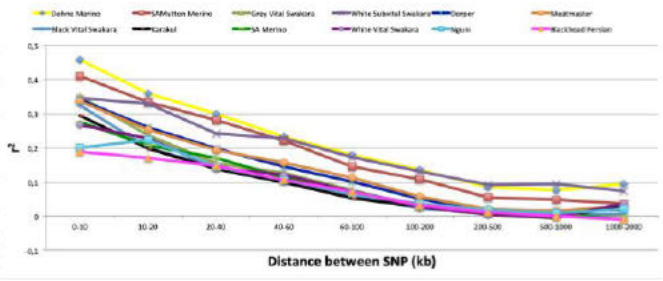
c: Dohne Merino



b: SA Mutton Merino



d: Afrino



e: Meatmaster

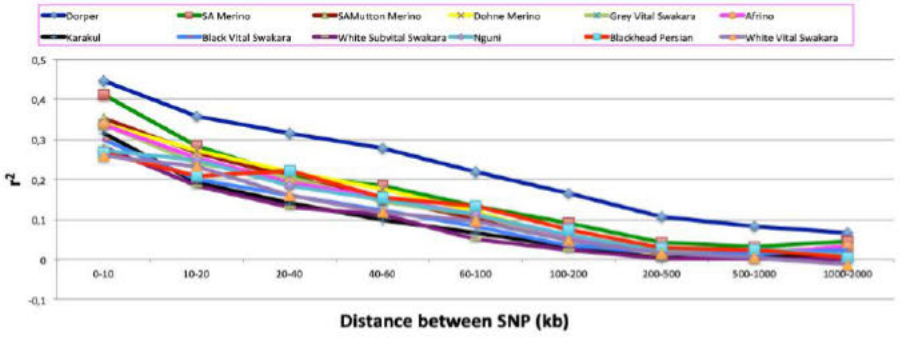
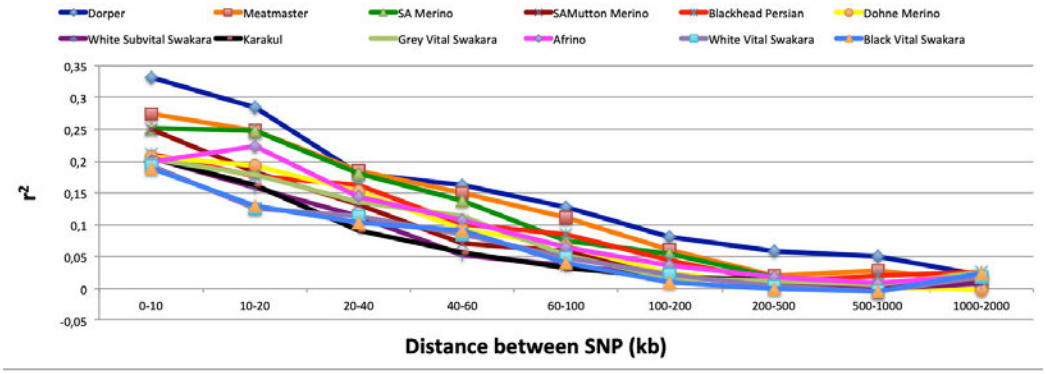
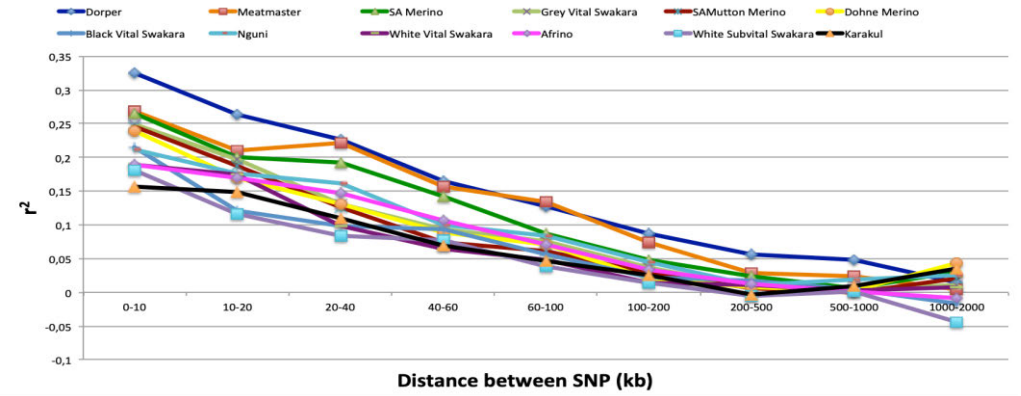


Figure 6.3 a-e: Pearson's correlations of signed r-values between all breed pairs at given distances of Merino and Merino derived breeds

f: Nguni



g: Blackhead Persian



h: Dorper

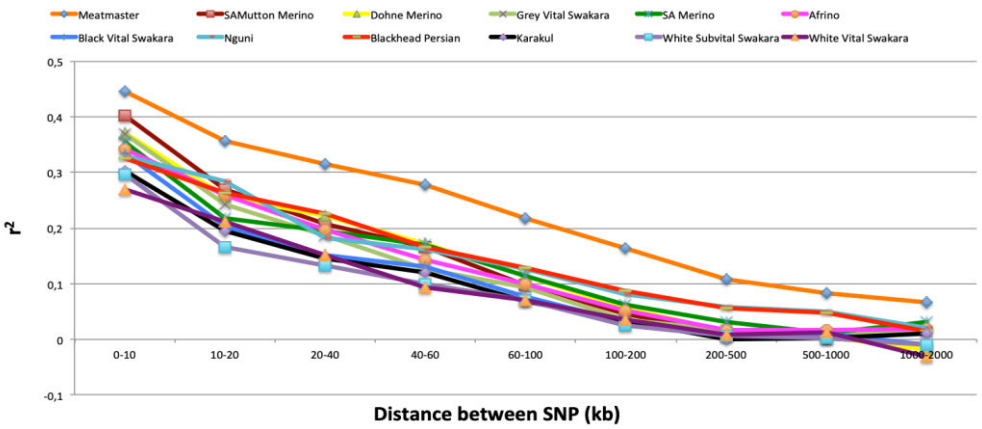
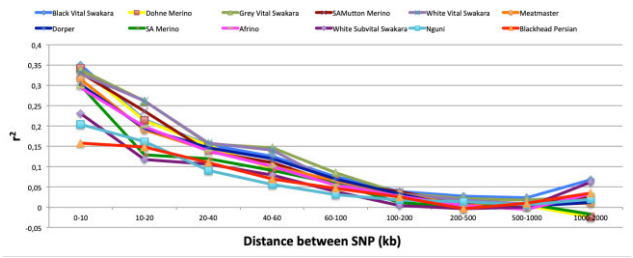
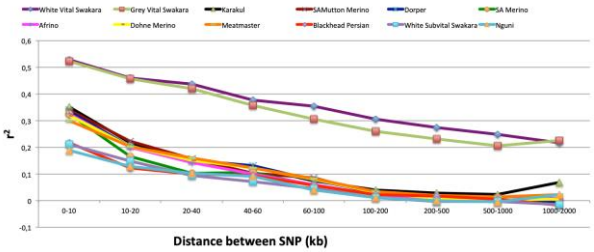


Figure 6.4 f-h: Pearson's correlations of signed r-values between all breed pairs at given distances of Nguni, Dorper and Blackhead Persian breeds

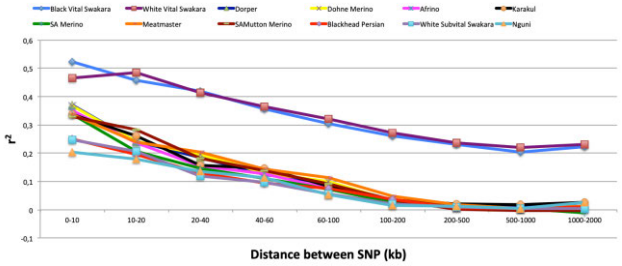
i: Karakul



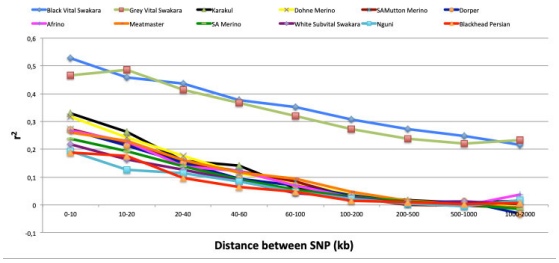
j: Black Vital Swakara



k: Grey Vital Swakara



l: White vital Swakara



m: White Subvital Swakara

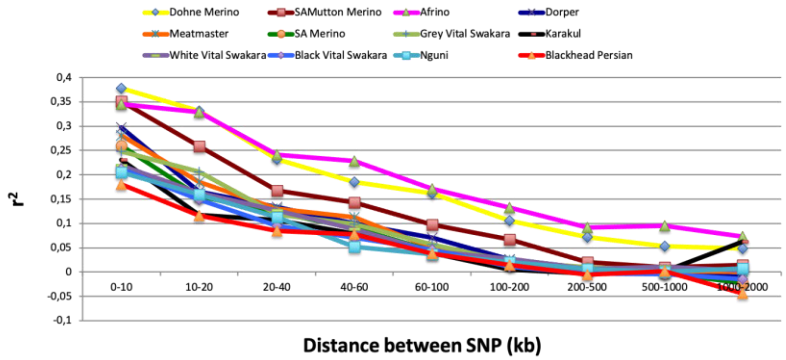


Figure 6.5 i-m: Pearson's correlations of signed r-values between all breed pairs at given distances of Karakul and Swakara subpopulations

6.4 Discussion

Linkage disequilibrium is an important parameter crosscutting in many applications of genomics. The extent and pattern of genome-wide LD influences the accuracy and precision of Genomewide association studies and QTL mapping of production traits in association studies. In essence, the strength of LD between markers and QTL provides the statistical power to detect associations in GWAS. The extent of LD of SNPs determine the optimal SNP density of a reliable SNP chip on a given population. The success of multi-breed reference population based genomic selection programs depend on the correlation in gametic phase between the targeted populations. LD was studied in the South African sheep population in order to make inferences on (i) the utility of the OvineSNP50 Beadchip for the different genomic applications, (ii) make inferences on population genetic diversity and effective population sizes and (iii) the possibility of multi-breed reference population and genomic selection in the SA sheep breeds.

Overall, high LD was observed in small breed such as the Blackhead Persian, Black Vital Swakara, White Subvital Swakara, Namaqua Afrikaner, the Nguni and the South African Mutton Merino. The high LD in the small SA sheep breeds is similar to what has been reported in similarly raised Southern African chickens (Khanyile *et al.,* 2015; Cattle (Makina *et al.,* 2015) and goats (Mdladla *et al.,* 2016) and is attributed to the low effective population sizes and high inbreeding levels in such populations as observed in Chapters 3 and 5. Ghoreishifar *et al.* (2018) and Gusev *et al.* (2011) reported that high LD is common within sub structured populations and in closely related individuals that share longer haplotypes. As reported in Chapter 3, The Nguni, Blackhead Persian, Namaqua Afrikaner and the Swakara subpopulations are small sized populations raised as fragmented flocks and characterised by low genetic diversity, high inbreeding levels and divergence from other populations, parameters that would explain the high average LD in these populations. Al-Mamun *et al.* (2015) observed high and persistent LD in Border Leicester and Poll Dorset, small sheep breeds that were associated with low genetic diversity. Breeds such as the Dohne Merino, Afrino and Meatmaster are composite breeds developed from the Merino and indigenous breeds such as the Damara, Ronderib Afrikaner and Ngunis. Crossbreeding and combining of multiple genetic pools when composite breeds are formed contributes to the low LD observed in these breeds (Prieur *et al.,* 2017). The Dohne Merino for example had the lowest LD of 0.09 ± 0.13. Fu *et al.* (2015) observed low LD in crossbred chickens compared to purelines. Overall, lowest LD below 0.10 was observed in Dohne Merino and Merino sheep. These are generally commercially raised populations well

managed with effective population sizes and inbreeding levels monitored to sustain such breeds (Chapter 3). Similar results of low LD were observed with other Merino and Merino derived breeds (Kijas *et al.,* 2014; Al-mamun *et al.,* 2015). Across breeds, LD estimates were associated with high standard errors which could be either due to the diversity within populations as reported in Chapters 3, 4 and 5 or due to relatively smaller sample sizes compared to ther studies. Compared to other livestock species, the South African sheep populations had low levels of LD. Overall, the average LD observed in the South African sheep populations was lower than in other livestock species, which is consistent with observations made by Kijas *et al.,* (2014) and attributed to sheep domestication that involved multiple and genetically diverse wild ancestors. In addition sheep experienced less severe bottleneck events during domestications when compared to other domesticated animals (Kijas *et al.,* 2014). In addition to the population demographic factors in sheep in South Africa and globally, contributing to the relative lower short distance LD, the design of the OvineSNP50 Beadchip and the minimum inclusion of content from African breeds could have also contributed to low LD through ascertainment bias. Ascertainment bias increases the frequency of the most commonly polymorphic loci and eliminates markers for loci that are less polymorphic in the screening panel,which will as a result affects the estimates of population genetic parameters, allele frequency distribution and linkage disequilibrium (Albrechtsen *et al.,* 2010). As observed in Chapter 3, majority of the SA sheep particularly the small indigenous breeds lost a number of SNPs during quality control because of low MAF. Such exclusion of a significant number of SNPs (ranging from 7500 to 15 000) could result in the final panel of SNPs used being sparsely distributed across the genome thereby contributing to low and short distance LD.

The rate of LD decay was greater in Merino and Merino derived breeds of Dohne Merino, Afrino and Meatmaster relative to the indigenous breeds of Nguni, Blackhead Persian, Swakaras and Namaqua Afrikaner. Al-Mamun *et al.* (2015) observed that the higher rate of LD decay in Merino derived breeds confirms its high level of genetic diversity. In Iranian Zandi sheep, Ghoreishifar *et al.* (2019) observed an average LD of 0.21 for SNP 0-10 Kb apart which decayed rapidly with increasing marker distance. In the Zandi breed, LD remained at a moderate level (r^2 = 0.26 ± 0.30) only up to 10 Kb and decreased rapidly to 0.10±0.16 when the average SNP pair distance was 40–60 Kb. Garcia-Gamez *et al.* (2012), Kijas *et al.* (2012) and Al-mamun *et al.* (2015) observed faster LD decay at much shorter genomic distances. Short-range LD was observed in all populations but the rate of LD decay differed between

populations. For breeds such as Blackhead Persian. Nguni, Namaqua Afrikaner and Swakara subpopulations LD decayed to >0.20 at 1000-2000Kb relative to an LD <0.11 for Merino and Merino derined breeds within the same window. This was a similar trend to that observed by Al-Mamun *et al.* (2015) in an analysis of Merino based and other Australian sheep breeds.

The amount of LD and extend of its decay has direct implications on use of SNPs chips for applications such as QTL mapping, GWAS and genomic selection (Qanbari *et al.,* 2010). Lower levels of LD which rapidly decays with increased marker distances implies that a denser SNP panel is required for genomic applications. As reporter earlier (Kijas *et al.,* 2014; Al-Mamun *et al.,* 2015; Alvarenga *et al.,* 2018), LD in sheep persists for relatively shorter genomic distances when compared to other species such as cattle, chickens and goats. Similarly in South Africa, LD persisted over relatively longer distances in chickens (Kanyile *et al.,* 2015), cattle (Makina *et al.,* 2015) and goats (Mdladla *et al.,* 2016) relative to that observed in sheep. Kijas *et al*. (2014) indicated that sheep have been less intensively selected when compared to other domestic species and coupled to being domesticated for a larger initial gene pool, would explain the low range LD. The low LD levels that persist over very short SNP distances required a higher density SNP panel for genomic applications i.e. diversity studies, GWAS, QTL mapping and genomic selection. Kijas *et al.* (2014) recommended the use of the commercially available Ovine 600K SNP panel particularly for highly diverse breeds. Very limited studies have however used this panel to provide a basis for comparison.

Coupled to breed differences, variations were observed in LD between chromosomes across breeds. Overall, a high LD was observed on chromosome 10 with an average of 0.06 ± 0.12 relative to a chromosomal average of 0.04 ± 0.08. In breeds such as the Blackhead Persian, SAMM and Nguni, chromosome 10 displayed high LD of 0.28 ± 0.30; 0.29 ± 0.30 and 0.24 ± 0.27, respectively. Liu *et al.* (2017) also observed at high LD of 0.303 ± 0.346 at 0-10 Kb window on Chromosome 10 in Chinese Merino (Xinjiang type) sheep which was consistent with haplotype block structure and the ROH distribution. As reported in Chapter 4, Chromosome 10 had the highest frequencies of ROH across the SA sheep breeds and is associated with genomic regions under selection for horn size and shape amongst a number of sheep breeds (Johnston *et al.,* 2011; Kijas *et al.,* 2012; Wiedemar *et al.,* 2015). Similarly, maximum average LD between adjacent SNPs was observed on chromosome 10 in Border Leicester, Poll Dorset and Merino Breeds also associated with long haplotype blocks (Al-

Mamun *et al.*, 2015). Alvarenga *et al.* (2018) observed high LD on chromosomes 2, 10 and 16 in Brazilian Santa Inês sheep breed.

The N_e analysis indicated an effective population size of <500 for all the populations 66 generations ago and decreasing to <250 23 generations ago and well below 200 15 generations ago. Breeds such as the Dohne Merino, Dorper, SA Merino had relative higher Ne across generations relative to the indigenous and smallholder breeds of the Nguni, Blackhead Persian, Namaqua Afrikaner and Karakul. The Ne values were overall lower than was reported by Kijas *et al.* (2012) in indigenous breeds such as the Brazilian Santa Inês breed which had an effective population size averaging 520. Al-Mamun *et al.* (2015) reported values of Ne ranging from 140 in indigenous breeds such as Border Leicester to 348 in Merino breeds. García-Gámez *et al*. (2012) reported Ne values ranging from to 83 to 159 in Spanish Churra sheep. The Ne values reported in the South African sheep breeds reflected the dynamics of low population sizes, closed production systems, population fragmentation and artificial and natural selection. The FAO (http://www.fao.org/3/AD013E/AD013E04.htm) and Franklin (1980) suggested that a minimum effective size of 500 is considered viable and is needed to preserve useful genetic variation. In the current study, populations/ breeds such Namaqua Afrikaner, Nguni, Blackhead Persian and Swakara subpopulations are associated with very low effective population sizes and seem to be at risk of extinction. Previous reports on these breeds have corroborated the findings of this study (Synman, 2014a-e). For example, there are only 28 members farming Blackhead Persian in South Africa with only ±20 000 sheep amongst them (Synman, 2014a). The Namaqua Afrikaner on the other hand is considered an endangered breed with an estimate of less than 2000 animals in South Africa (Synman, 2014f), whilst Swakara breeders have been declining in the country to a number of less than 54 breeders in South Africa and Namibia farming less than 40 000 animals (Synman, 2014g).

The persistence of LD phase demonstrate relatedness of populations is essential for effective genomic selection (Goddard *et al*., 2006). Populations with high correlations in gamete phase can easily be combined into single reference populations that can yield accurate multi-breed genomic selection (Larmer *et al*., 2014). In the current study high correlations were observed amongst (i) Merino and Merino derived breeds *i.e.* (ii) Karakul and Swakara based subpopulations. Weak correlations were observed amongst the indigneous sheep of Nguni, Blackhead Persian, Namaqua Afrikaner and Dorper as well as between any of these breeds and other South African breeds. The observed correlations are supported by the PCA based

populations clusters (Chapter 3) and aligned with breed history and production systems. Similar results were reported by Brito *et al*. (2017) who observed that the more distant the relationship between individuals, the shorter the genomic distance over which the phase will be consistent. When compared to other studies *e.g.* Brito *et al.* (2017), the correlations observed in this study were generally lower even for breeds that are considered to be highly related *e.g.* the Merinos and Swakaras. This implies and supports the considerable breed divergence reported in Chapter 3 and Chapter 5 and overall implies that it will be challenging to merge breeds into a multi-breed reference populations. This will further disadvantage the initiatives to implement genomic selection in such a highly diverse sheep sector that is associated with multiple breeds, breeding goals and production systems. Alvarenga *et al*. (2018) reported that the low correlations are also aluded to the low utility of the SNP50K over SNP chips of higher density such as the SNP 600K.

6.5 Conclusions

The study reported, considerable LD persistent over short distance in the South African sheep breeds, LD was relatively high in small and fragmented breeds such as the Blackhead Persian, Nguni, Namaqua Afrikaner and Swakara subpopulations when compared to Merino and Merino derived breeds with the exception of SA Merino. The South African sheep breeds are of considerable effective population sizes that has however been decreasing with time. The utility of the OvineSNP50 Beadchip panel whilst demonstrated for a number of applications in the SA breeds, might be challenged by the low short range LD, which will affect applications such as GWAS and QTL mapping and by the low correlations in gamete phase between breeds which will affect applications of SNP chip in genomic selection and other applications.

7 GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

7.1 General discussion

The broad objective of the study was to use the OvineSNP50 BeadChip array to investigate the genomic architecture of the South African sheep population. The South African sheep population consist of multiple breeds raised under different production systems and for diverse set of goals ranging from mutton, wool, pelt and dual-purpose functions. The current gene pool is a mixture of indigenous and exotic animals some of which have been raised in South Africa for so long that it is hypothesised they have evolved and adapted to the local conditions. Examples of such breeds that were introduced into South Africa and now considered indigenous include the Blackhead Persian, Damara, Namaqua Afrikaner and Merino breeds. In addition to these are composite breeds that have been developed through crossbreeding of merino breeds to other indigenous populations with examples such Afrino, Meatmaster and Dormer. The South African sheep sector shares its gene pool with the outside world. Breeds such as the Dohne Merino locally developed in South Africa are farmed in countries such as Russia, South America and Australia (https://dohnemerino.com/the-dohne-merino-breed-holds-its-first-world-conference-in-july-2016/). The Karakul sheep have been imported from Uzbekistan, Central Asia (https://www.farmersweekly.co.za/animals/sheep-goats/the-rise-of-swakara-farming/), into southern African countries including South Africa, where they have been selected, crossbred with local breeds and rebranded into the then Swakara subpopulations.

Whilst being one of the few countries in Africa that have well established sheep breeds, South Africa also hosts a number of non-descript ecotype populations raised by smallholder farmers. Nguni sheep that are distributed as four ecotypes of Landim, Pedi, Swazi and Zulu (Kunene *et al*., 2015); Selepe *et al*., 2018) are such examples. Local ecotypes are predominantly uncharacterised, and farmed under low input production systems where they are exposed to diverse natural selection pressures ranging from climatic, nutritional and exposure to diseases and disease pathogens. The non-descript ecotype gene pool is hypothesised to be harbouring unique genes that enables them to survive and produce under extreme production conditions. Overall, the South African sheep genetic diversity is complex and shaped by multiple evolutionary forces that have resulted into a unique and diverse gene pool and phenotypes providing the nation with food security, improved livelihoods and national economy through their contribution to GDP.

The OvineSNP50 Beadchip is a very comprehensive genomic tool that demonstrated robustness to a number of applications in studying sheep. The tool has been used in a number of sheep populations including commercially raised mutton and wool breeds and indigenous and non-descript communal populations (Selepe *et al*., 2018; Molotsi *et al*., 2017; Ciani *et al*., 2020) and for a number of applications such as analysis of population genetic diversity (Kijas *et al*., 2018), investigation of population divergence (Edea *et al*., 2017; Ciani *et al*., 2020) and signatures of selection (Fariello *et al*., 2014; Paim *et al*., 2018), genomewide association studies, QTL mapping and candidate gene profiling (Xu *et al*., 2018; Cao *et al*., 2020) and genomic predictions (Auvray *et al*., 2014; Borlomaa *et al*.,2017).

This study used the OvineSNP50 Beadchip panel for analysis of population diversity and genetic structures (Chapter 3), investigation of population divergence and selection signatures in composite breeds (Chapter 4), inference into population history and production system driven selection using ROH analysis (Chapter 5) and lastly linkage disequilibrium and consistency of gamete phase to infer on utility of the OvineSNP50 Beadchip in the SA population (Chapter 6). These analysis were conducted to test the hypothesis that (i) the South African sheep sector is genetically diverse and sub structured according to production systems (ii) breed specific production goals and selection pressures, resulted in divergence of breeds from their founding populations (iii) production system and within-breed selection pressure had pronounced effects on the genome architecture of South African sheep, resulting in reduced genetic diversity and frequency of runs of homozygosity, and (iv) the OvineSNP50 Beadchip genotypes of the South African sheep population demonstrates high and persistent linkage disequilibrium and correlated consistence in gamete phase between breeds implying utility of this genomic tool for applications such as multi-breed genomic predictions.

The South African sheep population does not exist in isolation and shares its gene pool with other countries in Africa and globally. It therefore became a prerequisite to compare the diversity and architecture of the SA sheep gene pool in the context of global populations. Chapter 3 and 5 therefore included genotypes from global sheep populations and made available by the International Sheep Genomics Consortium (http://www.sheephapmap.org). The 623 genotypes from this global resource were purposely selected from breeds and populations assumed to have some connection with the SA sheep breeds either through founder population effect as in the case of the Merino type breeds and Karakas, or either being produced under similar production systems and therefore experiencing similar selection pressures as well

as other forces of evolution as in the case of the Ethiopian Menz and Bangladesh Galore. Overall, the 623 genotypes represented diversity from Africa, Asia and Europe.

Chapter 3 was designed to address the first hypothesis of the existence of a highly diverse but sub-structured South African sheep population and in so doing investigated genetic diversity, population genetic structure and divergence between South African sheep breeds representing mutton, pelt and mutton and wool dual-purpose breeds as well as non-descript village sheep. Overall the results demonstrated low genetic diversity in the small indigenous sheep populations that are either raised within communal farming systems (Nguni sheep) or a small and localised breed (Blackhead Persian, Namaqua Afrikaner; Swakara subpopulations) relative to commercial breeds of Merino, Dohne Merino, Afrino, Meatmaster with the exception of SA Mutton Merino. The low diversity observed in the SAMM reflects the intense selection of this breed from its founder the German Merino (Neser, 2000). Higher diversity in the composite breeds of Dohne Merino, Afrino and Meatmaster on the other had could be a founder effect were the breeds were developed from a combination of two or more genetic populations. Comparing them to the global populations saw the alignment in genetic diversity of the indigenous African breeds of Ethiopian Menzi and the Nguni which are raised under similar production systems. Other smaller and localised breeds that demonstrated similar genetic diversity with the Blackhead Persian and Namaqua Afrikaner are the Bangladesh Galore and the McArthur Merino breeds. Overall the population diversity analysis showed that the within population diversity is a function of population founder effects and production/management systems which is in agreement with reports by a number of studies including Kijas *et al.* (2012) and Edea *et al.* (2017). The population clustering reported in Chapter 3 demonstrated several insights. Firstly the distinction between fat tailed and Merino and merino derived breeds as reported by a number of studies (Mwacharo *et al.*, 2017; Edea *et al.*, 2017; Ahbara *et al.*, 2019) and secondly to the clustering of population according to population history and production systems as confirmed by Soma *et al* (2012) and Selepe *et al.*, (2018).

With such a strong sheep industry, serving both the smallholder food security needs and the commercial sector's contribution to GDP, South Africa is one of the few countries with well-established breeds some that are farmed globally. The worldwide farmed Merino breeds are prevalent in the country and South Africa went on like other countries to develop other breeds from a combination of Merino and indigenous breeds. These composite merino derived breeds such as the Afrino, Dohne Merino and Meatmaster are characterised by high productivity and

robustness to the challenging production systems of diseases, parasites and extreme climatic conditions. Chapter 4 evaluated genetic diversity, population structure and breed divergence of South African Merino-derived sheep breeds of Dohne Merino, Meatmaster and Afrino relative to the presumed ancestors of Merino, SA Merino, SA Mutton Merino, Damara, Ronderib Afrikaner and Nguni. In addition to Afrino, Meatmaster and Dohne Merino is Dormer a composite mutton breed developed from Dorset and Merino that could adapt to conditions of cold and wet winters of South Africa. The analysis in Chapter 4, however, exempted the Dormer because of unavailability of SNP genotypes for the Dormer. However, because of the similarities in the nature in which these composite breeds were developed extrapolations can be made on the Dormer and other composite breeds based on results observed with the Afrino, Meatmaster and Dohne Merino. Results from this study can actually inform future initiatives on crossbreeding and breed development, which are inevitable in the face of continuous changes in consumer preferences (Olynik *et al*., 2012) and climate change (Rojas-Downing *et al*., 2017).

Clustering based on principal component analyses showed the Meatmaster to be more aligned with its non-Merino ancestors of the Damara, Nguni *etc*, which was attributed to the fact that according to breed standards Meatmaster should contain 50 % Damara (Synman, 2014). It is therefore the breeding goals and breed standards that determine how much of each of the presumed ancestor the breed will be aligned to. In developing these composite breeds, breeders have to taken a lot of factors into consideration. Whilst productivity is paramount, other traits such as adaptation to local concern, disease resistance/tolerance and animal welfare concerns are key, which explain why often composite breeds are derived from a mixture of exotic breeds (high merit in productive traits) and indigenous breeds that carry the adaptation and animal welfare traits.

Whilst they maintained some of their genomic architecture from the presumed ancestors *i.e.*, the three Merino derived breeds also showed some significant level of divergence from the ancestral populations yielding selection sweeps in genomic regions associated with both production traits *i.e.* hair/wool traits, development of the skeleton and mammary glands and metabolic genes as well as immune response genes associated with adaptive traits. Overall, findings in Chapter 4 demonstrated that whilst derived breeds carry the genes and maintain some of the traits from their ancestral breeds, they also diverge in a direction driven by breed-specific selection goals and production factors.

Production system and within-breed selection pressure have pronounced effects on the genome architecture and may cause reduced genetic diversity and frequency of runs of homozygosity islands (Szmatoła *et al*., 2014). Based on the results in Chapter 3, it was hypothesised that the different production systems under which South African sheep are raised i.e. mutton, wool, pelt, dual-functions or absence of selection programs in non-descript breeds kept by smallholder farmers influences the occurrence and distribution of ROH in this gene pool. Screening for ROH and analysing their distribution in the different breeds was therefore going to shed more light and confirm the breed history and selection/evolutionary forces the South African sheep are under. Chapter 5, showed the prevalence of short (1-6 Mb) ROH in the SA sheep breeds with only few breeds of the Nguni and the Blackhead Persian that had a relatively higher frequency of long ROH >48 Mb. As comprehensively discussed in Chapter 5, the frequency and distribution of ROH implied some populations *i.e.* (Blackhead Persian) experienced ancient bottlenecks probably due to founder effects whilst other breeds are characterised by more recent events of reduced population sizes and inbreeding. Some populations *e.g.* the Nguni were characterised by both long and short ROHS implying both ancient and recent events contributing to stretches of homozygous regions are at play in these populations. Again as in Chapter 3 on diversity parameters, the McArthur Merino displayed a ROH pattern similar to that of Blackhead Persian and Nguni. The annotation of genomic regions characterised by ROH, showed ancient selection pressures i.e. that for horn presence or absence at chromosome 10 as well as more recent selection pressures predominantly for metabolic and immune response traits. Overall, the frequency and patterns of distribution of ROH observed in this study corresponds to the breed history and implied selection pressures exposed to the sheep populations under study.

The last chapter of the experiments (Chapter 6), conducted an analysis that was looking for information that would inform the utilisation of the OvineSNP50 Beadchip panel in the South African sheep population, focusing on linkage disequilibrium and correlation in consistence of gamete phase amongst breeds. There is growing interest to use genomics to genetically improve sheep in South Africa across production systems and multiple breeding goals. The accuracy of genome-wide association studies, genomic selection and other applications is dependent on the level of LD across the genome and for multi-breed genomic applications the correlations in consistence in gamete phase. Based on the previous chapter, it was evident that the South African sheep populations is substructured along breed history and category (*i.e.* fat tailed vs non-fat tailed sheep) and production system (*i.e.* commercial breeds, indigenous breeds and

non-descript populations). Similarly LD patterns differed along the clusters with higher LD observed in the Blackhead Persian, Namaqua Afrikaner and unimproved Nguni sheep and low in SA Merino and the Merino derived breeds of Dohne Merino, Afrino, and Meatmaster, which according to Chapter 3 were highly diverse and accumulated less ROH. In agreement with ROH distribution, Chromosome 10 had the highest LD with r^2 values ranging from 0.10±0.15 (SA Merino) to 0.28±0.30 in Blackhead Persian and 0.29±0.30 (SA Mutton Merino). What this implies is that genomic regions on chromosome 10 experienced intensive selection pressures leading to the accumulation of ROHs, LD block or extended haplotypes that give rise to elevated i|HS| or Rsb scores. The analysis of trends in N_e across generations confirmed the low populations in Blackhead Persian, Nguni, Namaqua Afrikaner, Swakara subpopulations that were implied by results in Chapter 3 and Chapter 5 and supported by literarure (Synman, 2014 a-h). High correlations in gametic phase were only observed within the 0-10kb window between pairs of Merino and merino derived breeds which even implied that applications such as multi-breed genomic predictions which might be a solution for genomic selection in small population sized South African sheep breeds will be challenging. As for the indigenous breeds, the highest correlation observed with Nguni sheep was with Dorper sheep (0.33) within the 0-10kb window, which was similar to that observed with Blackhead Persian sheep and Dorper (0.32) again within the same window. A correlation of 0.33 is low and does not meet the required threshold for combing breeds in multi-breed genomic evaluation programs. Overall, Chapter 6 reported, considerable LD persistent only over short distance in the South African sheep breeds with low-moderate correlations in consistence in gamete phase amongst breeds, highlighting the limited utility for such a panel in this populations. Although Kijas *et al.*, 2012 suggested the use of denser SNP panels such as the Ovine 600K, these are expensive and might not be affordable for resource limited sectors. Other alternatives will be to customise the existing 50K with markers fit for the SA populations and minimise the effects of ascertainment bias.

Overall, the whole study managed to investigate the genomic architecture of the South African sheep populations, highlighting the considerable diversity emanating from multiple sub-structured breeds. In addition, the study determined the genetic relatedness of the SA sheep gene pool to the global sheep population and the effects of crossbreeding and composite breed development on the genomic architecture of derived breeds. SNP genotyping is an expensive undertaking particularly in developing countries were resources are limiting. Although the study did not capture all the breeds in South Africa, it sampled representative breeds and production systems from which inferences can be made with caution to other breeds excluded

in the study. Indigenous populations such as the Pedi and Zulu sheep could have similar dynamics as those of the studied Nguni sheep. Other sheep excluded include commercial breeds such as Dormer which is a composite breed developed from crossing the Dorset Horn and the SA Mutton Merino, Van Rooy Sheep which are crosses of Ronderib Afrikaner sheep and Rambouillet and exotic breeds of Suffolk and the Ile de France. The ISGC genotypes also genotyped on the OvineSNP50 Beadchip were a valuable resource and highlights the importance of sharing genomic data across projects at least for academic purposes. Of cause the technology of arrayt based genomewide SNP genotyping and the development of the OvineSNP50 Beadchip as a genomic tool made this possible as results can easily be shared between research groups and laboratories.

7.2 Conclusions

Based on the analysis conducted in the four experimental chapters, the following conclusions are made:

- A broad overview of the genomic architecture of South African breeds, with insights on within and between breed diversity as well as relations with global populations was presented using sheep populations sampled from a broad range of production systems as well as geographic regions. The genetic differences between fat tailed sheep population, Merino type breeds and the English Dorsets was highlighted as well as the low levels of genetic diversity in small and indigenous breeds such as the Nguni, Namaqua Afrikaner and Blackhead Persian that differed significantly from the commercial breeds of South African Mutton Merino and Dorset Horn. Significant population differences were observed in genomic regions associated with the traits growth, meat/carcass, and reproduction as well as immune response mechanisms.

- Significant genetic variation is still maintained in South African Merino and Merino-derived sheep breeds and the study demonstrated some genetic similarities between the Merino and Merino-derived breeds emanating from common founding populations as well as some divergence driven by breed-specific selection goals.

- The frequency and patterns of distribution of ROH observed in this study corresponds to the breed history and implied selection pressures exposed to the sheep populations under study. Domestication driven ancient selection for traits such as horn presence, which tends to be universal across breeds was highlighted as well as recent and ongoing

selection for production traits and adaptive and welfare traits which tended to be breed specific.

- Using the OvineSNP50 Beadchip the study showed moderate LD that only persisted over short genetic distances. This LD decayed faster with increased marker distance across all breeds studied. There will be challenges in implementing applications such as multi-breed genomic prediction is SA sheep particularly for those less related breeds as suggested by the low correlations in consistence in gamete phase.

7.3 Recommendations and future studies

Genomewide SNP data present opportunities to interrogate the genomes and evolutionary processes that will be manifesting within populations. This study generated 400 genotypes for South African sheep breeds and merged it with an additional 623 genotypes from the ISGC. With such a data set, there were many possibilities of investigating the genomic architecture of SA sheep populations which were only constrained by time within the scope of a PhD program. Beyond this study, additional analysis can be undertaken to further understand South African sheep populations. Such analyses include but are not limited to:

(i) ***Structural variant analysis:*** Whilst a comprehensive set of analysis was conducted, it is recommended that additional analysis using other genomic variants, *i.e.* CNVs, could provide additional information on the genomic architecture of the South African sheep population. Many studies have demonstrated the role of CNV in traits of economic importance in a number of livestock species including sheep. However, little is known about the role of CNVs in South African sheep breeds.

(ii) ***Maternal and paternal origins of SA sheep population:*** The current study focused on the autosomal markers and excluded mitochondrial and Y chromosomal markers which could add valuable information particularly in understanding the history of local sheep populations and their introduction into South Africa from their presumed wild ancestors.

(iii) ***A finer analysis of subpopulations within the South African sheep population:*** The analysis of Merino and Merino-derived breeds reported interesting results on genomic changes that take place when breed are formed. In addition to Merino type breeds, South Africa hosts a number of other categories of sheep that share history and are

worthy of in-depth study. Examples of such populations are the Swakara subpopulations which were established from crossing of Karakul breeds and other fat-tailed sheep breeds and belong to a unique production system in South Africa. The Zulu ecotypes, including the Nguni breed that was included in this study, are another cluster of sheep populations characterised by a common ancestral population but raised in different geographical and ethnic systems.

(iv) The comprehensive analysis of genomewide SNP data of South African sheep populations undertaken in this study could be complemented by genomewide association analysis (GWAS) and other functional genomic studies to further elucidate the functionality of SNPs and genes associated with specific populations. Such analyses could make it more feasible to apply genomics in practical breed selection and improvement programs.

8 REFERENCES

Ahbara A., Bahbahani H., Almathen F., Al Abri M., Agoub M.O., Abeba A., Kebede A., Musa H.H, Mastrangelo S., Pilla F., Ciani E., Hanotte O., Mwacharo J.M. (2019) Genome-Wide Variation, Candidate Regions and Genes Associated With Fat Deposition and Tail Morphology in Ethiopian Indigenous Sheep. Frontiers in Genetics. 9, 699. https://doi:10.3389/fgene.2018.00699

Akanno E.C., Chen L., Abo-Ismail M.K. et al. (2018) Genome-wide association scan for heterotic quantitative trait loci in multi-breed and crossbred beef cattle. Genetic Selection and Evolution 50**,** 48. https://doi.org/10.1186/s12711-018-0405-y

Akey J.M., Ruhe A.L., Akey D.T., Wong A.K., Connelly C.F., et al. (2010) Tracking footprints of artificial selection in the dog genome. Proceedings of the National Science Academy 107: 1160–1165.

Alam M., Lee Y.M., Park B.L., Kim J.H., Lee S.S., Shin H.D., Kim K.S., Kim N.S. and Kim J.J. (2011) A whole genome association study to detect single nucleotide polymorphisms for body conformation traits in a Hanwoo population. Asian-Australian Journal of Animal Science 24, 322–329.

Al-Atiyat R., Flood W., Franklin I., Kinghorn B. and Ruvinsky A. (2016) Microsatellite-based genetic variation and differentiation of selected Australian Merino sheep flocks. Small Ruminant Research. 136, 137-144. https://doi.org/10.1016/j.smallrumres.2016.01.018.

Al-Atiyat R., Salameh N., and Tabbaa M. (2014) Analysis of genetic diversity and differentiation of sheep populations in Jordan. Electronic Journal of Biotechnology 17, 10.1016/j.ejbt.2014.04.002.

Al-Atiyat R.M., Aljumaah R.S., Alshaikh M.A. and Abudabos A.M. (2018) Microsatellite-Based Genetic Structure and Diversity of Local Arabian Sheep Breeds. Frontiers in Genetics9, 408. doi: 10.3389/fgene.2018.00408.

Alberto F.J., Boyer F., Orozcoter Wengel P., Streeter I., Servin B., de Villemereuil P. and Benjelloun B. et al. (2018) Nature Communications. 9, 813. https://doi: 10.1038/s41467-018-03206-y.

Albrechtsen A., Nielsen F.C., Nielsen R. (2010) Ascertainment biases in SNP chips affect measures of population divergence. Molecular Biology and Evolution 27: 2534–2547.

Al-Mamun H. A. A., Clark S., Kwan P. et al. (2015) Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genetic Selection and Evolution 47**,** 90 (2015). https://doi.org/10.1186/s12711-015-0169-6.

Almeida A.M. (2011) The Damara in the Context of Southern Africa Fat-Tailed Sheep Breeds. Tropical Animal. Health and Production. 43, 1427–1441.

Alvarenga A. B., Rovadoscki G. A., Petrini J. et al. (2018) Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries.* Scientific Reports 8, 8851 (2018). https://doi.org/10.1038/s41598-018-27259-7.

Álvarez I., Fernández I., Traoré A. et al. (2020) Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. Scientific Reports; 10, 2824. https://doi.org/10.1038/s41598-020-59839-x

Álvarez I., Fernández I., Traoré A. et al. (2020) Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. Scientific Reports 10, 2824 https://doi.org/10.1038/s41598-020-59839-x

An X., Ma H., Han P. et al. (2018) Genome-wide differences in DNA methylation changes in caprine ovaries between oestrous and dioestrous phases. Journal of Animal Science and Biotechnology 9, 85. https://doi.org/10.1186/s40104-018-0301-x.

Aramburu O., Ceballos F., Casanova A., Le Moan A., Hemmer-Hansen J., Bekkevold D., Bouza C., Martínez P. (2020) Genomic Signatures After Five Generations of Intensive Selective Breeding: Runs of Homozygosity and Genetic Diversity in Representative Domestic and Wild Populations of Turbot (*Scophthalmus maximus*). Frontiers in Genetics 11, 296. https://doi: 10.3389/fgene.2020.00296

Atlija M., Arranz J. J., Martinez-Valladares M. et al. (2016) Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. Genetic Selection and Evolution 48, 4. https://doi.org/10.1186/s12711-016-0182-4.

Bahbahani H., Clifford H., Wragg D. et al. (2015) Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. Scientific Reports 5, 11729. https://doi.org/10.1038/srep11729.

Bahrami A., Miraei-Ashtiani S.R., Mehrabani-Yeganeh H. (2012) Associations of growth hormone secretagogue receptor (GHSR) genes polymorphisms and protein structure changes with carcass traits in sheep. Gene. 505, 379-383. https://doi:10.1016/j.gene.2012.06.009

Bakhtiarizadeh M. R., Salehi A., Alamouti A. A. et al. (2019) Deep transcriptome analysis using RNA-Seq suggests novel insights into molecular aspects of fat-tail metabolism in sheep. Scientific Reports 9**,** 9203. https://doi.org/10.1038/s41598-019-45665-3

Banos G., Bramis G., Bush S. J., Clark E. L., McCulloch M. E. B., Smith J., Schulze G., Arsenos G., Hume D. A. and Psifidi A. (2017) The genomic architecture of mastitis resistance in dairy sheep. BMC Genomics 18, 624. DOI 10.1186/s12864-017-3982-1

Bertolini, F., Servin, B., Talenti, A. et al. (2018) Signatures of selection and environmental adaptation across the goat genome post-domestication. Genetic Selection and Evolution 50**,** 57 https://doi.org/10.1186/s12711-018-0421-y.

Bezuidenhout R. Bringing back the Afrino. Farmers weekly. (2012) https://www.farmersweekly.co.za/animals/cattle/bringing-back-the-afrino/

Biscarini F., Cozzi P., Gaspa G. and Marras G. (2018) detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network).

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

Bolormaa S., Chamberlain A. J., Khansefid M. et al. (2019) Accuracy of imputation to whole-genome sequence in sheep. Genetic Selection and Evolution 51, 1. https://doi.org/10.1186/s12711-018-0443-5.

Bolormaa S., Hayes B. J., van der Werf J. H. et al. (2016) Detailed phenotyping identifies genes with pleiotropic effects on body composition. BMC Genomics. 17, 224 https://doi.org/10.1186/s12864-016-2538-0

Brisbin A., Bryc K., Byrnes J. et al. (2012) PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Human Biology 84, 343-364. doi:10.3378/027.084.0401

Brito L.F., Jafarikia M., Grossi D.A. et al. (2015) Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. BMC Genetics 16, 67 https://doi.org/10.1186/s12863-015-0220-1.

Brito L.F., McEwan J.C., Miller S.P. et al. (2017) Genetic diversity of a New Zealand multi-breed sheep population and composite breeds' history revealed by a high-density SNP chip. BMC Genetics 18, 25 https://doi.org/10.1186/s12863-017-0492-8

Browning S.R. and Browning B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet.; 81(5):1084-97. doi: 10.1086/521987.

Bruford M., and Townsend, S. J. (2006). Mitochondridal DNA diversity in modern sheep. In M. A.Zeder, D. G. Bradley, E. Emshwiller, and B. D. Smith (Eds.), Documenting domestication: New genetic and archaeological paradigms (pp. 306-3 1 6). Berkeley: University of California Press.

Buduram P.O. (2004) Genetic characterization of Southern African sheep breeders using DNA markers. MSc thesis. 2004. Department of Animal, Wildlife and Grassland Sciences. University of the Free State.

Campbell A.W., Bain W.E., McRae A.F., Broad T.E., Johnstone P.D., Dodds K.G., Veenvliet B.A., Greer G.J., Glass B. C. and Beattie A. (2003) Bone density in sheep: genetic variation and quantitative trait loci localization. Bone 33, 540-548.

Cavanagh C.R., Jonas E., Hobbs M., Thomson P.C., Tammen I. and Raadsma H.W. (2010) Mapping Quantitative Trait Loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. Genetic Selection and Evolution 42, 36 doi: 10.1186/1297-9686-42-36.

Chen M., Pan D., Ren H. et al. (2016) Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. Genetic Selection and Evolution. 48(1): 76. doi:10.1186/s12711-016-0254-5.

Cheruiyot E.K., Bett R.C., Amimo J.O., Zhang Y., Mrode R., Mujibi F.D.N. (2018) Signatures of Selection in Admixed Dairy Cattle in Tanzania. Frontiers in Genetics 9:607. doi: 10.3389/fgene.2018.00607.

Chessa B., Pereira F., Arnaud F., Amorim A., Goyache F., et al. (2009) Revealing the history of sheep domestication using retrovirus integrations. Science. 324:532–536.

Ciani E., Mastrangelo S., Da Silva A. et al. (2020) On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. Genetic Selection and Evolution 52, https://doi.org/10.1186/s12711-020-00545-7.

Ciani E., Lasagna E., D'Andrea M., Alloggio I., Marroni F., Ceccobelli S., Bermejo J. V. D., Sarti F.M., Kijas J., Lenstra J. A. and Pilla F. and the International Sheep Genomics Consortium. (2015) Merino and Merino-derived sheep breeds: a genome-wide intercontinental study Genetics Selection Evolution. 47:64 https://doi.org/10.1186/s12711-015-0139-z.

Clark E.L.., Bush S.J.., McCulloch M.E.B., Farquhar I.L., Young R., Lefevre L., et al. (2017) A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS Genet 13, e1006997. https://doi.org/10.1371/journal.pgen.1006997.

Cloete S.W.P., Olivier J.J., Sandenbergh L. and 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 Science. http://dx.doi.org/10.4314/sajas.v44i4.1.

Cloete J.J.E., Cloete S.W.P., Scholtz A.J. and Matthee S. (2013) The Effect of Breed, Ewe Age and Season on Tick Counts of Indigenous and Commercial Sheep in South Africa. Proceedings for the Association of the Advancement of Animal Breeding and Genetics 20, 187–190.

Cloete S.W.P. and Olivier J.J. (2010) South African Sheep and Wool Industry. In: The International Sheep and Wool Handbook. Ed: Cottle, D.J., Nottingham University Press, Manor Farm, Thrumpton, Nottingham NG11 0AX, United Kingdom. pp. 95-112.

Cloete S.W.P., Olivier J.J. (2016) South African industry. International Sheep and Wool Handbook. 2016; 99-106.

Cole J.B. and da Silva M.V.G.B. (2016) Genomic selection in multi-breed dairy cattle populations. R. Bras. Zootec., 45:195-202, 2016

Conservation Priority List", livestockconservancy.org, The Livestock Conservancy, retrieved 2013-09-03.

Crawford A. M., Paterson K.A., Dodds K.G., Diez Tascon C., Williamson P.A., Roberts Thomson M., Bisset S.A., Beattie A.E., Greer G.J. and Green R.S. (2006) Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. BMC Genomics 7, 178. doi:10.1186/1471-2164-7-17.

Crispim B.A., Grisolia A.B., Seno L.O., Egito A.A., Vargas Junior F.M., Souza M.R. (2013) Genetic diversity of locally adapted sheep from Pantanal region of Mato Grosso do Sul. Genetic Molecular Research. 12, 5458–66.

Curik I., Ferenčaković M., Sölkner J. (2014) Inbreeding and runs of homozygosity: A possible solution to an old problem. Livestock Science. 166, 26-34. https://doi.org/10.1016/j.livsci.2014.05.034

DAFF, 2018. Abstracts of Agricultural Statistics, 2019. https://www.daff.gov.za/Daffweb3/Portals/0/Statistics%20and%20Economic%20Analysis/Statistical%20Information/Abstract%202019.pdf 2018. Abstracts of Agricultural Statistics, Directorate Statistics and Economic Analysis. (2013) Abstract of Agricultural Statistics; Department of Agriculture, Forestry and Fisheries: Pretoria, South Africa, 2013.

D'Ambrosio J., Phocas F., Haffray P., et al. (2019) Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. Genetics Selection and Evolution 51, 26. https://doi: 10.1186/s12711-019-0468-4.

Deb G.K., Choudhury M.P., Kabir M.A.., Khan M.Y.A., Ershaduzzaman M., Nahar T.N., Hossain S.M.J., Alam M.S., Alim M.A. (2019) Genetic relationship among indigenous sheep population of Bangladesh. Bangladesh Journal of Animal Science 48, 17-22.

Deniskova T. E., Dotsev A.V., Selionova M.I., Kunz E., Medugorac I., Reyer H., Wimmers K., Barbato M., Traspov A.A., Brem G. and Zinovieva N.A. (2018) Population structure and genetic diversity of 25 Russian sheep breeds based on whole genome genotyping. Genetics Selection Evolution 50, 29 https://doi.org/10.1186/s12711-0180399

Deniskova T., Dotsev A., Lushihina E., Shakhin A., Kunz E., Medugorac I., Reyer H., Wimmers K., Khayatzadeh N., Sölkner J., Sermyagin A., Zhunushev A., Brem G. and Zinovieva N. (2019) Population Structure and Genetic Diversity of Sheep Breeds in the Kyrgyzstan. Frontiers in Genetics10:1311. doi: 10.3389/fgene.2019.01311

Department Agriculture, Forestry and Fisheries–South Africa (DAFF). A profile of the South African wool market value chain; 2015. E-print.

Dieckgraefe B.K., Stenson W.F.., Korzenik J.R., Swanson P.E., and Harrington C.A. (2000) Analysis of mucosal gene expression in inflammatory bowel disease by parallel

oligonucleotide arrays. Physiological Genomics 4, 1-11. https://doi.org/10.1152/physiolgenomics.2000.4.1.1

Dlamini N.M., Visser C., Snyman G., Soma P. and Muchadeyi F. (2019) Genomic evaluation of resistance to *Haemonchus contortus* in a South African Dohne Merino flock. Small Ruminant Research 175, 10.1016/j.smallrumres.2019.04.020.

do Amaral Crispim B., Banari A. C., de Oliveira J.A., dos Santos Fernandes J. and Grisolia A. B. (2017) Naturalized Breeds in Brazil: Reports on the Origin and Genetic Diversity of the Pantaneiro Sheep. DOI: 10.5772/66019.

Domestic Animal Diversity Information System (DAD-IS): http://www.fao.org/dad-is/browse-by-country-and-species/en/

Dong K., Yang M., Han J., Ma Q., Han J., Song Z., Luosang C., Gorkhali N.A., Yang B., He X., Ma Y. and Jiang L. (2020) Genomic analysis of worldwide sheep breeds reveals PDGFD a major target of fat-tail selection in sheep. BMC Genomics. doi: https://doi.org/10.1101/2020.05.08.085431

Du Toit DJ. (2008) The indigenous livestock of Southern Africa. E-print. Available from: http://www.damaras.com/newsletters/du-toit.pdf

Edea Z.., Dessie T., Dadi H., Do K-T. and Kim K-S. (2017) Genetic Diversity and Population Structure of Ethiopian Sheep Populations Revealed by High-Density SNP Markers. Frontiers in Genetics 8, 218. doi: 10.3389/fgene.2017.00218

El-Halawany N., Zhou X., Al-Tohamy A.F., El-Sayd Y.A., Shawky A.E., Michal J.J. and Jiang Z. (2016) Genome-wide screening of candidate genes for improving fertility in Egyptian native Rahmani sheep. Animal Genetics 47, 513 https://doi.org/10.1111/age.12437.

Epstein H. (1960) History and Origin of the Ronderib and Namaqua Afrikaner Sheep. Zeitschrift für Tierzüchtung und Züchtungsbiologie 74, 267-292. doi:10.1111/j.1439-0388.1960.tb00132.x

Excoffier L. and Lischer H.E.L. (2009) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10, 564-567. DOI: 10.1111/j.1755-0998.2010.02847.x

FAO. (2007) State of the world's animal genetic resources for food and agriculture

Fariello M-I., Servin B., Tosser-Klopp G., Rupp R., Moreno C., International Sheep Genomics Consortium, et al. (2014) Selection Signatures in Worldwide Sheep Populations. PLoS ONE 9, e103813. https://doi.org/10.1371/journal.pone.010381

Fontanesi L., Beretti F., Martelli P.L., Colombo M., Dall'olio S., Occidente M., Portolano B., Casadio R., Matassino D., Russo V. (2011) A first comparative map of copy number variations in the sheep genome. Genomics 97, 158–165.

Fu W., Dekkers J.C., Lee W.R. et al. (2015) Linkage disequilibrium in crossbred and pure line chickens. Genetic Selection and Evolution 47, 11 https://doi.org/10.1186/s12711-015-0098-4

Garcia-Gamez E., Sahana G., Gutierrez-Gil B. and Arranz J. J. (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC Genetics. 13:43.

Gautier M. and Vitalis (2012) R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure, Bioinformatics 28, 1176–1177, https://doi.org/10.1093/bioinformatics/bts115

Gebretsadik Z.T. and Anal A.K. (2014) Indigenous sheep breeds of North Ethiopia: characterization of their phenotype and major production system. Tropical Animal Health Production. 46, 341-7. doi: 10.1007/s11250-013-0494-0.

Ghoreishifar, S. M., Moradi-Shahrbabak, H., Parna, N., Davoudi, P. and Khansefid, M. (2019) Linkage disequilibrium and within-breed genetic diversity in Iranian Zandi sheep. Archives animal breeding *62*, 143–151. https://doi.org/10.5194/aab-62-143-2019

Gifford-Gonzalez D., and Hanotte O. (2011). Domesticating animals in Africa: Implications of genetic and archaeological findings. Journal of World Prehistory 24,1–23. doi: 10.1007/s10963-010-9042-2.

Gizaw S., Abegaz S., Rischkowsky B., Haile A. and Mwai A.O. (2013) Dessie T. *Review of Sheep Research and Development Projects in Ethiopia.* Nairobi: International Livestock Research Institute (ILRI). https://cgspace.cgiar.org/handle/10568/35077.

Goddard M. (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245–57.

Grasso A.N., Goldberg V.N., Elly A.I., Wanda G., Diego A., Medrano I., Rincón J., Gonzalo F. and Gabriel C. (2014) Genomic variation and population structure detected by single nucleotide polymorphism arrays in Corriedale, Merino and Creole sheep. Genetic Molecular Biology 37, 389-395. https://doi.org/10.1590/S1415-47572014000300011

Grobler H. (2010) Evaluation Survey at Extensive Farms in the Central Karoo. Agriprobe 7, 6–9.

Groeneveld L.F., Lenstra J.A., Eding H., Toro M.A., Scherf B., Pilling D., Negrini R., Finlay E.K., Jianlin H., Groeneveld E., Weigend S. et al. (2010) Genetic diversity in farm animals - a review. Anim. Genetics 6-31. https://doi.org/10.1111/j.1365-2052.2010.02038.x

Grossman S.R., Shlyakhter I., Karlsson E.K., Byrne E.H., Morales S., Frieden G. et al. (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327, 883–886. doi: 10.1126/science.1183863.

Gudbjartsson D., Walters G., Thorleifsson G., *et al.* (2008) Many sequence variants affecting diversity of adult human height. Nature Genetics 40, 609–615. https://doi.org/10.1038/ng.122

Gwala P.E., Kunene N.W., Bezuidenhout C.C. and Mavule B.S. (2015) Genetic and phenotypic variation among four Nguni sheep breeds using random amplified polymorphic DNA (RAPD) and morphological features. Tropical Animal Health and Production 47, 1313-1319. doi: 10.1007/s11250-015-0865-9.

Haile A., Mészáros G., Rischkoiwsky B., Huson H.J., Gizaw M., Wurzinger M., Mwai A. O. and Sölkner J. (2019) Genetic diversity, population structure and runs of homozygosity in Ethiopian short fat-tailed and Awassi sheep breeds using genome-wide 50k SNP markers. Livestock Science 232, https://doi.org/10.1016/j.livsci.2019.103899.

Halim B. N., Nagara M., Regnault B., Hsouna S., Lasram K., Kefi R., Azaiez H., Khemira L., Saidane R., Ammar S.B., Besbes G., Weil D., Petit C., Abdelhak S. and Romdhane L. (2015) Estimation of Recent and Ancient Inbreeding in a Small Endogamous Tunisian Community Through Genomic Runs of Homozygosity. Annals of Human Genetics 79, 402-417. https://doi:10.1111/ahg.12131

Hassan D. I., Mbap S.T. and Naibi S.A. (2014) Socio-economic characteristics of Yankasa sheep and West African dwarf goat's farmers and their production constraint in Lafia Nigeria. International Journal of Food, Agriculture and Veterinary Sciences. 2 http://www.cibtech.org/jfav.htm2015Vol.5(1) January-April, pp. 82-93/Hassan et al

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

Hayes B.J., Bowman P.J., Chamberlain A.C. et al. (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetic Selection and Evolution 41, 51. https://doi.org/10.1186/1297-9686-41-51 Hedge N.G. 2019. Livestock Development for Sustainable Livelihood of Small Farmers. Asian Journal of Research in Animal and Veterinary Sciences 3, 1-17.

Hlophe S. R. (2011) Genetic variation between and within six selected South African sheep breeds using random amplified polymorphic DNA and protein markers PhD Thesis. 2011. Faculty of Science and Agriculture University of Zululand. https://www.semanticscholar.org/paper/Genetic-variation-between-and-within-six-selected-Hlophe/8fbfb389f50090df8b0d5f655f3ff24015b73c65

Hoffman I. and Scherf B. (2006) Animal genetic resources- Time to worry. FAO Livestock Report. http://www.fao.org/3/a0255e/a0255e05.pdf

Islam R., Li Y., Liu X., Berihulay H., Abied A., Gebreselassie G., Ma Q., Ma Y. (2019) Genome-Wide Runs of Homozygosity, Effective Population Size, and Detection of Positive Selection Signatures in Six Chinese Goat Breeds. Genes. 10, 938. https://doi.org/10.3390/genes10110938

Jemaa S.B., Thamri N., Mnara S., Rebours E., Rocha D. and Boussaha M. (2019) Linkage disequilibrium and past effective population size in native Tunisian cattle. Genetics and molecular biology. 42, 52–61. https://doi.org/10.1590/1678-4685-GMB-2017-0342

Jiang R., Cheng J., Cao X., Ma Y., Chaogetu B., Huang Y., Lan X., Lei C.Z., Hu L.Y. and Chen H. (2019) Copy Number Variation of the SHE Gene in Sheep and Its Association with Economic Traits. Animals 9, 531; https://doi.org/10.3390/ani9080531

Joaquim L.B., Chud T.C.S., Marchesi J.A.P., Savegnago R.P., Buzanskas M.E. and Zanella R. et al. (2019) Genomic structure of a crossbred Landrace pig population. PLoSONE 14, e0212266. https://doi.org/10.1371/journal.pone.0212266

Johnston S.E., McEwan J.C., Pickering N.K., Kijas J.W., Beraldi D., Pilkington J.G., Pemberton J.M. and Slate J. (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Molecular Ecology 20, 2555-2566. https://doi: 10.1111/j.1365-294X.2011.05076.x.

Jones P., Chase K.., Martin A, Davern P. and Ostrander E.A. (2008) Single- nucleotide-polymorphism-based association mapping of dog stereotypes. Genetics 179, 1033–1044.

Jordan W. (2013) Enhancing the breed analysis of the DM by accounting for heterogeneous variances and phantom parents. MSc Thesis. Stellenbosch University.

Karamichou E., Richardson R.I., Nute G.R., Gibson K.P. and Bishop S.C. (2006) Genetic analyses and quantitative trait loci detection, using a partial genome scan, for intramuscular fatty acid composition in Scottish Blackface sheep. Journal of Animal Science 84, 3228-38.

Khanyile K.S., Dzomba E.F. and Muchadeyi F.C. (2015) Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. Frontiers in Genetics 6, 13. doi: 10.3389/fgene.2015.00013

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H, McEwan J, Dalrymple B. and International Sheep Genomics Consortium. (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. PLoS Biology 10, https://doi: 10.1371/journal.pbio.1001258.

Kijas J.W., Porto-Neto L., Dominik S., Reverter A., Bunch R., McCulloch R., Hayes B.J., Brauning R., McEwan J. and International Sheep Genomics Consortium. (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal Genetics 45, 754-7. doi: 10.1111/age.12197.

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. and International Sheep Genomics Consortium. (2009) A Genome Wide Survey of SNP Variation

Reveals the Genetic Structure of Sheep Breeds Plos One https://doi.org/10.1371/journal.pone.0004668.

Kim E.S., Elbeltagy A., Aboul-Naga A. et al. (2016) Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity 116, 255–264 https://doi.org/10.1038/hdy.2015.94.

Kim E.S., Sonstegard T.S., Van Tassell C.P., Wiggans G., Rothschild M.F. (2015) The Relationship between Runs of Homozygosity and Inbreeding in Jersey Cattle under Selection. PLoS One 10, e0129967. https://doi:10.1371/journal.pone.0129967

Kirin M., McQuillan R., Franklin C.S., Campbell H., McKeigue P.M. and Wilson J.F. (2010) Genomic Runs of Homozygosity Record Population History and Consanguinity. PLoS ONE 5, e13996. https://doi.org/10.1371/journal.pone.0013996.

Kotzé J.J.J. (1951) The development of a mutton woolled sheep for the sour-grassveld area. Farming in South Africa 28, 110-113.

Kristensen T.N., Hoffmann A.A., Pertoldi C. and Stronen A.V. (2015) What can livestock breeders learn from conservation genetics and vice versa? Frontiers in Genetics 6, 38. doi:10.3389/fgene.2015.00038.

Kunene N., Nesamvuni E.A., Fossey A. (2007) Characterisation of Zulu (Nguni) Sheep using Linear Body Measurements and Some Environmental Factors Affecting these Measurements. South African Journal of Animal Science 37, 11–20.

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

Kunene N.W., Ceccobelli S., Di Lorenzo P., Hlophe S.R., Bezuidenhout C.C. and Lasagna E. (2014) Genetic diversity in four populations of Nguni (Zulu) sheep assessed by microsatellite analysis. Italian Journal of Animal Science 13, 3083.

Kunene N.W., Nesamvuni A.E., Nsahlai I.V. (2009) Determination of prediction equations for estimating body weight of Zulu (Nguni) sheep. Small Ruminant Research, 84, 41-46, https://doi.org/10.1016/j.smallrumres.2009.05.003

La Y., Zhang X., Li F., Zhang D., Li C., Mo F. and Wang W. (2019) Molecular characterization and expression of *SPP1*, *LAP3* and *LCORL* and their association with growth traits in sheep. Genes 14, 616. https://doi.org/10.3390/genes10080616.

Lancioni H., Di Lorenzo P., Ceccobelli S., Perego U.A., Miglio A. et al. (2013) Phylogenetic relationships of three Italian Merino-derived sheep breeds evaluated through a complete mitogenome analysis. Plos One 8, e73712.

Liu J., Zhang L., Xu L., et al. (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BMC Genomics 14, 229. doi:10.1186/1471-2164-14-229

Liu N., Li H., Liu K. *et al.* (2014) Identification of skin-expressed genes possibly associated with wool growth regulation of Aohan fine wool sheep. BMC Genetics 15, 144. https://doi.org/10.1186/s12863-014-0144-1

Liu H., Zhou H., Wu Y., Li X., Zhao J., et al. (2016). The Impact of Genetic Relationship and Linkage Disequilibrium on Genomic Selection. PLOS ONE 11(8): e0162057. https://doi.org/10.1371/journal.pone.0162057

Liu S., He S., Chen L. et al. (2017) Estimates of linkage disequilibrium and effective population sizes in Chinese Merino (Xinjiang type) sheep by genome-wide SNPs. Genes Genomes 39, 733–745 https://doi.org/10.1007/s13258-017-0539-2.

Liu X., Ong R.T-H., Pillai E.N., Elzein A.M., Small K.S., Clark T.G., Kwiatkowski D.P. and Teo Y-Y. (2013) Detecting and characterizing genomic signatures of positive selection in global populations. American Journal of Human Genetics 92, 866–881. DOI: 10.1016/j.ajhg.2013.04.021

Liu Z., Ji Z., Wang G.., Chao T, Hou L. and Wang J. Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. BMC Genomics. 17, 863. doi: 10.1186/s12864-016-3212-2.

Lv X., Chen W., Sun W., Hussain Z., Wang S. and Wang J. (2020) Analysis of lncRNAs Expression Profiles in Hair Follicle of Hu Sheep Lambskin. Animals 10, 1035. https://doi.org/10.3390/ani10061035.

Maijala K. (1997) Genetic aspects of domestication, common breeds and their origin. In: Piper, L. and A. Ruvinsky (eds.). The genetics of sheep. CABI

Makina S.O., Taylor J.F., van Marle-Köster E., Muchadeyi F.C., Makgahlela M.L., MacNeil M.D. and Maiwashe A. (2015) Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. Frontiers in Genetics 6, 337. doi: 10.3389/fgene.2015.00337

Malesa M.T. (2015) Population genetics of Swakara sheep inferred using genome-wide SNP genotyping. MSc Thesis. University of Kwazulu Natal.

Manunza A., Cardoso T., Noce A. et al. (2016) Population structure of eleven Spanish ovine breeds and detection of selective sweeps with BayeScan and hapFLK. Scientific Reports 6**,** 27296. https://doi.org/10.1038/srep27296.

Manzari Z., Mehrabani-Yeganeh H., Nejati-Javaremi A., Moradi MH and Gholizadeh M., (2019) Detecting selection signatures in three Iranian sheep breeds. Animal Genetics 50, 298-302. doi: 10.1111/age.12772.

Maqhashu A., Mapholi N.O., O'Neill H.A., Nephawe K.A., Ramukhithi F.V., Sebei J.P., Nxumalo K.S. and Nedambale T.L. (2020) Assessment of genetic variation in Bapedi sheep using microsatellite markers. South African Journal of Animal Science 50, http://dx.doi.org/10.4314/sajas.v50i2.15

Marshall K., Mugambi J. M., Nagda S., Sonstegard T. S., Van Tassell C. P., Baker R. L. and Gibson J. P. (2013) Quantitative trait loci for resistance to Haemonchus contortus artificial challenge in Red Maasai and Dorper sheep of East Africa. Animal genetics 44, 285-295.

Marshall K., Maddox J.F., Lee S.H., Zhang Y., Kahn L., Graser H.U., Gondro C., Walkden-Brown S.W. and van der Werf J.H. (2009) Genetic mapping of quantitative trait loci for resistance to Haemonchus contortus in sheep. Animal Genetics 40, 262-272.

Mason I.L. (1996) A world dictionary of livestock breeds, types and varieties. 4th Edition. CAB International Wallingford UK.

Mastrangelo S., Ciani E., Sardina M.T., Sottile G., Pilla F., Portolano B. Bi.Ov. Ita Consortium. (2018) Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Animal Genetics 49, 71-81. doi: 10.1111/age.12634. Epub 2018 Jan 15.

Matika O., Riggio V., Anselme-Moizan M., Law A.S., Pong-Wong R., Archibald A.L. Bishop S.C. (2016) Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs. Genetic Selection and Evolution 8, 48:11. doi: 10.1186/s12711-016-0191-3.

Mavule B.S., Sarti F.M., Lasagna E. and Kunene N.W. (2016) Morphological differentiation amongst Zulu sheep populations in KwaZulu-Natal, South Africa, as revealed by multivariate analysis. Small Ruminant Research 140, 50–56.

McRae A. F., McEwan J. C., Dodds K. G., Wilson T., Crawford A. M. and Slate J. (2002) Linkage Disequilibrium in Domestic Sheep. Genetics 160, 1113-1122.

McRae K.M., McEwan J.C., Dodds K.G. et al. (2014) Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. BMC Genomics 15, 637 https://doi.org/10.1186/1471-2164-15-637.

Mdladla K., Dzomba E.F. and Muchadeyi F.C. (2018) Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. Heredity 120, 369–378. https://doi.org/10.1038/s41437-017-0044-z.

Mdladla K., Dzomba E.F., Huson H.J. and Muchadeyi F.C. (2016) Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. Animal Genetics 47, 471-482. https://doi.org/10.1111/age.12442

Meadows J. R. S., Cemal I., Karaca O., Gootwine E. and Kijas J. W. (2007) Five ovine mitochondrial lineages identified from sheep breeds of the Near East. Genetics 175, 1371-1379.

Meadows J. R., Li K., Kantanen J., Tapio M., Sipos W. and Pardeshi V. et al. (2005) Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. Journal Heredity 96, 494–501.

Meadows J. R. S., Hanotte O., Drögemuller C., Calvo J., Goffrey R., Coltman D., et al. (2006) Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. Animal Genetics 37, 444-453.

Meadows J., Hiendleder S. and Kijas J. (2010) Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. Heredity 106, 700–706 (2011). https://doi.org/10.1038/hdy.2010.122

Mengesha M. and Tsega W. (2012) Indigenous sheep production in Ethiopia: a review. Iranian Journal of Applied Animal Science 2, 311-318.

Metzger J., Schrimpf R., Philipp U. and Distl O. (2013) Expression levels of LCORL are associated with body size in horses. PLoS ONE. 8, https://doi.org/10.1371/journal.pone.0056497

Meyermans R., Gorssen W., Wijnrocx K., J. Lenstra A., Vellema P., Buys N. and Janssens S. 2019. Unravelling the genetic diversity of Belgian Milk Sheep using medium-density SNP genotypes. Animal Genetics. doi: 10.1111/age.12891.

Molotsi A., Dube B., Oosting S., Marandure., Mapiye C., Cloete S. and Dzama K. (2017) Genetic Traits of Relevance to Sustainability of Smallholder Sheep Farming Systems in South Africa. Sustainability 9, 1225; https://doi.org/10.3390/su9081225.

Molotsi A., Dube, B., Oosting, S., Marandure, T., Mapiye, C., Cloete, S. and Dzama, K., 2017. Genetic traits of relevance to sustainability of smallholder sheep farming systems in South Africa. Sustainability, 9, 1225; doi: https://doi.org/10.3390/su9081225

Molotsi A.H., Taylor J.F., Cloete S.W., Muchadeyi F., Decker J.E., Whitacre L.K., Sandenbergh L. and 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. DOI: 10.1007/s11250-017-1392-7

Molotsi A.H., Taylor J.F., Cloete S.W.P., Muchadeyi F., Decker J.E., Sandenbergh L. and Dzama K. (2017) Preliminary genome-wide association study for wet-dry phenotype in smallholder ovine populations in South Africa. South African Journal of Animal Science http://dx.doi.org/10.4314/sajas.v47i3.9.

Molotsi A.H., Taylor J.F., Cloete S.W.P., Muchadeyi F., Decker J.E., Sandenbergh L. and Dzama K. (2017) Preliminary genome-wide association study for wet-dry phenotype in smallholder ovine populations in South Africa. South African Journal of Animal Science. http://dx.doi.org/10.4314/sajas.v47i3.9

Mooney J.A., Huber C.D., Service S., Sul J.H., Marsden C.D., Zhang Z., Sabatti C., Ruiz-Linares A., Bedoya G. Freimer N., Lohmueller KE. and Costa Rica/Colombia Consortium for Genetic Investigation of Bipolar Endophenotypes. (2018) Understanding the Hidden Complexity of Latin American Population Isolates. American. Journal of Human Genetics 103, 707–726. https://doi.org/10.1016/j.ajhg.2018.09.013

Muchadeyi F.C., Malesa M.T., Soma P. and Dzomba E.F. (2015) Runs of Homozygosity in Swakara Pelt producing Sheep: Implications on Sub-vital performance. Proceedings for Association for the Advancement of Animal Breeding and Genetics. 21, 310-313 http://www.aaabg.org/aaabghome/AAABG21papers/Muchadeyi21310.pdf

Muigai A.W.T. and Hanotte O. (2013) The Origin of African Sheep: Archaeological and Genetic Perspectives. African Archaeology Reviews 30, 39-50, DOI 10.1 007/s 1 0437-0 13-91 29-30.

Mukhongo D.M., Mwai O., Tapio M. and Muigai A. (2014) Genetic diversity and population structure of the indigenous sheep in Kenya based on microsatellite analysis: implications for their conservation. Journal of Livestock Science 5, 65–78.

Murdoch B. M. (2019) The functional annotation of the sheep genome. Journal of Animal Science 97, 16, https://doi.org/10.1093/jas/skz122.029.

Mwangi J.G. (2013) Developing a vibrant livestock industry in East Africa through market driven research. Journal of US-China Public Administration 10, 608-617.

Naidoo P., Cloete S.W.P. and Fossey A. (2005) South African Merinos divergently selected for multiple rearing ability: a preliminary study of divergence based on RAPD markers. Proceedings for Association for the Advancement of Animal Breeding and Genetics. 16, 254-257.

Nandolo W., Utsunomiya Y.T., Mészáros G., Wurzinger M., Khayadzadeh N., Torrecilha R. B. P., Mulindwa H. A., Gondwe T. N., Waldmann P., Ferencakovic M., Garcia J. F., Rosen B. D., Bickhart D., Tassell C. P., Curik I. and Solkner J. (2018) Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. Genetic Selection and Evolution. 50, 43. https://doi.org/10.1186/s12711-018-0414-x.

Naval-Sanchez M., Nguyen Q., McWilliam S., Porto-Neto L.R., Tellam R., Vuocolo T., et al. (2018) Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. Nature Communications 9, 859.

Nxumalo K.S., Grobler J. P., Ehlers K., Banga C. B., Mapholi N. O. (2018) Genomic diversity and population structure of four South African indigenous Sheep. Proceedings of the World Congress of Genetics Applied to Livestock Production 2018. http://www.wcgalp.org/system/files/proceedings/2018/genomic-diversity-and-population-structure-four-south-african-indigenous-sheep.pdf

Olynk N. J., Assessing changing consumer preferences for livestock production processes, Animal Frontiers 2, 32–38, https://doi.org/10.2527/af.2012-0046

Paim, Tiago do Prado et al. Detection and evaluation of selection signatures in sheep. Pesq. agropec. bras. [online]. 2018, vol.53, n.5 [cited 2021-01-19], pp.527-539. https://doi.org/10.1590/s0100-204x2018000500001.

Peripolli E., Stafuzza N.B, Munari D.P. et al. (2018) Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. BMC Genomics. 19, 34. https://doi:10.1186/s12864-017-4365-3.

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, 101-108.

Pickrell J.K., Pritchard J.K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genetics 8. https://doi.org/10.1371/journal.pgen.1002967

Piper L. The genetics of sheep, Ed by Piper L, Ruvinsky A, Ramsay K, Harris L, Kotzé A., Landrace breeds: South Africa's indigenous and locally developed farm animals. 1998. Farm Animal Conservation Trust.

Pollott G. and Wilson R. T. (2009) Sheep and goats for diverse products and profits. FAO Diversification Booklet 9.

Porto-Neto L. R., Kijas J.W., Reverter A. (2014) The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genetic Selection and Evolution 46, 22.

Pridans C. (2016) A Csf1r-EGFP Transgene Provides a Novel Marker for Monocyte Subsets in Sheep. in Journal of immunology (Baltimore, Md. : 1950)

Prieur V., Clarke S.M., Brito L.F. et al. (2017) Estimation of linkage disequilibrium and effective population size in New Zealand sheep using three different methods to create genetic maps. BMC Genetics 18, https://doi.org/10.1186/s12863-017-0534-2.

Prieur V., Clarke, S.M., Brito, L.F. et al. (2017) Estimation of linkage disequilibrium and effective population size in New Zealand sheep using three different methods to create genetic maps. BMC Genetics 18, https://doi.org/10.1186/s12863-017-0534-2

Pryce J. E., Hayes B. J., Bolormaa S., Goddard M. E. (2011) Polymorphic Regions Affecting Human Height Also Control Stature in Cattle. Genetics 3, 981–4.

Purcell S.N., Todd-Brown K., Thomas L., Ferreira M., Bender D., Maller J., Sklar P., De Bakker P., and Pc D.M.S. (2007). PLINK: a toolset for. http://pngu.mgh.harvard.edu/purcell/plink/

Purfield D. C., McParland S., Wall E. and Berry D. P. (2017) The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS ONE 12(5): e0176780. https://doi.org/10.1371/journal.pone.0176780.

Purfield D.C., McParland S., Wall E., Berry D.P. (2017) The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS ONE. 12, e0176780. https://doi.org/10.1371/journal.pone.0176780.

Qanbari S. (2020) On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. Frontiers in Genetics. 10, 1304. https://doi: 10.3389/fgene.2019.01304

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

Qwabe S.O., van Marle-Köster E. and Visser C. (2013) Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. Tropical Animal Health and Production 45, 511-516. doi: 10.1007/s11250-012-0250-x

Ramsay K., Harris L. and Kotze A. (2001) Landrace Breeds: South Africa's Indigenous and Locally developed Farm Animals. Publication Farm Animal Conservation Trust. ISBN: 0-620-25493-9.

Randhawa I. A. S., Khatkar M. S., Thomson P. C. and Raadsma H. W. (2016) A Meta-Assembly of Selection Signatures in Cattle. PLoS ONE 11, e0153013. https://doi.org/10.1371/journal.pone.0153013

Rege J.E.O., and Gibson J. P. (2003) Animal genetic resources and economic development: issues in relation to economic valuation. Ecological Economics 45, 319-330.

Ren X., Yang G., Peng W. et al. (2016) A genome-wide association study identifies a genomic region for the polycerate phenotype in sheep *(Ovis aries).* Scientific Reports 6, 21111 https://doi.org/10.1038/srep21111.

Rezaei H. R., Naderi S., Chintauan-Marquier I. C., Jordan S., Taberlet P., Virk A. T., Naghash H. R., Rioux D., Kaboli M., Luikart G., Pompanon F. (2010) Molecular Phylogenetics Evolution. 54, 315-326.

Rochus C. M., Jonas E. and Johansson A.M. (2020) Population structure of five native sheep breeds of Sweden estimated with high density SNP genotypes. BMC Genetics 21, 27 (2020). https://doi.org/10.1186/s12863-020-0827-8.

Rojas-Downing M. M., Nejadhashemi A. P., Harrigan T., Woznicki S. A. (2017) Climate change and livestock: Impacts, adaptation, and mitigation, Climate Risk Management, 16, 145-163, ISSN 2212-0963. https://doi.org/10.1016/j.crm.2017.02.001.

Ruiz-Larrañaga O., Langa J., Rendo F. et al. (2018) Genomic selection signatures in sheep from the Western Pyrenees. Genetic Selection and Evolution. 2018; 50, 9. https://doi.org/10.1186/s12711-018-0378-x

Ruiz-Larrañaga O., Langa J., Rendo F., Manzano C., Iriondo M. and Estonba A. (2018) Genomic selection signatures in sheep from the Western Pyrenees. Genetics Selection Evolution 50, 9 DOI: 10.1186/s12711-018-0378-x

Saatchi M., Schnabel R.D., Taylor J. F., Garrick D. J. (2014) Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics. 15, 442.

Sabeti P. C. et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature. 419, 832–837.

Sams A. J., Boyko A. R. (2019) Fine-Scale Resolution of Runs of Homozygosity Reveal Patterns of Inbreeding and Substantial Overlap with Recessive Disease Genotypes in Domestic Dogs. G3 (Bethesda, Md.). 9, 117–123. https://doi.org/10.1534/g3.118.200836.

Sandenbergh L. (2015) Identification of SNPs associated with robustness and greater reproductive success in the South African Merino sheep using SNP chip technology. 2015. Doctoral dissertation, Stellenbosch: Stellenbosch University.

Sandenbergh L., Cloete S.W.P., Roodt-Wilding R., Snyman M.A., Bester-van der Merwe A.E. (2016) Evaluation of the OvineSNP50 Beadchip for use in four South African sheep breeds. South African Journal of Animal Sciences. 46, 89-93. http://dx.doi.org/10.4314/sajas.v46i1.11.

SAS Institute Inc. Cary N.C., US. Statistical Analysis Systems user's guide (SAS). 2013; v 9.4

Scheet P., Stephens M. A. (2006) Fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. American Journal of Human Genetics. 78, 629–644. DOI: 10.1086/502802

Scherf, B.D. World watch list for domestic animal diversity. 3rd ed. Roma: FAO; 2000. 726 p

Schoeman S. J., Cloete S.W.P. and Olivier J.J., (2010) Returns on investment in sheep and goat breeding in South Africa. Livestock. Science 130, 70-82.

Schönherz A. A., Szekeres B. D., Nielsen V.H. and Guldbrandtsen B. (2020) Population structure and genetic characterization of two native Danish sheep breeds. Acta Agriculture Scandinavica. 69, 53-67. https://doi.org/10.1080/09064702.2019.1639804

Selepe M. M., Ceccobelli S., Lasagna E. and Kunene N.W. (2018) Genetic structure of South African Nguni (Zulu) sheep populations reveals admixture with exotic breeds. PLoS ONE. 13, e0196276. https://doi.org/10.1371/journal.pone.0196276

Signer-Hasler H., Burren A., Ammann P., Drögemüller C and Flury C. (2019). Runs of homozygosity and signatures of selection: a comparison among eight local Swiss sheep breeds. Animal Genetics 50, 512-525. https://doi.org/10.1111/age.12828

Sim Z. and Coltman D.W. (2019) Heritability of Horn Size in Thinhorn Sheep. Frontiers in Genetics 10, 959. https://doi:10.3389/fgene.2019.00959

Slatkin M. (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9, 477–485. https://doi:10.1111/age.12309.

Snyman G. and Fisher A.D. (2019) Genetic parameters for traits associated with resistance to Haemonchus contortus in a South African Dohne Merino sheep flock. Small Ruminant Research 176, 76-88.

Snyman M. A. (2014a) South African sheep breeds: Persian sheep. Grootfontein Agricultural Development Institute. Info-pack ref. 2014/026

Snyman M.A. (2014b) South African sheep breeds: Afrino sheep. Info-pack ref. 2014/013. Grootfontein Agricultural Development Institute. http://gadi.agric.za/InfoPacks/2014013%20South%20African%20Sheep%20breeds%20-%20Afrino.pdf

Snyman M.A. (2014c) South African sheep breeds: Damara sheep. Info-pack ref. 2014/021. Grootfontein Agricultural Development Institute. http://gadi.agric.za/InfoPacks/2014014%20South%20African%20Sheep%20breeds%20-%20Damara.pdf

Snyman M.A. (2014d) South African sheep breeds: Dohne Merino sheep. Info-pack ref. 2014/015. Grootfontein Agricultural Development Institute. http://gadi.agric.za/InfoPacks/2014015%20South%20African%20Sheep%20breeds%20-%20Dohne%20Merino.pdf

Snyman M.A. (2014e) South African sheep breeds: Meatmaster sheep. Info-pack ref. 2014/021. Grootfontein Agricultural Development Institute. http://gadi.agric.za/InfoPacks/2014021%20South%20African%20Sheep%20breeds%20-%20Meatmaster.pdf

Snyman M.A. (2014f) South African sheep breeds: Namaqua Afrikaner sheep. Info-pack ref. 2014/021. Grootfontein Agricultural Development Institute. http://gadi.agric.za/InfoPacks/2014021%20South%20African%20Sheep%20breeds%20-%20NamaquaAfrikaner.pdf

Snyman M.A. (2014g) South African sheep breeds: Karakul sheep. Info-pack ref. 2014/021. Grootfontein Agricultural Development Institute. http://gadi.agric.za/InfoPacks/2014021%20South%20African%20Sheep%20breeds%20-%20Karakul.pdf

Snyman M.A. (2014h) South African sheep breeds: Nguni sheep. Info-pack ref. 2014/021. Grootfontein Agricultural Development Institute. http://gadi.agric.za/InfoPacks/2014024%20South%20African%20Sheep%20breeds%20-%20Nguni.pdf

Snyman M.A., van Marle-Koster E., Qwabe S.Q., Visser C. (2013) Genetic and Phenotypic Profile of Three South African Namaqua Afrikaner Sheep Flocks. Grootfontein Agriculture 13, 1–10.

Soma P., Kotze A., Grobler J.P. and Van Wyk J.B. (2012) South African sheep breeds: Population genetic structure and conservation implications. Small Ruminant Research 103, 112-119.

Stephan W., Song Y.S., Langley C.H. (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics 172, 2647–2663. pmid:16452153

Steyn Y., Lourenco D. A. L., Misztal I. (2019) Genomic predictions in purebreds with a multibreed genomic relationship matrix1. Journal of Animal Science 4, 4418-4427. doi: 10.1093/jas/skz296.

Szmatoła T., Gurgul A., Jasielczuk I., Ząbek T., Ropka-Molik K., Litwińczuk Z., Bugno-Poniewierska M. (2019) A Comprehensive Analysis of Runs of Homozygosity of Eleven Cattle Breeds Representing Different Production Types. Animals 9, 1024. https://doi.org/10.3390/ani9121024.

Szpiech Z. A., Xu J., Pemberton T. J., Peng W., Zöllner S., Rosenberg N. A and Li J.Z. (2019) Long runs of homozygosity are enriched for deleterious variation. American journal of human genetics. 93, 90 – 102. https://doi.org/10.1016/j.ajhg.2013.05.003.

Taberlet P., Coissac E., Pansu J., Pompanon F. (2011) Conservation genetics of cattle, sheep, and goats. Comptes Rendus Biologies 334, 247–254. 10.1016/j.crvi.2010.12.007

Taberlet P., Valentini A., Rezaei H. R., Naderi S., Pompanon F., Negrini R., et al. (2008) Are cattle, sheep, and goats endangered species? Molecular Ecology (2008) 17: 275–284. pmid:17927711

Takadi K., Lemmer W. and de Jager P. (2020) Global mutton market dominated by China. https://www.bizcommunity.com/PDF/PDF.aspx?l=1andc=358andct=1andci=185616

Tang K., Thornton K.R. and Stoneking M. (2007) A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biology 5, e171. doi: 10.1371/journal.pbio.0050171.

Tapio M., Marzanov N., Ozerov M., Cinkulov M., Gonzarenko G. et al. (2006) Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. Molecular Biology and Evolution. 23, 1776–1783.

Tapio M., Tapio I., Grislis Z., Holm L. E. and Jeppsson S. et al. (2005) Native breeds demonstrate high contributions to the molecular variation in northern European sheep. Molecular Ecology 14, 3951–3963.

Tijjani A., Utsunomiya Y.T., Ezekwe A.G., Nashiru O. and Hanotte O. (2019) Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypanotolerant West African Muturu Cattle. Frontiers in Genetics 10, 442. doi: 10.3389/fgene.2019.00442

Tisdell C. (2003) Socioeconomic causes of loss of animal genetic diversity: analysis and assessment. Ecological Economics 45, 365-376.

Uemoto Y., Takeda M. and Ogino A. (2020) et al. Genetic and genomic analyses for predicted methane-related traits in Japanese Black steers. Animal Science Journal. 91, e13383. https://doi.org/10.1111/asj.13383.

Upadhyay M., Chen W. and Lenstra J. et al. (2017) Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. Heredity 118, 169–176. https://doi.org/10.1038/hdy.2016.79.

Usai M. G., Casu S., Sechi T. et al. (2019) Mapping genomic regions affecting milk traits in Sarda sheep by using the OvineSNP50 Beadchip and principal components to perform combined linkage and linkage disequilibrium analysis. Genetic Selection and Evolution 51, 65 https://doi.org/10.1186/s12711-019-0508-0.

Vali U., Einarsson A., Waits L. and Ellegren H. (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Molecular Ecology 17, 3808–3817.

Van der Werf J. H. J. (2009) Potential Benefit of Genomic Selection in Sheep. Proceedings for the Association for Advancement Animal Breeding and Genetics 18, 38–41.

Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. (2006) A map of recent positive selection in the human genome. PLoS Biol. 4, e72. doi: 10.1371/journal.pbio.0040072.

Wang W., Zhang X., Zhou X., Zhang Y., La Y., Zhang Y., Li C., Zhao Y., Li F., Liu B. and Jiang Z, (2019) Deep Genome Resequencing Reveals Artificial and Natural Selection for Visual Deterioration, Plateau Adaptability and High Prolificacy in Chinese Domestic Sheep. Frontiers in Genetics 10, 300. doi: 10.3389/fgene.2019.00300

Wang X.L., Zhou G.X., Li Q., Zhao D.F. and Chen Y.L. (2014) Discovery of SNPs in RXFP2 related to horn types in sheep. Small Ruminant Research 116, 133–136. https://doi.org/10.1016/j.smallrumres.2013.10.022.

Wang Z., Zhang H., Yang H., Wang S., Rong E., Pei W., Li H. and Wang N. (2014) Genome-wide association study for wool production traits in a Chinese Merino sheep population. PloS One 9, e107101.

Weedon M.N., Lango H., Lindgren C.M., Wallace C., Evans D.M. Mangino M. et al. (2008) Genome-wide association analysis identifies 20 loci that influence adult height. Nature Genetic 40, 575–83.

Wei C., Wang H., Liu G., Wu M., Cao J., Liu Z. et al. (2015) Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. BMC Genomics 16, 194. PMID: 25888314 doi:10.1186/s12864-015-1384-9.

Wiedemar N. and Drögemüller C. A. (2015) 1.8-kb insertion in the 3′-UTR of RXFP2 is associated with polledness in sheep. Animal Genetics 46, 457–61. DOI: 10.1111/age.12309.

Wiedemar N., Drögemüller C. A. (2015) 1.8-kb insertion in the 3'-UTR of RXFP2 is associated with polledness in sheep. Animal Genetics 46, 457-461. https://doi:10.1111/age.12309.

Wientjes Y.C., Veerkamp R.F.., Calus M.P. 92013) The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193, 621-631. doi:10.1534/genetics.112.146290

Wilson R.T. (2011) Populations and production of fat-tailed and fat-rumped sheep in the Horn of Africa. Tropical Animal Health and Production 43, 1419–1425. https://doi.org/10.1007/s11250-011-9870-9.

Woad K. J. Hammond A. J. Hunter M. Mann G. E. Hunter M. G. Robinson R. S. FGF2 is crucial for the development of bovine luteal endothelial networks in vitro. Reproduction. 2009; 138:581-588 doi: 10.1530/REP-09-0030.

Wood A. R., Esko T., Yang J., Vedantam S., Pers T. H., Gustafsson S. et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics. 46, 1173– 1186.

Xu L., Zhao G., Yang L., et al. (2019) Genomic Patterns of Homozygosity in Chinese Local Cattle. Scientic Reports 9**,** 16977. https://doi.org/10.1038/s41598-019-53274-3.

Yan L., Xu Y., Zhou Y., Liu M., Wang L. and Kijas J.W. (2017) Diversity of copy number variation in a worldwide population of sheep. Genomics 110, 143-148. https://doi.org/10.1016/j.ygeno.2017.09.005.

Yang GL, Fu DL, Lang X, et al. (2013) Mutations in MC1R gene determine black coat color phenotype in Chinese sheep. Scientific World Journal: 675382. doi:10.1155/2013/675382

Yang., Wen-Rong Li., Feng-Hua Lv., San-Gang He., Shi-Lin Tian., Wei-Feng Peng., Ya-Wei Sun., Yong-Xin Zhao., Xiao-Long Tu., Min Zhang., Xing-Long Xie., Yu-Tao Wang., Jin-Quan Li., Yong-Gang Liu., Zhi-Qiang Shen., Feng Wang., Guang-Jian Liu., Hong-Feng Lu. and Juha Kant. (2016) Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. Molecular Biology and Evolution 33, 2576–2592.

Yeomans L., Martin L. and Richter T. (2017) Expansion of the known distribution of Asiatic mouflon (Ovis orientalis) in the Late Pleistocene of the Southern Levant. Royal Society Open Science 4, 170409. doi:10.1098/rsos.17040.

Yurchenko, A.A., Deniskova T.E., Yudin N.S. et al. (2019) High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. BMC Genomics 20**,** 294.

Zhang Q., Guldbrandtsen B., Bosse M., Lund M.S., Sahana G. (2015) Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics 16, 542.

Zhang R., Liu F., Hunt P., et al. (2019) Transcriptome analysis unravelled potential mechanisms of resistance to *Haemonchus contortus* infection in Merino sheep populations bred for parasite resistance. Veterinary Research 50, 7 https://doi.org/10.1186/s13567-019-0622-6.

Zhang S., Lin H., Kong S., et al. (2013) Physiological and molecular determinants of embryo implantation. Molecular Aspects Medicine 34, 939-980. doi:10.1016/j.mam.2012.12.011.

Zhang L., Mousel M.R., Wu X., Michal J. J., Zhou X., et al. (2013) Genome-Wide Genetic Diversity and Differentially Selected Regions among Suffolk, Rambouillet, Columbia, Polypay and Targhee Sheep. PLos One. 8, e65942. https://doi.org/10.1371/journal.pone.0065942

Zhao F., Wei C., Zhang L. Liu J., Wang G., Zeng T., Du L (2018). A genome scan of recent positive selection signatures in three sheep populations. Journal of Integrative Agriculture. 15, 162-174. https://doi.org/10.1016/S2095-3119(15)61080-2

Zhao Q., Sun H., Zhang Z., Xu Z, Olasege B. S., Ma P., Zhang X., Wang Q. and Pa Y. (2019) Exploring the structure of haplotype blocks and genetic diversity in Chinese indigenous pig populations for conservation purpose. Evolutionary Bioinformatics 15, 1-8.

Zhi D., Da L., Liu M., Cheng C., Zhang Y., Wang X., Li X., Tian Z., Yang Y., He T., Long X., Wei W. and Cao G. (2018) Whole genome sequencing of Hulunbuir short-tailed sheep for identifying candidate genes related to the short-tail phenotype. G3: Genes, Genomes, Genetics 8, 377-383; https://doi.org/10.1534/g3.117.300307.

Zhu C., Fan H., Yuan Z., Hu S., Zhang L., Wei C., Zhang Q., Zhao F., and Du L. (2015) Detection of selection signatures on the X chromosome in three sheep breeds. International Journal of Molecular Science 16, 20360-20374; doi:10.3390/ijms160920360.