Genomic characterization of autozygosity in South African Merino sheep

by

Micah Forsythe

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae Agriculturae

MSc (Agric) Animal Science: Animal Breeding and Genetics

In the Faculty of Natural and Agricultural Sciences

Department of Animal Science

University of Pretoria

Pretoria

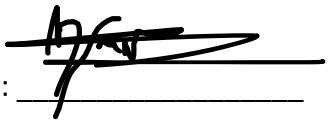
Date: November 2024

Supervisor: Prof Carina Visser

Co-supervisor: Dr Simon Frederick Lashmar

Declaration

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

Signature: 

Name: Micah Forsythe

Date: November 2024

Pretoria

Acknowledgements

A special thanks to SA Stud Book and University of Pretoria for providing the data and facilities to make this study possible.

I would like to express my deepest gratitude to my supervisors Prof C. Visser and Dr S. Lashmar for their guidance, patience and support throughout the duration of my Masters. I am truly thankful for the knowledge they shared with me, and dedication to helping me develop my writing and critical thinking skills.

I would also like to thank my family for the sacrifices they made in order for me to be able to pursue my passions and education. To my friends, thank you for staying up late with me and providing much-needed companionship during my writing process.
Lastly, I want to acknowledge my own dedication, late nights and hard work I put into creating a dissertation I can be proud of.

Abstract

The South African (SA) Merino is a medium-framed, white wool breed, and is one of the most prominent breeds in the SA sheep industry. Because of its local importance, the SA Merino is the only SA sheep breed with sufficient single nucleotide polymorphism (SNP) genotypes to implement genomic selection (GS), and this provides comprehensive sample size (i.e. number) and composition (i.e. representation) of genotypic data to study the breed's diversity and inbreeding. These estimates play a vital role as tools for breeders to maintain genetic progress whilst conserving diversity within flocks. This study aimed to perform a homozygosity-focused SNP investigation of the past and present inbreeding and autozygosity levels in the SA Merino sheep breed. A total of 1 738 animals (1 567 females and 171 males), genotyped using different versions of the Illumina® Ovine 50K SNP panel, were used to estimate various proxy-indicator parameters of genomic variability, runs of homozygosity (ROH) and heterozygosity-rich region (HRR) profiles, and conserved ROH-based selection signatures for the SA Merino breed. The reported observed (H_O=0.351± 0.018) and expected (H_E=0.365± 0.0001) heterozygosity, and their relationship to one another (i.e. H_O<H_E), as well as SNP-based inbreeding coefficient (F_{SNP}=0.037±0.05) and effective population size (most recent Ne=305) indicated a slight loss in within-population genetic diversity, and low inbreeding levels. Additionally, high MAF values (0.275±0.14) and LD (0.365±0.180) were observed within the SA Merino population. The autozygosity analysis identified a total of 43 147 ROH with a mean chromosome-wide ROH length per chromosome of 6.53Mb, and a corresponding mean F_{ROH} of 0.074±0.04 that further supported low inbreeding. Conserved ROH revealed selection signatures harbouring a total of 275 protein-coding genes (within a threshold of 16.4%) associated with various functions in sheep involving reproduction, fibre formation, and inflammatory responses. Overall, the results from this study conveyed that mating programs and intense selection are well managed within the breeding objectives of SA Merino sheep, and that sufficient levels of genetic diversity exists to facilitate sustainable production and genetic improvement in the future.

Table of Contents

List of abbreviations

A	Afrino
AM	Australian Merino
ARC	Agricultural Research Council
BL	Border Leicester
BLUP	Best Linear Unbiased Prediction
BP	Blackhead Persian
C	Corriedale
CM	Chinese Merino
CR	Consecutive Runs
CW	Coopworth
DALRRD	Department of Agriculture, Land Reform and Rural Development
D	Dorper
DAR	Damara
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DM	Dohne Merino
DNA	Deoxyribonucleic Acid
DR	Dormer
EBV	Estimated Breeding Value
F	Inbreeding Coefficient
F_{GRM}	Genomic Relationship Matrix-based Inbreeding Coefficient
F_{IS}	Genomic Inbreeding Coefficient
F_{SNP}	SNP-based Inbreeding Coefficient
F_{PED}	Pedigree-based Inbreeding Coefficient
F_{ROH}	ROH-based Inbreeding Coefficient
F_{ST}	F-statistic Genetic Differentiation
F_{UNI}	F-statistic Uniting Gametes
GEBV	Genomic Estimated Breeding Values
GO	Gene Ontology
GRM	Genomic Relationship Matrix
GS	Genomic Selection
H_E	Expected Heterozygosity
HI	Global Humidity Index
H_O	Observed Heterozygosity
HRR	Heterozygosity-rich Regions
HWE	Hardy-Weinberg Equilibrium
IBD	Identical-by-Descent
IBS	Identical-by-State
INTERGIS	Integrated Registration and Genetic Information System
ISGC	International Sheep Genomics Consortium
K	Karakul
Kb	Kilobase
KEGG	Kyoto Encyclopaedia of Genes and Genomes

LD	Linkage Disequilibrium
LSU	Large Stock Unit
MAF	Minor Allele Frequency
Mb	Megabase
MM	Meatmaster
N	Non-Descript Nguni Sheep
NA	Namaqua Afrikaner
Ne	Effective Population Size
NSSIS	South African National Small Stock Improvement Scheme
OAR	Ovine Chromosome
P	Pedi
PCR	Polymerase Chain Reaction
PD	Poll Dorset
QC	Quality Control
QTL	Quantitative Trait Loci
r^2	R-squared/squared correlation for Linkage Disequilibrium
RA	Ronderib Afrikaner
RNA	Ribonucleic acid
ROH	Runs of Homozygosity
RP	Redhead Persian
S	Swakara
SAL	South African Landsheep
SAM	South African Merino
SAMM	South African Mutton Merino
SNP	Single Nucleotide Polymorphisms
SPM	Spanish Merino
SSU	Small Stock Unit
SW	Swazi
SW	Sliding Window
WGS	Whole-genome sequencing
WS	White Suffolk

List of tables

Addendum

List of figures

Chapter 1: Introduction

1.1 Introduction

The domestication of sheep (*Ovis aries*) occurred approximately 11 000 years ago. All modern sheep breeds are descendants of the Asiatic wild mouflon (*Ovis orientalis*; Zohary *et al.,* 1998). A pivotal moment in the development of commercial sheep breeds occurred when humans moved from being nomadic hunter-gatherers to farmers in settlements approximately 4 000 years ago (Kijas *et al.,* 2012; Florian *et al.,* 2018). Domestic sheep originated in southwest Asia (the Fertile Crescent) and spread from the Middle East throughout south-eastern Europe and the rest of Europe during the Neolithic period (Zeder, 2008; Vigne, 2011). Zeder (2008) suggested that the entry of sheep into Africa was due to colonists pushing indigenous hunter-gatherers outwards from Northern territories and through the horn of Africa. Since its introduction to the continent, archaeological records show that the southward migration of domesticated sheep towards southern Africa occurred approximately two millenniums ago (Vigne, 2011).

Although the exact number of modern sheep breeds globally has not officially been confirmed, there is a general agreement that it is at least several hundred (Rasali *et al.,* 2006; Handley *et al.,* 2007), if not thousands (Gorkhali & Yuehui, 2015). In South Africa (SA), there are 20 recognized breeds (DALRRD, 2006) and these include mutton-type (e.g., Dorper, White Dorper, and Meat Master; Cloete & Olivier, 2010), dual-purpose (e.g., Afrino, Dohne Merino, Dormer, Merino, and SA Mutton Merino; Cloete & Olivier, 2010), and indigenous (e.g., Damara, Namaqua Afrikaner, Ronderib Afrikaner, Pedi, Swazi, and Zulu; Molotsi *et al.,* 2020a) breeds. The SA Merino is one of SA's most prominent sheep breeds (Van der Merwe *et al.,* 2019), with 9 million Merino sheep in SA commercial farms (DALRRD, 2022).

The SA Merino is a medium-framed, white wool sheep breed introduced into South Africa by Spanish travellers in the nineteenth century (Granero *et al.,* 2022). Over time, Spanish, Rambouillet, Saxony, Australian, and American Merino sheep populations all contributed to

the genetic composition of the South African Merino population (Mason, 1996). The South African Stud Book Association was established in 1904 and facilitated pedigree recording and selection of sheep using phenotypes that were used for research in genetic resource flocks according to criteria for stud breeding at the time (Schoeman *et al.,* 2010; Cloete *et al.,* 2015). Thereafter, in 1956, the former Department of Agriculture and Technical Services commenced performance recording (Schoeman *et al.,* 2010). Fleece testing became a requirement when the National Fleece Testing Centre was established at Grootfontein Experimental Farm in 1965 along with the establishment of the Wool Testing Bureau in Port Elizabeth (Cloete & Olivier, 2010; Cloete *et al.,* 2015). With the influx of data, the need for an integrated pedigree and data recording system was fulfilled by the South African National Small Stock Improvement Scheme (NSSIS), which was established in 1964 and officially launched in 1990 (Cloete & Olivier, 2010; Schoeman *et al.,* 2010; Cloete *et al.,* 2015). Breeders actively participate in animal and performance recording using databases such as Logix Small Stock and the Integrated Registration and Genetic Information System (INTERGIS), which are facilitated by SA Stud Book and the Agricultural Research Council (ARC), respectively (Schoeman *et al.*, 2010).

The general aim of small ruminant animal recording facilities is to improve the total income per animal by placing selection pressure on increased reproduction rates, shorter production cycles, and optimized fiber traits (Cloete & Olivier, 2010). Animal recording facilities enabled improvements in growth, reproduction, and wool traits, demonstrating that there was sufficient genetic variation within the resource flocks to make genetic progress and improve subsequent generations (Schoeman *et al.,* 2010). Traditional genetic gain relied on phenotypic data and pedigrees; however, genetic evaluations and selection decisions have since been based on conventional best linear unbiased prediction (BLUP) based estimated breeding values (EBVs) incorporated into selection indices (Goddard*,* 2012).

Since the discovery of molecular markers, the types, and methodologies of their typing and analyses have evolved over the decades. The marker(s) of choice evolved over time from PCR-based protein markers in the 1990s (Williams *et al.*, 1990) to microsatellites in the late 1980s (Rebelato & Caetano, 2018) but the widespread use of DNA-based technology in livestock only began in the 2000s (Blasco & Toro, 2014). In conjunction with the introduction and

advancements in whole-genome sequencing (WGS) technologies, single nucleotide polymorphisms (SNPs) started to replace microsatellites as the marker of choice from 2009 onwards (Goddard & Hayes, 2007; Kijas *et al.*, 2009). The discovery of SNPs allowed for automated and more affordable genotyping of thousands of markers distributed evenly across the genome (Magee *et al.,* 2010; Dodds *et al.,* 2018).

In 2010, the first fully sequenced reference genome for sheep (Ovar Ari v1.0), based on two unrelated Texel sheep and assembled using short-read WGS technologies available at the time, was released (Archibald *et al.,* 2010). Owing to the introduction and evolvement of long-read WGS technologies in the decade that followed, an improved ovine reference genome assembly, *ARS-UI_Ramb_v2.0.*, was released in 2022, which allowed for in-depth functional annotation of the sheep genome (Davenport *et al.,* 2022). The initial reference genome assembly, as well as its progressively updated versions, allowed Illumina® in collaboration with the International Sheep Genomics Consortium (ISGC) to develop the first commercial SNP-based genotyping panel for sheep, namely the Illumina® 50K SNP Beadchip in 2010 (Archibald *et al.,* 2010). Higher-density panels such as the Illumina® *Ovine* SNP 600k Beadchip and the Affymetrix Ovine 600K HD SNP array (Anderson, 2014; Wang *et al.*, 2020), have since become available for higher-resolution genome-wide studies.

The availability of high-density SNP data has made it possible to estimate, amongst others, diversity parameters as well as inbreeding and autozygosity levels (Mastrangelo *et al.,* 2018; Nel *et al.,* 2022a) in sheep (e.g., Ferenčaković, 2015; Nosrati *et al.,* 2021) and other livestock populations (e.g. cattle, Lashmar *et al.*, 2022). Factors that may potentially influence autozygosity levels include consanguineous mating in the distant or recent past, utilization of the same sires across flocks, historic population bottleneck events, and within-breed selection pressure (Kijas *et al.,* 2012; Purfield *et al.,* 2012; Szmatoła *et al.,* 2019; Dzomba *et al.,* 2021).

Several studies have shed light on the potential negative effects of inbreeding within populations. Inbreeding may result in a decrease in functionality, performance, and profitability resulting in production losses, as reported for various breeds across continents (Rebelato & Caetano, 2018; Nosrati *et al.*, 2021). Inbreeding may also result in a reduction in the population mean for traits closely related to fitness and reproductive performance as well

as a loss in genetic variability from a reduced effective population size (Ne), which has been reported in various studies (Charlesworth & Willis, 2009; Leroy, 2014; Rebelato & Caetano, 2018; Xu *et al.*, 2019). Characterizing and monitoring autozygosity will play a crucial role in conserving population genetic diversity, as well as maintaining the future sustainability of animal breeding programs (Ferenčaković, 2015; Zavarez *et al.,* 2015a; Peripolli *et al.,* 2018; Xu *et al.,* 2019).

1.2 Aim of the study

This study aimed to perform a homozygosity-focused single nucleotide polymorphism (SNP) investigation of the past and present inbreeding and autozygosity levels in the South African Merino sheep breed.

The aim was achieved by attaining the following objectives:

1. To quantify the genetic variability present within the SA Merino breed through the estimation of several genomic diversity indicators, including the gain or loss of heterozygosity.
2. To characterize the runs of homozygosity (ROH) and heterozygosity-rich-region (HRR) profiles of the SA Merino population.
3. To estimate the level of inbreeding in the SA Merino population, by calculating both SNP-based (F_{SNP}), and ROH-based (F_{ROH}) inbreeding coefficients.
4. To investigate within-population selection signatures, using a conserved-ROH approach.

Chapter 2: Literature Review

2.1 Introduction

The small frame, docile nature, and versatility of its products (i.e., meat, wool, and pelt) contributed to sheep being one of the first farm animal species to be domesticated (Mazinani & Rude, 2020). Initially, sheep were kept for their milk and meat (Mazinani & Rude, 2020), and later human-mediated selection occurred for wool (Kijas *et al.,* 2012). The domestication of sheep ultimately led to a change in behavioural and morphological traits, resulting in a decrease in the size of their horns, body, and brain, a change in coat colouration, and an increased breeding period (Zohary *et al.,* 1998).

The domestication process, followed by intense directional selection, greatly affected the genetic diversity of global sheep populations. Sheep founder flocks were reproductively isolated from their original wild counterparts (Zohary *et al.,* 1998), which resulted in reduced effective population sizes (Ne) (Rebelato & Caetano, 2018) and, subsequently, a loss of founder alleles. During the initial Merino breed development in SA, breed purity needed to be maintained which restricted the gene flow between the diverging breeds (Nel *et al.,* 2022b) and contributed to a reduction in genetic variability (Xu *et al.,* 2019). A high number of related mating within Merino flocks later caused production losses (Zavarez *et al.,* 2015b; Rebelato & Caetano, 2018). The existing genetic diversity in Merino and Merino-derived sheep breeds is essential for evolution and breeding as it is the catalyst for natural and artificial selection (Erhardt & Weimann, 2007; Qanbari & Simianer, 2014), which fuels genetic progress within breeds.

Intense directional selection, amongst other factors (e.g. genetic drift, related mating, population bottlenecks, and natural selection) resulted in changes in autozygosity levels (Ferenčaković, 2015; Xu *et al.,* 2019). Autozygosity refers to homozygosity in which two alleles are identical by descent (IBD) (Howrigan *et al.,* 2011; Xu *et al.,* 2019). To maintain genetic progress whilst conserving diversity, breeding programs need to consider autozygosity levels and control consanguineous mating (Rebelato & Caetano, 2018). Previously, pedigree

information was used to measure inbreeding, based on the statistical probability that an individual has inherited alleles identical by descent (Wright, 1978, Rebelato & Caetano, 2018).

The advent of genome-wide SNP information has allowed genomic estimates of inbreeding to replace pedigree-based estimates as the statistic of choice. At present, runs of homozygosity (ROH) are widely used as a measure of inbreeding by estimating the level of autozygosity in the genome (Kim *et al.,* 2013, 2016; Ferenčaković, 2015), which contributes to understanding inbreeding depression of traits (Xu *et al.,* 2019). Other statistical parameters used in the estimation of homozygosity, inbreeding levels, and genome-wide genetic diversity in sheep and other livestock populations are inbreeding coefficients (e.g. the F-statistics F_{SNP}, and F_{ROH}), effective population size (Ne), linkage disequilibrium (LD), as well as observed and expected heterozygosity (H_O and H_E, respectively). By investigating these statistical parameters, breeding programs can be adapted to maintain diversity for selection while at the same time avoiding high consanguinity rates to avoid production losses from inbreeding depression (Rebelato & Caetano, 2018).

This literature review aimed to provide an overview of inbreeding, and more specifically autozygosity, as well as its causes and effects with specific reference to sheep populations, focussing on the SA Merino breed.

2.2 Overview of the South African sheep production environment and industry

South Africa has an area of 63.4 million hectares of available farming land, which is utilized for animal production (DALRRD, 2022). Sheep are predominantly farmed in extensive production systems in the dry western and central regions of the country (Cloete & Olivier, 2010). This is because more than 80% of farmed land is situated in semi-arid and arid climatic regions (Ramsay *et al.,* 2021) and is mostly suitable for extensive livestock utilization (Livestock Development Strategy for South Africa, 2006). The carrying capacity in SA varies depending on the region and can range between 1-2 ha/Large stock unit (LSU) in western and central areas of SA to 35-40 ha/LSU in north-eastern parts of SA (1 LSU = 6 Small stock units,

SSU) (Cloete *et al.,* 2014). Although intensive sheep production systems do exist in SA (Cloete & Olivier, 2010), they form a small part of the sheep industry (12.1%).

The global humidity index (HI) classifies South Africa as a semi-arid to arid country (South African Weather Services, 2022). The climate ranges from desert and semi-desert in the dry north-western regions to sub-humid and wet along the eastern coastal areas (Benhin, 2006). South Africa is generally warm, consisting of sunny days and cool nights (Benhin, 2006). The mean annual temperature in South Africa is 18.23°C (South African Weather Services, 2022). Temperatures in SA depend largely on elevation and distance from the sea; the range of temperatures in different areas can be seen in Figure 2.1 which shows the Köppen-Geiger climate type map as well as the average summer and winter temperatures in SA (Beck *et al.,* 2018). The average annual rainfall in SA is 450mm (South African Weather Services, 2022), and only 10% of the country receives more than 750mm of annual precipitation (Benhin, 2006).

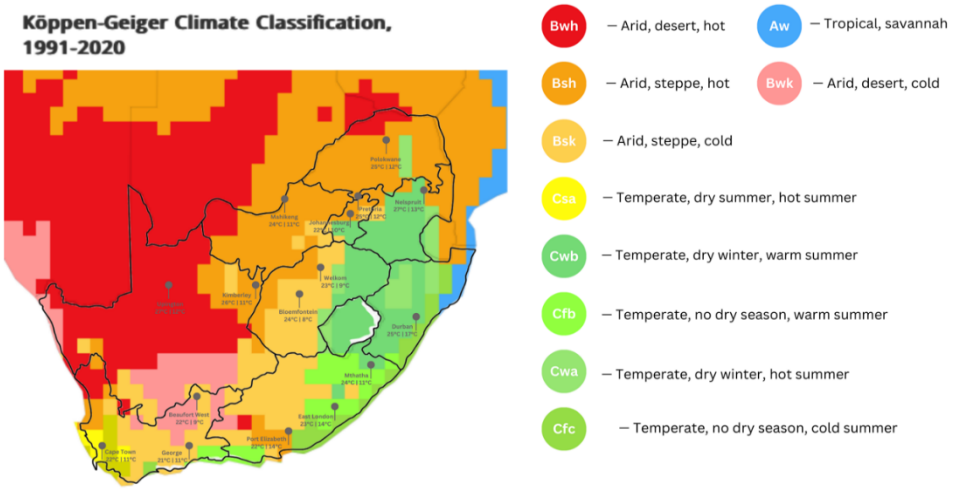


Figure 2.1 Köppen-Geiger map of the climatic regions of South Africa (Beck *et al.,* 2018).

In 2021, the estimated number of commercial sheep in SA was approximately 18.7 million, of which 9.8 million were Merino-type sheep (DALRRD, 2022). The SA sheep industry consists of 20 recognised breeds with various breeding objectives and production systems (DALRRD, 2006). In South Africa, the most prevalent breeds farmed for mutton and dual purposes are the South African Mutton Merino and the Dohne Merino, respectively (National Agricultural

Research and Development Strategy, 2008; Dzomba, 2021). Non-descript breeds are also farmed by smallholders for multi-purpose goals, as well as Karakul-derived breeds for pelt (Qwabe *et al.,* 2012; Malesa, 2015; Molotsi *et al.,* 2020). Sheep are farmed throughout all nine provinces of SA; however, the highest percentage of sheep can be found in the Eastern Cape (6 417 000), followed by the Northern Cape (5 177 000) and the Free State (4 262 000) (DALRRD, 2023). Figure 2.2 is a visual representation of the spread of sheep farming across South Africa.

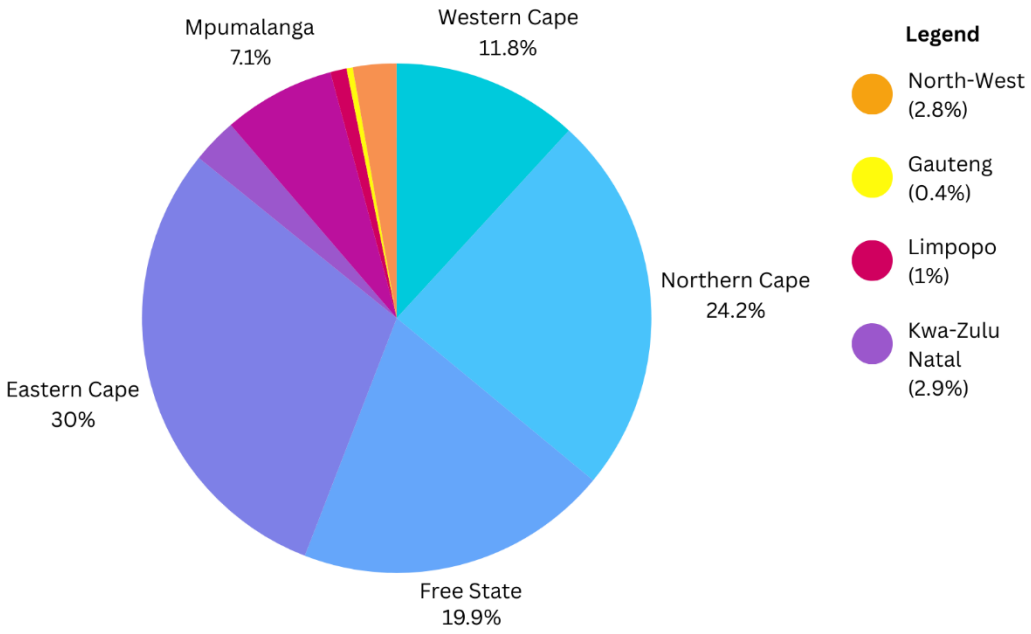


Figure 2.2. A pie chart representing the number of sheep located in each province across South Africa.

South Africa earned a total income of R417.1 billion from agriculture and related services in 2021, of which animals and animal products generated the largest sales within the agricultural sector amounting to R151.93 billion (Statistics SA, 2021). In 2021, sheep products contributed 11.6% of total agricultural sales in this sector (Statistics SA, 2021). The bulk of this value was derived from mutton and lamb (60.6%) followed by wool (31.4%), and lastly karakul pelts (0.2%) (DALRRD, 2022). Commercial sheep numbers in South Africa have declined from 29.9 million (29 979 000) in 1990 to 18.7 million in 2021 (DALRRD, 2022). There are several reasons to explain this decline including unstable product prices, stock theft, predators, and increased human activities, such as mining, in farm areas (Vink and Kirsten, 2002; Herselman, 2006; Bila *et al*., 2022). Along with the decline in commercial sheep numbers over the last three decades,

the per capita consumption of sheep, lamb, and goats in SA has also declined from 5.8kg in 1990 to 2.8kg in 2021 (DALRRD, 2022). The demand for mutton depends on various economic and non-economic factors (Taljaard *et al.,* 2003). Economic factors include income and price whereas non-economic factors include issues concerning consumer preference, convenience, quality, animal welfare, the environment, health, and safety (Taljaard *et al.,* 2003; Delport *et al.,* 2017).

Despite the steady decline in sheep numbers and per capita consumption, as well as the diminishing contribution of the sheep industry to the national economy (Delport *et al.,* 2017), the South African sheep industry is still of importance to the livelihoods of a large percentage of the population (Cloete & Olivier, 2010). The hardiness and ability of sheep to produce in extreme conditions are the main motivating factors for sheep farming in South Africa (Sandenbergh, 2015). Sheep farming allows for sustainable production in extensive pastoral areas, such as the Karoo (Benhin, 2006), where no alternative farming systems are practiced (Dzomba, 2021). In addition, sheep farming complements the crop industry by utilizing crop residues and by-products of crop production (Cloete & Olivier, 2010). Many rural communities depend on informal, communal sheep farming for food security, as well as socio-economic and cultural needs (Rust & Rust, 2013; van Marle-Kőster *et al.,* 2015; Dzomba, 2021).

2.3 The history, origin, and distribution of the SA Merino breed

The Merino was introduced to South Africa in the eighteenth century by Spanish travellers who brought two Spanish Merino rams and four ewes that were donated to the Dutch government (Granero *et al.,* 2022). From 1891 onwards, American Vermont, Australian Wanganella, and Peppin Merinos were introduced into SA (Mason, 1996). The SA Merino breed has undergone extensive natural and artificial selection to be considered a locally developed South African breed today (Dzomba *et al.,* 2021). The modern South African Merino breed is medium-framed, white, and classified as a dual-purpose (wool-meat type) sheep.

The SA Merino has become a major dual-purpose breed in South Africa, owing to their hardiness and ability to maintain productivity under extreme conditions (Schoeman *et al.,* 2010; Cloete *et al.,* 2014; Sandenbergh *et al.,* 2015). Dual-purpose breeds such as the Merino are also popular since these production systems may produce a greater income than farming solely with single-purpose breeds (Louw, 2019). South African Merino genetic resources have been introduced and imported into other countries globally, and due to ancestral relationships and using sires of international origin, the SA Merino breed has strong genetic links with international Merino populations (Dzomba *et al.,* 2021; Nel *et al.,* 2022a).

South Africa is one of the few countries in Africa with distinct sheep breeds and organized breed improvement programs (Snyman, 2014; Dzomba *et al*., 2021). Sheep farmers that breed with stud and commercial sheep breeds are represented in established breeders' societies, of which Merino SA is one of the most prominent (Dzomba *et al.,* 2021). Merino SA, established in 1937, is the official organization for the management, representation, and communication of the Merino breed and its derivatives in South Africa and has represented the industry for more than eighty years. The organization is, furthermore, affiliated internationally and is a member of the World Federation of Merino Breeders (van Heerden, 2022).

Research conducted on designated resource flocks contributed to the breed improvement of sheep in SA (Schoeman *et al.,* 2010). There are five Merino resource flocks namely, Tygerhoek, Elsenburg, and Langgewens which are in the Western Cape, as well as Klerefontein, and the Cradock fine-wool flock, which are located in the Northern Cape and Eastern Cape provinces, respectively (Table 2.1; Schoeman *et al.,* 2010). Research that was conducted on these flocks focused on several different breeding objectives (and, hence, selection criteria) specific to the Merino namely, 1) increased fleece weight solely, 2) improved conformation (Erasmus *et al.,* 1990; Snyman *et al.,* 1996a, 1998), and 3) increasing fleece weight while simultaneously keeping fiber diameter constant (Heydenrych *et al.,* 1984; Cloete *et al.,* 1998). From 1986 to 2010, breeding objectives for the Elsenburg and Grootfontein Merino flocks involved divergent selection for reproduction focusing on the number of lambs weaned and lamb survival (Cloete *et al.,* 2004). A reduction in fiber diameter was the main selection objective between 1998 and 2010 (Cloete *et al.,* 2001). Table 2.1 provides detailed information and a

timeline for the selection objectives, selection lines, breeds, and research flocks used in the selection experiments. The breeding objectives set for the resource flocks were achieved using phenotypic selection based on pedigree records (Matebesi-Ranthimo *et al.,* 2018; Bonato *et al.,* 2021).

Table 2.1 A summary and timeline for the selection objectives, selection lines, breed, and research flocks used in various selection experiments performed on the South African resource flocks (Schoeman *et al.,* 2010)

Flock	Breed	Time span	Location (Province)	Selection Objectives	Selection lines/types	References
Klerefontein Merino Flock	Merino	1962-1983	Northern Cape	Increasing Fleece weight and Improving conformation	1. Control line 2. Fleece weight line 3. Visual appraisal line	*(Erasmus et al., 1990)* *(Snyman et al., 1996b)* *(Snyman et al., 1998)*
Koopmansfontein flock	Dorper	1966-1982	Northern Cape	Selection for growth under different scenarios	1. Weaning weight 2. Weaning weight in ewes and post- weaning feedlot gain in rams 3. Subjective selection	*(Neser et al., 1995)*
Tygerhoek flock	Merino	1969-2010 1995-2016	Western Cape	Increasing fleece weight without changing fibre diameter Faecal worm egg count objectively measured in wool traits	1. Control Line 2. Clean fleece weight line 3. S/P line	*(Heydenrych et al., 1984)* *(Cloete et al., 1998)* *(Matebesi-Ranthimo et al., 2014; Mpetile, 2019)*
Upington flock	Karakul	1970-2010	Northern Cape	Improving pelt quality	1. Control line 2. Hair length line 3. Pattern line 4. Hair quality line 5. Curl development line	*(Greeff et al., 1991)*
Klerefontein Namaqua flock	Namaqua Afrikaner	1982-2010	Northern Cape	Conservation of indigenous fat-tailed breeds	1. Live weights and reproduction traits recorded	*(Snyman et al., 1993)*
Elsenburg flock	Merino	1986-2010 2010 - 2019	Western Cape	Divergent selection for reproduction (number of lambs weaned) Lamb survival following long term selection for number of lambs weaned	1. H line (Selected for) 2. L line (selected against)	*(Cloete et al., 2004)* *(Nel et al., 2021) (Burger, 2019)*
Jansenville fine-mohair herd	Angora	1988-2010	Eastern Cape	Selection for reduced fibre diameter	1. Control line 2. Fine-mohair line	*(Snyman, 2001)*
Tygerhoek fine-wool flock	Merino	1998-2010	Western Cape	A reduced fibre diameter	1. Control line 2. Fine-wool line	*(Cloete et al., 2001)*
Klerefontein Dorper flock	Dorper	1993-2000	Northern Cape	Comparison of lines within Dorper breed	1. Hairy type 2. Woolly type	*(Snyman & Olivier, 2002)*
Grootfontein Merino Stud		1986-2022		Number of lambs weaned, conception rate, litter size and ewe rearing ability	1. Control line 2. Number of lambs weaned line 3. Litter size line	*(Oliver et al., 1995; Snyman et al., 1998)*
Cradock fine wool Merino stud	Merino	1986-2022	Eastern Cape	Number of lambs weaned, conception rate, litter size and ewe rearing ability	1. Control Line 2. Fine-wool line	*(Schoeman et al., 2010; Oliver, 2014; Nel et al., 2022; Nel et al., 2023)*
		2002-2014		Increase birth weight, maintain fleece weight & reduce fibre diameter		*(Oliver et al., 2014)*
Langgewans Merino flock	Merino	1988-2022	Western Cape	Number of lambs weaned, conception rate, litter size and ewe rearing ability	1. Reproduction '+' line (selected for) 2. Reproduction '-' line (selected against)	*(Cloete et al., 2001)*
Elsenburg flock	Dormer	1943-2019 2008-2018	Western Cape	Body weight, weaning weight, lamb survival Yearling weight, clean yield, clean fleece weight, staple length, staple strength and fibre diameter	1. H line (Selected for) 2. L line (selected against)	*(Cloete et al., 2004; Muller, 2020)*

2.4 Traditional tools for management and selection

Traditionally, genetic diversity within and between breeds was estimated by comparing phenotypic data (e.g., performance and pedigree information) between different individuals in a breed (Sabir *et al.,* 2014; Woolliams & Oldenbroek, 2017). The presence of phenotypic variation is indicative of potential genetic variation that can be used to change the mean trait values in a population (Dekkers & Hospital, 2002).

Pedigree information can be used to determine inbreeding and relatedness by using identical-by-descent (IBD) information (Ferenčaković, 2015). Crow (1954) defined IBD as two haplotypes that have been inherited from a common paternal or maternal ancestral allele that has not undergone any mutation or recombination. Therefore, inbreeding calculated from pedigree information is based on correlations and is determined by the pedigree pattern (Ferenčaković, 2015). However, in practice, comprehensive inbreeding estimation using IBD values is near-impossible as not all IBD segments can be identified for every shared ancestor especially over a long period, i.e. dependent on pedigree depth (Cockerham & Weir, 1968). Pedigree-based inbreeding can be estimated per individual through the calculation of Wright's (1922) inbreeding coefficient (*F*), which relies on the availability of deep pedigree records, as well as genetic relationships with ancestors (Biscarini *et al.,* 2020).

Limitations of pedigree-based inbreeding estimation include the overdependence on the precision (accuracy) and consistency (depth) of on-farm pedigree recording, which often influences the reliability of these estimators through the introduction of human error (Ablondi *et al.,* 2022). The improvements in the availability of genomic data for larger numbers of animals per population, however, allowed for more accurate, DNA-level estimations of inbreeding and relatedness, with pedigree information being supplementary (Saif-ur-Rehman *et al.*, 2023).

2.5 The development of molecular technology

Molecular markers are single-site or stretches of deoxyribonucleic acid (DNA) that are associated with parts of a genome presenting different detectable variants (Toro *et al.,* 2009). Since the inception of molecular markers, they have evolved over the past four decades from blood-type markers to polymerase chain reaction (PCR) based markers, such as microsatellite markers (Vekemans *et al.,* 2002; Bruford *et al.,* 2003; Caballero *et al.,* 2008) to single nucleotide polymorphisms (SNPs) (Goddard & Hayes, 2007). Traditional phenotype-based selection could thus be complemented by including information based on molecular markers and advanced statistical analyses (Toro *et al.,* 2009). Therefore, the discovery of these molecular markers allowed for genome-level investigations of population genetic characterization, including, for example, within-breed diversity.

In livestock genetics studies, single nucleotide polymorphisms have for the most part replaced microsatellite markers as the marker of choice. As markers, SNPs are highly abundant, bi-allelic point mutations distributed across the genome and are located in coding, non-coding, and regulating regions (Vignal *et al.,* 2002). An advantage of SNPs is that due to their whole-genome coverage, they offer the potential to detect both neutral and rare genetic variations (Toro *et al.,* 2009). The availability of genome-wide SNP marker information has made it possible to comprehensively characterize the genetic diversity of livestock populations including sheep (Nel *et al.,* 2022a). Additionally, SNPs have allowed for the extrapolation of breed history, evolution, and population genetic structure through parameters such as effective population size and inbreeding coefficients (Williams *et al.,* 2016). Table 2.2 lists a few studies focusing on the characterization of Merino sheep populations worldwide using various marker types. Initially, the cost of SNP genotyping panels was a limiting factor in their use (Van Tassell *et al.,* 2008), however, the availability of low-density panels and improvements in sequencing and genotyping technologies have reduced the per-SNP cost of genotyping for all livestock species (van Marle-Köster *et al.*, 2015). Another drawback in the utilization of SNP-based genotyping is the possible introduction of ascertainment bias, which occurs when small sample sizes, not fully representative of all the breeds within a species, are used in the discovery phase during SNP panel development (Albrechtsen *et al.*, 2010). This

bias in SNP discovery may lead to reduced sensitivity for detecting genome-wide genetic diversity levels (Dzomba, 2021).

Table 2.2 A non-comprehensive summary of genomic studies on the characterization of Merino sheep populations worldwide

Breed	Article title	Markers used	Reference
20 SA breeds (Including SA Merino)	Genetic characterization of South African Sheep breeds using DNA markers	31 Microsatellite markers	(Buduram, 2004)
29 European breeds (including Merino)	Genetic structure of European sheep breeds	23 Microsatellite markers	(Handley *et al.,* 2007)
5 Australian breeds (including Merino)	Genome-wide linkage equilibrium and genetic diversity in five populations of Australian domestic sheep.	Illumina® OvineSNP50 BeadChip	(Al-mamum *et al.,* 2015)
SA Merino	Identification of SNPs associated with robustness and greater reproductive success in the South African Merino sheep using SNP chip technology	Illumina® OvineSNP50 BeadChip	(Sadenbergh, 2015)
37 Merino or Merino-derived breeds	Merino and Merino-derived sheep breeds: a genome-wide intercontinental study	Illumina® OvineSNP50 BeadChip	(Ciani *et al.,* 2015)
11 Merino or Merino-derived breeds	A combined multi-cohort approach reveals novel and known genome-wide selection signatures for wool traits in Merino and Merino-derived sheep breeds	Illumina® OvineSNP50 BeadChip	(Megdiche *et al.,* 2019)
14 SA breeds (including SA Merino)	Runs of homozygosity analysis of South African sheep breeds from various production systems investigated using OvineSNP50k data	Illumina® OvineSNP50 BeadChip	(Dzomba *et al.,* 2021)
63 global breeds (including Merino)	The pattern of runs of homozygosity and genomic inbreeding in worldwide sheep populations	Illumina® OvineSNP50 BeadChip	(Nosarati *et al.,* 2021)
14 SA and Australian sheep breeds	The genomic structure of isolation across breed, country, and strain for important South African and Australian sheep populations	Illumina® OvineSNP50 BeadChip	(Nel *et al.,* 2022)
403 Spanish Merino sheep	Genomic population structure of the main historical genetic lines of Spanish Merino sheep	Axiom™ Ovine 50K SNP Genotyping array	(Granero *et al.,* 2022)
10 SA Mutton Merino, 39 Australian and Chinese Merino	Detection of selection signatures in South African Mutton Merino sheep using whole-genome sequencing data	Whole Genome sequencing (WGS)	(Liu *et al.,* 2022)
42 Merino and Merino-derived sheep breeds	A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds	Illumina® OvineSNP50 BeadChip	(Ceccobelli *et al.,* 2023)

2.6 Genomic parameters for the estimation of autozygosity levels in sheep populations

Genetic diversity and the maintenance thereof play a crucial role in populations as it allows for adaptive flexibility within a population (Ovaska *et al.*, 2021). In populations with little genetic variability or where multiple alleles are no longer present, or are present in low frequencies, adaption to a shift in selection pressure may be more difficult (Oldenbroek & van der Waaii, 2015). In addition, inbreeding (i.e., increased homozygosity levels), results in an increase in the frequency and/or fixation of alleles with deleterious effects causing recessive disorders to become expressed (Oldenbroek & van der Waaii, 2015). Maintaining genetic diversity can be achieved by maximizing the effective population size (Ne) or by minimizing the consequences of inbreeding (de Cara *et al.,* 2011; Gómez-Romano *et al.,* 2016).

The quantitative effects of inbreeding can be measured by assessing the autozygosity levels and inbreeding depression of a population (Lashmar *et al.,* 2018). Intense directional selection over several generations, the overuse of a small number of superior sires, and the use of genomic selection practices (through a reduction in the generation interval) may contribute to an increase in the occurrence of inbreeding-related genetic consequences (Nicholas & Smith, 1983; Verrier *et al.,* 1993; Robertson, 2008). Changes in autozygosity levels within a population impact several aspects, for instance, the redistribution of genetic variation within and between populations (Fernández *et al.,* 1995), with potentially decreasing within-population genetic diversity in finite populations (Ferenčaković, 2015). Inbreeding depression affects additive genetics by reducing the population mean for traits closely related to fitness (Charlesworth & Willis, 2009), for example, Selvaggi *et al.* (2010) reported a 20.47% reduction in inbreeding coefficients in in-bred sheep fertility traits compared to non-inbred sheep. Inbreeding depression causes a higher incidence of homozygous recessive defects caused by single genes (Arcos-Burgos & Muenke, 2002; Alvarez *et al.,* 2009), for example, the occurrence of spider-lamb syndrome (Analla *et al.*, 1998) and consequently a decrease in homeostasis in a population (Lerner, 1954). Other impacts of autozygosity include changes in the effective population size and population structure, genetic drift, and deviations from Hardy-Weinberg equilibrium (HWE) (Ferenčaković, 2015).

At present, runs of homozygosity (ROH) are widely used to estimate the level of autozygosity in the genome (Kim *et al.,* 2016; Peripolli *et al.,* 2018). Other statistical parameters currently used for the estimation of homozygosity levels in livestock populations include pedigree-based (F_{PED}) (which will not be discussed in depth in this review) and SNP- and/or ROH-based (F_{SNP} & F_{ROH}) F-statistics, respectively. Proxy indicators of diversity include linkage disequilibrium (LD), observed and expected heterozygosity (H_O and H_E), and effective population size (Ne).

2.6.1 Proxy-indicators of inbreeding

2.6.1.1 Minor Allele Frequency (MAF)

Minor allele frequency (MAF) refers to the frequency at which the less common allele occurs in a population (Keller *et al.,* 2011), and relates to genetic diversity. A higher MAF explains a higher level of polymorphic markers, which corresponds to higher levels of H_E and consequently indicates a greater genetic diversity (Engelsma *et al.,* 2014). The opposite is also true, a low MAF indicates a lower genetic diversity and hence more frequent fixed alleles (Engelsma *et al.,* 2014) and a higher probability of homozygous genotypes (for the fixed allele).

2.6.1.2 Observed and Expected heterozygosity (H_O and H_E)

Mean expected heterozygosity (H_E) and observed heterozygosity (H_O) are both measures of genetic diversity that are estimated from molecular information (marker-based summary statistics) (Eusebi *et al.,* 2020). The H_E parameter is defined as the probability that an individual's two alleles chosen randomly from the population are different at a specific locus (Nei*,* 1973; Mäki-Tanila *et al.,* 2010). The H_O parameter is defined as a measure of the frequency of heterozygous genotypes in a set of markers (Mäki-Tanila *et al.,* 2010). The H_O value is negatively correlated with inbreeding depression (Mäki-Tanila *et al.,* 2010). In nature one cannot assume that a finite population is in HWE, therefore, the relationship between H_E

and H_O reveals a gain or loss of genetic diversity, which translates to the estimation of a population's inbreeding coefficient (Eusebi *et al.,* 2020). A population with high H_E values is assumed to be more genetically diverse than populations with lower H_E values and thus are capable of adapting to changes and responding to selection pressures (Allendorf *et al.,* 2013). Therefore, genotypes showing low observed heterozygosity may indicate inbreeding (Anderson *et al.,* 2010).

2.6.1.3 Linkage disequilibrium (LD)

Linkage disequilibrium (LD) is a genetic diversity parameter that shows the extent of non-random association between two alleles located at different loci within a population (Al-Mamun *et al.,* 2015; Dzomba, 2021). Linkage disequilibrium may be the result of several different genetic forces such as selection, mutation, genetic drift, non-random mating, as well as non-genetic causes (Sabatti & Risch, 2002). The degree of LD within a population is reflective of breeding systems, population history, and the pattern of geographical subdivision (Goddard, 2012).

The degree of LD within a population can be measured by the estimation of |D'|, which in small sample sizes, and the presence of rare alleles, is overestimated (Bohmanova *et al.,* 2010). Alternatively, LD can be estimated using r^2, which is the correlation between two loci and is mostly used in association studies (Bohmanova *et al.,* 2010). The r^2 measure depends on the time and rate of recombination, as well as the population size (Hill, 1974). Due to the correlation between LD and Ne, to avoid LD estimation biases from small sample sizes, a minimum sample size of 55 is required for accurate estimation of r^2 (Bohmanova *et al.,* 2010). The r^2 measure's formula, as proposed by Hill and Robertson (1968), is as follows:

$$r^2_{(P_a,P_b,P_{ab})} = \left(\frac{(P_{ab} - P_aP_b)^2}{P_a(1-P_a)P_b(1-P_b)}\right),$$

where P_{ab} represents the frequency of the haplotype consisting of 2 SNPs; P_a and P_b represent the frequency of allele *a* at the first locus and allele *b* at the second locus, respectively.

Linkage disequilibrium directly reflects the recombination rate; regions associated with low recombination rates will have a higher extent of LD and the opposite is also true (Zavattari, 2000). Inter-SNP distance is inversely proportional to the level of LD, i.e. SNPs that are near one another are often inherited as a unit and are usually found in high LD (Zhai *et al.*, 2004). Sheep typically exhibit low genome-wide LD due to their origin from a genetically diverse pool (Kijas *et al.*, 2012), i.e. smaller haplotypes are shared within the population because of increased recombination of large haplotypes in the past. Higher LD captured through dense SNP panels allows for increased chances that markers are in LD with quantitative trait loci (QTL) and, therefore, the range of LD affects the power and precision of mapping (Goddard & Hayes, 2007). High levels of LD and more relatedness result in higher levels of specific allelic combinations which are inherited over generations than what can be expected by chance (Eusebi *et al.,* 2020). Therefore, livestock species that make use of genomics-driven selection, indirectly rely on the level of LD within the population (Meadows *et al.,* 2006; Kijas *et al.,* 2012).

2.6.1.4 Effective population size (Ne)

The effective population size (Ne) of an existing population can be defined as the ideal population size that will result in the same amount of genetic diversity as that which is present in the current existing population (Wright, 1974; Gholizadeh & Ghafouri-Kesbi, 2016). It is possible that Ne can be a smaller number than the actual population and this is due to the disproportionate contribution of genes from individuals to the next generation (Goddard & Hayes, 2007). The Ne parameter determines a population's performance based on genetic variation and inbreeding over long periods (Fernández *et al.,* 2005). In the past, Ne has been estimated using pedigree information (Falconer and Mackay, 1996) and this is still used in many cases where there is no or insufficient genomic information available. More recently, SNP data or SNP-based methods can estimate both historical and recent Ne, but these

methods are limited by the extent of LD captured throughout the genome i.e., the SNP panel density and the number of animals genotyped (Goddard & Hayes, 2007; Visser *et al.,* 2023). As the distance between markers increases, the LD among them decreases (Liu *et al.,* 2017), which allows for the estimation of ancient effective population sizes (Ne) using LD at short distances and the evaluation of recent Ne using LD at longer distances (Hayes *et al.,* 2003).

The SNeP software tool, for example, uses genome-wide SNP data to estimate recent Ne trends as proposed by Barbato *et al.* (2015), with the following formula employed:

$$N_{T(t)} = (4f(c_t))^{-1} (E[r^2_{adj} \mid c_t]^{-1} - \propto)$$

where, $N_{T(t)}$ is the effective population size t generations ago calculated as $t = (2f(c_t)) - 1$ (Hayes *et al*., 2003); c_t is the recombination rate for a specific physical distance between SNPs estimated using Sved and Feldman approximation (Sved & Feldman, 1973), 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 (Ohta & Kimura, 1971).

Most livestock species have small effective population sizes which can mainly be attributed to the intense selection of a small number of animals for the parent generation (Leroy, 2014). As for small stock breeds, larger Ne estimates are possible due to less intense selection in certain breeds (Groeneveld *et al.,* 2010). Frankham *et al.* (2002) suggested that to avoid inbreeding depression in the short term, a Ne of at least 50 should be maintained whereas other studies have suggested a Ne of 500 should be maintained (Holt *et al.,* 2005; Meuwissen, 2009). Falconer & Mackay (1996) noted that the number of years that have passed since a common ancestor can be calculated depending on the length of the generation interval (*L*) as the rate of recombination is proportional to the number of generations.

2.6.1.5 Runs of Homozygosity (ROH) and Heterozygosity-rich areas (HRR)

Runs of homozygosity (ROH) are long stretches of homozygous regions throughout the genome that reflect autozygosity due to consanguineous mating (Ferenčaković, 2015).

Broman & Weber (1999) were the first to recognize that these long stretches of homozygous segments in a human population most likely reflect autozygosity and may have major consequences for human health (Ku *et al.,* 2011; Ferenčaković, 2015). Thereafter, the concept was extrapolated to various livestock species as illustrated in Figure 2.3.

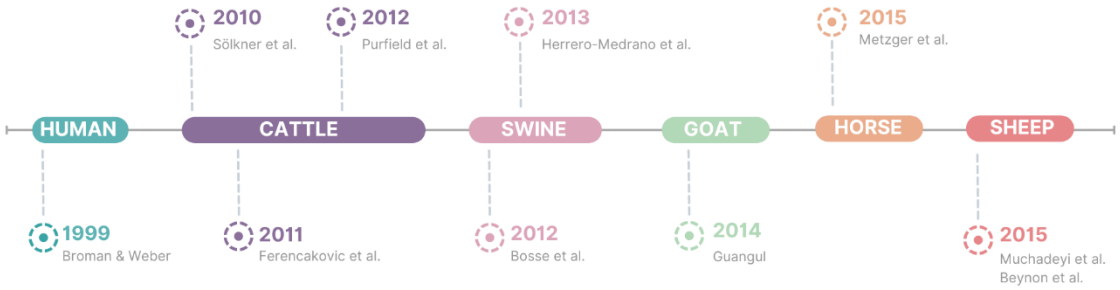


Figure 2.3. Timeline showing the extrapolation of the ROH concept from humans to different livestock animals.

The availability of dense SNP panels has allowed for determining the genome-wide ROH profile of animals (Mastrangelo *et al.,* 2018). Pemberton *et al.* (2012) characterized ROH lengths in humans, showing that the length of the ROH segment is determined by recombination events i.e., the number of generations from a common ancestor. Longer segments (~ 10Mb) translate to more recent inbreeding whereas shorter segments (~ 1Mb) translate to ancestral inbreeding (Pemberton *et al.,* 2012; Ferenčaković, 2015).

Contrary to ROH, runs of heterozygosity (ROHet), more appropriately referred to as heterozygosity-rich regions (HRRs), are consecutive stretches of heterozygous SNP genotypes (Williams *et al.,* 2016). An animal's HRR profile provides information regarding a population's evolutionary history and diversity (Samuels *et al.,* 2016). A population undergoes a continuous process of loss and creation of heterozygous-rich areas due to the selection changes in the population (Mulim *et al.,* 2022). Biscarini *et al.* (2020) showed that HRR can be utilized to detect specific genome segments where preserving greater genetic diversity could be more advantageous. Therefore, the estimation of HRR within a population is important for the identification and maintenance of sufficient genetic diversity (Purfield *et al.,* 2012; Peripolli *et al.,* 2017).

Two popular software, namely PLINK (Purcell *et al.*, 2007) and the *detect*RUNS R package (Biscarini *et al.*, 2018), are used to calculate ROH, with *detect*RUNS additionally calculating HRR. The *detect*RUNS sliding window approach is comparable to the sliding window software in PLINK (Dixit *et al.*, 2020). This method for the detection of both ROH and HRR uses an algorithm that involves a sliding window size of SNPs that slides along the chromosome (50 as a default), and at every sliding window position, every SNP within that window is determined to be homozygous or not (Purcell *et al.*, 2007; Bjelland *et al.*, 2013). The sum for each SNP in the number of fully homozygous windows and the total number of windows is calculated (Bjelland *et al.*, 2013). An ROH will be called if a specified minimum number of consecutive SNPs are determined to have a homozygous window threshold of more than five percent (Biscarini *et al.*, 2020). In the case of HRR, the characteristics of the sliding window are used to determine if a homozygous SNP is not present in a run (Biscarini *et al.*, 2020).

The *detect*RUNS R package (Biscarini *et al.*, 2018) presents an alternative method for detecting ROH and HRR, namely the consecutive runs method (Marras *et al.*, 2015). This method lacks a specified "window" and rather directly scans the genome on a SNP-by-SNP basis (Marras *et al.*, 2015), in other words, averages over single-point estimates (Dadousis *et al.*, 2022).

2.6.2 Direct measures of inbreeding: F-statistics

The inbreeding coefficient (*F*) is a measure of an individual's level of inbreeding and is defined as the probability that two alleles are IBD relative to a base population that segregates following HWE (Wright, 1922). Several different methods to estimate inbreeding coefficients exist, depending on the data type utilized. The F_{PED} estimator is estimated using pedigree data (Malécot, 1948) whereas the F_{IS} estimator uses genomic observed and expected heterozygosity levels to estimate genetic variance within a population (Williams et al., 2016). The level of inbreeding can also be calculated by examining the marker-by-marker identical-by-state (IBS) information based on an SNP-derived genomic relationship matrix (GRM) and estimating inbreeding as F_{GRM} (Van Raden, 2008). This dissertation will only focus on the F_{SNP} and F_{ROH} inbreeding estimators.

2.6.2.1. Single nucleotide polymorphism (SNP) based F_{SNP}

The F_{SNP} inbreeding coefficient estimates the excess in homozygosity using SNP-by-SNP calculations (Rodríguez-Ramilo *et al.,* 2020). The inbreeding coefficients F_{SNP} and F_{ROH} are often highly correlated due to a large proportion of SNPs used to capture F_{SNP} also being present in ROH (Ghoreishifar *et al.,* 2023). In addition, Keller *et al.* (2011) suggested that as the effective population size decreases, the correlation between F_{ROH} and F_{SNP} increases. The formula to calculate F_{SNP} is as follows (Keller *et al.,* 2011):

$$F_{SNP} = \frac{O - E}{N - E}$$

where O is the observed number of homozygous SNPs per individual, E is the expected number of homozygous SNPs under the HWE calculated based on the estimated allele frequencies of the sample, and N is the total number of SNPs.

The level of inbreeding can also be calculated by examining IBS information on a marker-by-marker basis using an SNP-derived genomic relationship matrix (GRM) and estimating inbreeding as F_{GRM} (Van Raden, 2008), which goes beyond the scope of this study. The GRM shows the realized proportion of the genome that is shared by two individuals (Goddard *et al.,* 2011).

2.6.2.2 Runs of homozygosity (ROH) based F_{ROH}

The F_{ROH} coefficient, defined as the proportion of the autosomal genome having ROH above a specific length threshold, is a genomic measure of individual autozygosity (McQuillan *et al.*, 2008). The F_{ROH} inbreeding estimator is calculated from molecular information and, unlike F_{PED} (which is based on probability), F_{ROH} is based on observed homozygosity (Eusebi *et al.*, 2020). McQuillan *et al.* (2008) described the methodology used to calculate F_{ROH} by dividing the sum of an individual's ROH lengths by the total length of the genome that has SNP coverage. Due to this, F_{ROH} estimates are commonly higher than F_{SNP} measures because they cannot be

negative values. Pemberton *et al.* (2012) characterized ROH lengths and related the length to the time of inbreeding. Therefore, deriving inbreeding coefficients (F_{ROH}) from ROH not only provides information on the degree of inbreeding but also the "age" thereof (Ferenčaković, 2015). An advantage of ROH is the ability to distinguish between identical by descent (IBD) and identical by state (IBS), therefore showing which haplotypes are inherited and which ones are common haplotypes (Pryce *et al.*, 2012). Previous studies confirmed that F_{ROH} estimates are more accurate than pedigree-based estimates, which is advantageous because it accounts for the stochastic nature of inheritance on an individual marker level as well as a chromosome marker level (Sölkner *et al* 2010; Ferencakovic *et al.,* 2011; Purfield *et al.,* 2012). The F_{ROH} inbreeding coefficient can be calculated as described by McQuillan *et al.* (2008);

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

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

ΣL_{AUTO} = the length of the genome covered by SNPs

Aspects that affect the accuracy of ROH estimates include the varying parameters set to define an ROH between studies, which vary in both amount and distribution through the genome. The minimum length, minimum SNP density, minimum number of SNPs, and maximum distance between SNPs have the greatest effect on F_{ROH} values (Rodríguez-Ramilo *et al.,* 2020). The density of SNP panels is important because an increase in shorter lengths (<2Mb) may result in a higher rate of false positives (Zhang *et al.,* 2015), i.e. low SNP panel density results in genomic regions where it is impossible to detect short length ROHs (Meyermans *et al.*, 2020). Another factor that influences the accuracy of ROH is its association with effective population size (Ne). Populations with low Ne have more ROH present whereas the opposite is true for populations with larger Ne (Abied *et al.,* 2020). Due to the lack of uniformity of recombination rates across the genome, ROH with identical lengths located in genomic regions with differing recombination rates exposes a different inbreeding history. However, these ROHs will be considered to have equal F_{ROH} estimates (Eusebi *et al.,* 2020).

2.7 Within-population selection signatures

The genome-wide distribution and patterns of ROH are not random but may be shared amongst individuals and could, therefore, provide information on past selection events (Zhang *et al.*, 2015). Therefore, by using the conserved ROH approach, selection signatures can be identified (Kim *et al.*, 2013). Within-breed selection signatures (or selection sweeps) are characterized by genomic regions with reduced genetic variability (i.e., increased favourable allele frequencies within the genome) (Purfield *et al.,* 2017).

Selection signatures can be identified using different methods including the SNP-in-run approach and the conserved ROH approach. The SNP-in-run approach identifies genomic regions that are most associated with ROH by counting the proportion of times an SNP appears in an ROH and dividing it by the number of animals (Biscarini *et al.*, 2020). The conserved ROH approach identifies regions of reduced nucleotide diversity that have been subject to selection, otherwise known as ROH islands (Szmatola et al., 2016; Purfield et al., 2017). This study will focus on detecting conserved ROH regions that indicate genetic differentiation.

Almeida *et al.* (2019) demonstrated that functional analysis of identified genes within conserved ROH regions indicates the biological mechanisms that may be involved with performance and adaption traits. A limitation of using the conserved ROH approach is the difficulty in determining which proportion of ROH is due to genetic drift and this may result in false positives (Almeida *et al.,* 2019). However, this can be overcome by detecting selection signatures that are overlapping in regions with QTL and using ROH regions in common with at least two animals (Ruiz-Larrañaga *et al.,* 2018).

2.8 Previous autozygosity studies on sheep

Studies on the inbreeding and autozygosity levels of SA sheep breeds are scarce (van Marle-Köster *et al.,* 2021). The SA Merino breed is the only South African sheep breed with sufficient genotypes to implement genomic selection (GS), thus there is a comprehensive data set of

sheep with SNP-genotypic information available to study the genomic inbreeding and autozygosity levels in this breed. It is difficult to compare results to previously published studies on a global level, due to the different SNP panels used, the varying ROH defining criteria, and the different sample sizes between studies (Mastrangelo *et al.*, 2018). Table 2.3 shows a summary of some recent autozygosity studies in global sheep populations.

Table 2.3 A non-comprehensive summary of recent autozygosity studies on sheep

Breed[a]	**Sample size**	**SNP density**	**Parameters**[b]	**References**
CM	635	54241	r^2, Ne	Liu *et al.,* 2017
CM	635	54241	F_{ROH}	He *et al.,* 2020
NA, N, BP, SAM, DM, AM, CM, S, RA, S, MM	400	36975	MAF, F_{ST}, r^2, H_O, H_E, F_{IS}	Dzomba *et al*., 2020
SAMM, DM, MM, BP, NA, SAM, S, D, A, N	400	43556	F_{ROH}	Dzomba *et al.,* 2021
European sheep breeds (incl. MacArthur Merino)	2536	60301	F_{ROH}, H_O, H_E, F_{GRM}, F_{UNI}	Nosrati *et al.,* 2021
BP, DAR, NA, P, SAM, DM, D	319	54241	MAF, F_{ROH}, r^2, F_{ST}	van Marle-Köster *et al.,* 2021
SPM	403	56520	H_O, H_E, F_{ROH}	Granero *et al.,* 2022
AM, DM, D, SAMM, BL, CW, C, PD, WS, SAM, DR	3509	47789	F_{ST}, r^2, H_O, H_E, Ne	Nel *et al.,* 2022

[a] SAMM= South African Mutton Merino, DM= Dohne Merino, MM= Meat-master, BP= Blackhead Persian, NA= Namaqua Afrikaner, S= Swakara, SAM= South African Merino, D= Dorper, A= Afrino, N= non-descript Nguni sheep, AM= Australian Merino, BL= Border Leicester, CW= Coopworth, C= Corriedale, PD=. Poll Dorset, WS= White Suffolk, DR= Dormer, CM= Chinese Merino, RA= Ronderib Afrikaner, RP= Redhead Persian, K= Karakual, DAR= Damara, P= Pedi, SW= Swazi, SAL= South African Landsheep, SPM= Spanish Merino

[b] H_O= observed heterozygosity, H_E= expected heterozygosity, MAF= Minor allele frequency, F_{ROH}= F-statistic Runs of homozygosity, F_{GRM} = F-statistic Genomic relationship matrix, F_{UNI}= F-statistic Uniting gametes, F_{ST}= F-statistic Genetic differentiation, r^2 = linkage disequilibrium, Ne= effective population size

Among the afore-mentioned ROH studies, there was a consensus that ROH size fragments in sheep populations are mostly short and indicative of ancient inbreeding (He *et al.*, 2020; Dzomba *et al.*, 2021, Nosrati *et al.*, 2021; van Marle-Köster *et al.*, 2021). The level of genomic inbreeding in the studies was generally low, with all studies reporting F_{ROH} values below 0.037 (He *et al.*, 2020; Nosrati *et al.*, 2021; van Marle-Köster *et al.*, 2021; Granero *et al.*, 2022). A

common observation in these studies was the sufficiency of genetic diversity levels in sheep populations, with H_E ranges between 0.318 and 0.374 (van Marle-Köster *et al.*, 2021; Granero *et al.*, 2022), H_O ranging between 0.327 and 0.380 (van Marle-Köster *et al.*, 2021; Dzomba *et al.*, 2021), and a general tendency toward only small losses in genetic diversity (i.e., $H_O > H_E$). Furthermore, relatively low genome- and chromosome-wide LD (e.g., r^2-value range: 0.09-025) were mostly reported, indicating high recombination rates (Liu *et al.*, 2017; Dzomba *et al.*, 2021; Nel *et al.*, 2022).

Chapter 3: Materials and Methods

3.1 Introduction

This study aimed to investigate the autozygosity and inbreeding levels of the South African Merino sheep breed using single nucleotide polymorphism (SNP) genotypic data. A total of 1 738 animals (1 567 females and 171 males) were available for analyses. The genotypes were first analyzed to quantify proxy-indicator estimators of genomic diversity, including heterozygosity levels, minor allele frequency (MAF), and linkage disequilibrium (LD). Autozygosity was then investigated by characterizing the genome-wide runs of homozygosity (ROH) and heterozygosity-rich region (HRR) profiles, as well as estimating several inbreeding statistics. Lastly, within-population selection signatures were identified using the conserved ROH approach. All genotypic data was used with the consent of the SA Merino Breeders' Association and provided by SA Stud Book. Ethics approval (NAS126/2023) for the use of all data in this study was obtained from the Ethics Committee of the Faculty of Natural and Agricultural Sciences at the University of Pretoria.

3.2 Materials

A total of 1 738 animals (1 567 females and 171 males), genotyped using different versions of the Illumina® Ovine 50K SNP panel, were included in the study. A descriptive summary of the panels used, as well as the number of animals genotyped with each, is given in Table 3.1.

The SNP-calling was done according to the standard genotyping protocols of each individual service provider. The resulting genotype input files were converted into PLINK-compatible input files (i.e., .map, and .ped) using a plug-in of Illumina® Genome Studio software v2.0 (Illumina, San Diego, California 92122 U.S.A).

Table 3.1 Summary of single nucleotide polymorphism (SNP) genotyping panels and the number of animals genotyped with each panel.

Array Type	SNPs available per array[a]	Number of Females	Number of Males	Total Number of animals
Illumina® v1	52452	258	31	289
Illumina® v2	51719	1170	49	1219
Illumina® v3	50199	139	62	201
Unistel v3	61954	0	29	29
Total		1567	171	1738

[a] SNP distributed over 27 autosomes and two sex chromosomes

3.3 Methods

For standardization, the autosomal SNPs in common between the panels were identified and extracted from each data set in PLINK using the *--extract* command. The data from the four panels were then merged to form one dataset using the *--merge-list* command. Thereafter, duplicate SNPs were removed from the merged dataset using the *--list-duplicate-vars ids-only suppress-first* and the *--exclude* commands. A subset of 50 169 SNPs was then subjected to quality control.

Standard quality control (QC) procedures were performed on the merged dataset using PLINK software (Purcell *et al.,* 2007). Sample-based as well as animal-based pruning was performed, to remove both non-informative SNPs and individuals from the dataset. Individual animals were removed if their genotyping call rate was less than 95% (using *--mind 0.05*), while SNPs were removed if their call rate was below 98% (using *--geno 0.02*). As suggested by Meyermans (2020) for ROH detection, neither minor allele frequency (MAF) nor linkage disequilibrium (LD) filtering was applied. This is because pruning for LD before ROH analysis results in a negative reduction of <70% of SNP detection, underestimated F_{ROH} estimates and a reduction in the number of ROH detected (Meyermans *et al.,* 2020). Furthermore, MAF pruning during quality control results in decreased ROH detection especially in high homozygous regions (Meyermans *et al.,* 2020). The MAF was, however, used as a QC

parameter for HRR detection. Table 3.2 summarizes the available number of SNPs and animals after quality control was applied.

Table 3.2 A summary of animal- and marker-based quality control, indicating the number of animals and SNPs removed for each parameter.

Quality Control	**PLINK** (Purcell *et al.*, 2007)	**Number of SNPs after QC**	**Number of animals after QC**
Common SNPs[a]	--extract	50169	1738
Marker-based QC	--geno 0.02[b]	46952	-
Animal-based QC	--mind 0.05[c]	-	258

[a] unique SNPs i.e. after duplicate SNPs were removed from the dataset.

[b] SNPs with 98% genotyping rate (2% missingness) were included.

[c] Individuals with more than 5% missing genotypes were excluded.

3.3.1 Genetic diversity

Once QC procedures were applied, marker-based summary statistics including the proxy-indicators for genetic diversity (namely H_O & H_E, MAF, LD, Ne) were estimated using PLINK (Purcell *et al.,* 2007) as follows:

3.3.1.1 Observed and expected heterozygosity (H_O, and H_E)

The mean observed and expected heterozygosity estimates were produced using the *--het* command. The resulting .het file was then imported into Microsoft Excel version 16.59 (Microsoft Corporation, 2018) where H_O and H_E were calculated using the following formulas:

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

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

where: N(NM) = the number of non-missing genotypes; O(Hom) = the observed number of homozygous genotypes; E(Hom) = the expected number of homozygous genotypes.

3.3.1.2 Minor allele frequency (MAF)

The MAF values were calculated using the *--freq* command. The resulting .frq file was imported into Microsoft Excel version 16.59 (Microsoft Corporation, 2018) where the mean MAF was calculated. The proportion of MAF per loci was visualized in a bar graph using the same software.

3.3.1.3 Linkage disequilibrium (LD)

The level of LD was calculated by adjacent pairwise analysis (measured as r^2) using the PLINK command *--r2*. To allow for all possible pairwise SNP comparisons per autosome, no minimum r^2 and inter-SNP distance restrictions were applied (i.e., *--ld-window-r2 0*, and *–ld-window-kb 99999* were applied). The resulting .ld file was imported into Microsoft Excel version 16.59 (Microsoft Corporation, 2018) where the mean r^2 estimate, mean inter-SNP distance, and minimum and maximum inter-SNP distance were calculated for all autosomes.

3.3.1.4 Effective population size (Ne)

The effective population size (Ne) was calculated using the SNeP v1.1 software tool (Barbato *et al.,* 2015). Microsoft Excel version 16.59 (Microsoft Corporation, 2018) was used to construct a line graph to visualize the Ne of the population over generations. The SNeP software predicts Ne using the following equation described by Corbin *et al.* (2012).

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

Where: $N_{T(t)}$ = the effective population size t generations ago calculated as $t = (2f(c_t\ )) - 1$ (Hayes *et al*., 2003), c_t is the recombination rate for a specific physical distance between SNPs

estimated using Sved and Feldman approximation (Sved & Feldman, 1973), r^2_{adj} is the LD value adjusted for sample size and α is a correction factor for the occurrence of mutations.

3.3.2 Runs of homozygosity (ROH) and heterozygosity-rich areas (HRR)

Both runs of homozygosity (ROH) and heterozygosity-rich areas (HRR) were detected using the *detect*RUNS R package (Biscarini *et al.,* 2018) through the execution of both the sliding window and consecutive runs approaches (SW, CR; Marras *et al.,* 2015). For a more comprehensive report on ROH profiles, both approaches were tested in this study, however, only the ROH detected through the SW approach were included in the calculation of the F_{ROH} inbreeding coefficient. The ROH and HRR defining parameters used in the SW and CR approaches are specified in Table 3.3 and Table 3.4, respectively. The ROH segments were assigned to five length categories, namely <4Mb, 4≤ROH<8Mb, 8≤ROH<16Mb, 16Mb≤ROH<32Mb, or ≥32Mb. The HRRs were classified into three length categories, namely 0.5-1Mb, 1-2Mb, and >2Mb. The number and chromosome distribution of the ROH and HRR were counted.

Table 3.3. ROH and HRR defining parameters and code for the SW approach

Parameter	Code	ROH value used	HRR value used
Size of the sliding window (number of SNPs).	windowSize	50	10
The threshold of overlapping windows in the same state to call a SNP in a run.	Threshold	0.05	0.05
Minimum number of SNPs in a run.	minSNP	49	20
Maximum number of opposing genotypes in the sliding window.	maxOppWindow	0	2
Maximum number of missing SNP in the sliding window.	maxMissWindow	2	1
Maximum distance between consecutive SNP in a window to be still considered a potential run.	maxGap	10^6 bps	10^6 bps
Minimum length of a run in bps.	minLengthBps	1000000	10000
Minimum number of SNP per kbps.	minDensity	0.001	0.000001
Maximum number of opposite genotype SNPs in the run.	maxOppRun	NULL	NULL
Maximum number of missing SNPs in the run.	maxMissRun	NULL	NULL

Table 3.4. ROH and HRR defining parameters and code for the CR approach.

Parameter	Code	ROH value used	HRR value used
Minimum number of SNPs in a run.	minSNP	49	20
Maximum distance between consecutive SNP in a window to be still considered a potential run.	maxGap	10^6 bps	10^6 bps
Minimum length of a run in bps.	minLengthBps	1000000	10000
Maximum number of opposite genotype SNPs in the run.	maxOppRun	1	2
Maximum number of missing SNPs in the run.	maxMissRun	1	1

To reduce the number of SNPs that occur in a run due to chance, the following formula was used to calculate the minimum number of SNPs in a run (Purfield *et al.,* 2012). Using the formula below the minimum number of SNPs allowed in a run was set to 49 for ROH detection and 20 for HRR.

$$\ell = \frac{log_e \frac{\alpha}{n_s . n_i}}{log_e(1 - h\bar{e}t)}$$

where: n_s is the number of SNPs per individual, n_i is the number of individuals, α is the percentage of false positive ROH (set to 0.05 in this study) and $h\bar{e}t$ is the mean SNP heterozygosity across all SNPs. For HRR, (1-$h\bar{e}t$) was replaced with $h\bar{e}t$ in the formula.

3.3.3 Inbreeding estimation

Two methods were utilized to estimate inbreeding, namely 1) F_{SNP} and 2) F_{ROH}. Box plots were generated for each inbreeding coefficient in R using the function boxplot (Chambers *et al*., 2018). Additionally, R was also used to calculate Pearson correlations between coefficients using the function cor.test (Best & Roberts, 1975).

3.3.3.1 SNP-based inbreeding coefficient (F_{SNP})

The SNP-by-SNP based inbreeding statistic (F_{SNP}) was calculated using the following formula defined by Keller *et al.* (2011):

$$F_{SNP} = \frac{O - E}{N - E}$$

where O is the observed number of homozygous SNPs per individual, E is the expected number of homozygous SNPs in Hardy-Weinberg equilibrium (HWE), calculated based on the estimated allele frequencies of the sample (Keller *et al.,* 2011), N is the total number of SNPs. The F_{SNP} inbreeding estimator can be negative when the observed homozygosity is lower than the expected homozygosity or be positive when the opposite is true (Ghoreishifar *et al.,* 2023).

3.3.3.2 ROH-based inbreeding coefficient (F_{ROH})

The ROH-based inbreeding coefficient (F_{ROH}) was estimated per individual as per the following formula defined by McQuillan *et al.* (2008);

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

where ΣF_{ROH} represented the length of ROH in one individual, and ΣL_{AUTO} represented the length of the genome covered by SNPs.

The inbreeding coefficients relevant to each length category (<4Mb, 4≤ROH<8Mb, 8≤ROH<16Mb, 16Mb≤ROH<32Mb, or ≥32Mb) were calculated in R using the *detect*RUNS package (Biscarini *et al.,* 2018). The SW approach in the literature is most used for ROH detection across livestock species and hence a resource for F_{ROH} detection (Peripolli *et al.,* 2017), hence for comparability with other studies, the SW approach was used for F_{ROH}.

3.3.4 With-in population selection signatures

The *detect*RUNS (Biscarini *et al.,* 2018) package was used to identify population-wide overlaps in ROH as well as the candidate genes harboured within these overlapping regions. The ROH detected in the 99.7th percentile (threshold of 16.4%) were considered potential ROH islands, as suggested in previous studies (Baazaoui *et al.*, 2021; Ceccobelli *et al.*, 2023). In addition, to include more candidate genes, a threshold calculated from the mean percentage of SNPs in ROH plus three standard deviations was considered (Grilz-Seger *et al.*, 2019).

The chromosome-wide proportion of times each SNP resided within a detected ROH was generated using Manhattan plots (Wickham, 2009). The *Ensembl* BioMart online tool (Cunningham *et al.,* 2022) was used to investigate candidate genes located within the overlapping ROH and HRR segments. Gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8 software tool (Huang et al., 2009). The DAVID software tool calculates a significance *p-value* for annotated genes, which uses a Fisher's Exact test that is adopted to measure the level of gene enrichment in annotation terms. Its biological interpretation is that the smaller the *p-value*, the more enriched the genes detected are for those ontology terms.

Chapter 4: Results

4.1 Introduction

Genome-wide single nucleotide polymorphism (SNP) genotypic data of 1 738 SA Merino sheep was analyzed to evaluate their degree of genomic diversity, as well as uniformity in the form of inbreeding and autozygosity levels. Several within-breed population parameters relating to the population's diversity and inbreeding were estimated, followed by detailed analyses of both runs of homozygosity (ROH) and heterozygosity-rich areas (HRR). The identified ROH was further interrogated to identify conserved segments that may harbour candidate genes. Lastly, the identified genes were annotated, and their functional role was investigated.

4.2 Quality control

Before quality control (QC) was performed, a subset of 50 169 common autosomal SNPs was extracted from the genomic data generated and available through four variable SNP genotyping panels and merged into a single genotyping dataset. After QC, 46 952 SNPs and 1 738 animals were available for downstream analyses.

4.3 Genetic diversity parameters

The mean minor allele frequency (MAF) across all autosomes was 0.275±0.14 with a median value of 0.287. The OAR 11 and OAR 23 autosomes had the lowest (0.245) and highest (0.284) chromosome-wide mean MAF values, respectively. The chromosome-wide mean MAF is presented in Figure 4.1. The highest percentage of monomorphic SNPs (MAF=0%) was observed on OAR 1 (11.25%) and the lowest on OAR 25 (1.23%). The highest percentage of SNPs had an MAF in the range of 0-0.19 (99.97%), with the lowest percentage of SNPs having an MAF of 0.5 (0.04%).

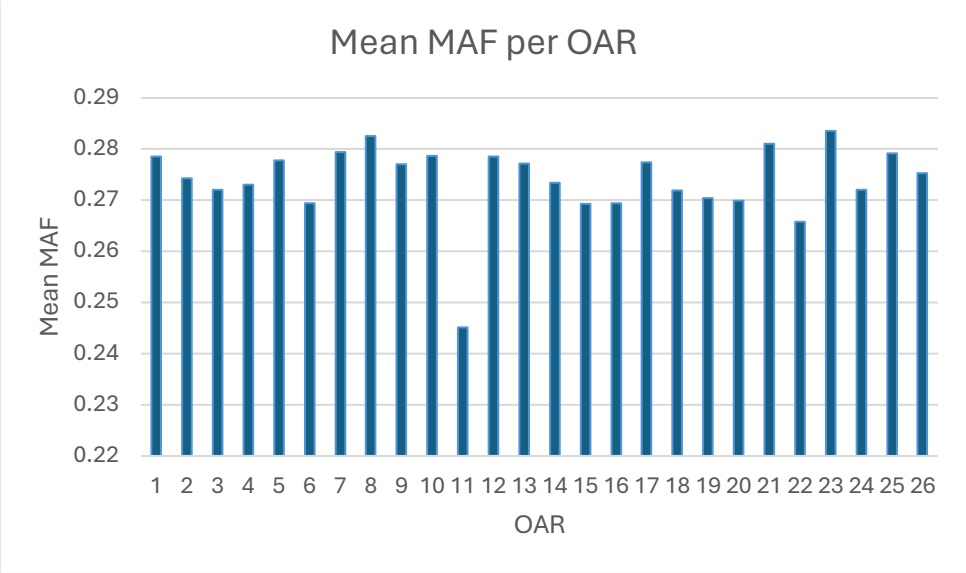


Figure 4.1 A graphic representation of the mean MAF per chromosome in SA Merino sheep.

The observed (H_O) and expected heterozygosity (H_E) levels were estimated for the population to identify a possible gain or loss in genetic diversity. The results indicated a loss in genetic diversity with a mean±standard deviation H_O of 0.351± 0.018 compared to a mean±standard deviation H_E of 0.365± 0.0001. The Ho values ranged from a minimum of 0.236 to a maximum of 0.390, whilst the H_E values had a range of 0.362 to 0.366.

To calculate the chromosome-wide mean LD between all pairwise SNP comparisons per autosome, no restrictions were set on the minimum r^2 (default: r^2=0.2) and inter-SNP distance (default=1000kb) allowed for LD estimation. The mean±standard deviation r^2-value obtained across all autosomes was 0.365±0.180. Table 4.2 shows the mean r^2 value per chromosome for different length windows namely 0-50kb, 0-100kb, and 0-1Mb. The average inter-SNP distance across all chromosomes was calculated to be 188.72kb, while the minimum and maximum inter-SNP distance across all chromosomes was 5.29kb and 998.07kb, respectively.

Table 4.1 Summary of the mean r^2 values and inter-SNP distance per chromosome for different length windows.

OAR	Mean r^2 (50kb)	Mean r^2 (100kb)	Mean r^2 (1Mb)	Mean inter-SNP distance
1	0.472	0.437	0.364	185.584
2	0.484	0.454	0.377	178.812
3	0.460	0.435	0.367	191.859
4	0.477	0.449	0.370	188.339
5	0.485	0.440	0.361	178.443
6	0.437	0.419	0.355	206.076
7	0.458	0.425	0.358	170.726
8	0.467	0.428	0.359	174.648
9	0.491	0.433	0.368	181.837
10	0.498	0.456	0.394	199.474
11	0.468	0.431	0.361	233.932
12	0.455	0.419	0.355	176.988
13	0.484	0.457	0.378	181.823
14	0.457	0.426	0.357	192.856
15	0.468	0.425	0.355	191.300
16	0.453	0.416	0.344	195.062
17	0.435	0.417	0.356	209.292
18	0.485	0.439	0.369	190.476
19	0.512	0.469	0.377	202.203
20	0.484	0.443	0.358	190.830
21	0.468	0.435	0.355	181.334
22	0.502	0.457	0.371	175.042
23	0.424	0.389	0.333	201.111
24	0.456	0.427	0.344	225.198
25	0.508	0.456	0.376	163.959
26	0.423	0.402	0.356	178.976

4.4 Runs of homozygosity and heterozygosity-rich regions

4.4.1 Identified runs of homozygosity (ROH)

Both the sliding window (SW) and consecutive runs (CR) methods were employed to estimate the ROH profile of the SA Merino population. The descriptive statistics of the identified ROH per method are summarised in Table 4.2. The CR detection method identified more (n=50291 ROH) homozygous runs compared to the SW approach (n=43147 ROH). The mean (± standard deviation) ROH length per chromosome was 6.53Mb±0.47Mb using the SW approach, compared to 6.36Mb ± 0.44Mb for the CR approach. The mean ROH length was the longest for OAR 1 in both the SW and CR approaches (7.24Mb and 6.99Mb). The shortest mean ROH length was found on OAR 19 (5.52Mb) in the SW approach and OAR 13 for the CR approach (5.49Mb), respectively.

Table 4.2 Summary statistics of runs of homozygosity (ROH) identified for the SA Merino population using two ROH detection methods.

	Sliding window	**Consecutive runs**
Number of ROH	43147	50291
Mean number of ROH / OAR	1659.5	1934.269
Mean ROH length / OAR (Mb)	6.529	6.361
Median ROH length / OAR (Mb)	6.598	6.455
Minimum ROH length / OAR (Mb)	5.523	5.49
Maximum ROH length /OAR (Mb)	7.237	6.99
Mean number of ROH / animal	29.153	33.98
Minimum ROH number / animal	5	5
Maximum ROH number / animal	79	106

The highest number of ROH was detected on OAR 3 using both the SW and CR methods (4565 and 5205, respectively), whilst both methods identified OAR 24 as having the lowest number of ROH (611 and 723, respectively). Table 4.3 shows the proportion of detected ROH within different segment size categories for each method.

Table 4.3 Summary of the percentage of ROH detected per class size in the SA Merino population.

	Sliding window	**Consecutive runs**
Class (Mb)	**Percentage (%)**	**Percentage (%)**
<4	36.38	38.17
4≤ROH<8	40.55	40.37
8≤ROH<16	17.10	16.03
16≤ROH<32	5.15	4.65
≥32	0.81	0.78

The size category with the highest percentage of ROH was 4≤ROH<8Mb (40.55%, and 40.37%, respectively) for both methods. The ROH coverage (%) decreased with decreasing autosome length (i.e. roughly from OAR 1 to OAR 26). The peak percentage of ROH coverage (sum of ROH per chromosome divided by each chromosome length) was found on OAR 3 (10.5%) for the SW approach and OAR 2 (10.3%) for the CR approach. The lowest overall percentage of autosomal ROH coverage was recorded on OAR 24 (1.41% and 1.43%) for both the SW and CR approaches.

The number of ROH per animal was additionally identified for each method and the results are represented in Figure 4.2. The SW method identified less ROH per individual than the CR approach. The mean number of ROH per individual was 29.15 runs for the SW approach and 33.98 runs for the CR approach.

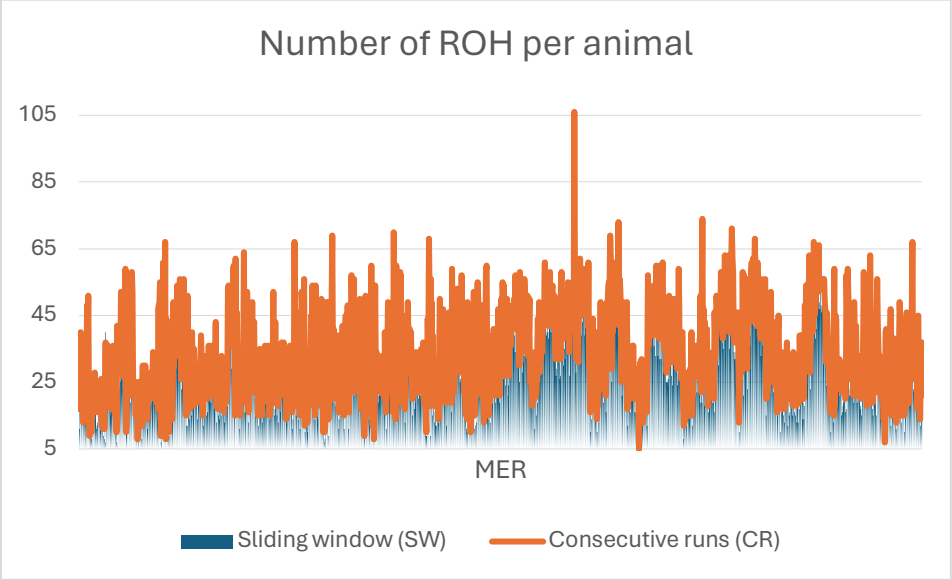


Figure 4.2 A graph representing the number of ROH detected per individual SA Merino sheep for both methods.

An increasing trend was observed regarding the length of ROH segments (in megabase pairs; Mb) and the number of ROH per individual, as shown in Figure 4.3. The figure shows that most individuals have detected ROHs that are short i.e. the sheep with a higher total ROH coverage had a higher number of short ROHs detected.

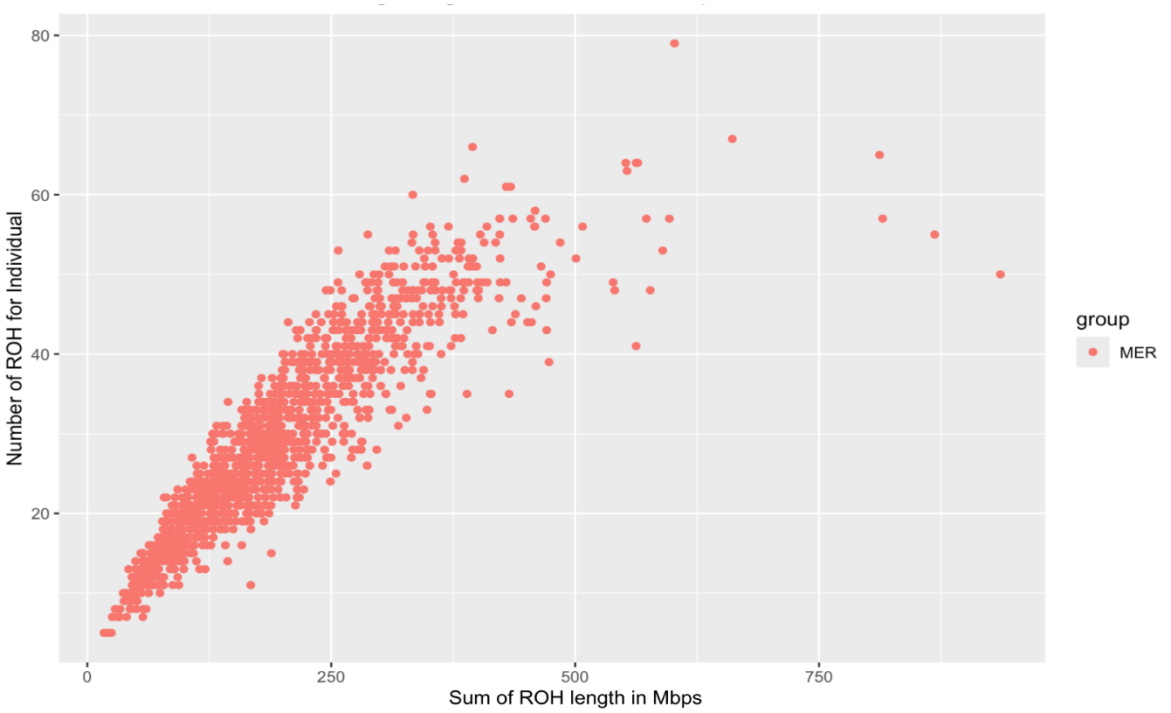


Figure 4.3. A plot showing the sum of ROH lengths (in Mb) against the number of ROH per individual.

4.4.2 Identified heterozygosity-rich regions (HRR)

Similar to ROH, HRR was identified using both the SW and CR methods. The SW detection method identified more heterozygous runs compared to the CR approach. The mean (± standard deviation) HRR length per chromosome was 1.3Mb± 0.14Mb for the SW approach and 1.1Mb± 0.17Mb for the CR approach. The statistics of the identified HRR per method are summarised in Table 4.4.

Table 4.4. Summary statistics of heterozygosity-rich areas identified for the SA Merino population using two HRR detection methods.

	Siding window	**Consecutive runs**
Number of HRR	11744	911
Mean number of HRR / OAR	451.692	39.609
Mean HRR length / OAR (Mb)	1.3	1.129
Median HRR length / OAR (Mb)	1.269	1.142
Minimum HRR length / OAR (Mb)	1.095	0.818
Maximum HRR length / OAR (Mb)	1.653	1.695
Mean number of HRR / animal	7.93	1.33
Minimum HRR number / animal	1	1
Maximum HRR number / animal	21	4

The OAR 2 autosome had the longest mean length of HRR in both the SW and CR approaches (1.65Mb and 1.69Mb, respectively). Additionally, the highest HRR count was also detected on OAR 2 in both methods (n=1647, and n=164, respectively). The shortest mean length of HRR was found on OAR 4 (1.09Mb) in the SW approach and OAR 10 for the CR approach (0.81Mb), whilst the lowest HRR count was found on OAR 11 for both the SW and CR methods (n=76, and n=3, respectively). Table 4.5 shows the percentage of HRR detected per class size for both methods.

Table 4.5 Summary of the percentage of HRR detected per class size in the SA Merino population.

	Sliding window	**Consecutive runs**
Class (Mb)	**Percentage (%)**	**Percentage (%)**
0.5-1	22.98	32.49
>1-2	71.65	63.67
>2	5.37	3.84

The size class with the highest percentage of HRR detected was 1-2Mb (71% and 63% for the SW and CR methods, respectively). The number of HRR detected per class size showed a

general decreasing trend from OAR 1 to 26 for the SW method, whereas the CR method showed a decreasing trend from OAR 1 to 25 with peaks on OAR 10 and OAR 13. The highest HRR coverage was detected on OAR 2 for both the SW and CR approaches (14.02% and 18%), respectively. The lowest overall HRR coverage was recorded on OAR 11 (0.64% and 0.32%) for the SW and CR approaches, respectively.

The number of HRR per animal was additionally identified for each method and the results are represented in Figure 4.4. A higher number of HRRs were identified per individual in the SW approach than in the CR approach. The mean number of HRRs per individual was 7.93 runs for the SW approach and 1.33 runs for the CR approach.

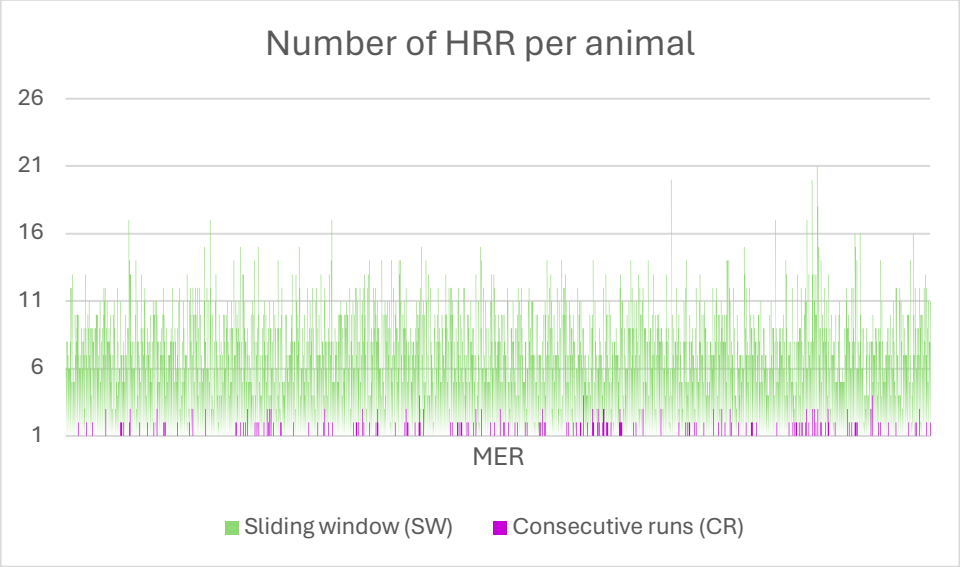


Figure 4.4 The number of HRR detected per individual SA Merino sheep for both methods.

The average length of HRR segments (Mb) showed a positive relationship with the average number of HRR per individual as shown in Figure 4.5. The midsection of the figure indicates that most individuals had HRR lengths of moderate size compared to the average number of runs per individual i.e. few sheep had extremely short and extremely long lengths of HRRs.

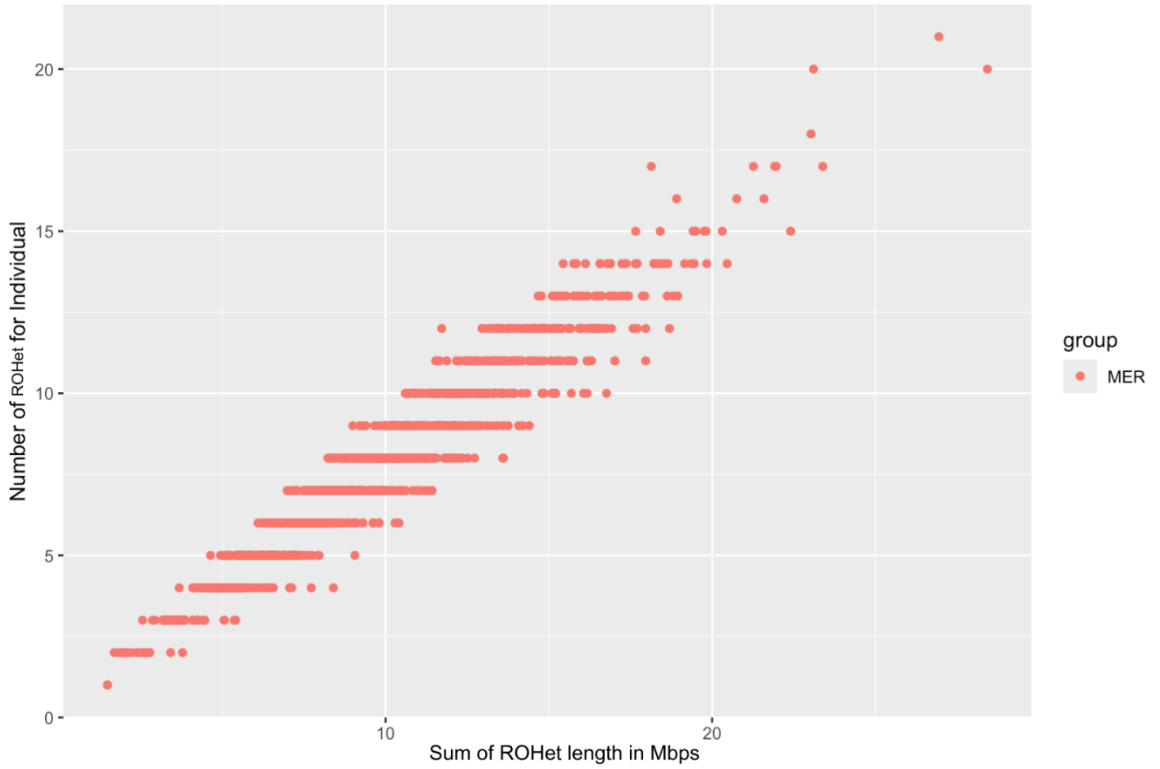


Figure 4.5. A scatter plot indicating the relationship between the sum of HRR lengths (in Mb) and the number of HRR per individual.

4.5 Inbreeding estimation

4.5.1 F-statistics

The mean±standard deviation genome-wide F_{ROH} was estimated as 0.074±0.04. The chromosome-wide F_{ROH} mean values ranged between 0.076 (OAR 1) and 0.352 (OAR 21). The mean F_{ROH} values per chromosome are reported in Table A1. The mean F_{ROH} values per ROH length category (in Mb) were $F_{ROH<4}$=0.074, $F_{\leq ROH<8}$= 0.061, $F_{8\leq ROH<16}$ =0.037, $F_{16\leq ROH<32}$ =0.023, and $F_{ROH\geq 32}$ =0.022, respectively. The mean±standard deviation F_{SNP} across the genome was marginally lower at 0.037±0.05. The mean F_{SNP} values estimated had a minimum value of -0.075 and a maximum value of 0.353. There was a strong, positive relationship between the genomics-based F-statistics (i.e., F_{SNP} and F_{ROH}), with a Pearson correlation of 0.95. Figure 4.6 shows boxplots, illustrating the mean, median, and outliers for the F_{SNP} and F_{ROH} estimators. Characteristically, the F_{SNP} coefficient showed more variation than the F_{ROH} coefficient.

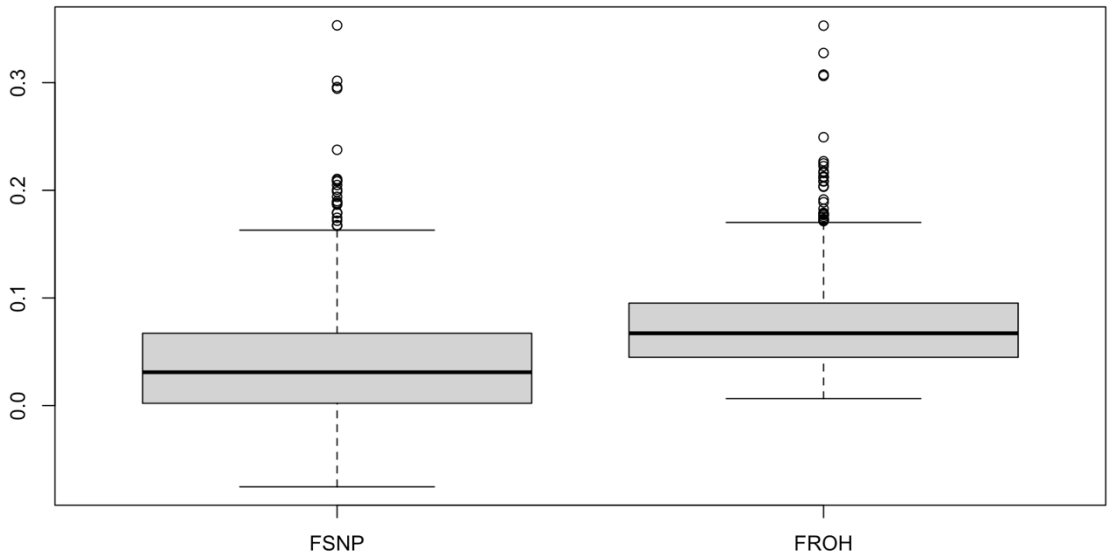


Figure 4.6 Boxplots for the F_{SNP} and F_{ROH} inbreeding estimators showing the mean, median, maximum, minimum, and outliers of each.

4.5.2 Effective population size (Ne)

The SNP-based Ne estimates were computed from 912 generations ago until 12 generations ago. Due to limitations in the panel density, the most recent estimate that the software could calculate was 12 generations ago, and this relates to ~38 and ~2900 years ago based on an estimated generation interval of 3.2 years (Cloete *et al.*, 2009). The Ne of the SA Merino breed decreased over time, as represented in Figure 4.7. The Ne declined from 3062 for the oldest generation (912 generations ago) to 305 for the most recent generation (12 generations ago).

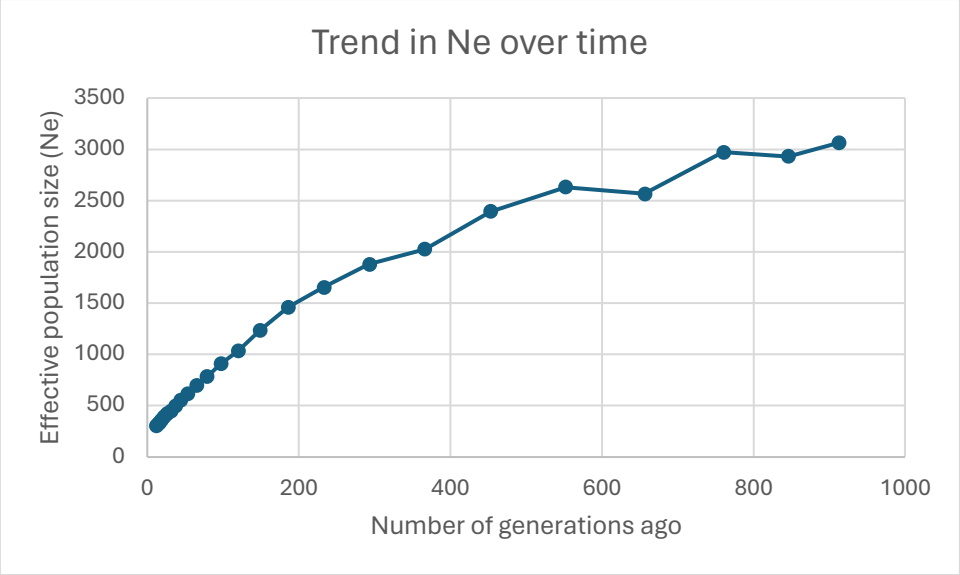


Figure 4.7 The decline in Ne from generation 912 to generation 12 of the SA Merino sheep population.

4.6 Within-population selection signatures

The chromosome-wide proportion of times an SNP resided within a detected ROH using the SW method is summarised in Table 4.6. A total of 275 genes (within a threshold of 16.4%) and 374 SNPs across four autosomes were detected.

Table 4.6 A summary of the number of genes located within each ROH region detected using a threshold of 16.4%.

OAR	n SNP[a]	Position (from)	Position (to)	Length (Mb)	n Genes[b]
2	60	13707962	16493535	2785573	15
2	41	17960258	20898157	2937899	12
3	1	47595933	47595933	-	1
3	5	47755179	48428503	673324	2
3	18	48950632	50074939	1124307	3
3	22	50229700	51353599	1123899	1
3	7	51607921	51922042	314121	1
3	1	63916843	63916843	-	0
6	57	39124095	43178901	4054806	6
6	6	53238259	53470583	232324	0
6	14	54029176	54762520	733344	0
11	142	23318858	33174792	9855934	234
Total	374	-	-	-	275

[a] Number of SNPs [b] Number of genes detected

A Manhattan plot of the proportion of times each SNP resides within detected ROH per chromosome is shown in Figure 4.8. The highest occurring consensus ROH haplotype was on OAR 11 (base pair position: 23318858-33174792), and this region contained 234 protein-coding genes. In descending order, the highest occurring consensus ROH regions were followed by OAR 2 with 27 protein-coding genes, OAR 3 with 8 protein-coding genes, and OAR 6 with 6 protein-coding genes.

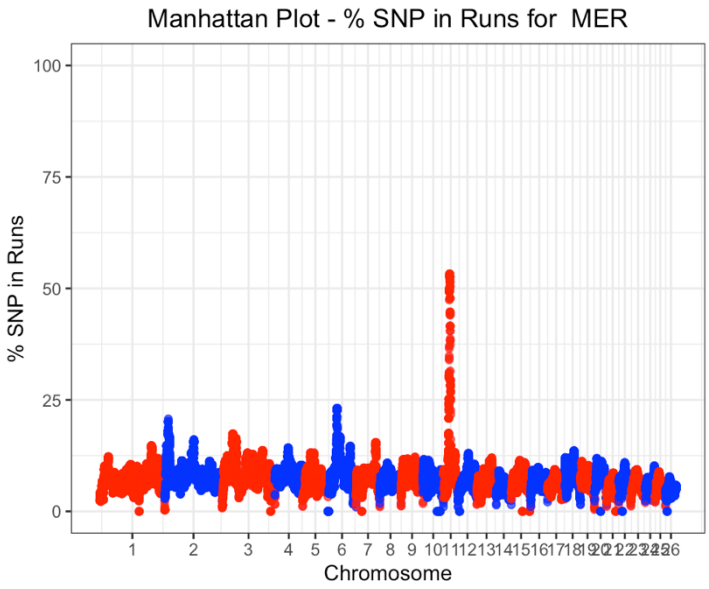


Figure 4.8 The chromosome-wide proportion of times each SNP resides within a detected ROH using the SW method.

Table 4.7 summarises the associated functions of each protein-coding gene located within the detected ROH regions. A more comprehensive version of the table is included in the addendum as Table A2. The protein-coding genes in ROH that were the most enriched (i.e., *p*-value of 1.70E-07) were associated with the lipoxygenase pathway and linoleic acid metabolic processes.

Table 4.7 A summary of the associated function of each protein-coding gene located within the defined ROH regions.

Associated function	No of genes	Significance level (p-value)
Arachidonic acid metabolic process	6	9.00E-08
Arachidonate 12-lipoxygenase activity	4	4.70E-06
Actin filament binding	9	4.70E-05
Arachidonate 15-lipoxygenase activity	3	3.30E-04
Myofibril	4	9.40E-04
Myosin complex	5	0.001
Olfactory receptor activity	15	0.019

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were constructed for genes within the detected ROH regions. The KEGG pathways for a) genes associated with Arachidonic acid metabolism, b) genes associated with motor proteins, and c) genes associated with olfactory transduction are indicated in Figure 4.9.

Figure 4.9a shows the genes that play a role in arachidonic acid metabolism are also associated with linolenic acid metabolism, namely, *ALOX12B*, *ALOX12* (Arachidonate 12-lipoxygenase, OAR 11), *ALOX15* (Arachidonate 15-lipoxygenase, OAR 11) and *ALOXE3* (Arachidonate lipoxygenase 3, OAR 11).

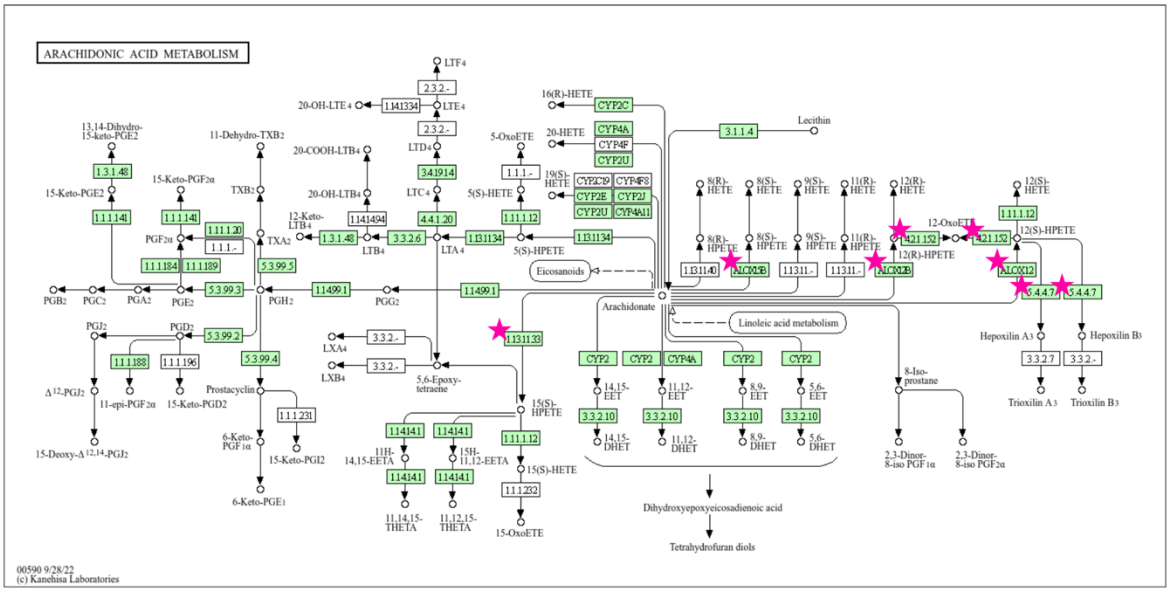


Figure 4.9a. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway for genes associated with Arachidonic acid metabolism in sheep.

Figure 4.9b shows genes that are associated with motor proteins and play a role in myofibril formation, namely, Genes *DNAH2* (dynein axonemal heavy chain 2, OAR 11), *DNAH9* (dynein axonemal heavy chain 9, OAR 11), *KIF1C* (kinesin 3, OAR 11), *MYH1, MYH2 MYH8* and *MYH10* (OAR 11).

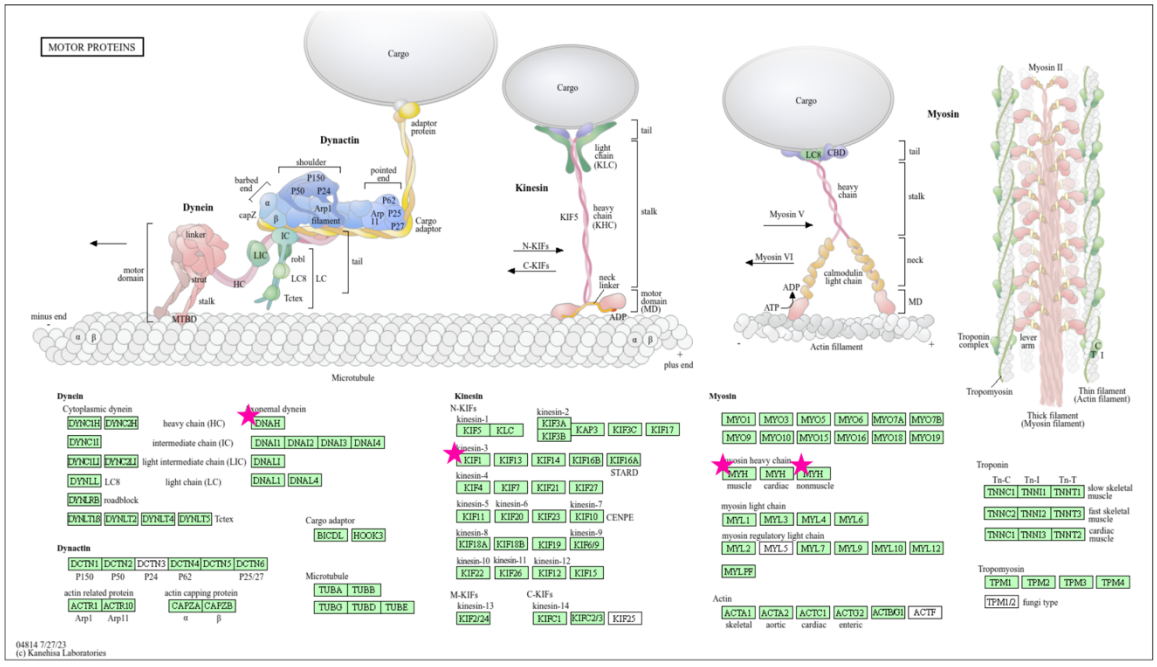


Figure 4.9b. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway for genes associated with functioning of motor proteins in sheep.

Figure 4.9c shows genes that are associated with olfactory transduction, namely, *ARRB2* (Arrestin beta 2, OAR 11), *GUCY2D* (Guanylate cyclase 2D retinal, OAR 11) and *DTMT* (Olfactory receptor-like protein, OAR 11).

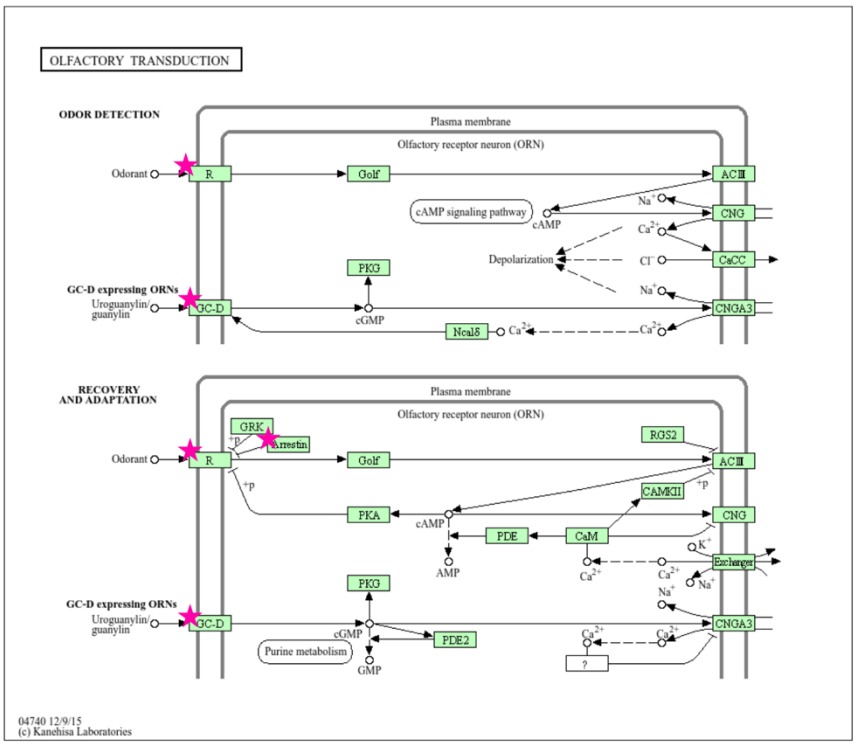


Figure 4.9c. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway for genes associated with olfactory transduction in sheep.

Figure 4.9 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for genes associated with a. Arachidonic acid metabolism, b. functioning of motor proteins, and c. olfactory transduction within enriched ROH regions.

Chapter 5: Discussion

5.1 Introduction

Merino and Merino-derived breeds are dispersed globally and are arguably the most abundant sheep breed and primary producer of the world's wool supply (Ciani *et al.*, 2015). Along the evolutionary timeline from the domestication of sheep to the development of commercial flocks, several factors (e.g. intense directional selection, genetic drift, related mating, population bottlenecks, and natural selection) have influenced the autozygosity levels of sheep flocks (Ferenčaković, 2015; Xu *et al.,* 2019).

This study aimed to perform an SNP-based investigation of the current autozygosity levels in the South African Merino sheep breed. Several population parameters that serve as proxy indicators of genetic diversity and/or inbreeding, including observed and expected heterozygosity- rates and effective population size (Ne) were used to quantify the genetic variability present within the SA Merino breed. Additionally, the genome-wide ROH profiles and associated F-statistics were used to reflect the autozygosity levels in the SA Merino sheep breed. Although previous local and global studies have explored within and across-population autozygosity levels in numerous sheep breeds, the present investigation provides a the first comprehensive overview of autozygosity levels in the SA Merino breed, specifically.

5.2 Genetic Diversity

Mean expected heterozygosity (H_E) and observed heterozygosity (H_O) are both measures of genetic diversity that are estimated from molecular information and the relationship between the two statistics reveals a gain or loss of genetic diversity (Eusebi *et al.,* 2020). The mean observed heterozygosity of the SA Merino, estimated in the current study, was moderate (H_O=0.35) and consistent with previously reported heterozygosity levels for the breed. Previous studies, of notably smaller sample sizes (e.g. 10 and 33 SA Merino sheep, respectively), reported similar observed heterozygosity values of 0.37 (e.g. Ceccobelli *et al.*,

2023; Dzomba *et al.* 2023). For a larger sample size (n=697 SA Merino sheep), Nel *et al.* (2022), for example, obtained an average observed heterozygosity value of 0.36 for SA Merino sheep. In the current study, including a larger and more representative sample of the breed, the expected (H_E=0.365) heterozygosity was slightly higher than the observed heterozygosity level and this agreed with previous studies where H_E= 0.378 (Ceccobelli *et al.*, 2023), 0.39 (Dzomba *et al.*, 2023), and 0.37 (Nel *et al.*, 2022) were reported for SA Merino sheep.

The small difference in H_O and H_E values indicates a negligible loss of within-population genetic diversity (Eusebi *et al.,* 2020), which could be due to maintaining breed purity or breed standards by the breed society. The moderateheterozygosity levels reported here were expected as the SA Merino is a well-managed commercial sheep breed with moderate to high population sizes and well-defined breeding programs (Dzomba *et al.*, 2021).

Minor allele frequency (MAF) relates to genetic diversity and indicates the level of informative and segregating markers within a population (Keller *et al.,* 2011). Highly polymorphic markers are associated with high MAF values which correspond to higher levels of H_E and consequently indicate a greater genetic diversity (Engelsma *et al.,* 2014). On the contrary, low MAF indicates possible fixation of one allele in the population and, hence, lower genetic diversity (Engelsma *et al.,* 2014). The high MAF values (highly informative SNPs) obtained for the Illumina® OvineSNP50 Beadchip genotyping panel are expected as this genotyping panel was specifically designed to include SNPs validated in the most economically- and globally-important *Ovis aries* breeds, including 127 samples from Merino-type breeds (Kijas *et al.,* 2009).

The mean MAF (0.275±0.14) reported in this study is similar to Ovine 50K-based values previously reported for smaller sample sizes of the SA Merino breed; these MAF values ranged from 0.26 to 0.27 (Dzomba, 2021; Sandenbergh, 2015; Sandenbergh *et al.*, 2016). The mean MAF values estimated are also comparable to previous Ovine 50K-based estimates for global Merino populations that ranged from 0.29 to 0.3 (Al-Mamun *et al.*, 2015; Nosrati *et al.*, 2021).

Linkage disequilibrium (LD) is a genetic diversity parameter that shows the extent of non-random association between two alleles located at different loci and is a direct reflection of the recombination rate within a population (Zavattari, 2000; Kijas *et al.*, 2014). The level of LD in a population reflects how closely SNPs are related to one another within the genome (Qanbari *et al.,* 2010). In this study, LD was estimated using r^2 as proposed by Bohmanova *et al.* (2010) because of the inverse relationship between r^2 and the sample size required to detect associations between SNPs (Bohmanova *et al.,* 2010). High levels of LD and, hence, more relatedness result in higher levels of specific allelic combinations inherited over generations than can be expected by chance (Eusebi *et al.,* 2020).

The average r^2-value calculated in this study was 0.365, which is consistent with other studies on similar sample sizes of Merino sheep. Liu *et al.* (2017); Dzomba *et al.* (2021), and Nel *et al.* (2022) reported r^2-values in a wide range of 0.14 to 0.31.

The effective population size (Ne) of an existing population can be defined as the ideal population size that will result in the same amount of genetic diversity as that which is present in the current existing population (Wright, 1974). The Ne size directly influences a population's inbreeding level (Gholizadeh & Ghafouri-Kesbi, 2016). In concordance with the low estimates of inbreeding reported here, the most recent estimate of Ne falls within the recommended range of 50 (Frankham *et al.,* 2002) and 500 (Holt *et al.,* 2005; Meuwissen, 2009), proposed for the survival of a breed by the FAO.

In this study, the most recent genomics-based Ne estimate achievable (considering the SNP density and, hence, capturable LD level) was for 12 generations ago. There was a decline in the SA Merino's estimated Ne; over the span of ~2800 years (based on a generation interval of 3.2 years), the Ne decreased from 3 062 animals for the oldest generation (912th generation) to 305 animals for the most recent generations ago (i.e., 12). This decreasing trend observed for Ne over generations is in concordance with previous studies on SA Merino sheep (Dzomba, 2021; van Marle-Köster *et al.*, 2021; Nel *et al.*, 2022; Ceccobelli *et al.*, 2023).

There were however discrepancies between previously reported Ne estimates, and those reported here. For example, Dzomba (2021) reported a decline from 2809 (ancestral) to 167

animals (most recent). Furthermore, van Marle-Köster *et al.* (2021), Nel *et al.* (2022) and Ceccobelli *et al.* (2023) reported a decline in Ne population sizes from 697 (ancestral) to 33 (most recent) for smaller sample sizes of SA Merino sheep. The fairly large discrepancies may, however, be attributable to the sample size and composition. Higher relatedness within a sample could overestimate various of the proxy-indicator parameters studied here, for example, the inbreeding coefficients and the degree of LD. Furthermore, the density of the SNP genotyping panel can be a limiting factor in achievable estimation of Ne in more recent generations, as it determines the chromosome-wide LD that can be captured (Goddard & Hayes, 2007; Visser *et al.,* 2023).

5.3 Runs of homozygosity and Heterozygosity-rich regions

Runs of homozygosity (ROH) are long stretches of homozygous regions throughout the genome that reflect autozygosity due to consanguineous mating (Ferenčaković, 2015). The lengths of ROH reflect recombination events, with longer segments referring to recent inbreeding and shorter segments referring to ancestral inbreeding (Pemberton *et al.,* 2012).

For ease of comparison, only the descriptive statistics of the ROH identified (and corresponding F_{ROH} calculations) by means of the SW approach will be referred to in this discussion. In this study, a total of 43 147 ROH with a mean per chromosome length of 6.53Mb were identified. The OAR 1 and OAR 19 autosomes had the longest and shortest mean ROH length, respectively. The number of genome-wide ROH present and identifiable depends on the population dynamics; for instance, Nosrati *et al.* (2021) identified an average number of 16.8 ROH per individual ROH from a sample size of 356 animals from Chinese Merino sheep populations. On the contrary, Dzomba *et al.* (2021) identified a total number of 5931 ROH from a sample size of 56 SA Merino sheep and total number of 4071 ROH from a sample size of 50 Australian Merino sheep.

The majority of ROH detected in this study, were classified into the second shortest length category (i.e. 4-8Mb), reflecting the predominance of distant inbreeding effects. In support of this observation, a positive relationship (or trend) was observed between the number of ROH

per individual and its summed ROH length (i.e. the sheep with higher total ROH coverage), implying that SA Merino animals generally had more ROH per individual albeit smaller ROH in size.

Due to discrepancies in ROH defining criteria across studies, the comparison of reported ROH descriptive statistics to previously published ROH results is limited, however, general trends can be deduced. Two previous studies have investigated ROH in the SA Merino; in van Marle-Köster *et al*. (2021), for example, the majority (69%) of identified ROH were classified into the shortest length category (ROH<4Mb). Ceccobelli *et al.* (2023) found a similar mean per-individual ROH length (6.93Mb) to this study, with most (45.51%) of the identified ROH falling within a 5-10Mb length category. In international Merino studies, similar trends in ROH descriptive statistics were reported. Al-Mamun *et al.* (2015) and Wanjala *et al.* (2023), for example, observed the majority of ROH in a 1-10Mb length category. All the afore-mentioned studies reported a general decreasing trend of the number of ROH per increasing length category, hence, reiterating the suggestion of less recent inbreeding in local and global Merino populations.

Heterozygosity-rich regions (HRR) are consecutive stretches of heterozygous SNP genotypes (Williams *et al.,* 2016), which provide information regarding a population's evolutionary history and diversity (Samuels *et al.,* 2016). The HRR can be used to identify specific genomic areas where it would be beneficial to maintain a higher level of genetic diversity in specific areas (Biscarini *et al.,* 2020), typically associated with adaptive traits (Mc Parland *et al.*, 2009; Selli *et al.*, 2021).

In this study, a total of 11 744 HRR were detected, most (71%) of which were classified as short in length (1-2Mb); the mean HRR length per chromosome was 1.3Mb. These findings suggest that HRRs are small and frequent within the SA Merino sheep genome. Studies investigating HRR in livestock species are scarce in literature, thus making it difficult to find comparable results. In a study by Selli *et al.* (2021), however, a mean of 139.59 HRR (with a mean length of 459.894kb) per animal was identified across populations. Higher density SNP genotyping panels are expected to contribute more knowledge to the HRR profiles in populations.

In HRR-focused studies on other livestock species, the average number of HRR regions detected per animal was lower than the average number of ROH regions detected per animal (HRR_{TURKEY}=57.80, ROH=126.21, Marras *et al.,* 2018; HRR_{CATTLE}=9.87, ROH=22.27, Biscarini *et al.*, 2020; HRR_{HORSE}=52.17, ROH= 75.63, Bizarria dos Santos *et al.*, 2021), which is a similar trend observed in this study. In contrast, Selli *et al.* (2021) investigated HRR for international Merino sheep and detected a higher total average number of HRR (165 558) than the total average number of ROH (80 639). This could be due to differences in defining parameters when calculating HRR.

5.4 Inbreeding estimation

The F_{SNP} inbreeding coefficient estimates the excess in homozygosity (Rodríguez-Ramilo *et al.,* 2020). The genomic inbreeding coefficients, i.e. F_{SNP} and F_{ROH}, are highly correlated, with a strengthening correlation as the effective population size decreases, and this is because F_{ROH} defined by short ROH (ancient inbreeding) correlates poorly with homozygous mutation load when Ne is small (Keller *et al.,* 2011). Dzomba (2021) recommends a safe threshold of F_{SNP}<0.1 as sufficient to allow population diversity.

In this study, the mean F_{SNP} (0.037) was comparable with mean values reported in other studies on SA Merino sheep. Sandenbergh *et al.* (2015) and Dzomba *et al.* (2023), for example, both reported mean F_{SNP} values closest in comparison to this study (i.e., F_{SNP}=0.060 and 0.040, respectively). van Marle-Köster *et al.* (2021) reported an F_{SNP} value remarkably lower than that of other studies (F_{SNP} =0.009). Overall, the tendency towards small F_{SNP} values suggests that that there is sufficient genetic diversity within the breed to avoid the effects of inbreeding depression.

The F_{ROH} coefficient is a genomic measure of individual autozygosity (McQuillan *et al.*, 2008). Due to the nature of the formula to calculate F_{ROH}, estimates are commonly higher than those of F_{SNP} measures because they cannot include negative values. The added benefit of deriving inbreeding coefficients from ROH (F_{ROH}) is that it provides additional information to the

degree of inbreeding; the length characteristics also provides an indication of the "age" of inbreeding effects (Ferenčaković, 2015).

In this study, the mean genome-wide F_{ROH} was estimated at 0.074. The average F_{ROH} values per ROH length category (in Mb) were furthermore produced. The highest F_{ROH} value (0.074) was found within the short length category (1-4Mb) and showed an inverse relationship with ROH length (i.e. the F_{ROH} values decreased for increasing ROH length). These results are in close agreement with previous results, with reported F_{ROH} ranges between 0.055 and 0.14 for smaller sample sizes of SA Merino (Dzomba, 2021; van Marle-Köster *et al.*, 2021; Ceccobelli *et al.*, 2023). As expected, the F_{ROH} values were all higher than F_{SNP} values, conveying sufficient autozygosity at the individual level to avoid negative inbreeding effects.

5.5 Within-population selection signatures

Selection signatures are characterized by genomic regions with reduced genetic variability (i.e., increased allele frequencies within the genome) (Purfield *et al.,* 2017). Studying the distribution of ROH patterns across the genome allows the detection of selection signatures (Zhang *et al.*, 2015), which most likely reflect selection events (Kim *et al.*, 2013). Therefore, the functional analysis of identified genes within selection signatures may indicate traits under selection pressure (Almeida *et al.,* 2019).

In this study, a total of 275 genes had an incidence of more than 16.4% in SNP-in-ROH regions. Ceccobelli *et al.* (2023) similarly identified a total of 166 genes which had an incidence above a 99.9th percentile threshold. Biscarini *et al.* (2020) identified 58 genes using a 99.9 quantile of the SNP-in-ROH frequency distribution (corresponds to a frequency threshold of 0.41). The highest occurring consensus ROH haplotype was on OAR 11 which covered many base pairs (9855934bp) and this region contained 234 protein-coding genes. This indicates that genes within this region of OAR 11 have been under extensive selection pressure. The protein-coding genes found within the selection signatures were associated with the arachidonic acid metabolic process, motor proteins, and olfactory transduction.

Arachidonic acid metabolic process (which includes lipoxygenase pathways and linoleic acid metabolism) are important in sheep as they play an important role in inflammatory responses (Selli *et al.*, 2021) as well as wool production (Wang *et al.*, 2020). The *ALOX12B* and *ALOX12* (Arachidonate 12-lipoxygenase, OAR 11) genes are associated with body size and feed efficiency in sheep (Kominakis *et al.*, 2017). The *ALOX15* (Arachidonate 15-lipoxygenase, OAR 11) gene plays a role in *H. contortus* resistance in sheep (Guo *et al.*, 2016). The *ALOXE3* (Arachidonate lipoxygenase 3, OAR 11) gene encodes the enzyme arachidonate lipoxygenase 3 (Dzomba *et al.,* 2021).

Detected genes were found to play a role in the functioning of motor proteins (i.e. myosin complex) and involved in myofibril formation, which is important in fibre formation (Megdiche *et al.*, 2019). The *DNAH2* (dynein axonemal heavy chain 2, OAR 11) and *DNAH9* (dynein axonemal heavy chain 9, OAR 11) genes are involved in MAP2K4-metabolic functioning in the liver (Dzomba *et al.*, 2023). The *KIF1C* (kinesin 3, OAR 11) gene is involved in the functioning of motor proteins (Megdiche *et al.*, 2019). The *MYH1, MYH10* and *MYH8* genes (all located on OAR 11) collectively facilitate diverse roles in energy metabolism, including muscular contraction, cytokinesis, and phagocytosis (Miyagi *et al.*, 1999). The *MYH2* gene (OAR 11) plays a role in motor activity, actin cytoskeleton, actin binding, contractile fibre and calmodulin binding (Selli *et al.*, 2021).

Several genes associated with olfactory transduction were in selection signatures, mostly located on OAR 11. The *ARRB2* (Arrestin beta 2, OAR 11) gene is involved in regulating anti-inflammatory responses in the lungs (Walker *et al.*, 2003). The *GUCY2D* (Guanylate cyclase 2D retinal, OAR 11) gene encodes a transmembrane GC receptor protein expressed in the murine olfactory epithelia (Yuge, 2021). The *DTMT* (Olfactory receptor-like protein, OAR 11) gene is involved in the binding of odour molecules for olfaction (Muñoz *et al.*, 2019).

The notable genes within ROH regions associated with metabolic and immune response pathways (*DNAH2, DNAH9, MYH1, MYH10, MYH8* and *ARRB2*) relate to selection pressure for traits necessary for survival in harsh, arid environments of South Africa (Mdladla *et al.,* 2018).

5.6 Conclusion

It is important to characterize and monitor autozygosity levels within a population to conserve population genetic diversity and ensure the sustainability of future breeding programs. In this study, genome-wide SNP data of 1 738 SA Merino sheep were utilized to characterize their autozygosity and inbreeding levels.

Observed and expected heterozygosity as well as MAF, LD, and Ne estimates were quantified as proxy-indicators of the genetic variability present within the SA Merino breed. The reported H_O and H_E were indicative of a gain in within-population genetic diversity, and conveys that mating programs and intense selection are well managed within the breeding objectives of SA Merino sheep. Additionally, high MAF values were reported along with low levels of LD within the SA Merino population. The relationship between MAF and LD presents opportunities to further develop genomic applications such as imputation. The Food and Agriculture Organization (FAO) (1998) proposed a threshold value of 50<Ne>500 animals per herd to prevent the loss of genetic diversity. The values reported in this study were within the recommendations provided by the FAO for a healthy genetic diversity status, and indicate that genetic diversity is maintained at a reasonable level for the SA Merino.

Runs of homozygosity (ROH) and genomic estimators of inbreeding, i.e. genomic F-statistics, were used to reflect the autozygosity and inbreeding levels in the SA Merino sheep breed. The ROH length profiles were indicative of more ancient inbreeding and may reflect historic population bottlenecks. Protein-coding genes found within the selection signatures were associated with various functions in sheep involving reproduction, fibre formation, inflammatory responses, wool production, lactate metabolism, DNA, lipid, and carbohydrate synthesis.

The sufficient levels of genetic diversity retained in the current study population indicate that the SA Merino sheep breed has the potential for future sustainable production. Accounting for the genetic diversity within the SA Merino sheep breed allows for more widespread exploration of genomic applications such as genomic selection (GS) within the breed, and if

successfully applied, can pave the way for (and be transferable to) GS in other SA sheep breeds.

References

Abied, A., Xu, L., Sahlu, B. W., Xing, F., Ahbara, A., Pu, Y., Lin, J., Berihulay, H., Islam, R., He, X., Mwacharo, J. M., Zhao, Q., & Ma, Y. 2020. Genome-wide analysis revealed homozygosity and demographic history of five Chinese sheep breeds adapted to different environments. Genes. 11, 1–16 https://doi.org/10.3390/genes11121480.

Ablondi, M., Sabbioni, A., Stocco, G., Cipolat-Gotet, C., Dadousis, C., Kaam, J. T. van, Finocchiaro, R., & Summer, A. 2022. Genetic diversity in the Italian Holstein dairy cattle based on pedigree and SNP data prior and after genomic selection. Front. Vet. Sci. 8, 773985. https://doi.org/10.3389/fvets.2021.773985.

Agricultural survey. 2021. Department of Statistics South Africa, Africa. https://www.statssa.gov.za/publications/P1101/P11012022.pdf. 27 October 2023.

Aguilar, I., & Misztal, I. 2008. Technical note: Recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. J. Dairy Sci. 91, 1669–1672 https://doi.org/10.3168/jds.2007-0575.

Alberto, F. J., Boyer, F., Orozco-Terwengel, P., Streeter, I., Servin, B., De Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., Barbato, M., Zamani, W., Alberti, A., Engelen, S., Stella, A., Joost, S., Ajmone-Marsan, P., Negrini, R., Orlando, L., Rezaei, H. R., Naderi, S., Clarke, L., Flicek, P., Wincker, P., Coissac, E., Kijas, J., Tosser-Klopp, G., Chikhi, A., Bruford, M. W., Taberlet, P., & Pompanon, F. 2018. Convergent genomic signatures of domestication in sheep and goats. Nat. Commun. 9. https://doi.org/10.1038/s41467-018-03206-y.

Allendorf, F., Gordon, H., Luikart, S., & Aitken, N. 2013. Conservation and the Genetics of Populations. Second Edition. Wiley-Blackwell: Hoboken, NJ, USA. 533-557.

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

Almeida, O. A. C., Moreira, G. C. M., Rezende, F. M., Boschiero, C., De Oliveira Peixoto, J., Ibelli, A. M. G., Ledur, M. C., De Novais, F. J., & Coutinho, L. L. 2019. Identification of selection signatures involved in performance traits in a paternal broiler line. BMC Genomics 20, 449. https://doi.org/10.1186/s12864-019-5811-1.

Alvarez, G., Ceballos, F. C., & Quinteiro, C. 2009. The role of inbreeding in the extinction of a European royal dynasty. PLoS One 4, 5174. https://doi.org/10.1371/journal.pone.0005174.

Anderson, E. C., & Garza, J. C. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172, 2567–2582 https://doi.org/10.1534/genetics.105.048074.

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., & Zondervan, K. T. 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–1573 https://doi.org/10.1038/nprot.2010.116.

Anderson, R. 2014 Development of a high density (600K) Illumina Ovine SNP Chip and its use to fine map the yellow fat locus. In Plant and Animal Genome XXII Conference (San Diego, CA)(Plant and Animal Genome, 2014).

Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., Maddox, J. F., McEwan, J. C., Hutton Oddy, V., Raadsma, H. W., Wade, C., Wang, J., Wang, W., & Xun, X. 2010. The sheep genome reference sequence: A work in progress. Anim Genet 41, 449–453 https://doi.org/10.1111/j.1365-2052.2010.02100.x.

Arcos-Burgos, M., & Muenke, M. 2002. Genetics of population isolates. Clin. Genet. 61, 233–247 https://doi.org/10.1034/j.1399-0004.2002.610401.x.

Baes, C. F., Makanjuola, B. O., Miglior, F., Marras, G., Howard, J. T., Fleming, A., & Maltecca, C. 2019. Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. J. Dairy. Sci. 102, 2807–2817. https://doi.org/10.3168/jds.2018-15520.

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. 2015. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in Genetics. 6. https://doi.org/10.3389/fgene.2015.00109.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. 2018. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci. Data. 5, 214. https://doi.org/10.1038/sdata.2018.214.

Benhin, J. K. 2006. Climate change and South African agriculture: Impacts and adaptation options. Centre for Environmental Economics and Policy in Africa (CEEPA), 21, 113.

Best, D. J., & Roberts, D. E. 1975. Algorithm AS 89: The Upper Tail Probabilities of Spearman's Rho. Appl. Stat. 24, 377. https://doi.org/10.2307/2347111.

Beynon, S. E., Slavov, G. T., Farré, M., Sunduimijid, B., Waddams, K., Davies, B., Haresign, W., Kijas, J., MacLeod, I. M., Newbold, C. J., Davies, L., & Larkin, D. M. 2015. Population structure and

history of the Welsh sheep breeds determined by whole genome genotyping. BMC Genet. 16, 65. https://doi.org/10.1186/s12863-015-0216-x.

Bila, L., Tyasi, T. L., & Khetsha, Z. P. 2022. Trends estimation in annual numbers of Merino sheep and wool production in South Africa, from 1980 to 2017. Online Journal of Animal and Feed Research. 12, 255–260. https://doi.org/10.51227/ojafr.2022.34.

Biscarini, F., Cozzi, P., Gaspa, G., and Marras, G. 2018. *detectRUNS:* an R package to detect runs of homozygosity and heterozygosity in diploid genomes. Available at: https://cran.rproject.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references.

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G., & Ciampolini, R. 2020. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. Animals. 10, 1–17. https://doi.org/10.3390/ani10122285.

Bizarria dos Santos, W., Pimenta Schettini, G., Fonseca, M. G., Pereira, G. L., Loyola Chardulo, L. A., Rodrigues Machado Neto, O., Baldassini, W. A., Nunes de Oliveira, H., & Abdallah Curi, R. 2021. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. Journal of Animal Breeding and Genetics 138, 161–173 https://doi.org/10.1111/jbg.12508.

Bjelland, D. W., Weigel, K. A., Vukasinovic, N., & Nkrumah, J. D. 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J. Dairy. Sci. 96, 4697–4706 https://doi.org/10.3168/jds.2012-6435.

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

Bohmanova, J., Sargolzaei, M., & Schenkel, F. S. 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics 11, 421. http://www.biomedcentral.com/1471_2164/11/421.

Bonato, M., Cloete, J. J. E., Kruger, A. C. M., & Cloete, S. W. P. 2021. Behavioural reactivity of two lines of South African Merino sheep divergently selected for reproductive potential. Appl. Anim. Behav. Sci. 234, 105-209. https://doi.org/10.1016/j.applanim.2020.105209.

Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A. F., Schook, L. B., Crooijmans, R. P. M. A., & Groenen, M. A. M. 2012. Regions of Homozygosity in the Porcine Genome: Consequence

of Demography and the Recombination Landscape. PLoS Genet. 8, 100-103. https://doi.org/10.1371/journal.pgen.1003100.

Broman, K. W., & Weber, J. L. 1999. Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Etude du Polymorphisme Humain. Am. J. Hum. Genet. 65,1493–1500.

Bruford, M. W., Bradley, D. G., & Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics. 4, 900–910. https://doi.org/10.1038/nrg1203.

Buduram, P. 2004. Genetic characterization of Southern African sheep breeds using DNA markers.

Burger, M. 2019. Studies on the genetics of ovine behaviour in a Western Cape resource flock aimed at improving animal welfare. Diss. Stellenbosch University. Faculty of AgriSciences.

Caballero, A., Quesada, H., & Rolán-Alvarez, E. 2008. Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. Genetics. 179, 539–554. https://doi.org/10.1534/genetics.107.083246.

Carothers, A. D., Rudan, I., Kolcic, I., Polasek, O., Hayward, C., Wright, A. F., Campbell, H., Teague, P., Hastie, N. D., & Weber, J. L. 2006. Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. Ann. Hum. Genet. 70, 666–676. https://doi.org/10.1111/j.1469-1809.2006.00263.x.

Ceccobelli, S., Landi, V., Senczuk, G., Mastrangelo, S., Sardina, M. T., Ben-Jemaa, S., Persichilli, C., Karsli, T., *et al.,* 2023. A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds. Genetics Selection Evolution. 55, 24. https://doi.org/10.1186/s12711-023-00797-z.

Charlesworth, D., & Willis, J. H. 2009. The genetics of inbreeding depression. Nat. Rev. Genet. 10, 783–796. https://doi.org/10.1038/nrg2664.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. 2018. Graphical Methods for Data Analysis. 1st ed. Chapman and Hall/CRC.

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

Climate Change Knowledge Portal: South Africa Current Climate - Climatology. 2021. The world bank group.

Cloete, S. W. P., Gilmour, A. R., Olivier, J. J., & Van Wyk, J. B. 2004. Genetic and phenotypic trends and parameters in reproduction, greasy fleece weight and liveweight in Merino lines divergently selected for multiple rearing ability. Australian Journal of Experimental Agriculture. 44, 745–754. https://doi.org/10.1071/EA02225.

Cloete, S., & Olivier, J. J. 2010. South African sheep and wool industry. International Sheep and Wool Handbook.95-112.

Cloete, S. W. P., Olivier, J. J., Sandenbergh, L., & Snyman, M. A. 2015. The adaption of the South Africa sheep industry to new trends in animal breeding and genetics: A review. S. Afr. J. Anim. Sci. 44, 307–321. https://doi.org/10.4314/sajas.v44i4.1.

Cloete, S. W. P., Olivier, J. J., Snyman, M. A., & du Toit, E. 1998. Genetic parameters and trends in a selection experiment for increased clean fleece weight involving South African Merinos. Australian Journal of Experimental Agriculture. 38, 427. https://doi.org/10.1071/EA97071.

Cloete, S. W. P., Olivier, J. J., & du Toit, E. 2001. Blowfly strike of Merino sheep in relation to selection strategy, as well as to objective and subjective wool traits. Proceedings of the FLICS Conference, Launceston.

Cockerham, C., & Weir, B. 1968. Sib mating with two linked loci. Genetics 60, 629-640.

Cole, J. B., Lewis, R. M., Maltecca, C., Newman, S., Olson, K. M., & Tait, R. G. 2013. Breeding and genetics symposium: Systems biology in animal breeding: Identifying relationships among markers, genes, and phenotypes 1. J. Anim. Sci 91, 521–522 https://doi.org/10.2527/jas2012-6166.

Corbin, L. J., Blott, S. C., Swinburne, J. E., Vaudin, M., Bishop, S. C., & Woolliams, J. A. 2010. Linkage disequilibrium and historical effective population size in the Thoroughbred horse.Pages 8–15 in Animal Genetics.

Crow, J. 1954. Breeding structure of populations. II. Effective population number. Statistics and mathematics in biology 543, 556.

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Giron, C. G., Genez, T., Martinez, J. G., Guijarro-Clarke, C., Gymer, A., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Juettemann, T., Kaikala, V., Kay, M., Lavidas, I., Le, T., Lemos, D., Marugán, J. C., Mohanan, S., Mushtaq, A., Naven, M., Ogeh, D. N., Parker, A., Parton, A., Perry, M., Pilizota, I., Prosovetskaia, I., Sakthivel,

M. P., Salam, A. I. A., Schmitt, B. M., Schuilenburg, H., Sheppard, D., Perez-Silva, J. G., Stark, W., Steed, E., Sutinen, K., Sukumaran, R., Sumathipala, D., Suner, M. M., Szpak, M., Thormann, A., Tricomi, F. F., Urbina-Gómez, D., Veidenberg, A., Walsh, T. A., Walts, B., Willhoft, N., Winterbottom, A., Wass, E., Chakiachvili, M., Flint, B., Frankish, A., Giorgetti, S., Haggerty, L., Hunt, S. E., Iisley, G. R., Loveland, J. E., Martin, F. J., Moore, B., Mudge, J. M., Muffato, M., Perry, E., Ruffier, M., Tate, J., Thybert, D., Trevanion, S. J., Dyer, S., Harrison, P. W., Howe, K. L., Yates, A. D., Zerbino, D. R., & Flicek, P. 2022. Ensembl 2022. Nucleic Acids Res. 50, 988–995. https://doi.org/10.1093/nar/gkab1049.

Dadousis, C., Ablondi, M., Cipolat-Gotet, C., van Kaam, J. T., Marusi, M., Cassandro, M., Sabbioni, A., & Summer, A. 2022. Genomic inbreeding coefficients using imputed genotypes: Assessing different estimators in Holstein-Friesian dairy cows. J. Dairy Sci. 105, 5926–5945 https://doi.org/10.3168/jds.2021-21125.

DAFF. 2019. Abstract of Agricultural Statistics 2019, Department of Agriculture, Forestry and Fisheries, Africa. https://www.daff.gov.za/Daffweb3/Portals/0/StatisticsandEconomicAnalysis/StatisticalInformation/Abstract2017.pdf. 29 June 2023.

DALRRD. 2022. Abstract of Agricultural statistics. Department of Agriculture, Land Reform & Rural Development, Africa. https://www.dalrrd.gov.za/phocadownloadpap/Statistics_and_Economic_Analysis/Statistical_Information/Abstract%202022.pdf. 26 October 2023.

Da Silva, J. M., Giachetto, P. F., Campos Da Silva, L. O., Cintra, L. C., Paiva, S. R., Caetano, A. R., & Yamagishi, M. E. B. 2015. Genomic variants revealed by invariably missing genotypes in Nelore cattle. PLoS One. 10, 0136035. https://doi.org/10.1371/journal.pone.0136035.

Davenport, K. M., Bickhart, D. M., Worley, K., Murali, S. C., Salavati, M., Clark, E. L., Cockett, N. E., Heaton, M. P., Smith, T. P. L., Murdoch, B. M., & Rosen, B. D. 2022. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. Gigascience. 11, 1-9. https://doi.org/10.1093/gigascience/giab096.

de Cara, M. A. R., Fernández, J., Toro, M. A., & Villanueva, B. 2011. Using genome-wide information to minimize the loss of diversity in conservation programmes. J. Anim. Breed. Genet. 128, 456–464. https://doi.org/10.1111/j.1439-0388.2011.00971.x.

Dixit, S. P., Singh, S., Ganguly, I., Bhatia, A. K., Sharma, A., Kumar, N. A., Dang, A. K., & Jayakumar, S. 2020. Genome-wide runs of homozygosity revealed selection signatures in Bos indicus. Front. Genet. 11. https://doi.org/10.3389/fgene.2020.00092.

Dekkers, J. C. M., & Hospital, F. 2002. The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3, 22–32. https://doi.org/10.1038/nrg701.

Delport, M., Louw, M., Davids, T., Vermeulen, H., & Meyer, F. 2017. Evaluating the demand for meat in South Africa: an econometric estimation of short-term demand elasticities. Agrekon. 56, 13–27. https://doi.org/10.1080/03031853.2017.1286249.

Dodds, K. G., Auvray, B., Pickering, N., & Mcewan, J. C. 2018. Quality control for Ovine SNP50 Beadchip genotypes. Biotechnology Tools and Challenges. 18, 296–299.

Dzomba, E. F., Chimonyo, M., Snyman, M. A., & 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.

Dzomba, E. F. 2021. Genetic diversity and differentiation of pelt, mutton, and wool sheep breeds of South Africa using genome-wide single nucleotide polymorphisms. Th. University of KwaZulu-Natal. Faculty Agriculture, Engineering and Science.

Dzomba, E. F., Chimonyo, M., Pierneef, R., & Muchadeyi, F. C. 2021. Runs of homozygosity analysis of South African sheep breeds from various production systems investigated using OvineSNP50k data. BMC Genomics. 22, 7. https://doi.org/10.1186/s12864-020-07314-2.

Dzomba, E. F., Van Der Nest, M. A., Mthembu, J. N. T., Soma, P., Snyman, M. A., Chimonyo, M., & Muchadeyi, F. C. 2023. Selection signature analysis and genome-wide divergence of South African Merino breeds from their founders. Front Genet 13 https://doi.org/10.3389/fgene.2022.932272.

Engelsma, K. A., Veerkamp, R. F., Calus, M. P. L., & Windig, J. J. 2014. Consequences for diversity when animals are prioritized for conservation of the whole genome or of one specific allele. J. Anim. Breed. Genet. 131, 61–70. https://doi.org/10.1111/jbg.12052.

Erasmus, G. J., De Lange, A. O., Delport, G. J., & Olivier, J. J. 1990. Genetic and phenotypic parameter estimates of production traits of Merino sheep in an arid environment. S. Afr. J. Anim. Sci 20, 1.

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

Eusebi, P. G., Martinez, A., & Cortes, O. 2020. Genomic tools for effective conservation of livestock breed diversity. Diversity. 12, 8. https://doi.org/10.3390/d12010008.

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

Ferenčaković, M. 2015. Molecular dissection of inbreeding depression for semen quality traits in cattle. Diss. University of Zagreb. Faculty of Agriculture.

Ferenčaković, M., Hamzic, E., Gredler, B., Curik, I., & Sölkner, J. 2011. Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. Agriculturae Conspectus Scienti cus. 76, 325–328.

Fernández, A., Toro, M. A., & Lopez-Fanjul, C. 1995. The effect of inbreeding on the redistribution of genetic variance of fecundity and viability in Tribollum castaneum. Heredity. 75, 376-381.

Fernández, J., Villanueva, B., Pong-Wong, R., & Toro, M. Á. 2005. Efficiency of the use of pedigree and molecular marker information in conservation programs. Genetics. 170, 1313–1321. https://doi.org/10.1534/genetics.104.037325.

Frankham, R., Briscoe, D., & Ballou, J. 2002. Introduction to Conservation Genetics. C. R. Biologies. 326, 22-29.

Freeman, A. R., Bradley, D. G., Nagda, S., Gibson, J. P., & Hanotte, O. 2006. Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. Anim. Genet. 37, 1–9 https://doi.org/10.1111/j.1365-2052.2005.01363.x.

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

Ghoreishifar, M., Vahedi, S. M., Salek Ardestani, S., Khansefid, M., & Pryce, J. E. 2023. Genome-wide assessment and mapping of inbreeding depression identifies candidate genes associated with semen traits in Holstein bulls. BMC Genomics. 24, 230. https://doi.org/10.1186/s12864-023-09298-1.

Goddard, M. E., & Hayes, B. J. 2007. Genomic selection. J. Anim. Breed. Genet. 124, 323–330. https://doi.org/10.1111/j.1439-0388.2007.00702.x.

Goddard, M. E. 2012. Uses of genomics in livestock agriculture. Anim. Prod. Sci. 52, 73–77 https://doi.org/10.1071/AN11180.

Gómez-Romano, F., Villanueva, B., Sölkner, J., de Cara, M. A. R., Mészáros, G., Pérez O'Brien, A. M., & Fernández, J. 2016. The use of coancestry based on shared segments for maintaining genetic diversity. J. Anim. Breed. Genet. 133, 357–365. https://doi.org/10.1111/jbg.12213.

Gorkhali, N., & Yuehui, M. 2015. Genetic diversity, differentiation, and relationship of Nepalese sheep genetic resources: using DNA markers. Th. Chinese Academy of Agricultural sciences, Institute of Animal Science. Faculty Animal Genetics, Breeding and Reproduction, (Beijing China). 22-35.

Granero, A., Anaya, G., Demyda-Peyrás, S., Alcalde, M. J., Arrebola, F., & Molina, A. 2022. Genomic population structure of the main historical genetic lines of Spanish Merino Sheep. Animals. 12, 1327. https://doi.org/10.3390/ani12101327.

Greeff, J. C., Faure, A. S., Minnaar, G. J., & Schoeman, S. J. 1991. Genetic and phenotypic parameters of pelt traits in a Karakul control flock. S.Afr.Tydskr.Veek. 21, 3.

Grilz-Seger, G., Neuditschko, M., Ricard, A., Velie, B., Lindgren, G., Mesarič, M., Cotman, M., Horna, M., Dobretsberger, M., Brem, G., & Druml, T. 2019. Genome-wide homozygosity patterns and evidence for selection in a set of European and near Eastern horse breeds. Genes. 10. https://doi.org/10.3390/genes10070491.

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. 2010. Genetic diversity in farm animals - A review. Anim. Genet. 41, 6–31 https://doi.org/10.1111/j.1365-2052.2010.02038.x.

Guangul, S. 2014. Design of community-based breeding programs for two indigenous goat breeds of Ethiopia. Diss. University of Natural Resources and Life Sciences. Department of Sustainable Agricultural Systems.

Gurgul, A., Szmatoła, T., Topolski, P., Jasielczuk, I., Żukowski, K., & Bugno-Poniewierska, M. 2016. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. J. Appl. Genet. 57, 527–530. https://doi.org/10.1007/s13353-016-0337-6.

Gutiérrez, J. P., Cervantes, I., Molina, A., Valera, M., & Goyache, F. 2008. Individual increase in inbreeding allows estimating effective sizes from pedigrees. Genetics Selection Evolution. 40, 359–378. https://doi.org/10.1051/gse:2008008.

Handley, L. J., Byrne, K., Santucci, F., Townsend, S., Taylor, M., Bruford, M. W., & Hewitt, G. M. 2007. Genetic structure of European sheep breeds. Heredity. 99, 620–631. https://doi.org/10.1038/sj.hdy.6801039.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92, 433–443. https://doi.org/10.3168/jds.2008-1646.

Hayes, B. J., Visscher, P. M., McPartlan, H. C., & Goddard, M. E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research. 13, 635–643. https://doi.org/10.1101/gr.387103.

He, S., Di, J., Han, B., Chen, L., Liu, M., & Li, W. 2020. Genome-wide scan for runs of homozygosity identifies candidate genes related to economically important traits in Chinese Merino. Animals. 10, 524. https://doi.org/10.3390/ani10030524.

Herrero-Medrano, J., Megens, H.-J., Groenen, M. A., Ramis, G., Bosse, M., Pérez-Enciso, M., & PMA Crooijmans, R. 2013. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. BMC Genetics. 14, 1-13.

Herselman MJ., 2006. A survey on the effectiveness of different problem animal control methods in the small stock producing areas of South Africa. In *Research Report of the Grootfontein Agricultural Development Institute*. Department of Agriculture, 23-24. Private bag X529, Middelburg, 5900, Eastern Cape, South Africa.

Heydenrych, H. J., Du Plessis, J. J., & Cloete, S. W. P. 1984. Increasing the wool production of Merino sheep in the Southwestern Districts of South Africa by direct and indirect selection. In: Hofmeyr Jh, Meyer EHH (eds.), Proceedings of the second World Congress on Sheep and Beef Cattle Breeding, 16-19 April, Pretoria. South African Stud Book and Livestock Improvement Assoc., Bloemfontein. 399-412.

Hill, W. G., & Robertson, A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226-231.

Hill, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. Heredity. 33, 229–239.

Holt, M., Meuwissen, T., & Vangen, O. 2005. The effect of fast created inbreeding on litter size and body weights in mice. Genetics Selection Evolution. 37, 523–537. https://doi.org/10.1051/gse:2005014.

Howrigan, D. P., Simonson, M. A., & Keller, M. C. 2011. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. BMC Genomics. 12, 460. https://doi.org/10.1186/1471-2164-12-460.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. https://doi.org/10.1038/nprot.2008.211.

Hudson, N. J., Porto-Neto, L., Kijas, J. W., & Reverter, A. 2015. Compression distance can discriminate animals by genetic profile, build relationship matrices and estimate breeding values. Genetics Selection Evolution. 4, 78. https://doi.org/10.1186/s12711-015-0158-9.

International Sheep Genomics Consortium, Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., Maddox, J. F., McEwan, J. C., Hutton Oddy, V., Raadsma, H. W., Wade, C., Wang, J., Wang, W., & Xun, X. 2010. The sheep genome reference sequence: A work in progress. Anim. Genet. 41, 449–453. https://doi.org/10.1111/j.1365-2052.2010.02100.x.

Kaseja, K., Mucha, S., Smith, E., Yates, J., Banos, G., & Conington, J. 2023. Including genotypic information in genetic evaluations increases the accuracy of sheep breeding values. J. Anim. Breed. Genet. 140, 462–471. https://doi.org/10.1111/jbg.12771.

Keller, M. C., Visscher, P. M., & Goddard, M. E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189, 237–249. https://doi.org/10.1534/genetics.111.130922.

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Neto, L. R., Cristobal, M. S., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., & Dalrymple, B. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*. 10, 1001258. https://doi.org/10.1371/journal.pbio.1001258.

Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., Wilson, P., Ingersoll, R. G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V. H., Nicholas, F. W., Raadsma, H., Barris, W., Bishop, S. C., Coltman, D., Crawford, A., Eggen, A., Erhardt, G., Forage, R., Hanotte, O., Hunt, P., Jianlin, H., Li, K., Marsan, P. A., Miller, J. E., Pemberton, J., & Schibler, L. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One, 4, 1-13. https://doi.org/10.1371/journal.pone.0004668.

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassel, C. P., Crooker, B. A., Liu, G., Da, Y., & Sonstegard, T. S. 2013. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS One 8, 80813. https://doi.org/10.1371/journal.pone.0080813.

Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., & Rothschild, M. F. 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.

Kominakis, A., Hager-Theodorides, A. L., Zoidis, E., Saridaki, A., Antonakos, G., & Tsiamis, G. 2017. Combined GWAS and 'guilt by association'-based prioritization analysis identifies functional candidate genes for body size in sheep. Genetics Selection Evolution 49, 41 https://doi.org/10.1186/s12711-017-0316-3.

Kristensen, T. N., & Sørensen, A. C. 2005. Inbreeding – lessons from animal breeding, evolutionary biology, and conservation genetics. Anim. Sci. 80, 121–133. https://doi.org/10.1079/ASC41960121.

Ku, C. S., Naidoo, N., Teo, S. M., & Pawitan, Y. 2011. Regions of homozygosity and their impact on complex diseases and traits. Hum. Genet. 129, 1–15. https://doi.org/10.1007/s00439-010-0920-6.

Lashmar, S. F., Visser, C., van Marle-Köster, E., & Muchadeyi, F.C. 2018. Genomic diversity and autozygosity within the SA Drakensberger beef cattle breed. Livest. Sci. 112, 111–119.

Lashmar, S. F., Visser, C., Okpeku, M., Muchadeyi, F. C., Mapholi, N. O., & van Marle-Köster, E. 2022. A within- and across-country assessment of the genomic diversity and autozygosity of South African and eSwatini Nguni cattle. Trop Anim Health Prod 54 https://doi.org/10.1007/s11250-022-03373-7.

Lerner, I. M. 1954. Genetic homeostasis. John Wiley and Sons, New York. 62-67.

Leroy, G. 2014. Inbreeding depression in livestock species: Review and meta-analysis. Anim. Genet. 45, 618–628. https://doi.org/10.1111/age.12178.

Liu, Z., Bai, C., Shi, L., He, Y., Hu, M., Sun, H., Peng, H., Lai, W., Jiao, S., Zhao, Z., Ma, H., & Yan, S. 2022. Detection of selection signatures in South African Mutton Merino sheep using whole-genome sequencing data. Anim. Genet. 53, 224–229. https://doi.org/10.1111/age.13173.

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

Livestock development strategy for South Africa (LDS). 2006. Agricultural Information Services, 21-23. Private Bag X144, Pretoria 0001, South Africa.

Louw, M. 2019. Dual purpose sheep breeds in South Africa. *Sheep Farming in South Africa.* https://southafrica.co.za/dual-purpose-sheep-breeds-south-africa.html 20 May 2023.

Lush, J., Chapman, A., Shrode, R., & Crow, J. 1994. The Genetics of Populations. Iowa Agriculture and Home economics experiment station, College of Agriculture, Iowa State University. 32-34.

Magee, D. A., Park, S. D. E., Scraggs, E., Murphy, A. M., Doherty, M. L., Kijas, J. W., & Machugh, D. E. 2010. High fidelity of whole-genome amplified sheep (Ovis aries) deoxyribonucleic acid using a high-density single nucleotide polymorphism array-based genotyping platform. J. Anim. Sci. 88, 3183–3186. https://doi.org/10.2527/jas.2009-2723.

Mäki-Tanila, A., Fernandez, J., Toro, M., Meuwissen, Hiemstra, S., de Haas, Y., Mäki-Tanila, A., & Gandini, G. 2010. Assessment and management of genetic variation. Wageningen Academic Publishers. Netherlands, EU. 76-79.

Malesa, M. 2015. Population genetics of Swakara sheep inferred using genome-wide SNP genotyping. Diss. University of KwaZulu-Natal. Faculty of Agriculture, Engineering and Science.

Malécot, G. 1948. The mathematics of heredity. Masson and Cie, Paris. 63.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J. L., & Macciotta, N. P. P. 2015. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 46, 110–121. https://doi.org/10.1111/age.12259.

Marras, G.; Wood, B.J.; Makanjuola, B.; Malchiodi, F.; Peeters, K.; van As, P.; Baes, C.F.; Biscarini, F.; Turkeys, H. Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (Meleagris gallopavo) population. In Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand, 10–16 February 2018; p. 763.

Mason, I. L. 1996. A World Dictionary of Livestock Breeds, Types and Varieties. 4th ed. CAB International, Wallingford. UK. 1-17.

Mastrangelo, S., Ciani, E., Sardina, M. T., Sottile, G., Pilla, F., & Portolano, B. 2018. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Anim. Genet. 49, 71–81. https://doi.org/10.1111/age.12634.

Matebesi-Ranthimo, P. A. M., Cloete, S. W. P., Van Wyk, J. B., & Olivier, J. J. 2018. Relationship of ewe reproduction with subjectively assessed wool and conformation traits in the Elsenburg Merino flock. S. Afr. J. Anim. Sci. 48, 29. https://doi.org/10.4314/sajas.v48i1.4.

Mazinani, M., & Rude, B. 2020. Population, world production and quality of sheep and goat products. Am. J. Anim. Vet. Sci. 15, 291–299. https://doi.org/10.3844/ajavsp.2020.291.299.

McParland, S., Kearney, F., & Berry, D. P. 2009. Purging of inbreeding depression within the Irish Holstein-Friesian population. Genetics Selection Evolution 41, 16 https://doi.org/10.1186/1297-9686-41-16.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A. K., Farrington, S. M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S. H., Dunlop, M. G., Wright, A. F., Campbell, H., &

Wilson, J. F. 2008. Runs of Homozygosity in European Populations. Am. J. Hum. Genet. 83, 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007.

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

Meadows, J. R. S., Hanotte, O., Drögemüller, C., Calvo, J., Godfrey, R., Coltman, D., Maddox, J. F., Marzanov, N., Kantanen, J., & Kijas, J. W. 2006. Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. Anim. Genet. 37, 444–453. https://doi.org/10.1111/j.1365-2052.2006.01496.x.

Megdiche, S., Mastrangelo, S., Hamouda, M. Ben, Lenstra, J. A., & Ciani, E. 2019. A combined multi-cohort approach reveals novel and known genome-wide selection signatures for wool traits in merino and merino-derived sheep breeds. Front. Genet. 10, 1025. https://doi.org/10.3389/fgene.2019.01025.

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., Gut, I. G., & Distl, O. 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics. 16, 764. https://doi.org/10.1186/s12864-015-1977-3.

Meuwissen, T. 2009. Genetic management of small populations: A review. Acta Agriculturae Scandinavica A: Animal Sciences 59, 71–79. https://doi.org/10.1080/09064700903118148.

Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. 2020. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. BMC Genomics. 21. https://doi.org/10.1186/s12864-020-6463-x.

Microsoft Corporation. 2018. *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel (Accessed May 10, 2023).

Molotsi, A. H., Dube, B., & Cloete, S. W. P. 2020. The current status of indigenous ovine genetic resources in Southern Africa and future sustainable utilisation to improve livelihoods. Diversity. 12, 14. https://doi.org/10.3390/d12010014.

Mortimer, S. I., Atkins, K. D., Semple, S. J., & Fogarty, N. M. 2010. Predicted responses in Merino sheep from selection combining visually assessed and measured traits. Anim. Prod. Sci. 50, 976–982. https://doi.org/10.1071/AN10085.

Mpetile, Z., & Cloete, S. W. P. 2019. Genetics of ovine resistance to gastrointestinal nematodes. Diss. Stellenbosch University. Faculty of Science and Agriculture.

Muchadeyi, F. C., Malesa, M. T., Soma, P., & Dzomba, E. F. 2015. Runs of homozygosity in Swakara pelt producing sheep: Implications on sub-vital performance. Proc. Assoc. Advmt. Breed. Genet. 21, 310-313.

Mucina, L., & Rutherford, M. C. 2006. The Biomes and Vegetation of South Africa. The vegetation of South Africa, Lesotho, and Swaziland. South African National Biodiversity Institute. 807–821.

Mulim, H. A., Brito, L. F., Pinto, L. F. B., Ferraz, J. B. S., Grigoletto, L., Silva, M. R., & Pedrosa, V. B. 2022. Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. BMC Genomics. 23. https://doi.org/10.1186/s12864-022-08384-0.

Muñoz, M., Bozzi, R., García-Casco, J., Núñez, Y., Ribani, A., Franci, O., García, F., Škrlep, M., Schiavo, G., Bovo, S., Utzeri, V. J., Charneca, R., Martins, J. M., Quintanilla, R., Tibau, J., Margeta, V., Djurkin-Kušec, I., Mercat, M. J., Riquet, J., Estellé, J., Zimmer, C., Razmaite, V., Araujo, J. P., Radović, Č., Savić, R., Karolyi, D., Gallo, M., Čandek-Potokar, M., Fernández, A. I., Fontanesi, L., & Óvilo, C. 2019. Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. Sci Rep 9, 13-46. https://doi.org/10.1038/s41598-019-49830-6.

National agricultural research and development strategy. 2008. Department of Agriculture, Africa. https://www.dalrrd.gov.za/phocadownloadpap/Policy_Documents/National%20Agricultural%20Research%20and%20Development%20Strategy%202008.pdf. 13 November 2023.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. proceedings of the National Academy of Sciences. 70, 3321–3323.

Nel, C., Gurman, P., Swan, A., van der Werf, J., Snyman, M., Dzama, K., Olivier, W., Scholtz, A., & Cloete, S. 2024. Including genomic information in the genetic evaluation of production and reproduction traits in South African Merino sheep. Journal of Animal Breeding and Genetics. 141, 65–82. https://doi.org/10.1111/jbg.12826

Nel, C. L., Swan, A. A., Dzama, K., Olivier, W. J., & Snyman, M. A. 2022a. An across-flock genetic analysis of components of reproduction in four South African Merino resource flocks. In Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP). Wageningen Academic Publishers. 3081–3084.

Nel, C., Gurman, P., Swan, A., van der Werf, J., Snyman, M., Dzama, K., Gore, K., Scholtz, A., & Cloete, S. 2022b. The genomic structure of isolation across breed, country, and strain for important

South African and Australian sheep populations. BMC Genomics. 23, 23. https://doi.org/10.1186/s12864-021-08020-3.

Nel, C. L., Swan, A. A., Dzama, K., Scholtz, A. J., & Cloete, S. W. P. 2021. Genetic parameters and trends for lamb survival following long-term divergent selection for number of lambs weaned in the Elsenburg Merino flock. Anim. Prod. Sci. 61, 1965–1981. https://doi.org/10.1071/AN21198.

Neser, F. W. C., Konstantinov, K. V, & Erasmus, G. J. 1995. Estimated genetic trends for weaning weight in three Dorper lines with different selection criteria. S. Afr. J. Anim. Sci. 23, 65-69.

Nicholas, F. W., & Smith, C. 1983. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. Animal Science. 36, 341–353. https://doi.org/10.1017/S0003356100010382.

Nosrati, M., Asadollahpour Nanaei, H., Javanmard, A., & Esmailizadeh, A. 2021. The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. Genomics. 113, 1407–1415. https://doi.org/10.1016/j.ygeno.2021.03.005.

Nosrati, M., Nanaei, H. A., & Esmailizadeh, A. 2022. Estimation of runs of homozygosity reveals moderate autozygosity in Northern European sheep breeds. Journal of Livestock Science and Technologies. 9, 31–40. https://doi.org/10.22103/jlst.2021.17929.1373.

Oldenbroek, K., & van der Waaii, L. 2015. Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources, Netherlands. 67-71.

Oliver, J. 1995. Response to selection on BLUP of breeding values in the Grootfontein Merino stud. S. Afr. J. Anim. Sci. 25, 13.

Olivier, J. J., & Cloete, S. 2007. Genetic change in some South African Merino studs participating in the Merino plan. Proc. Assoc. Advmt. Anim. Breed. Genet. 17, 324–327.

Olivier, W. J., & Snyman, M. A. 2014. The evaluation of a South African fine wool genetic resource flock. Diss. Stellenbosch University. Faculty of AgriSciences.

Palmer, T., & Ainslie, A. 2002. South Africa: FAO Country Profile. Country Pasture/Forage Resource Profile, Rome: FAO. https://ees.kuleuven.be/eng/klimos/toolkit/documents/658_SouthAfrica_English.pdf, (Accessed on 25 January 2024).

Pearl, R. 1913. A Contribution Towards an Analysis of the Problem of Inbreeding. Am. Nat. 47, 577–614. https://www.jstor.org/stable/2455472.

Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., & Li, J. Z. 2012. Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. 91, 275–292. https://doi.org/10.1016/j.ajhg.2012.06.014.

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., & Baldi, F. 2017. Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 48, 255–271. https://doi.org/10.1111/age.12526.

Peripolli, E., Metzger, J., de Lemos, M. V. A., Stafuzza, N. B., Kluska, S., Olivieri, B. F., Feitosa, F. L. B., Berton, M. P., Lopes, F. B., Munari, D. P., Lôbo, R. B., Magnabosco, C. de U., Di Croce, F., Osterstock, J., Denise, S., Pereira, A. S. C., & Baldi, F. 2018. Autozygosity islands and ROH patterns in Nellore lineages: Evidence of selection for functionally important traits. BMC Genomics. 19, 680 https://doi.org/10.1186/s12864-018-5060-8.

Peter, C., Bruford, M., Perez, T., Dalamitra, S., Hewitt, G., & Erhardt, G. 2007. Genetic diversity and subdivision of 57 European and Middle Eastern sheep breeds. Anim. Genet. 38, 37–44. https://doi.org/10.1111/j.1365-2052.2007.01561.x.

Pirchner, F. 1985. Genetic structure of populations. In General and quantitative genetics. Elsevier, Amsterdam. 227–250.

Pryce, J. E., Hayes, B. J., & Goddard, M. E. 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. J. Dairy Sci. 95, 377–388. https://doi.org/10.3168/jds.2011-4254.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.

Purfield, D. C., Berry, D. P., Mcparland, S., & Bradley, D. G. 2012. Runs of homozygosity and population history in cattle. BMC Genetics. 13, 70–81.

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

Qwabe, S. O., van Marle-Köster, E., & Visser, C. 2012. Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. Trop. Anim. Health Prod. 45, 511–516. https://doi.org/10.1007/s11250-012-0250-x.

Ramsay, K. A., Hlatshwayo, M., & Mamabolo, M. J. 2021. The livestock development strategy (LDS) and its potential impact on the red meat marketing and development in South Africa. Directorate Animal and Aquaculture Production. Department of Agriculture. South Africa.

Rasali, D. P., B Shrestha, J. N., & Crow, G. H. 2006. Development of composite sheep breeds in the world: A review. Can. J. Anim. Sci. 86, 1–24.

Rebelato, A. B., & Caetano, A. R. 2018. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. Pesqui. Agropecu. Bras. 53, 975–984. https://doi.org/10.1590/s0100-204x2018000900001.

Retief, A. 2020. Whole genome investigation of the genetic structure of South African sheep breeds. Diss. University of Pretoria. Faculty of natural and Agricultural Sciences.

Robertson, A. 2008. Inbreeding in artificial selection programmes. Genet. Res. 89, 275–280. https://doi.org/10.1017/S0016672308009452.

Rodríguez-Ramilo, S. T., Reverter, A., Sánchez, J. P., Fernández, J., Velasco-Galilea, M., González, O., & Piles, M. 2020. Networks of inbreeding coefficients in a selected population of rabbits. J. Anim. Breed. Genet. 137, 599–608. https://doi.org/10.1111/jbg.12500.

Ron, M., Blanc, Y., Band, M., Ezra, E., & Weller, J. 1. 1996. Misidentification Rate in the Israeli Dairy Cattle Population and Its Implications for Genetic Improvement. J. Dairy Sci. 79, 676-681.

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

Rust, J. M. 2019. The impact of climate change on extensive and intensive livestock production systems. Anim. Front. 9, 20–25. https://doi.org/10.1093/af/vfy028.

Rust, J., & Rust, T. 2013. Climate change and livestock production: A review with emphasis on Africa. S. Afr. J. Anim. Sci. 43, 255. https://doi.org/10.4314/sajas.v43i3.3.

Sabatti, C., & Risch, N. 2002. Homozygosity and Linkage Disequilibrium. Genetics. 160, 1707-19. https://doi.org/10.1093/genetics/160.4.1707.

Sabir, J., Mutwakil, M., El-Hanafy, A., Al-Hejin, A., Sadek, M. A., Abou-Alsoud, M., Qureshi, M., Saini, K., & Ahmed, M. 2014. Applying molecular tools for improving livestock performance: From DNA markers to next generation sequencing technologies. Agriculture & Environment. 12, 541–553.

Samuels, D. C., Wang, J., Ye, F., He, J., Levinson, R. T., Sheng, Q., Zhao, S., Capra, J. A., Shyr, Y., Zheng, W., & Guo, Y. 2016. Heterozygosity ratio, a robust global genomic measure of autozygosity and its association with height and disease risk. Genetics. 204, 893–904. https://doi.org/10.1534/genetics.116.189936.

Sandenbergh, L., Cloete, S. W. P., Roodt-Wilding, R., Snyman, M. A., & Bester-van der Merwe, A. E. 2016. Evaluation of the ovineSNP50 chip for use in four South African sheep breeds. S. Afr. J. Anim. Sci. 46, 90–93. https://doi.org/10.4314/sajas.v46i1.11.

Sandenbergh, L., Cloete, S. W. P., Roodt-Wilding, R., Snyman, M. A., & Van Der Merwe, A. E. 2015. Genetic diversity and population structure of four South African sheep breeds. Proc. Assoc. Advmt. Genet., 294–297.

Sandenbergh, L. 2015. Identification of SNPs associated with robustness and greater reproductive success in the South African Merino sheep using SNP chip technology. Diss. Stellenbosch University. Department of Genetics.

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

Seidel, G. E. 2010. Brief introduction to whole-genome selection in cattle using single nucleotide polymorphisms. Reprod. Fertil. Dev. 22, 138–144. https://doi.org/10.1071/RD09220.

Selli, A., Ventura, R. V., Fonseca, P. A. S., Buzanskas, M. E., Andrietta, L. T., Balieiro, J. C. C., & Brito, L. F. 2021. Detection and Visualization of Heterozygosity-Rich Regions and Runs of Homozygosity in Worldwide Sheep Populations. Animals 11, 26-96. https://doi.org/10.3390/ani11092696.

Snyman, M. A. 2014. South African sheep breeds: Dohne Merino sheep. Info-pack ref. 22.

Snyman, M. A., & Olivier, W. J. 2002. Productive performance of hair and wool type Dorper sheep under extensive conditions. Small Ruminant Research. 45, 17–23. https://doi.org/10.1016/S0921-4488(02)00116-5.

Snyman, M. A. 2001. Evaluation of a genetically fine mohair producing herd. Small Ruminant Research. 43, 105–111. https://doi.org/10.1016?S0921-4488(01)00270-x.

Snyman, M. A., Cloete, S. W. P., & Oliver, J. J. 1998. Genetic and phenotypic correlations of total weight of lamb weaned with body weight, clean fleece weight and mean fibre diameter in three South African Merino flocks. Livest. Prod. Sci. 55, 157–162. https://doi.org/10.1016/S0301-6226(98)00119-5.

Snyman, M. A., Olivier, J. J., & Olivier, W. J. 1996. Variance components and genetic parameters for body weight and fleece traits of Merino sheep in an arid environment. S.Afr.J.Anim.Sci. 26, 1. https://doi.org/10.10520/EJC-1f0433c3d8.

Snyman, M. A., Olivier, J. J., & Cloete, J. A. N. 1993. Productive and reproductive performance of Namaqua Afrikaner sheep. Karoo Agriculture. 5, 21-24.

Sölkner, J., Ferencakovic, M., Gredler, B., & Curik, I. 2010. Genomic metrics of individual autozygosity. applied to a cattle population. Proceedings of the 61st Annual Meeting of the European Association of Animal Production. Heraklion, Greece, 23.-27.8. 2010, Session code G38.03.

Soma, P., Kotze, A., Grobler, J. P., & van Wyk, J. B. 2012. South African sheep breeds: Population genetic structure and conservation implications. Small Ruminant Research. 103, 112–119. https://doi.org/10.1016/j.smallrumres.2011.09.041.

Sved, J. A., & Feldman, M. W. 1973. Correlation and probability methods for one and two loci. Theor. Popul. Biol. 4, 129–132 https://doi.org/10.1016/0040-5809(73)90008-7.

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., & Li, J. Z. 2013. Long runs of homozygosity are enriched for deleterious variation. Am. J. Hum. Genet. 93, 90–102. https://doi.org/10.1016/j.ajhg.2013.05.003.

Taljaard, P. R., Jooste, A., & Asfaha, T. A. 2006. Towards a broader understanding of South African consumer spending on meat. Agrekon. 45, 214–224. https://doi.org/10.1080/03031853.2006.9523744.

Tier, B. 1990. Computing inbreeding coefficients quickly. Genetic Science Evolution. 22, 419–430.

Toro, M. A., Fernández, J., & Caballero, A. 2009. Molecular characterization of breeds and its use in conservation. Livest. Sci. 120, 174–195. https://doi.org/10.1016/j.livsci.2008.07.003.

van der Merwe, D. A., Brand, T. S., & Hoffman, L. C. 2019. Application of growth models to different sheep breed types in South Africa. Small Ruminant Research. 178, 70–78. https://doi.org/10.1016/j.smallrumres.2019.08.002.

van Heerden, W. J. 2022. Merino SA focus 2022. CADAR printers. Gqeberha, SA. 23-26.

van Marle-Köster, E., Visser, C., Makgahlela, M., & Cloete, S. W. P. 2015. Genomic technologies for food security: A review of challenges and opportunities in Southern Africa. Food Research International. 76, 971–979. https://doi.org/10.1016/j.foodres.2015.05.057.

van Marle-Köster, E., Lashmar, S. F., Retief, A., & Visser, C. 2021. Whole-Genome SNP Characterisation Provides Insight for Sustainable Use of Local South African Livestock Populations. Front. Genet. 12, 714194. https://doi.org/10.3389/fgene.2021.714194.

VanRaden, P. M. 1992. Genetics and breeding Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations. J. Dairy Sci. 75, 3136–3144.

van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., Haudenschild, C. D., Moore, S. S., Warren, W. C., & Sonstegard, T. S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat. Methods. 5, 247–252. https://doi.org/10.1038/nmeth.1185.

Vekemans, X., Beauwens, T., Lemaire, M., & Roldán-Ruiz, I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Mol. Ecol. 11, 139–151. https://doi.org/10.1046/j.0962-1083.2001.01415.x.

Verrier, E., Colleau, J. J., & Foulley, J. L. 1993. Long-term effects of selection based on the animal model BLUP in a finite population. Theor. Appl. Genet. 87, 446–454.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution. 34, 275–305 https://doi.org/10.1051/gse:2002009.

Vigne, J. D. 2011. The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere. C. R. Biol. 334, 171–181 https://doi.org/10.1016/j.crvi.2010.12.009.

Vink N and Kirsten J (2002) Pricing behaviour in the South African food and agricultural sector Final report commissioned by the National Treasury and conducted with additional technical support from the National Departments of Agriculture, Trade and Industries and the Competition Commission, June 2002. http://www.nda.agric.za/docs/food_prices_report.pdf (Accessed 10 October 2023).

Visser, C., Lashmar, S. F., Reding, J., Berry, D. P., & van Marle-Köster, E. 2023. Pedigree and genome-based patterns of homozygosity in the South African Ayrshire, Holstein, and Jersey breeds. Front. Genet. 14, 3389. https://doi.org/10.3389/fgene.2023.1136078.

Visser, C. 2019. A review on goats in southern Africa: an untapped genetic resource. Small Ruminant Research. 176, 11-16.

Visser, C., & van Marle-Köster, E. 2014. Strategies for the genetic improvement of South African Angora goats. Small Ruminant Research. 121, 89–95. https://doi.org/10.1016/j.smallrumres.2014.01.012.

Walker, J. K. L., Fong, A. M., Lawson, B. L., Savov, J. D., Patel, D. D., Schwartz, D. A., & Lefkowitz, R. J. 2003. β-Arrestin-2 regulates the development of allergic asthma. Journal of Clinical Investigation 112, 566–574 https://doi.org/10.1172/JCI17265.

Wang, Y., Niu, Z., Zeng, Z., Jiang, Y., Jiang, Y., Ding, Y., Tang, S., Shi, H., & Ding, X. 2020. Using high-density SNP array to reveal selection signatures related to prolificacy in Chinese and Kazakhstan sheep breeds. Animals 10, 1–15 https://doi.org/10.3390/ani10091633.

Wanjala, G., Astuti, P. K., Kichamu, N., Bagi, Z., & Kusza, S. 2023. Assessing Genomic Inbreeding and Homozygosity Patterns in Hungarian Merino Sheep and its Relatives. Scientific papers animal science and biotechnologies. 1, 47.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., & Tingey, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18, 6531–6535 https://doi.org/10.1093/nar/18.22.6531.

Williams, J. L., Hall, S. J. G., Del Corvo, M., Ballingall, K. T., Colli, L., Ajmone Marsan, P., & Biscarini, F. 2016. Inbreeding and purging at the genomic Level: The Chillingham cattle reveal extensive, non-random SNP heterozygosity. Anim. Genet. 47, 19–27. https://doi.org/10.1111/age.12376.

Woolliams, J. A., & Oldenbroek, J. K. 2017. Chapter 1. Genetic diversity issues in animal populations in the genomic era. In Genomic management of animal genetic diversity. Wageningen Academic Publishers, The Netherlands. 13–47.

Wright S. 1978. Variability within and amoung Natural Population, 4. University of Chicago Press. Evolution and the genetics of populations, Chicago. 98-110.

Wright, S. 1974. Evolution and the genetics of populations, 4. University of Chicago Press, Chicago. 101-112.

Wright, S. 1922. Coefficients of Inbreeding and Relationship. Am. Nat. 56, 330–338. https://www.jstor.org/stable/2456273.

Xu, Z., Sun, H., Zhang, Z., Zhao, Q., Olasege, B. S., Li, Q., Yue, Y., Ma, P., Zhang, X., Wang, Q., & Pan, Y. 2019. Assessment of autozygosity derived from runs of homozygosity in Jinhua pigs disclosed by sequencing data. Front. Genet. 10, 274. https://doi.org/10.3389/fgene.2019.00274.

Yan, J., Shah, T., Warburton, M. L., Buckler, E. S., McMullen, M. D., & Crouch, J. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One. 4, 8451. https://doi.org/10.1371/journal.pone.0008451.

Yuge, S. 2021. Handbook of Hormones: Comparative Endocrinology for Basic and Clinical Research.Pages 349–352 in Handbook of Hormones. 2nd ed. Elsevier.

Zavarez, L. B., Utsunomiya, Y. T., Carmo, A. S., Neves, H. H. R., Carvalheiro, R., Ferencakovic, M., Pérez O'Brien, A. M., Curik, I., Cole, J. B., Van Tassell, C. P., da Silva, M. V. G. B., Sonstegard, T. S., Sölkner, J., & Garcia, J. F. 2015. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. Front. Genet. 6, 5. https://doi.org/10.3389/fgene.2015.00005.

Zavattari, P. 2000. Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. Hum. Mol. Genet. 9, 2947–2957.

Zeder, M. A. 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proc. Natl. Acad. Sci. USA. 105, 11597–11604 https://doi.org/10.1073/pnas.0801317105.

Zhai, W., Todd, M. J., & Nielsen, R. 2004. Is haplotype block identification useful for association mapping studies? Genet Epidemiol 27, 80–83 https://doi.org/10.1002/gepi.20014.

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. https://doi.org/10.1186/s12864-015-1715-x.

Zohary, D., Tchernov, E., & Horwitz, L. K. 1998. The role of unconscious selection in the domestication of sheep and goats. J. Zool. 245, 129–135. https://doi.org/10.1111/j.1469-7998.1998.tb00082.x.

Addendum

Table A1. Average F_{ROH} values per chromosome.

OAR	Average F_{ROH}
1	0.0762
2	0.0885
3	0.0930
4	0.1096
5	0.1012
6	0.1117
7	0.1066
8	0.1145
9	0.1251
10	0.1128
11	0.1672
12	0.1139
13	0.1120
14	0.1262
15	0.1151
16	0.1214
17	0.1186
18	0.1355
19	0.1381
20	0.1502
21	0.1635
22	0.1497
23	0.1355
24	0.1797
25	0.1550
26	0.1438

Tables A2. Comprehensive summary of the associated functions of each protein-coding gene located within the detected ROH regions.

Associated function	No of genes	Significance level (p-value)	Gene names
Arachidonic acid metabolic process &	6	9.00E-08	ALOX12B, ALOX12, LOC101121185, ALOX15, ALOXE3
Actin filament binding	9	4.70E-05	
Inoleic acid metabolic process	5	1.70E-07	
Lipoxygenase pathway	5	1.70E-07	
Linoleate 13S-lipoxygenase activity	4	4.70E-06	
Lipoxin A4 biosynthetic process	3	3.30E-04	
Cytoplasm	37	6.90E-04	
Myofibril & Myosin complex (Motor	4	9.40E-04	DNAH2, DNAH9, KIF1C, MYH1, MYH10, MYH2, MYH8, LOC101113252, LOC101111980
Proteins)	5	0.001	
Response to heat	4	1.50E-03	
Oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	3	0.003	
Microtubule associated complex	3	4.10E-03	
Histone deacetylase binding	5	9.50E-03	
Microfilament motor activity	3	0.011	
ATP binding	25	0.012	
Negative regulation of transcription from RNA polymerase II promoter	10	0.014	
ATP metabolic process	3	0.018	
Olfactory receptor activity	15	0.019	ARRB2, GUCY2D, LOC101119277, LOC101106784, LOC101107296, LOC115611253, LOC114117013, LOC101120175, LOC101119916, LOC101120421, LOC101120929, LOC101106536, LOC101108065, LOC101120480, DTMT, LOC101108585

Flagellated sperm motility	4	0.02
Peptide secretion	2	0.021
Cellular response to interleukin-13	2	0.021
Linoleate 9S-lipoxygenase activity	2	0.021
Establishment of skin barrier	3	0.024
Positive regulation of telomerase activity	3	0.024
Iron ion binding	6	0.025
Ceramide biosynthetic process	3	0.026
Adult walking behaviour	3	0.028
Phosphatidylinositol 3-kinase complex, class IB	2	0.033
MLL1 complex	3	0.035
Calcium	11	0.039
1-phosphatidylinositol-3-kinase regulator activity	2	0.042
Cytoskeleton	7	0.043
G-protein coupled receptor activity	16	0.045
Sulfotransferase activity	3	0.048
ATPase activity	7	0.048
Motor activity	3	0.05
Positive regulation of cardiac muscle cell differentiation	2	0.052
Positive regulation of peroxisome proliferator activated receptor signalling pathway	2	0.062
Third ventricle development	2	0.062
Elongator holoenzyme complex	2	0.064
Protein self-association	3	0.064

Transcriptional repressor complex	3	0.065
Ubiquitin protein ligase binding	6	0.067
Protein lipidation	2	0.071
Extracellular ATP-gated cation channel activity	2	0.072
Leading edge membrane	2	0.074
Lateral plasma membrane	3	0.074
Cytosol	33	0.075
Neuromuscular junction	3	0.081
Bone marrow development	2	0.081
Regulation of calcium ion transport	2	0.081
Nuclear migration	2	0.081
Purinergic nucleotide receptor activity	2	0.082
Dendritic spine	3	0.084
External side of plasma membrane	5	0.085
Protein desumoylation	2	0.091
Positive regulation of production of MIRNAS involved in gene silencing by MIRNA	2	0.091
Negative regulation of production of MIRNAS involved in gene silencing by MIRNA	2	0.091
Behavioural response to pain	2	0.091
Transcription corepressor activity	4	0.093