GENOMIC ANALYSIS OF SWAKARA SHEEP SUB-POPULATIONS

by

Andiswa Njilo

BSc (Hons) Genetics

UNIVERSITY OF KWAZULU-NATAL ™
INYUVESI YAKWAZULU-NATALI

As the candidate's supervisor, I have approved this dissertation for submission.

Supervisor: Dr E.F Dzomba

Signature: [redacted]

Date: 12/02/2024

Co-supervisor: Dr F.C Muchadeyi

Signature: [redacted]

Date: 12 February 2024

PREFACE

The research contained in this dissertation was completed by the candidate from August 2021 to November 2023 while based in the Discipline of Genetics, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa, and the Agricultural Research Council-Biotechnology Platform. The research was financially supported by the National Research Foundation (NRF) and the Agricultural Research Council-Biotechnology Platform.

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

11/02/2024

Signed: Andiswa Njilo (Candidate) Date:

12/02/2024

Signed: Edgar F. Dzomba (Supervisor) Date:

12 February 2024

Signed: Farai C. Muchadeyi (Co-supervisor) Date:

DECLARATION OF PLAGIARISM

I, ANDISWA NJILO……., declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written but the general information attributed to them has been referenced

b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed 

…… ……

ABSTRACT

The Swakara, originally derived from the Karakul breed of Uzbekistan, is a fat-tailed sheep breed first imported into Namibia in the 1900's and then later spread throughout Southern Africa. With the aim of producing superior pelts, the breed was subjected to intense crossbreeding and selection with the white wooled indigenous Namaqua Afrikaner and Blackhead Persian breeds, with the aim of producing superior pelts possessing wavy hair with short, lustrous fibres and clear patterns. Swakara sheep can be found in four different coat colours (black, grey, brown, and white), with only the black being inherent to the original Karakul with the presence of coat colour variation thought to be the basis of some of the inbreeding and selection imposed on the breed. The sheep are prone to genetic disorders such as the subvital factor that causes the animals to die within 48 hours of birth and is prevalent in grey and white-wooled sheep. It is suggested that these genetic disorders are due to intensive selection imposed on a breed of limited population size. Little is known about the genomic architecture of the Swakara sheep and its divergence between the sub-populations and from the founding breeds. Such lack of information makes it difficult to understand the factors contributing to the appearance of genetic disorders. Furthermore, it imposes a challenge in the implementation of future breeding programs, trying to select for new traits as well as in the conservation of genetic resources. In this study, 244 sheep from 8 sub-populations representative of Swakara Sheep from Namibia (n = 171), and South Africa (n = 44) and founding Karakul sheep from Germany (n = 5) were sampled and genotyped using the OvineSNP50 beadchip. These sheep were of Brown (n = 25), Black (n = 51), Grey (n = 42), White Vital (n = 34), and White Subvital (n = 63) sheep. In addition, ancestral breeds of Namaqua Afrikaner (n = 10) and Blackhead Persian (n = 14) were obtained from previously published data. The first set of analysis investigated the genetic diversity and structure of Swakara sub-populations and its presumed founding breeds. Genetic diversity ranged from H_O = 0.29±0.15 for the White Subvital to H_O = 0.41±0.22 for the Karakul. The first principal component analysis (PCA) produced five clusters with PC1 explaining 27.35% of the total variation whilst PC2 accounted for 19.25% of the total variation. Cluster A consisted of only the Brown Swakara, Cluster B consisted of the Black, Grey, and White Swakara. Some of the White Vital and Black Vital also clustered together with the Karakul in Cluster C. The Namaqua Afrikaner and Blackhead Persian clustered separately in Clusters D and E, respectively. Per marker F_{ST} showed differentiated SNPs within QTLs associated with milk production, wool quality traits, and coat colour. Signatures of selection were identified utilising

three methods of within population (iHS), between population (XP-EHH) and across the global population (HapFLK). A total of 73, 619 and 1931 selective sweeps were detected from each of the analysis, respectively. A large number of significant selective sweeps (|*iHS*|>3.0) were identified within the ancestral populations across 16 different chromosomes (OAR1, OAR2, OAR3, OAR4, OAR5, OAR6, OAR7, OAR8, OAR12, OAR15, OAR17, OAR18, OAR20, OAR21, OAR22, OAR23) and 28 of the signals were detected within the coat colour groups and ancestral groups. Overlapping genes were associated with QTLS of body weight (*FBN2*, *CNTNAP5*) and milk fat percentage (*TMEM163*, *NDUFB3*). The strongest signals were observed between the coat colour sub-populations with *XP-EHH* values ranging from 3.93 to 5.27. Further analysis detected signatures of selection related to 56 candidate genes, including *PTGER3*, *ADAMTS3*, *TROAP*, *NEIL2*, *FBXO8*, *BTBD10*, *KLF13*, and *CCNT1* that are associated with hot carcass weight, testes weight and milk yield. HapFLK revealed strong selection signals on chromosome 2 and 3. Genes related to gastrointestinal parasites such as *GPC6*, *KAZN*, *MCTP1*, and *HS3ST3A1* were detected. Runs of homozygosity (ROH) analysis revealed the Brown Vital and Blackhead Persian to have ROHs of the longest mean lengths. Furthermore, F_{ROH} based inbreeding estimates were low across populations with the highest inbreeding (F_{ROH} = 0.26) identified in the Namaqua Afrikaner. This study sought to provide insight into the genomic architecture of Swakara sheep and the presumed ancestors to understand the prevalence of genetic disorders, guide future breeding programs and facilitate preservation of genetics.

ACKNOWLEDGEMENTS

I would like to express my immense gratitude to my supervisors Dr Farai Muchadeyi and Dr Edgar Dzomba. Your guidance and support as supervisors are deeply appreciated. The contributions made by Dr Muchadeyi in this research garners the most endless gratitude from me. You played a huge role in making me a better scientist. Thank you.

I am extremely grateful to the Gellap-Ost research station in Namibia for their contribution in this research. I would also like to express gratitude to the Agricultural Research Council-Biotechnology Platform and the National Research Foundation for the financial support which has allowed me to conduct this research.

To my fellow ghosts, thank you for the endless support and late nights working together ensuring that we always push each other forward. I could not imagine a better group of people to have walked this journey with. To Dr Nompilo Hlongwane and Dr Magriet Van Der Nest, thank you for always being there to lend a hand when I would struggle and being incredible mentors that I could always count on. I consider myself very lucky to have been in your space during this research.

To my friends and family, my dear parents whom I endlessly grateful for, thank you for bringing me into this world to be able to pursue these dreams and do what I love. I hope you are proud of me. To my aunt and sister, I could not be where I am today without you, I love you deep. To my friend, Andiswa Mgaga, whom I have been on this journey with from the beginning, thank you for always being an incredible friend to me throughout this journey.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Rationale for the research

The Swakara sheep are a fat-tailed breed known for high-quality pelts of lustrous hair fibres, and short hairs that display beautiful curl patterns (Nyoni *et al.*, 2020). The Swakara sheep were originally derived from the Karakul breed of Uzbekistan, Central Asia and introduced to Namibia and South Africa in the 1900s, where they were subjected to intensive selection and breeding. Karakul sheep were first imported into Africa in 1907.

In Namibia and South Africa, crossbreeding was done shortly after the first group of Karakul sheep, which consisted of only black coat coloured sheep that the original Karakul was found in. Crosses were carried out with Blackhead Persians, Afrikaners and Merinos, with the latter showing the poorest pelt quality (Campbell, 2007). Selection in Swakara mainly focused on fur traits as breeders aimed to improve the quality of the pelts for economic gain by breeding for short, strong hairs that depict wavy patterns and a lustrous texture. Years of intense selection and crossbreeding of the Karakul with various indigenous African breeds resulted in the quality improvement of the pelt to one of magnificent and significantly improved hair texture and patterns. Crossbreeding in the Karakul is believed to have been last done in 1979 (Schoeman, 1998). Since then, the breeding has evolved into specific improvement of the desired market traits through within sub-population selection. The current pool of these sheep is now referred to as Swakara (South West African Karakul), and considered a separate breed derived from the original Karakul but mixed with other indigenous breeds and selected to produce pelt of a unique and superior quality. Currently, the Swakara breed is farmed mainly in Namibia, South Africa and Botswana (Malesa, 2015). The pelts produced from Swakara lambs, are significantly different and unmatched in their lustrous features that have evolved beyond the characteristics of the initial Karakul that possessed long, curly hair fibres (BiénabeKirsten and Bramley, 2013).

The Swakara sheep are bred predominantly for their pelts, quickly adapt to arid lands due to their ability to tolerate harsh conditions as well as their resistance to most diseases and contribute significantly to the Namibian economy. A significant expansion of Swakara in South West Africa and arid western areas of South Africa led to more than 95% of approximately 4.4 million sheep in South West Africa being Karakuls by the early 1970s, and almost 5 million pelts were being exported to European markets by 1979. More than 200 000 Swakara sheep are found in Namibia and approximately 40 000 in South Africa, farmed mainly in the Northern

Cape. The industry has seen a significant decline in the number of Swakara stud breeders over the years, dropping from more than 700 in the 1970s to a staggering 54 in the early 2010s (Synman, 2014). This can be attributed to the decreased demand for pelt products within the fashion industry and the increased costs associated with the breeding of Swakara. This breed remains in poor smallholder farms with small herd sizes. In 2013, it was estimated that the Swakara flock in South Africa and Namibia was around 34, 254 and 261, 848, respectively (Kruger *et al.*, 2013). The evolving fashion trends play a significant role in the value of Swakara pelts resulting in a constant fluctuation. An increase of pelt prices has been seen over the years moving from around R62, 50 in 1994 to approximately R697 in 2013 (Itenge and Shipandeni, 2015).

Dominant coat colours in Swakara pelts are white, black, grey, and brown, with white being the most preferred colour as it can be dyed into different colours to suit trends in the fashion industry (Malesa, 2015). These white coloured pelts have a higher market value. The white colour is not inherent to Karakul sheep but was only introduced into white Swakara through crossbreeding of the Karakul with indigenous white-wooled Persian and Namaqua Afrikaner sheep (Schoeman, 1998) followed by intensive selection. Moreover, the first group of imported Karakul consisted of only black coloured individuals. Several studies have been conducted better to understand the mode of inheritance of these pelt colours. Greeff *et al.* (1984) showed five loci responsible for expressing these pelt colours in Swakara sheep. Other coat colours found in Swakara include spotted sheep ranging from small spots on white bodies individuals to black patches on white animals (Itenge and Shipandeni, 2015).

Swakara sheep are predominantly bred for pelts and selection is mainly focused on fur traits rather than the typical traits that conventional livestock are bred for (i.e., milk production, meat production). The pelts are harvested by well-trained personnel within 24 hours of birth due to the curl structure of the pelt deteriorating as the lamb ages (Nyoni *et al.*, 2020). The Swakara breed is plagued with poorly understood genetic disorders in the Grey, White and Brown sub-populations. The white sub-population in its homozygous form (WW) is subjected to a subvital effect due to a recessive mutation which causes the pure white sheep to die within 48 hours of birth (Muchadeyi *et al.*, 2015). The presence of a recessive mutation affects digestive organs and the ability to feed, thus hindering the growth and development of the lambs. Thus, the prevalence of the subvital trait in white pelt sheep poses a challenge when replacing flocks. The subvital also manifests in the grey Swakara although at a more severe degree, and only homozygous grey individuals are affected. Subvital sheep can be identified at birth by the lack

of pigmentation of the ears, tongue, and palate (Lundie, 2011). The genetic difference between these Swakara sheep is not clearly known. A deeper genomic analysis of the population genetic structure is vital in understanding the genetics behind the appearance of lethal variants.

The intense selection that the Swakara breed has been subjected to is hypothesised to have led to the prevalence of sub-lethal and sub vital genetic factors in some populations. Breeders have less interest in the other livestock production traits. The targeted selection on pelt traits can, consequently, lead to a shift in genetic variations distribution between and within populations of Swakara which is presenting itself in a decrease in performance and functionality, and an appearance of homozygous recessive defects. The initial twelve Karakul sheep imported into Southern Africa in 1907 and consisting of only seven ewes, three lambs and two rams (Campbell, 2007) highlights their small founder population and its effect on the appearance of distinct allele frequency patterns in the newly formed population and resulting genetic diversity loss. To date, small population sizes of Swakara have been reported with highly fragmented flocks that are geographically restricted. Consequently, farmers intensely select within closely related animals resulting in high inbreeding levels relative to other sheep breeds (Dzomba *et al.*, 2021).

The challenges associated mainly with the pure white, grey, and brown sheep significantly affect breed improvement programs and the industries that rely on the economic value of the sheep. Swakara sheep generally require very little maintenance; however, sheep displaying subvital characteristics increase management costs as they require special care from birth. These Swakara sheep require intensive care (e.g., specialised diet), and even then, it may not be enough as they may die before they reach sexual maturity thus cannot be utilised for breeding (Schoeman, 1998).

1.1 Justification

The Karakul breed has undergone crossbreeding followed by intensive selection and possibly inbreeding with the primary aim to improving the quantity and quality of pelt produced and is reflected in the Swakara sheep population of Southern Africa. The genetic impact of this crossbreeding and selection on the breed has not been extensively studied. In order to stay competitive in the market, Swakara breeders are consistently selecting and trying to improve the quality of pelt albeit facing challenges of genetic disorders. Genetic data on the population genetic structure and the divergence between the different coat-coloured sub-populations and other ancestral breeds is lacking and this hampers progress in managing these

genetic disorders whilst improving the quality of pelt sheep. For example, white sheep with no subvital effects have been identified. However, knowledge of the genomic differences between this population and other white, black, brown, and grey pelted populations is still lacking making it challenging to ascertain genetic mechanism/causes of the disorders. Furthermore, a deeper understanding of the genetic make-up of Swakara concerning its founding breeds can shed more light into the evolution of Swakara sheep and guide future production and improvement programs.

The appearance of lethal variants is believed to be due to the intensive selection that Swakara sheep has been subjected to which may have, inadvertently, promoted inbreeding. Genomic concepts to further analyse the population genetic structure and SNPs associated with the appearance of genetic disorders have been developed and applied to address similar problems.

The Swakara is known to be a blend of various sheep breeds due to the crossbreeding resulting in the admixed sub-populations. Admixture results in new lineages and alleles being introduced into a population. Looking at admixture will provide further insight into the genetic ancestry of the Swakara pelt colour sub-populations. Moreover, application of principal component analysis (PCA) will characterise the genetic structure and level of relatedness between the populations (Dzomba *et al*., 2021; Hlongwane *et al*., 2020).

Signatures of selection are genomic regions with high frequency haplotypes due to strong selection pressure (Onzima *et al*., 2018). With the detection of selection signatures, one can gain an insight into the mechanisms involved in diversification across populations, the similarities between the sub-populations and between sub-populations and other founding breeds (Zhao *et al*., 2015; Manzari *et al*., 2019) and allows one to make inferences on the selection pressures any breed could have been subjected to.

Runs of homozygosity (ROHs) for example, appear as contiguous lengths of homozygous genotypes due to identical haplotypes being transmitted to offspring from their parents. The continuum of homozygote segment lengths depends on the degree of shared parental ancestry and its age, and as such, recent or ancient inbreeding within the different coat-coloured sub-populations can be determined as well as the association with prevalent genetic disorders (Zhang *et al*., 2015; Peripolli *et al.*, 2018).

This study intends to apply genome-wide SNP genotyping through the use of the OvineSNP50 beadchip to gain more insight into the genetic structure of the Swakara breed as influenced by its ancestral founding breeds, intensive selection and improvement practises as

well as understand the genetics of the prevalent genetic disorders. The utility of the OvineSNP50 genome-wide genotyping array has been assessed and demonstrated in various studies (Dzomba *et al.*, 2021, Eydivandi *et al.*, 2021). The application of the OvineSNP50 beadchip to the Swakara breed will provide understanding of the breed's genetic diversity and genomic architecture. Furthermore, it will provide exploration and identification of genomic variants that could be associated with the genetic disorders in Swakara. The addition of the ancestral breeds of Swakara in this study, which have been neglected in previous studies, will reveal its history and basis of genetic origin. Additionally, the Brown Vital Swakara has remained largely ignored and its inclusion in this study can be pivotal in fully characterising the Swakara breed. Subvitality and sub-lethal genetic disorders have not been extensively studied in Swakara sheep. The appearance of the subvital disorder in the pure white sheep has severe financial/economic implications. Farmers cannot earn a higher profit since the white pelts are the most profitable, but the breeding for the pure white sheep is not ideal resulting in them settling for B- and C-white sheep that have black spots on white skin or large white patches on dark skin that fetch lower prices. Moreover, the prevalence of genetic disorders is reflective of inbreeding in this breed that needs to be better managed to not threaten its sustainability.

Utilising high-throughput genotyping methods for genetic analyses will provide more insight into the breed's ancestral origin and genetic structure thus allow the Swakara sheep industry to breed for white pelts better without encountering low pelt yields, high management costs, high mortality, and survivability of replacement.

1.2 Aims

Analyse genomic regions to characterise the genetic differences and similarities that exist within and between Swakara sheep and its ancestral breeds and make inferences on the influence on prevalent genetic disorders.

1.3 Objectives

The specific objectives of this study were to:

- Determine the population genetic structure of Swakara and its relationship with presumed founding breeds.
- Investigate the divergence and affected genomic regions between Swakara sub-populations and between the sub-populations and the presumed ancestors (Karakul, Blackhead Persian, Afrikaner).

- Investigate the genetics of Swakara genetic disorders in Grey, Brown and White subvital Swakara sheep utilising ROHs.

1.4 Outline of dissertation structure

The dissertation contains a general introduction and literature review for the study followed by three experimental chapters written in manuscript format. Conclusions and general discussions will be in chapter 6. The dissertation chapters are as follows:

Chapter 1: General introduction of the study containing the rationale, justification and aims and objectives.

Chapter 2: Literature review on the current knowledge on Swakara sheep. The methods and technologies for the characterisation of the breed and its founders are also outlined.

Chapter 3: Explores the genetic diversity and population structure of Swakara and its ancestral breeds.

Chapter 4: Signatures of selection analyses analysing genomic regions under positive selection and seeking to identify selective sweeps within, between and across Swakara sub-populations and founders.

Chapter 5: Investigates the prevalence and distribution of ROHs and makes inferences on the genetics of the subvital and lethal genetic disorders prevalent in Swakara sheep.

Chapter 6: General discussion and conclusion as well as the proposed further research.

2 LITERATURE REVIEW

2.1 Introduction

The Karakul pelt industry has long been a significant branch of the animal industry in Southern Africa and has been successfully farmed for many years in north-western areas of South Africa and the south of Namibia (Greeff *et al.*, 1991). In 1907, the first group of Karakul sheep were imported into Namibia, consisting of 10 ewes and two rams, with a further group of 23 rams and 251 ewes arriving in 1909. The Karakul industry was then slowly established as a key force in the agricultural and economic growth of the country (Campbell, 2007). This has been attributed mainly to the unique quality of hair produced by the breed that is characterised by the exceptional patterns and quality textures due to the intensive research and crossbreeding the Karakul has been subjected to (Nyoni *et al*., 2020). From 1930 to 1939, the numbers of Karakul sheep expanded in South West Africa and other arid western areas of South Africa. This development was credited mainly to the ability of the sheep breed to adapt to severe climatic conditions, thereby enabling them to survive during the prevailing droughts and a sharp decrease in wool prices due to the depression of the early thirties (Schoeman, 1998). By the early 1970s, more than 95% of approximately 4.4 million sheep in South West Africa were Karakuls, and by 1979 almost 5 million pelts were being exported to European markets. The pelts produced from the Namibian lambs are significantly different and unmatched in their lustrous features and unique quality compared to the indigenous breeds. As a result, they were renamed to Swakara (South West African Karakul).

The Swakara breed is farmed mainly in Namibia, South Africa and Botswana (Malesa, 2015). They are known for their adaptation to arid lands due to their resistance to most diseases and contributing significantly to the Namibian economy. Well-trained personnel harvest the pelts within 24 hours of birth due to the curl structure of the pelt deteriorating as the lamb ages (Figure 2.2) (Nyoni *et al.*, 2020). Dominant colours in Swakara pelts are white, black, grey, and brown (Figure 2.1), with white having the highest market value as it can be dyed into different colours to suit trending fashion styles (Malesa, 2015).

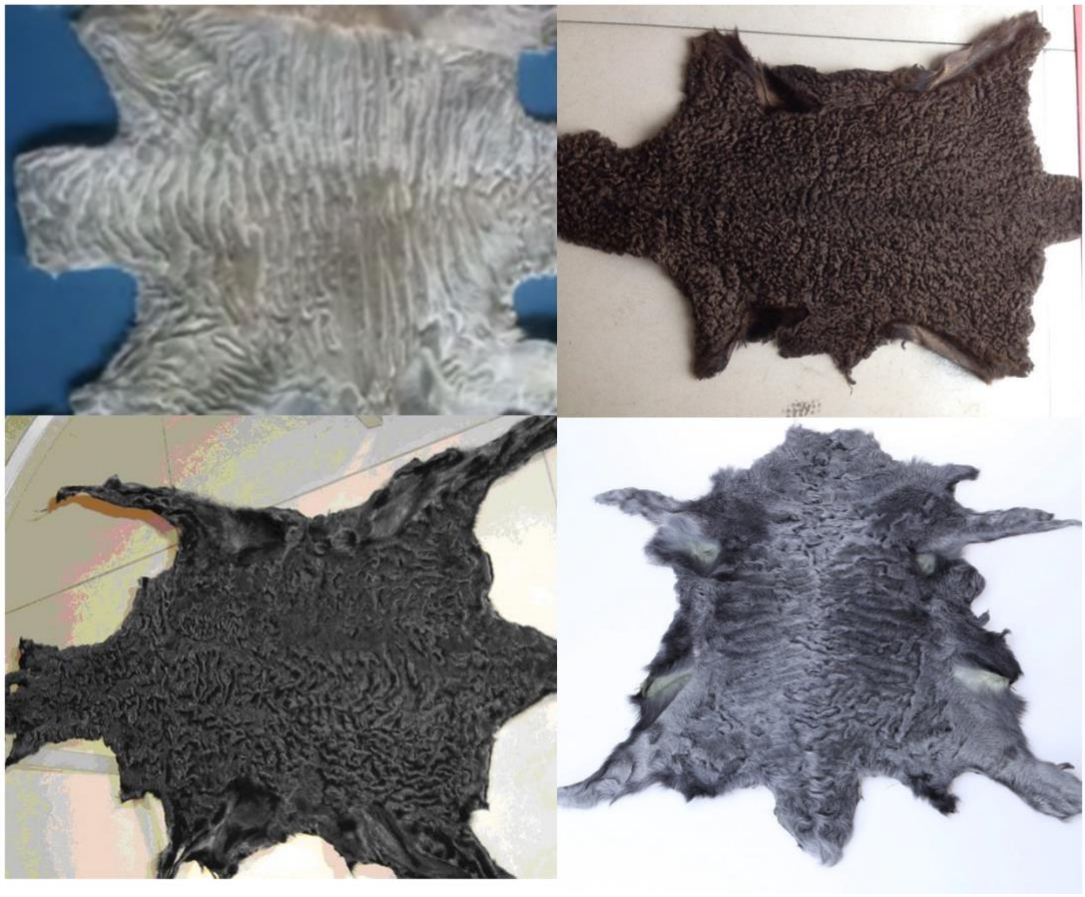


Figure 2.1 Swakara pelt colours Top: i) White pelt ii) Brown pelt Bottom: iii) Black pelt d) Grey pelt (Adapted from Campbell, 2007; Malesa, 2015)

Swakara crossbreeding was done immediately after the sheep were imported to South Africa and Namibia. Numerous breeds were utilised, namely the Namaqua Afrikaner, Blackhead Persian and Dohne Merino which, according to Theron (1966), the Merino was observed to show the poorest pelt quality. The Swakara breed is known for its unique pelt characteristics that make it economically important. Schoeman (1998) showed that pelt price varies according to different coat colours and predominantly by hair quality score, pattern score, hair length and curl type (Table 2.1). Many breeders prefer pelt patterns with more lyre patterns rather than drawn pattern formation (Campbell, 2007). Certain traits like hair thickness and pelt thickness have been determined to correlate to price, even though their contributions are minor. Furthermore, it was ascertained that the importance of the relative influence of pattern score decreases with a decrease in curl development, whereas hair quality score becomes more critical (Schoeman, 1998). Over the years, very few commercial farmers have continued producing Swakara sheep. Traditionally, Swakara sheep were not farmed for meat production

(Martins and Peters, 1992a). Breeders have since ventured into farming Swakara for meat, although it is less profitable due to its inferior carcass traits; thus, it is marketed mostly locally (Campbell, 2007).

Figure 2.2 The quality of Swakara pelts deteriorate as they age (Adapted from https://www.parcanimalierlabarben.com/en/animal/karakul-sheep and https://za.pinterest.com/pin/402157441696607625/)

In Southern Africa, the Swakara pelt industry is very vulnerable within the fashion industry due to its unpredictability although most farmers are dependent on it. Despite the strides made to produce more unique pelt patterns in the white Swakara, this has been met with its challenges as issues concerning this sub-population continue to threaten the industry. The continuous breeding/selection of the white Swakara has resulted in the appearance of subvital effects causing white homozygotes sheep to have higher mortality rates and lower reproductive rates (Malesa, 2015). This subvital disorder is also witnessed in the grey sheep.

Today, the Karakul breed is part of an extensive group of sheep genetic resources in Africa where there are more than 170 estimated sheep breeds (Muigai and Hanotte, 2013), and more than 80% being considered as indigenous and part of smallholder farming systems. The different sheep breeds in Southern Africa can be classified into two groups; thin-tailed which are found mainly in North Africa and fat-tailed sheep which are more widely distributed across the African continent. The Swakara is known to be a fat-tailed sheep breed. As with other domestic sheep, African sheep are believed to have descended from the Asiatic mouflon (*Ovis orientalis*) (Meadows *et al.*, 2011). Phenotypic traits are seen in some of these sheep today and their origins are not completely understood. Furthermore, minimal information can be found on the genetic structure of these breeds. The Swakara is one such poorly understood and understudied breed with a complex ancestral background and little knowledge on its full

genomic architecture. A molecular genetics study of the genetic ancestry, population diversity and structure will improve our understanding of these genetic resources.

Table 2.1 Traits of importance in Swakara pelts.

Trait	Description
Curl type	Degree of curl development (smooth/curly)
Hair quality	Texture and lustre of fibres
Pattern score/formation	Expresses attractiveness of the pelt
Fibre length	Pelts with short hair are more valuable

2.2 Crossbreeding and Genetic Diversity of Karakul

The Karakul breed is believed to be one of the oldest of domesticated sheep. Although it originated from central Asia, this fat-tailed sheep breed has been introduced into many other countries, mainly France, Austria, Argentina, Canada, Peru, Germany, and the USA (Schoeman, 1998). Importation of this breed into these countries led to the Karakul being crossbred with other indigenous breeds for breed improvement purposes.

The most prominent importation of Karakul to Africa led to its popularity due to increased fashionability of pelt products in the fashion industry(MusaviKhadimiyan and Azimi, 2022). Today, the breed is found all over the world. Its importation into the USA occurred in 1908 where it was immediately crossbred with indigenous breeds with the aim of increasing numbers as more rams were imported than ewes (Campbell, 2007). The Karakul was crossbred mainly with the Lincoln, Cheviot, Cotswold, and Merino. The industry quickly fell apart before it'd been properly established, as the results of the crossbreeding were not very successful, and the pelt production could not survive under the American conditions. In Germany, the Karakul was crossbred with the Zackel and Rambouillet (Theron, 1966). Since the first importation of Karakul in Africa in 1907, Namibia and South Africa have been breeding and producing the most Karakul. Shortly after the first group of Karakul sheep were imported, crossbreeding was carried out which resulted in the quality improvement of the pelt characterised by magnificent hair texture and patterns. Crosses were carried out with Blackhead Persian, Afrikaner and

Merino, and various other breeds that showed poor pelt quality (Campbell, 2007). The present-day product of this selection and crossbreeding is known as the Swakara, which is unique to Southern Africa and characterised by pronounced coat/pelt colour variants.

2.3 Coat colour variants in Swakara

Swakara sheep are found in four main coat colour variants: grey, black, brown, and white. The original Karakul from which the Swakara was developed was found in only the black coat colour (Schoeman, 1998). The introduction of the colour variation was first seen with the white because of the crossbreeding with other white-wooled indigenous sheep of Southern Africa like the Blackhead Persian and Namaqua Afrikaner. The appearance of the subvital factor in the pure white (A-white) sheep because of these crossbreedings and further selection posed many challenges for farmers who rely on the economic value of the sheep pelts. A long-standing practise for farmers has been to mate black and white sheep to mitigate the effects of the disorder. This in turn introduced further variants of B- and C-white sheep which are characterised by white bodies with broken black in the face and other times small black spots on the body (Schoeman, 1998). However, these calculated matings by farmers also resulted in the grey colour which can vary in intensity depending on the black:white fibres ratio in the pelt. When homozygous, the grey coat colour is also lethal, grey Swakara therefore exist in heterozygous form (Nel, 1960). The origin of the brown coat colour is not well known but it is believed to have resulted from other crossbreedings with undocumented breeds.

2.4 Inheritance and colour variation in Swakara sheep

A considerable variation of coat colours exists within different breeds of sheep. This may be due to both genetic and non-genetic factors. In wild breeds, a uniform phenotype is uncommon and shows species-specific patterns and colours (Koseniuk *et al.*, 2018). Alternatively, domesticated species display a higher range of coat colour variation. Consequently, for studies pertaining to coat colour variation, domesticated breeds are a more suitable model.

2.4.1 Colour variation

The amount and level of distribution of melanin that is secreted by mature melanocytes is primarily what determines coat colour variation (Yao *et al.*, 2019). Melanocytes are cells derived from the ectoderm, and differentiation is dependent on the regulation of specific genes (Koseniuk *et al.*, 2018). The different melanins can be divided into eumelanin (black/brown) and pheomelanin (red/yellow) (Han *et al.*, 2015). During embryonic development,

melanoblasts form from the neutral crest and, under the regulation of specific genes, will undergo specification and migration to the skin to form mature melanocytes (Figure 2.3) (Yao *et al.*, 2019; Koseniuk *et al.*, 2018). Studies conducted to effectively understand the genetics of pigmentation, albeit in mice, showed that majority of the genes code for factors found in melanocytes (Bennett and Lamoreux, 2003, Hoekstra and Nachman, 2003).

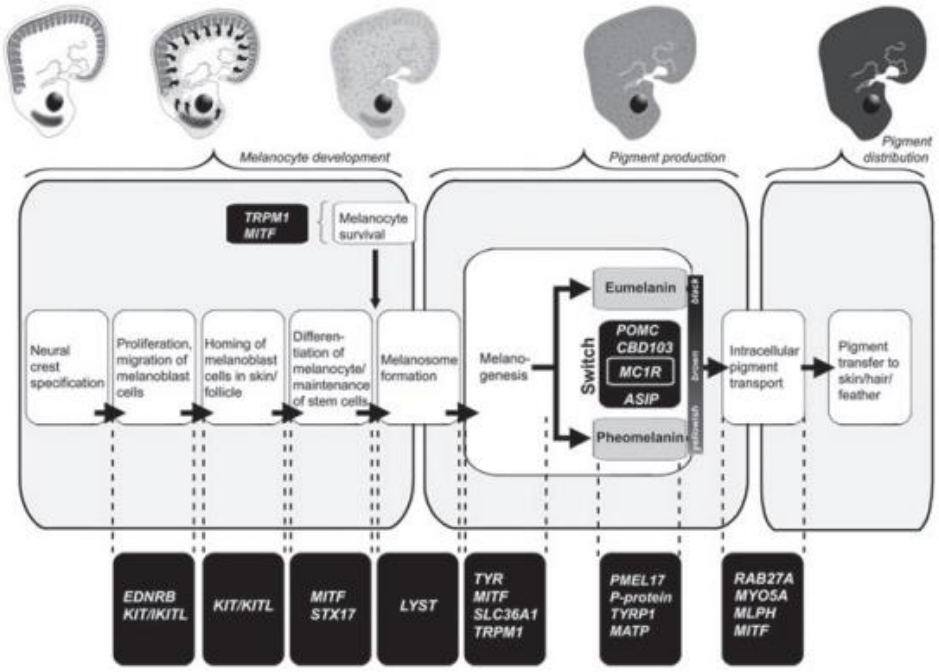


Figure 2.3 Progression of pigment cell development, pigment synthesis, intracellular pigment transport and pigment transfer (Cieslak *et al.*, 2011)

2.4.2 Gene factors

There are currently more than 150 genes that have been discovered to have some sort of influence on pigmentation and a further 300 genetic loci associated with coat colour (Rochus *et al.*, 2019). These various loci can be accountable for one or more pigment-associated traits either singularly or combined with more loci (Cieslak *et al.*, 2011). The primary genes function in pathways where the two essential pigments are produced (eumelanin and pheomelanin) or distributed. In sheep specifically, the candidate genes that code for the key factors involved in pigment synthesis are melanocortin 1 receptor (*MC1R*), agouti signalling protein (*ASIP*) and tyrosinase-related protein 1 (*TYRP1*).

2.4.2.1 Melanocortin 1 receptor (MC1R) and Agouti signalling protein (ASIP) genes

The fundamental coat colour variation is defined by the ratio of eumelanin to pheomelanin, which is controlled by the melanocortin 1 receptor (MC1R) ligand-receptor system.. The

MC1R is a 7-transmembrane G-protein-coupled receptor that is encoded by the extension loci. It is primarily expressed in melanocytes and melanoma cells (Muhaghegh and Habibizad, 2014, Rochus *et al.*, 2019). The size of the gene varies amongst different breeds, but it is believed to be 953 bp long in cattle (Koseniuk *et al.*, 2018). During signalling, the *MC1R* receptor will bind to melatonin (α-MSH) on the cell membrane and the adrenocorticotropic hormone (Yao *et al.*, 2019). Following the coupled G-protein's conversion to GTP, the membrane's adenylate ring will become active. (Rochus *et al.*, 2019). Cyclic adenosine monophosphate (cAMP) will then be activated, which will, in turn, activate tyrosine (TYR) kinase to engage in melanin synthesis (Yao *et al.*, 2019), where under normal conditions, melanocytes will produce pheomelanin rather than eumelanin under excessive activation of TYR kinase.

The *ASIP* gene is the other significant factor involved in the MC1R signalling pathway. The agouti signalling protein (ASIP) is encoded by the agouti locus.. The association of mutations found in *ASIP* to the variation in sheep coat colours has been of interest for many authors (Cieslak *et al*., 2011, Rochus *et al.*, 2019, Kijas *et al*., 2012, Lundie, 2011). The white allele in ASIP is known to be approximately 190kb containing *ASIP*, its two neighbouring genes; the promoter region *ITCH* and *AHCT* (Rochus *et al.*, 2019). The agouti signalling protein is known to antagonise *MC1R* actively. *ASIP* binds to MC1R and competes with α-MSH to block α-MSH's initiation signal, hence preventing the generation of cAMP (Yao *et al.*, 2019, Cieslak *et al.*, 2011). Ultimately, by blocking the *MC1R* signalling pathway, pheomelanin production is increased whilst eumelanin synthesis is increased (Jackson *et al.*, 2006).

2.4.2.2 Tyrosinase-related protein 1 gene (TYRP1)

Tyrosinase-related protein 1 is one of the candidate genes identified in coat colour variation. *TYRP1* is a candidate gene for the B locus in sheep mainly due to its black to brown eumelanin pigmentation effects in dogs, mice (Bennett and Lamoreux, 2003), and most significantly in Soay sheep (Gratten *et al.*, 2007). TYRP1 is a type 1 membrane-bound protein that is actively involved in the eumelanic pathway. It is expressed both in the retinal epithelium and melanocytes. Dopaquinone, a precursor of both eumelanin and pheomelanin, is converted into eumelanin during melanogenesis by the TYRP1 protein and DCT (Dopachrome tautomerase). Reports of oculocutaneous albinism in humans and mice due to mutations in the *TYRP1* have been made (Boissy *et al.*, 1996). *TYRP1* mutations are also known to result in brown phenotypes in cats sheep, and cattle (Posbergh and Huson, 2018, Cieslak *et al*., 2011). Additionally, it is believed that a single non-synonymous mutation in the *TYRP1* gene is the

reason why Soay sheep's ancestral dark brown coat has become less distinct. (Gratten *et al.*, 2007).

2.4.3 Genetic inheritance of coat colours

Sheep breeds can possess many different colours, unique markings, and patterns. Genetic studies focusing on the mechanism behind this variation began fervently in the 1920s with various researchers in many countries (Roberts, 1924, Vasin, 1928, Roberts and Jenkins, 1926). In Swakara sheep, the predominant colours are black, brown, grey and white (Nsoso & Madimbe, 2003), with black being the most common colour (Schoeman, 1998). Several studies have been conducted to ascertain the mode of inheritance of these coat colours (Adalsteinsson *et al.*, 1978, Greeff *et al.*, 1984, Adalsteinssons and Breed, 1980) and many others. Five loci have been identified to be responsible for the expression of pelt colours in Swakara sheep. The most common coat loci are the A locus (agouti) and the E locus (extension) (Koseniuk *et al.*, 2018). Although these loci can also be found in other mammals, the number of alleles is variable in different breeds for both loci.

The E locus is primarily responsible for the expression of black. Two common E alleles have been found: D (dominant) and e (recessive). The E^D allele mainly generates black phenotype individuals whilst recessive homozygotes (ee) display a red (tan) colour. A less frequent allele E^+ (wild type), was also identified (Adalsteinssons, 1980). Incidentally, the allele's neutrality was identified with its presence in genotypes and ability to be expressed with any other colour (e.g. E^D/E^+ - black) (Koseniuk *et al.*, 2018). The A allele generates brown coloured coats, and recessive homozygotes (aa) express black coloured coats. The E^D dominant allele allows for the complete inhibition of all alleles at the A-locus, and e recessive allele allows all alleles at the A-locus to be fully expressed (Schoeman, 1998). Two alleles have been identified at the B-locus responsible for the expression of brown colour. B dominant individuals express black coats, and b recessive individuals express a chocolate brown colour ("moorit"). Moreover, animals with the genotype bb at the B-locus express a chocolate brown coat colour due to the effect of the E^D allele. Nel (1966), in his early work, suggested a multi-allelic series at the E-locus for the expression of dark coffee brown, black, brown and Dobermann with an order of dominance of $E^C > E^D > e^b > e^t$, respectively. It was later suggested by Adalsteinssons (1980) that for the rare Dobermann colour, the A^w allele was responsible for the variation of colour and patterns, with A^{wh} top dominant being responsible for red and white coat colours in sheep breeds (Table 2.2). The red pigmentation can be eliminated by genes at other loci making the animal completely white (Schoeman, 1998). Significantly, white is not

inherent to Swakara sheep but was only introduced through severe crossbreeding with white-wooled Persian and black Karakul (Nel, 1966). The white coat colour allele (W) is partially epistatic to the brown and black base colours (B and E loci). The white Swakara sheep in its homozygous form (WW) is subjected to a lethal subvital effect. Similarly, the allele responsible for the expression of grey (or roan) is the W allele. Swakara sheep can express brown-roan or black-roan where the grey gene will counteract the base colour (brown or black) to produce a roan individual. Grey Swakara are typically heterozygous as the dominant grey is lethal when in homozygous form (Schoeman, 1998)

Table 2.2 Genetics of coat colour in sheep

Locus	Alleles	Description
Agouti – A	A^{wt} white/tan	Dominant. Only white/tan
	Wild – A^{+}	
Brown – B	Brown – B	Dominant. Black pigment
	Wild – B+	Recessive. Brown pigment
Extension – E	Dominant black – E^{D}	Inhibits agouti locus alleles
	Wild – E^{+}	Agouti alleles expressed
Roan – Rn	Lethal roan – Rn	Dominant. Homozygotes paler
	Wild – Rn^{+}	Recessive. No greying

Note. Data from Lundie, 2011, Adalsteinssons, 1980, Koseniuk *et al.*, 2018

Traditionally, phenotypic classification of breeds has been the ‘go to’ method of grouping and studying animals in breeding schemes. However, this method can be severely flawed as many phenotypes may be the result of the effects of many genes and alleles as well as the environment. Similarly, alleles of many genes can result in similar phenotypes, so an adequate understanding of genetic mechanisms is crucial.

2.4.4 Genomic and Statistical tools for DNA analysis

2.4.4.1 The OvineSNP50 Beadchip

The OvineSNP50 beadchip developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC) has become a crucial genomic tool in genomic studies and particularly in understanding and uncovering molecular information of the sheep genome. The beadchip is a genome-wide genotyping array containing more than 50K markers and was developed for the domestic sheep (Kharzinova *et al.*, 2015). It generates data that is crucial in

developing molecular breeding values, or identifying genomic regions that explain variation in monogenic or polygenic traits. Genotyping different breeds with the SNP50 beadchip can also be utilised to understand breed differentiation and signature sweeps (Johnston *et al.*, 2011) as well as identify quantitative trait loci. The Ovine 600K panel has been suggested by Kijas *et al.* (2014) for breeds of high diversity. However, this study opted for the 50K beadchip as very few studies have been conducted utilising the denser panel and with its high costs it is not feasible for studies with large samples and limited resources. Other studies investigating the genetic diversity and population structure of sheep have similarly utilised this panel *(Han et al.*, 2022, Edea *et al.*, 2017, Bedhiaf-Romdhani *et al.*, 2020, Molotsi *et al.*, 2017).

Table 2.3 Commercial SNP chips of some domestic species

Domestic species	Beadchip	No. SNP Loci
Cattle	BovineSNP50	50, 000
Dog	CanineSNP20	22, 362
Sheep	OvineSNP50	56, 000
Pig	PorcineSNP60	60, 000
Horse	EquineSNP50	54, 602
Chicken	ChickenSNP60	60, 000

2.4.4.2 GWAS (Genome-wide Association Studies)

The study and analysis of genomes to identify common causal genetic variants have been made widely possible due to genome-wide association studies (GWAS). Many species genomes have been studied to analyse genetic variants that possibly have genotype-phenotype associations (Tam *et al.*, 2019). The use of this technology has been monumental in studying traits of economic importance (Zhu *et al.*, 2021), and most significantly, providing a genetic understanding of common phenotypes of biomedical importance, including asthma, diabetes, and cancer (McCarthy *et al.*, 2008), and identifying new drug targets and disease biomarkers.

In the past decade, thousands of GWAS covering distinct diseases have been published, with more than 50,000 SNP-trait associations of significance identified (Manolio, 2010, Tam *et al.*, 2019). Identifying genetic variation for phenotypes at the genome-wide level was facilitated,

for many years, by mapping quantitative trait loci (QTL). This approach utilised microsatellite markers with low map resolution making it difficult to identify the significant genes for traits of economic importance (Zhang *et al.*, 2012). Large-scale SNP collections and the successful creation of affordable techniques for SNP data processing at a larger scale has made this technology not only feasible in humans but domestic animals as well (Hirschhorn and Daly, 2005).

Two types of gene mapping exist for common diseases and associated quantitative traits: candidate gene studies that use association or resequencing techniques, and genome-wide studies that utilise linkage mapping and GWAS. Traditionally, genome-wide linkage analysis has been the preferred method of identifying disease-related genes. The markers surrounding the illness gene must segregate with the disease in families for linkage mapping to be effective. As a result, segregation within each family within the same 10–20 cM chromosomal background will be identified. Furthermore, for most common diseases where linkage analysis has been utilised, there have been limited successes, and where genes have been identified, they have only been able to explain a minute fraction of the overall disease heritability (Wang *et al.*, 2015, Tam *et al.*, 2019, Hirschhorn and Daly, 2005). GWAS are far superior to traditional QTL mapping strategies, including its power to identify common causal variants with modest effects on disease and to define narrower genomic regions bearing genetic variants for economically important traits.

GWAS were first introduced and applied in the study of human diseases resulting in significant progress. The approach was extended to the field of domestic animal genetics. Although the rate of progress in both fields is not comparable, great studies have been made to produce extensive work exploring the successes of this approach in domestic livestock and other animals. These studies include horses, cattle, pigs, sheep and dogs (Zhang *et al.*, 2012). Moreover, in sequencing, many SNPs were discovered, and many commercial SNP chips have been made available as a result (Table 2.3) and have played a crucial role in genome analyses as a genome-wide genotyping tool.

Johnston *et al.* (2011) published the first account of GWAS usage in sheep. The authors sought to ascertain the genetic architecture for different inherited polymorphisms for horn morphology in wild Soay sheep. 36 000 SNPs were used in the genome-wide association investigation to identify the leading candidate for horns (single autosomal locus controlling horn morphology) to be an autosomal gene, the *RXFP2* gene, which is known to determine

primary sex in mice and humans. The authors were able to later show evidence of support for a new and simpler model of horn-type inheritance in Soay sheep. This study laid the foundation for investigating the relationship between horn genotype and phenotypic traits related to fitness. For many years only a handful of GWAS in sheep have been performed due to the limited availability of information on the sheep genome but the availability of the OvineSNP50 Beadchip has allowed for more extensive research.

2.4.4.3 Analysis of Population Structure

Understanding the structure of a population is imperative for any subsequent analyses that may follow. Different methods for doing so exist and utilised in humans (Elhaik *et al.*, 2014) and livestock (Edea *et al.*, 2017, Engelhardt and Stephens, 2010). Two widely used methods for analysing population structure exist: the population structure analysis is commonly the initial analysis of data investigation in most analyses of populations genetics (Engelhardt and Stephens, 2010). The genetic similarity, ancestry and demographic history of a group of individuals can be easily determined whilst gaining necessary knowledge on how to model the populations for further downstream analyses (Elhaik, 2022). Admixture based models are utilised in identifying genetic loci in a population of mixed ancestry that may elucidate on phenotypic differences. Furthermore, the number of underlying populations can be determined through cross-validation to estimate individual ancestry (Alexander and Lange, 2011). Although model-based methods are less robust compared to non-model tools, the results they produce are easier to interpret and can be applied much more efficiently (Wang, 2022).

Several other analyses such as pairwise F_{ST}, and identity by descent (IBD) have become the basis for appropriately estimating population structure. Wright's F_{ST} (Wright, 1949) is widely used in the quantification of genetic divergence in populations and describing the influences of genetic differentiation and describing the influences of genetic differentiation (Kitada *et al.*, 2020). Identifying segments of IBD can have numerous applications when detected from high-density genetic data. Identity by descent is the shared inheritance of an identical section of the genome between two individuals and can be essential in gaining information on the demographic history of a population (Sticca *et al.*, 2021). This can be significant in inferring the cause of genetic variation witnessed.

2.4.4.4 Runs of homozygosity

The quantification of inbreeding in livestock, humans and plants has long been a primary goal for agricultural scientists and evolutionary biologists. In the past century alone, methods

to do so have evolved drastically. Denser genome-wide microsatellite scans became available in the mid-1900s ushering in a new era by discovering uninterrupted long runs of homozygosity (ROH) genotypes (Ceballos *et al.*, 2018). ROHs are regions of homozygous genotypes that are homozygous for identical haplotypes passed down from parents to their progeny. (Purfield *et al.*, 2012) (Figure 2.4). Even in cases of a very distant relationship, identical chromosomal fragments can nevertheless be passed on. Therefore, the length of a homozygote segment might vary depending on the age and degree of shared parental ancestry. (Kirin *et al.*, 2010). Consanguinity may be shown by the presence of long ROHs indicative of recent inbreeding and relatively short recombination time permitted to separate the identical by descent segment (Dzomba *et al.*, 2021; Curik *et al.*, 2014; Kirin *et al.*, 2010; Li *et al.*, 2022; Purfield *et al.*, 2012). However, low recombination rates in specific regions of the genome may allow some extremely lengthy segments to remain., unusual mutation, or linkage disequilibrium. The presence of shorter ROH may provide information on the existence of more ancient relatedness, which can be unaccounted for in pedigree records due to limitations in the recording process (Purfield *et al.*, 2012). This might be because of limitations or the influence of population founders or breeds, especially in domestic animals. Numerous studies have been conducted on ROHs in human populations; however, in domestic animals, most have focused on cattle and pigs, with very few analysing other species such as sheep. Analyses focusing on a broader range of species will provide better characterisation of the genomic distribution of ROHs and their association with selective and neutral factors.

One of the most effective methods for identifying the consequences of inbreeding currently is the whole-genome inbreeding estimated from ROH (F_{ROH}) (Mastrangelo *et al.*, 2018a). Apart from recognising the difference between recent and ancient inbreeding, F_{ROH} also outperforms inbreeding coefficients derived from pedigree information as it provides more accurate ascertainment of the level of inbreeding (Cortes-Hernández *et al.*, 2021). Moreover, ROHs patterns are believed to be distributed non-randomly across the genome (Curik *et al.*, 2014). In a study by Bosse *et al.* (2012), they examined wild and commercial pig populations in Asia and Europe and found that animals within the same populations displayed similar ROH patterns, but ROH size and abundance differed variably among the breeds. They further showed that selection might play a more negligible effect on the genomic distribution of ROH compared to recombination and demography. However, another author (Zhang *et al.*, 2015) has, on the other hand, shown that ROH patterns are probably caused by selection events rather than just demography.

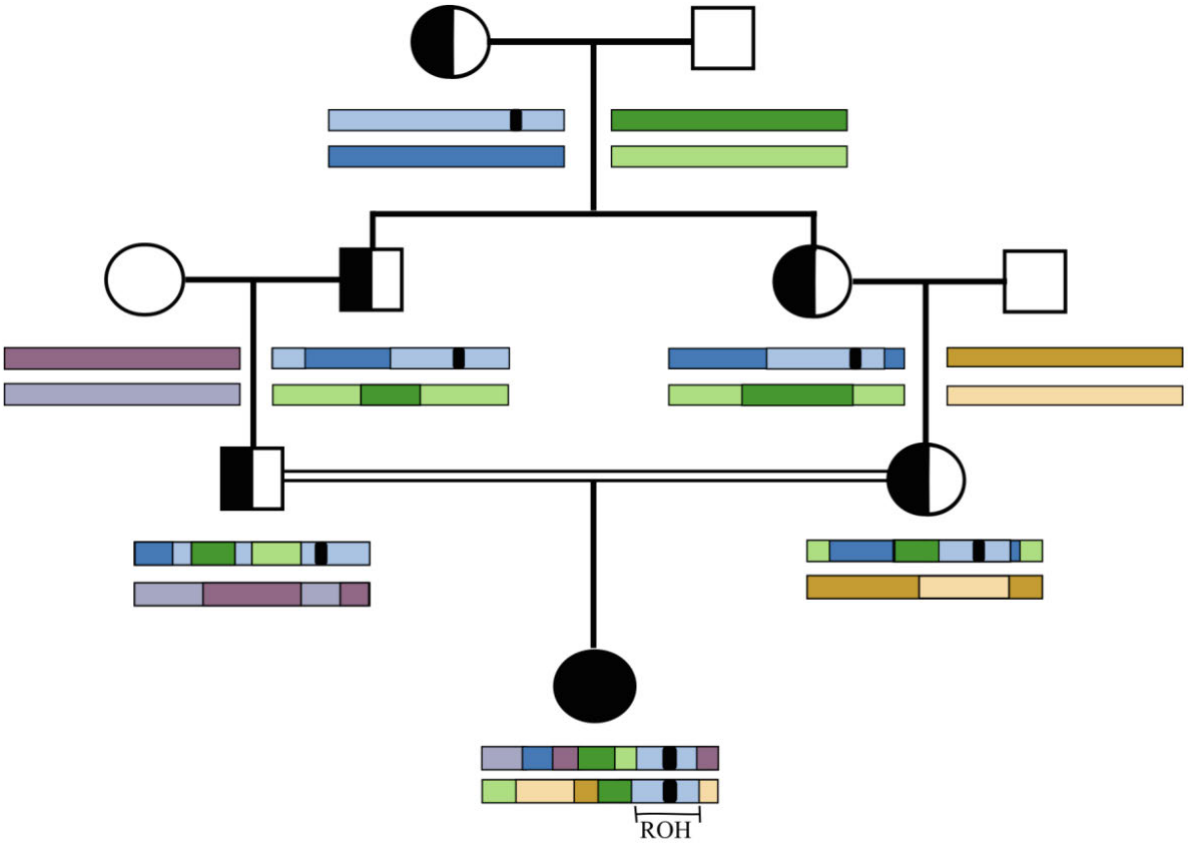


Figure 2.4 Schematic representation of a small consanguineous pedigree and ROH with a pathogenic variant (Adapted from (Oliveira *et al.*, 2018))

2.4.4.5 Selection signatures

Artificial selection pressures have resulted in the prevalence of SNP patterns in genomes, known as signatures of selection. In the traditional view of natural selection, favourable mutations can be beneficial and increase individuals' chances of survival. The process of evolution also includes a series of processes that are not primarily understood. The increased frequency of alleles for favourable adaptations and the associated alleles, which reduces variation in the linked sites when selecting for a trait, is referred to as ‘selective sweep’. This phenomenon is thoroughly described by Panigrahi *et al.* (2023). Genomic studies analysing signatures of selection for traits of interest have been performed in many livestock species. These studies traditionally used one method to identify loci that are potentially under positive selection, whether it be extended haplotype homozygosity, fixation index (F_{ST}) or runs of homozygosity (Avila *et al.*, 2018, MoioliPilla and Ciani, 2015). However, most methods are based on comparing the distribution of allelic frequencies, either directly or indirectly. In cases where the principle of 'selective sweep' has been sound is in a study by Moili *et al.* (2015) where they were able to identify genomic regions associated with fat deposition in sheep.

Genotyping data obtained using the OvineSNP50 beadchip detected a missense mutation in the *BMP2* and *VNRT* genes as the most probable genes involved in the fat-tail trait.

2.4.4.6 Copy number variation

Genetic variation can appear in different forms in the human genome, from significant chromosomal anomalies to single nucleotide polymorphisms. Insertion, deletion, and duplication of DNA can lead to copy number variants (CNVs) in many different species (Mace *et al.*, 2018). CNVs are DNA fragments that range from 1kb to larger mb sizes present at variable copy numbers compared to a reference genome (Stranger *et al.*, 2007). More significant segments of CNV are referred to as indels, and with the advancement of next-generation sequencing (NGS), the identification of CNVs as small as 500bp has been made possible. CNVs can occur at different frequencies in a specific population and are known to have serious phenotypic consequences due to altered gene dosage, disruption of coding sequences or perturbed long-range gene regulation (Fontanessi *et al.*, 2009; Mace *et al.*, 2018). When the frequency exceeds 1%, they are known as copy number polymorphisms (CNPs). Despite the fact that several CNVs are linked to complex or Mendelian genetic traits and human illnesses as well as other domestic livestock species, numerous other CNVs can have benign polymorphic variants (Fontanessi *et al.*, 2011).

CNVs are implicated in cases of disease susceptibility in humans, goats, and sheep. Therefore, correctly detecting CNVs and establishing their association with quantitative traits and phenotypes is essential in discovering disease aetiology (Mace *et al.*, 2018). However, their detection can be challenging. There are several approaches that may be used to identify CNVs across the whole genome. High-throughput sequencing and ultra-dense genotyping using SNP chips are two possible bases for this or hybridisation of DNA in oligonucleotide arrays (ClopVidal and Amills, 2012). The technique that is most frequently employed is array comparative genomic hybridization (aCGH). By observing the changes in fluorescence intensity of the test and reference genomes at each probe, one may find variations in copy number since aCGH is based on the competitive hybridisation in DNA arrays of one test and one reference fragmented and differently labelled genomes.. Relative to the reference genome, copy number polymorphisms can be described (Clop *et al.*, 2012).

Comparative analyses of the genomes of ruminants (goats and cattle) and primates (humans and chimpanzees) revealed the presence of recurrent CNVs in closely related species. This suggests that primordial segmented duplications may aid in the formation of CNVs. CNVs alter

genes related with several phenotypic characteristics. Additionally, it is thought that CNVs influence the agouti locus in sheep and goats, which in turn contributes to diversity in coat colour in both species (Norris and Whan, 2008, Fontanesi *et al.*, 2011, Fontanesi *et al.*, 2009).

2.5 Non-genetic factors influencing fur traits

Hair pattern and quality are considered to be the most economically important pelt traits. Hair pattern can be described as the design, extension, and delineation of the curls, according to Nel (1966). Although genetic factors significantly influence traits, various environmental effects, internal and external, can also influence fur traits. Internal influences such as the age of dam and sex of lamb are some factors (Schoeman, 1998). Several other factors such as month of birth, season of birth and type of birth also have an effect. The month of birth has been shown to significantly affect curl type, hair texture, and pattern score (Schoeman, 1998). Moreover, season of birth also affects hair quality score because as feeding conditions follow a seasonal pattern, fur traits will be affected. High feeding conditions can negatively affect fur traits. The assumption is that higher feeding conditions will result in heavier lambs at birth, with heavier and thicker pelts, leading to more curl development but ultimately a deterioration in pattern and hair quality scores (SchoemanGroeneveld and Albertyn, 1993, Schoeman, 1998). Age of dam has been shown to significantly affect curl type, pattern score, hair quality score, hair thickness, and hair length (Greeff *et al.*, 1991, Campbell, 2007).

The effects of non-genetic factors on coat colour have not been extensively studied in sheep. Although genetics determine the basic colour in animals, other factors can influence the intensity and variation of the colour in some species, such as cats and rabbits. In cats where the full pigmentation allele is (C/-), diets deficient in tyrosine can result in a change of hair colour from reddish to brown. It is also associated with a reduction in melanin in the hair (Akpomedaye, 2015). In Himalayan rabbits, the C gene controls pigmentation. The expression of this gene is regulated by temperature (Lobo, 2008).

2.6 Conclusion

Improving the traits of economic importance of domestic species via artificial selection has been a long-established practise for many decades. The economic gain from the Swakara breed industry prompted breeders to invest in the quality development of their pelts. The intense artificial selection and crossbreeding that the sheep breed has been subjected to has not gone without some challenges for farmers. The prevalence of the lethal subvital effect in white and grey sheep is one consequence of severe artificial selection, which has significant economic

impacts. This study intends to analyse the effects of artificial selection on the Swakara sub-populations. Genome-wide analyses of runs of homozygosity, genetic structure as well as identifying selection signatures will provide valuable resources in studying the diversity and genetic distribution in Swakara sheep and the resultant appearance of lethal disorders.

3 GENETIC DIVERSITY AND POPULATION STRUCTURE OF SWAKARA SHEEP SUB-POPULATIONS AND THEIR FOUNDERS

3.1 Abstract

Originally derived from the Karakul breed of Uzbekistan, the Swakara is a fat-tailed sheep breed known for its high-quality pelts. Karakul sheep were first introduced into Southern Africa in the 1900's and subjected to intense selection and crossbreeding with indigenous white breeds of Namaqua Afrikaner and Blackhead Persian to produce sheep whose pelt is more superior with unique features and increased numbers of white-wooled sheep that carry more economic value. Intense selection and crossing of Karakul to indigenous white-wooled breeds resulted in the development of a subvital and sub-lethal factors in the Brown, Grey, and White sheep. In the Grey and pure White sheep, the subvital factor results in the death of a lamb within 48 hours of birth. With a complex ancestral background and evolutionary history, there is a paucity of published scientific information on the genomic architecture, and divergence between the sub-populations making it difficult to understand the factors contributing to the occurrence of genetic disorders. A total of 113 sheep sampled in Namibia were genotyped using the OvineSNP50 beadchip and combined with a previous dataset of 131 animals sampled from Namibia, South Africa, and Germany. Sheep from 8 populations were represented: the Black Vital Swakara (n=51), Grey Vital (n=42), Brown Vital (n=25), Karakul (n=5), White Subvital (n=63) and White Vital (n=34), the Blackhead Persian (n=14) and Namaqua Afrikaner (n=10). We report in this study the genetic diversity and structure of Swakara sub-populations and its presumed founding breeds. Genetic diversity ranged from H_O = 0.29±0.15 for the White Subvital to H_O = 0.41±0.22 for the Karakul, and high inbreeding coefficients were observed in the Namaqua Afrikaner (F_{IS} = 0.13±0.10) and White Vital Swakara (F_{IS} = 0.08±0.09). The principal component analysis produced five clusters with PC1 explaining 27.35% of the total variation and grouping cluster D, which contained the Namaqua Afrikaner only, and cluster E with the Blackhead Persian, separately. Cluster A consisted of the Brown Swakara only, whilst clusters C and B had the Black Vital, Brown Vital, Grey Vital, White Subvital and Karakul. PC2 accounted for 19.25% of the total variation and separated cluster A of the Brown from all the other clusters. The clear separation of the ancestral breeds from the derived Swakara sub-populations was witnessed. Genetic variation of the Brown Swakara from the White, Grey, and

Black Swakara sheep was an indication of divergence of the Swakara sub-populations as a result of selection pressures. The PCA demonstrated the effect of geographic origin with the sheep breeds sampled from Germany clustering together and separate from the Southern African sheep. Pairwise F_{ST} was highest in amongst the Swakara breeds (0.356 for Brown vs White) when compared to the ancestral breeds (0.329 for White vs Namaqua Afrikaner) and other derived sub-populations. The Namaqua Afrikaner cluster was the most highly differentiated with $F_{ST} > 0.25$ when compared with all clusters. Per marker F_{ST} showed differentiated SNPs within QTLs associated with milk production (*RNF220*, *RAB3IP*), wool quality traits (*ICA1*, *IGFBP2*), and coat colour (*CDH13*). The detection of coat colour genes provided evidence to selection targets implemented because of a human led selection. This study gives insight into the genomic architecture of Swakara sheep and its presumed founding breeds towards a better understanding of the underlying causes of the prevalent genetic disorders.

Keywords: *PCA, divergence, genotype, breed diversity, genetic disorder*

3.2 Introduction

As one of the earliest known livestock to be domesticated, the distribution of sheep across the globe has created divergence of single breeds with the selection and crossbreeding practises imposed allowing for the manifestation of multiple breeds and further diversity within breeds. Much of the diversity between these breeds can be witnessed through the wide variety of phenotypes. The Swakara sheep is classified as a fat-tailed sheep breed originating from one of the oldest domesticated sheep of the Karakul (Itenge and Shipandeni, 2015) . Although the Swakara is mainly bred for pelt production, it is classified as a muti-purpose breed that can also be kept for milk, wool, and meat production (Nyoni *et al*., 2020). The Swakara pelt industry has become a cornerstone in the economy and agriculture of countries like Namibia and Botswana. Gross income from Swakara pelts in Namibia in 2013 was around R86, 315, 323 which was a huge increase from 1994 (R7, 607, 284) (Itenge and Shipandeni, 2015). The curl structure of Swakara pelt has resulted in the breed being highly sought after in the pelt industry. The unique pelt was developed through intense selection by breeding for desirable hair traits (colour, lustre, and curl type) which are now distinctly found in the Swakara sheep (Malesa, 2015). The ideal coat colour for farmers is pure white fur that has the flexibility to be easily dyed to other alternative colours.

Since the first importation of Karakul sheep into Africa, it was subjected to extreme levels of crossbreeding mainly with indigenous breeds such as the Namaqua Afrikaner and Blackhead Persian with the primary objective of improving pelt texture (Campbell, 2007). The crossbreeding was also aimed at increasing Karakul ewe numbers whilst breeding for improved reproductive rates (Schoeman, 1998), and various other indigenous breeds were included in the crossbreeding programs. The Namaqua Afrikaner and Blackhead Persian were from the onset considered best breeds for crossbreeding with the Karakul (Frolich *et al.*, 1954; Gouws *et al*., 1970). These two breeds were first selected for upgrading of the Karakul for specific reasons: the Blackhead Persian is characterised by a white body and a head covered with black, short shiny hair (Figure 3.1), and is considered to be hardy and was believed to be ideal for crossbreeding due to the black colour gene being recessive and limited it to the head and only a small proportion of spotted lambs would be produced (Lundie and Coloured Sheep Breeder's Association of New Zealand, 2004). The breed also has short hair and high fertility. Furthermore, similar to the Karakul, the Blackhead Persian consists of narrow primary to secondary hair follicle ratio (Campbell, 2007). The Namaqua Afrikaner (which was believed to produce the best results with the white karakul) has a creamy white wool body with a head that can be black or red coloured (Figure 3.1). Because the Afrikaner is dominantly white and consists of very short hairs, its idealness and suitability was considerably apparent. Unlike the white Swakara sheep, the white Afrikaner does not suffer from the subvital factor that plagues the Swakara.

The Swakara industry is extremely dependent and vulnerable to the demands of the fashion market. The popularity of the white pelt demanded further improvement of the white Karakul. Further breeding of the Karakul to produce this white pelt became problematic as homozygous white genotypes have higher mortality and lower reproductive rates compared to black sheep. Further crossbreeding experiments were carried out over the years with other breeds including the Gotland, and Romanov breeds (Greeff *et al*., 1984). However, these crossings were found

to produce inferior pelt and not further pursued.

Figure 3.1 Flocks of Blackhead Persian (left) and Namaqua Afrikaner sheep (right) adapted from (*https://breeds.okstate.edu/sheep/blackhead-persian-sheep.html*) and (https://www.farmersweekly.co.za/farming-basics/how-to-livestock/all-about-the-namaqua-afrikaner/)

The genetic diversity currently witnessed among the Swakara can be attributed to both artificial and natural selection pressures. Crossbreeding processes to upgrade breeds for preferred traits have resulted in both advantageous and negative attributes that are expressed at the phenotypic and genetic level in these breeds. These historical selection and crossbreeding practises that have defined the genetic composition of the Swakara sheep we see today are largely undocumented. Investigating the genetic contribution of the indigenous breeds in the improvement of the Karakul breed to today's Swakara sheep, and the contribution made by selection in the genomic architecture of the Swakara is key to understanding the observed phenotypes and resultant genetic diversity in these populations. The objective of this study is to fully characterise the genomic structure of Swakara sheep and investigate the genetic relationship/divergence from its presumed ancestors.

3.3 Materials and methods

3.3.1 Animal Material

Genotype data from 131 sheep was obtained from a previous study (Malesa, 2015). Tissue samples from a further 113 animals were collected from the Gellap-Ost research station in Namibia to give a total of 244 sheep samples which included the Brown Swakara that the previous data did not contain. Swakara sheep representing the four different colour sub-populations from pelt-producing farms were used. The Black Vital (n=51), Grey Vital (n=42) Swakara were all sampled from both the Northern Cape, SA and Keetmanshop, Namibia. The Brown Vital (n=25) was solely collected in Namibia. The Karakul (n=5) was sampled from Halle, Germany whilst the White Subvital (n=63) and White Vital (n=34) sheep were sampled

at the Gellap-Ost research station in Namibia. Additionally, the Blackhead Persian (n=14) and Namaqua Afrikaner (n=10) individuals were obtained from Northern Cape as representatives of the presumed founding breeds of Swakara sheep.

3.3.2 SNP genotyping and quality control analysis

The 113 animals were genotyped using the OvineSNP50 genotyping Beadchip (Illumina, United States) which contains more than 54 000 evenly distributed polymorphic SNPs with an average gap spacing of <43kb. Processing of the genotyped data was done at the Agricultural Research Council-Biotechnology Platform in South Africa using Illumina's GenomeStudio v 2.0 (Illumina, United States), pedigree and SNP panel files were generated for downstream analysis. The data was merged with the previous dataset using PLINK v1.9 (Purcell *et al.*, 2007). To ensure better data quality, rigorous standards were imposed on the whole population data. SNP pruning using PLINK v1.9 was performed with thresholds set for removing uninformative markers (MIND) > 0.1 removing animals with missing genotypes, SNPs with missing genotypes (GENO) > 0.01, SNPs with low minor allele frequencies (MAF) < 0.01 and those significantly violating the Hardy-Weinberg Equilibrium (HWE) > 0.0001. Further QC of each sub-population was conducted using the parameters. LD pruning was further applied to the data with parameters set to 50 5 0.5 using the CHM method found in Golden Helix SVS v8.8.1 (Golden Helix Inc, 2016).

3.3.3 Within population genetic diversity

To analyse genetic variation within the sub-populations the observed (H_O) and expected heterozygosity (H_E) were estimated as well as the frequency of each marker in the population using the --het function on PLINK v1.7. Additionally, the inbreeding coefficient (F_{IS}) was estimated on PLINK v1.7 by following the formula of calculating the difference between the expected and observed heterozygosity divided by the expected heterozygosity [$(H_E - H_O)/H_E$]. Minor allele frequency for each sub-population was estimated using the --freq command on PLINK v1.7.

3.3.4 Population structure analysis

An analysis of molecular variance was performed to study the level of relatedness among the individuals and sub-populations. The PGDSpider software (Lischer and Excoffier, 2012) was utilised for the conversion of PLINK files to Arlequin format. The Arlequin v3.5 software (ExcoffierLaval and Schneider, 2005) was employed to deduce the level of population substructure. The hierarchies investigated for included (i) the ancestral breeds of Namaqua

Afrikaner, Karakul and Blackhead Persian (ii) White Swakara (Vital and Subvital) (iii) Karakul and Black Swakara, and (iv) all Swakara sub-populations (White Subvital, White Vital, Brown, Grey and Black).

To assess population clustering Principal Component Analyses were conducted using Golden Helix SNP Variation Suite (SVS) v8.8.1 (Golden Helix Inc, 2016).

Furthermore, ADMIXTURE v1.3 (AlexanderNovembre and Lange, 2009) was used to determine the probable number of ancestral clusters/populations based on the SNP genotype data. ADMIXTURE was run from K = 1 to 15. The optimal K value was determined as that having the lowest cross-validation error (CV-error). FRAPPE v1.0 (Tang *et al.*, 2006) was used to determine the individual clustering and admixture proportions of each sub-population at the optimum number of clusters (K = 9) and it was ran with all 38 116 markers for 5 000 iterations.

3.3.5 Effective population size (Ne)

Ne was calculated using equation 1 of Sved (1971), implemented in SNeP (Barbato *et al.*, 2015), given the known connection between r^2, Ne, and the recombination rate (c) between two loci. By computing the number of generations (t) in the past as 1/2c, multiple estimates of Ne at different time periods were obtained using the different SNP marker distance bins utilised for r^2 analysis.

3.3.6 Pairwise per marker per population F_{ST}

Population pairwise F_{ST} values were estimated for the between pairs of highly differentiated ADMIXTURE based clusters using Golden Helix SVS v8.8.1 (Golden Helix Inc., 2016) calculated following the formula by Weir and Cockerham (1984). Based on the pairwise population values per marker F_{ST} was calculated between highly differentiated clusters ($F_{ST} \geq$ 0.30). Highly differentiating SNPs ($F_{ST} \geq 0.80$) were selected for a gene association analysis. Annotated genes of the significant SNPs including the associated QTLs were identified from ENSEMBL based on the OAR_Rambouillet_V 1.0 reference genome.

3.4 Results

3.4.1 Quality control statistics

Out of the 244 individuals genotyped, three animals had a call rate of <90% and were from the Black Vital, White Subvital and Namaqua Afrikaner groups. These animals were excluded for their low genotyping. Across all populations, SNP marker quality control removed 15 716 SNPs and of these SNPs 4 619 markers were excluded based on Hardy-Weinberg test ($P < 0.0001$), 6 072 failed the missingness test (Geno > 0.01), and 5 639 SNPs failed the frequency test (MAF < 0.01). The Brown Swakara had the highest number of SNPs (n=385) removed due to low minor allele frequency compared to the Grey Swakara which had the least variants excluded (n=144). Overall, a total of 38 116 (71%) SNPs from 241 sheep remained after quality control for downstream analysis. The Grey Vital Swakara maintained the highest proportion of polymorphic markers (71%) whilst the Brown Vital Swakara had the lowest proportion (64%).

3.4.2 Population genetic diversity

The genetic diversity parameters of the eight sub-populations are reported in Table 3.1. Within the sub-populations, genetic diversity ranged from $H_O = 0.29\pm0.15$ for the White Subvital to $H_O = 0.41\pm0.22$ for the Karakul. Expected heterozygosity values ranged from $H_E = 0.30\pm0.16$ for the White Subvital Swakara to $H_E = 0.36\pm0.12$ for the Karakul. High standard errors were observed in the Swakara sub-populations and the Karakul group indicating high variability of individuals with low and high genetic diversity and inbreeding. The Namaqua Afrikaner and White Vital Swakara sheep had the highest inbreeding coefficients of $F_{IS} = 0.13\pm0.10$ and $F_{IS} = 0.08\pm0.09$, respectively. Low inbreeding was observed in the Blackhead Persian and the Karakul with inbreeding values of $F_{IS} = -0.15\pm0.05$ and $F_{IS} = -0.12\pm0.02$. Except for the Blackhead Persian, Karakul, and Brown Vital Swakara, all F_{IS} values were positive suggesting some inbreeding, albeit at very low levels. Mean MAF was found to be the highest in the Karakul population (0.27±0.13) and lowest in the Brown Swakara (0.20±0.16). The ancestral breeds had higher average MAF compared to the Swakara sub-populations.

Table 3.1 Genetic diversity measures for each sub-population of sheep breeds and whole population

	N	SNP %	MAF	H_O	H_E	F_{IS}
BHP	14	64	0.24±0.14	0.36±0.20	0.32±0.15	0.12±0.07
BVS	50	66	0.23±0.15	0.30±0.17	0.31±0.16	0.04±0.08
BRVS	24	62	0.20±0.16	0.34±0.19	0.31±0.16	0.11±0.09
NQA	10	70	0.26±0.14	0.30±0.18	0.35±0.14	0.13±0.10
KAR	5	64	0.27±0.13	0.41±0.22	0.36±0.12	0.15±0.07
WSV	33	64	0.22±0.15	0.30±0.17	0.30±0.16	0.01±0.11
WVS	63	65	0.23±0.14	0.29±0.15	0.31±0.15	0.08±0.09
GVS	42	71	0.23±0.15	0.30±0.16	0.31±0.16	0.03±0.10
ALL	241	71	0.27±0.14	0.32±0.13	0.35±0.14	0.10±0.09

3.4.3 Estimation of N_e

The estimated Ne (t) at t generations in the past was explored. Based on the results shown in Figure 3.2, a sharp increase in the effective population size was seen after 149 generations for all sub-populations. A gradual decrease across generations was witnessed in the Namaqua Afrikaner, Karakul and Black Vital Swakara. An effective population size of <500 was indicated for all populations 65 generations ago and further decreased to <200 26 generations ago.

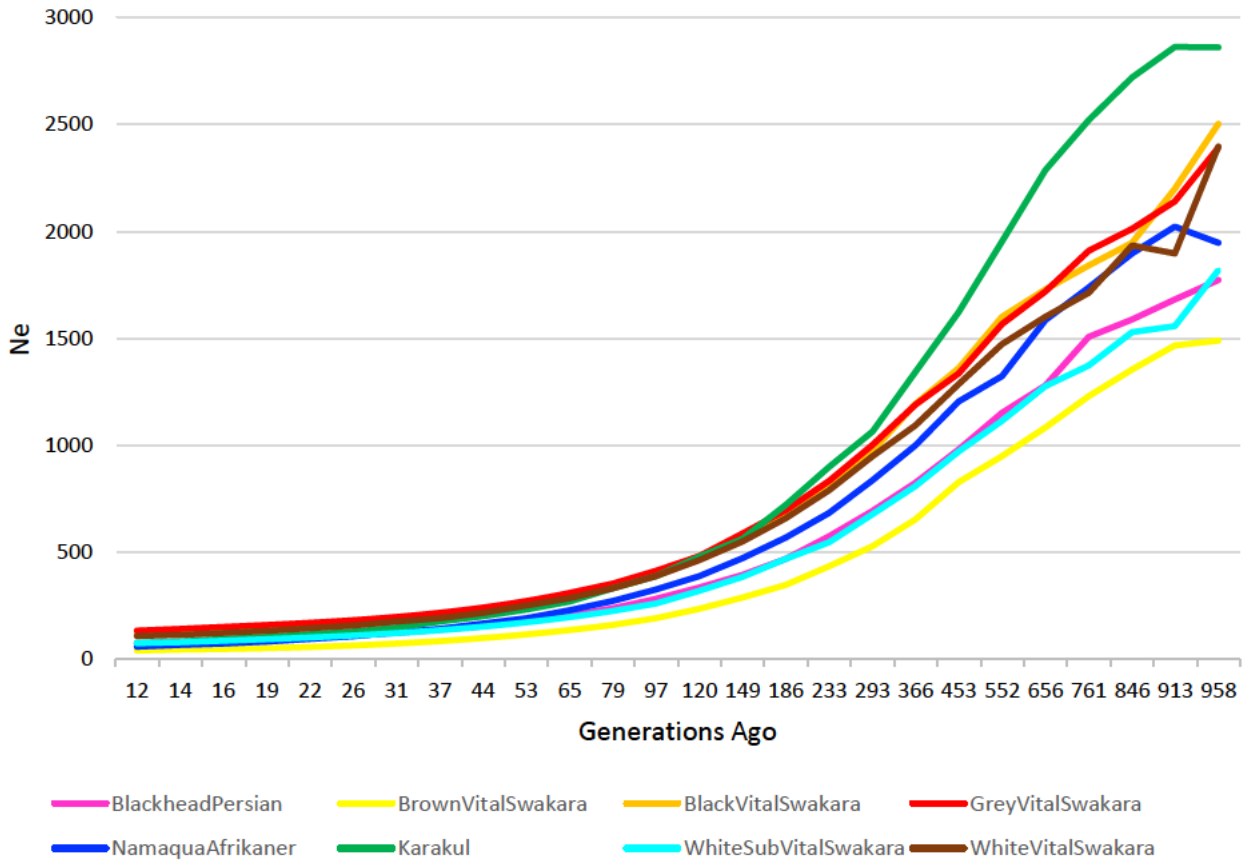


Figure 3.2 Effective population size over the past 958 generations

3.4.4 Analysis of molecular variance (AMOVA)

Molecular variation within the individuals of the White Swakara group was the highest at 98.65%, whilst the ancestral breeds of Namaqua Afrikaner, Blackhead Persian and Karakul recorded the highest variation amongst the different populations with the lowest variation within each of the individuals in the breeds (Table 3.2). The Swakara breed showed little variation amongst the four-coat colour sub-populations at 6.82%. As the original Karakul was black, the genetic variation between these groups was investigated and was recorded to be 13.52% between them. A higher degree of variation was witnessed within the individuals of the different population groups in comparison to among the population sets.

Table 3.2 Molecular variation in different population groups

Population group	Percentage variation (%)	
	Among populations	**Within individuals**
Ancestral breeds	24.162	75.838
White Swakara (Vital & Subvital)	1.346	98.654
Karakul & Black Swakara	13.521	86.479
All Swakara sub-populations	6.824	93.176

3.4.5 Population structure analysis

3.4.5.1 PCA based clustering

The PCA based clustering of animals resulted in five genetic clusters as presented in Figure 3.3. The first principal component (PC1) explained 27.35% of the total variation and separated cluster A (Brown Swakara) from cluster B (White, Black, Grey Swakara), C (White Vital, Black and Karakul), D (Namaqua Afrikaner) and E (Blackhead Persian). The second PCA (PC2) accounted for 19.25% of the total variation and separated cluster A (Brown Swakara) from all the other clusters. The Blackhead Persian (cluster E) and Namaqua Afrikaner (cluster D) clearly separated from the rest of the Swakara sub-populations. The White Vital and Black Vital were part of two different clusters: cluster B and C where they clustered with the Grey Vital and White Subvital in cluster B and with the Karakul in cluster C. Similar to the Blackhead Persian and Namaqua Afrikaner, the Brown Vital separated completely from all the other clusters forming cluster A.

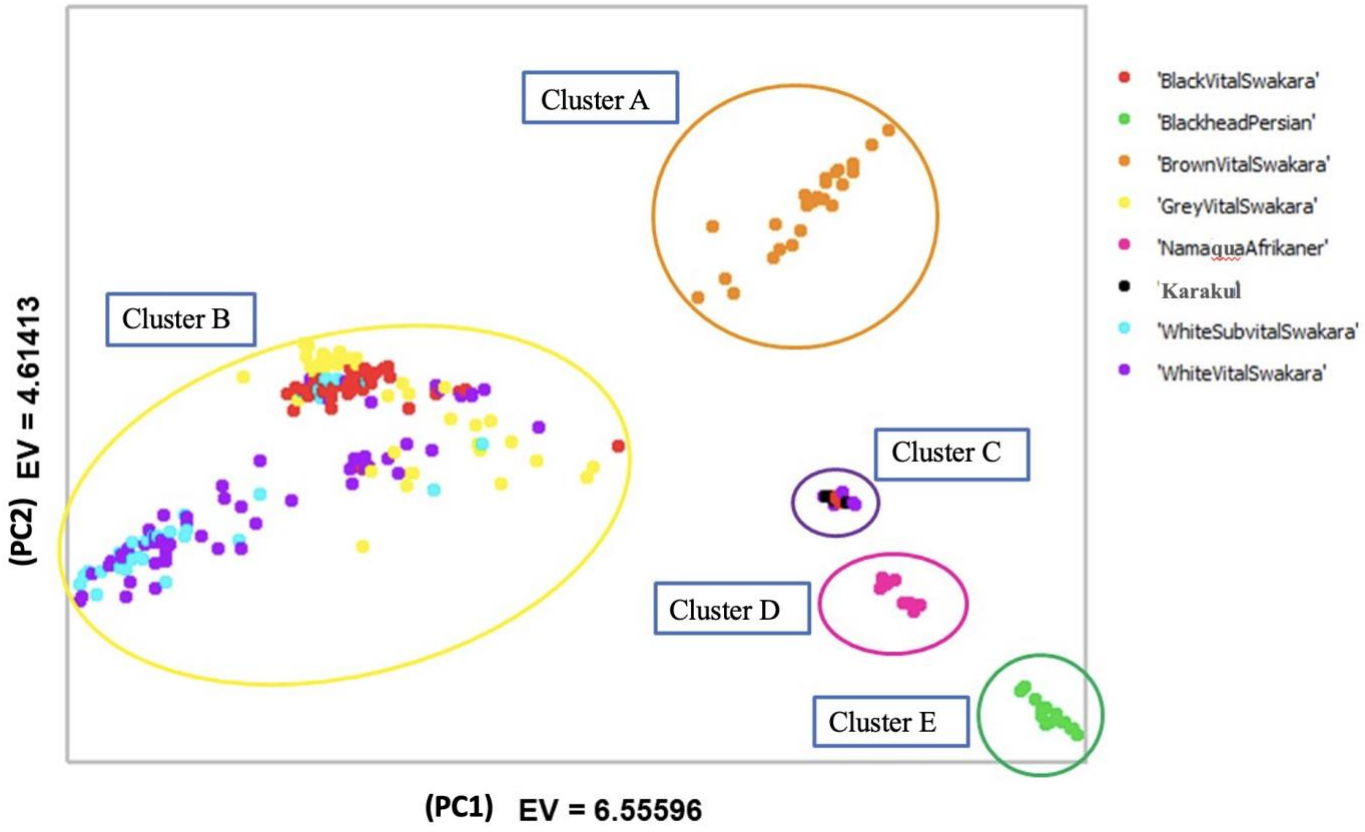


Figure 3.3 PCA based population clustering of sheep breeds from three different breeds of Blackhead Persian, Namaqua Afrikaner and Swakara subpopulations.

3.4.5.2 ADMIXTURE based population clustering.

The genetic structure and diversity of these sheep breeds was further investigated using ADMIXTURE. The cross-validation error was calculated, and the optimal K value was found to be K = 9 (Figure 3.4). The complete and clear separation of the Blackhead Persian from the rest of the populations at K = 2 is observed in Figure 3.5. The White Vital and White Subvital Swakara which clustered together from K = 1 to K = 2, started separating at K = 3. K = 3 and K = 4 saw the Black Vital, Swakara and White Vital clustering the same way as witnessed in the PCA. From K = 4 to K = 9 the White Vital showed high levels of admixture whereas the Karakul came out as less admixed. From K = 2 to K = 9, the Blackhead Persian maintained its homogeny forming its own cluster and with less diversity. Similar to the PCA, the Brown Vital Swakara maintained its clustering with very little admixture.

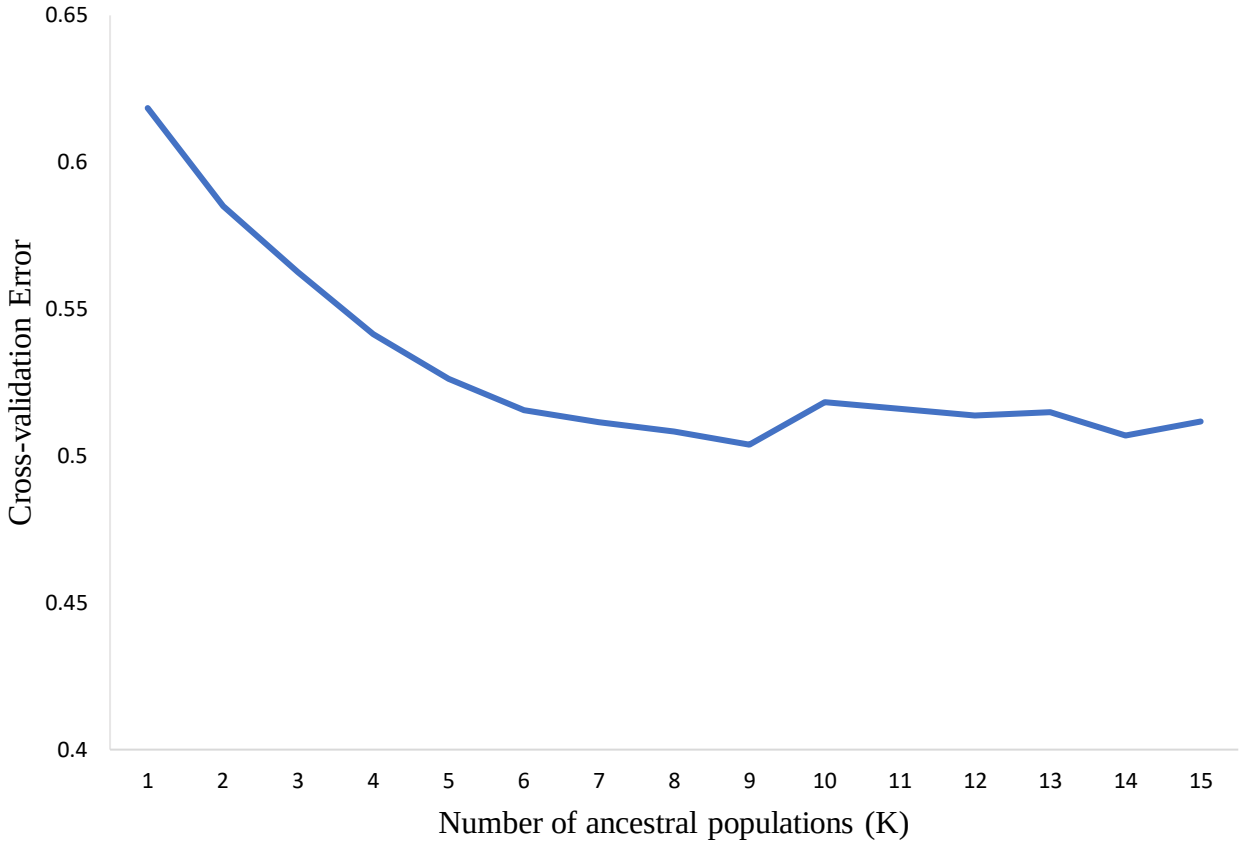


Figure 3.4 Cross-validation errors of K = 1 to K = 15

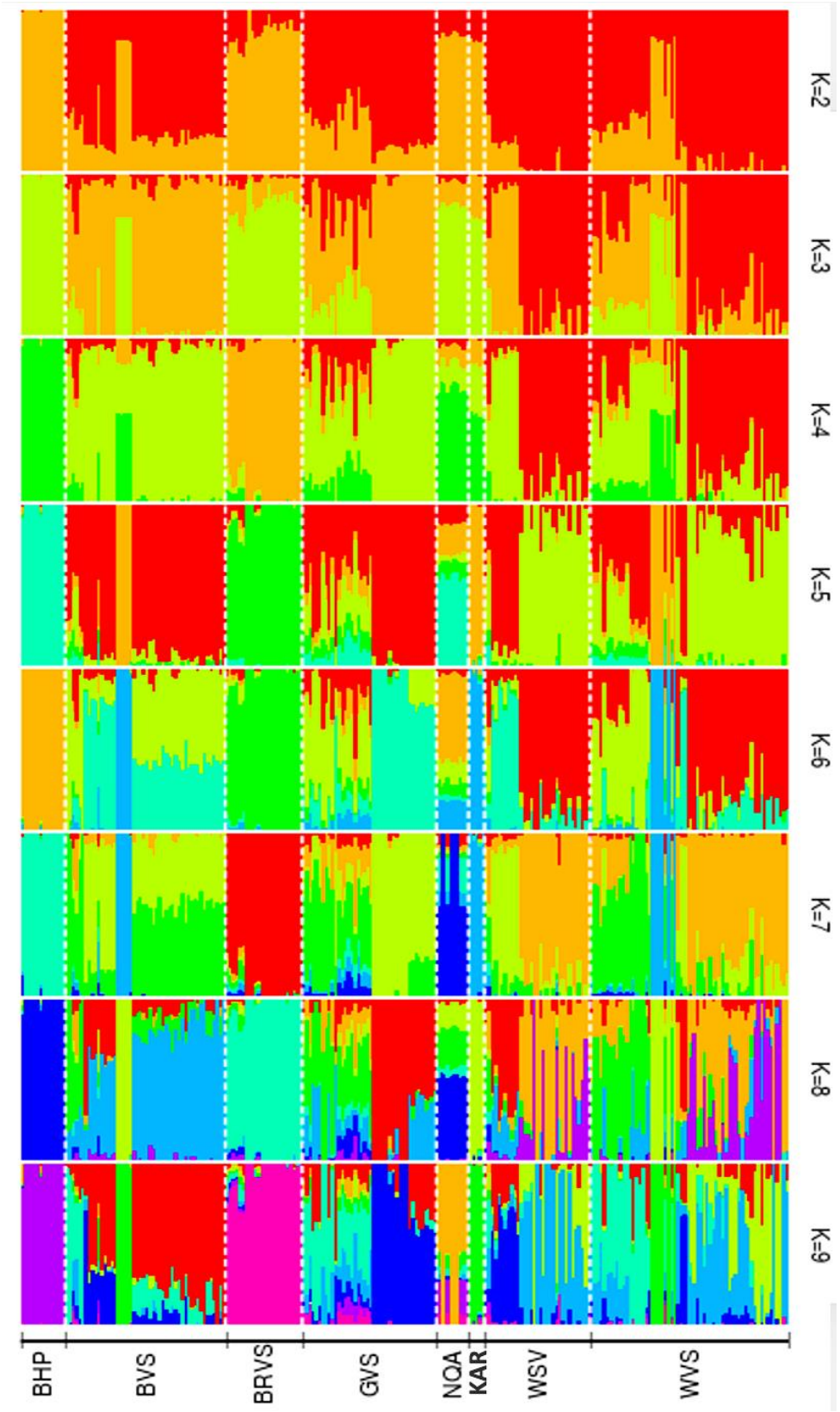


Figure 3.5 ADMIXTURE based clustering of sheep breeds from K = 2 to K = 9. BHP-Blackhead Persian; BVS-Black Vital Swakara; GVS-Grey Vital Swakara; NQ- Namaqua Afrikaner; KAR- Karakul; WSV- White Subvital Swakara; WVS-White Vital Swakara.

The optimal K with the lowest cross-validation error of 0.538 was K = 9. The clustering of individuals at this K was further investigated using FRAPPE which is a more computationally efficient software that allows for the analysis of more markers than other software. It implements an expectation-maximization algorithm for the estimation of individual membership in a group. The Blackhead Persian and Karakul populations were the least admixed with admixture proportions of 0.98±0.05 and 0.97±0.04 observed, respectively (Table 3.3). The Black, and White Vital Swakara shared some components with the Karakul (group 2) and interestingly, the White Subvital Swakara shared no genetic components with the Karakul (0.01±0.02). High admixture proportions were seen for both Black (0.57±0.30) in group 6 and White in group 9 which consisted of the Subvital (0.59±0.39) and Vital (0.42±0.39). The Namaqua Afrikaner seemingly clustered with the Blackhead Persian, albeit at very low proportion (0.11±0.11) in group 5 but was separated individually in group 3 (0.77±0.24).

Table 3.3 Admixture proportions of sheep breeds at K=9

Population	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9
BHP	0.00±0.00	0.00±0.00	0.01±0.03	0.00±0.00	0.98±0.05	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
BVS	0.11±0.14	0.11±0.30	0.00±0.01	0.17±0.22	0.00±0.01	0.57±0.30	0.00±0.02	0.02±0.07	0.02±0.03
BRVS	0.00±0.01	0.01±0.01	0.01±0.01	0.01±0.02	0.00±0.011	0.01±0.02	0.94±0.10	0.00±0.01	0.01±0.02
GVS	0.41±0.39	0.02±0.03	0.02±0.04	0.12±0.15	0.02±0.03	0.15±0.15	0.02±0.03	0.18±0.26	0.04±0.09
NQA	0.00±0.01	0.04±0.05	0.77±0.24	0.00±0.00	0.11±0.11	0.00±0.00	0.03±0.04	0.03±0.03	0.01±0.01
WSV	0.20±0.27	0.01±0.02	0.00±0.01	0.01±0.03	0.00±0.02	0.10±0.12	0.02±0.04	0.06±0.09	0.59±0.39
WVS	0.06±0.12	0.10±0.30	0.00±0.01	0.11±0.26	0.00±0.01	0.08±0.11	0.01±0.02	0.22±0.32	0.42±0.39
KAR	0.00±0.01	0.97±0.04	0.00±0.01	0.00±0.01	0.01±0.01	0.00±0.00	0.01±0.01	0.00±0.01	0.00±0.00

Using the admixture proportion of each individual in the population, each animal with >80% genetic contribution was put into specific clusters that corresponded to the highest membership co-efficient. 123 individuals were subsampled based on this threshold for further analyses. Group 1 (n = 10) consisted of Grey Vital only. Group 2 (n=16) was the most diverse cluster

consisting of the White Vital, Black Vital and Karakul. The least diverse clusters were group 3 (n=5), group 5 (n= 14) and group 7 (n=21) comprising of only the Namaqua Afrikaner, Blackhead Persian, and Brown Vital, respectively. The White Vital was the common sub-population in all the groups mainly being found in four different groups. The group with the lowest number of individuals was group 3 which was made up of only just 5 of the Namaqua Afrikaner.

A second PCA was run to determine the clustering of the sheep based on their ADMIXTURE proportions and determine significant genomic regions in the sub-populations. Based on the FRAPPE results (Figure 3.6) the genetic contribution of each individual in each cluster was utilised. Individuals in group 1, 4, 6 and 8 gathered together in cluster B which contained the Black Vital, White Vital and Grey Vital. Expectedly, some animals from the same cluster were grouped together as witnessed in cluster D (all from group 2) and cluster C (all from group 9). As witnessed in the first PCA and Admixture plot, the Blackhead Persian, Namaqua Afrikaner and Brown Vital remained separated from the rest of the groups/clusters.

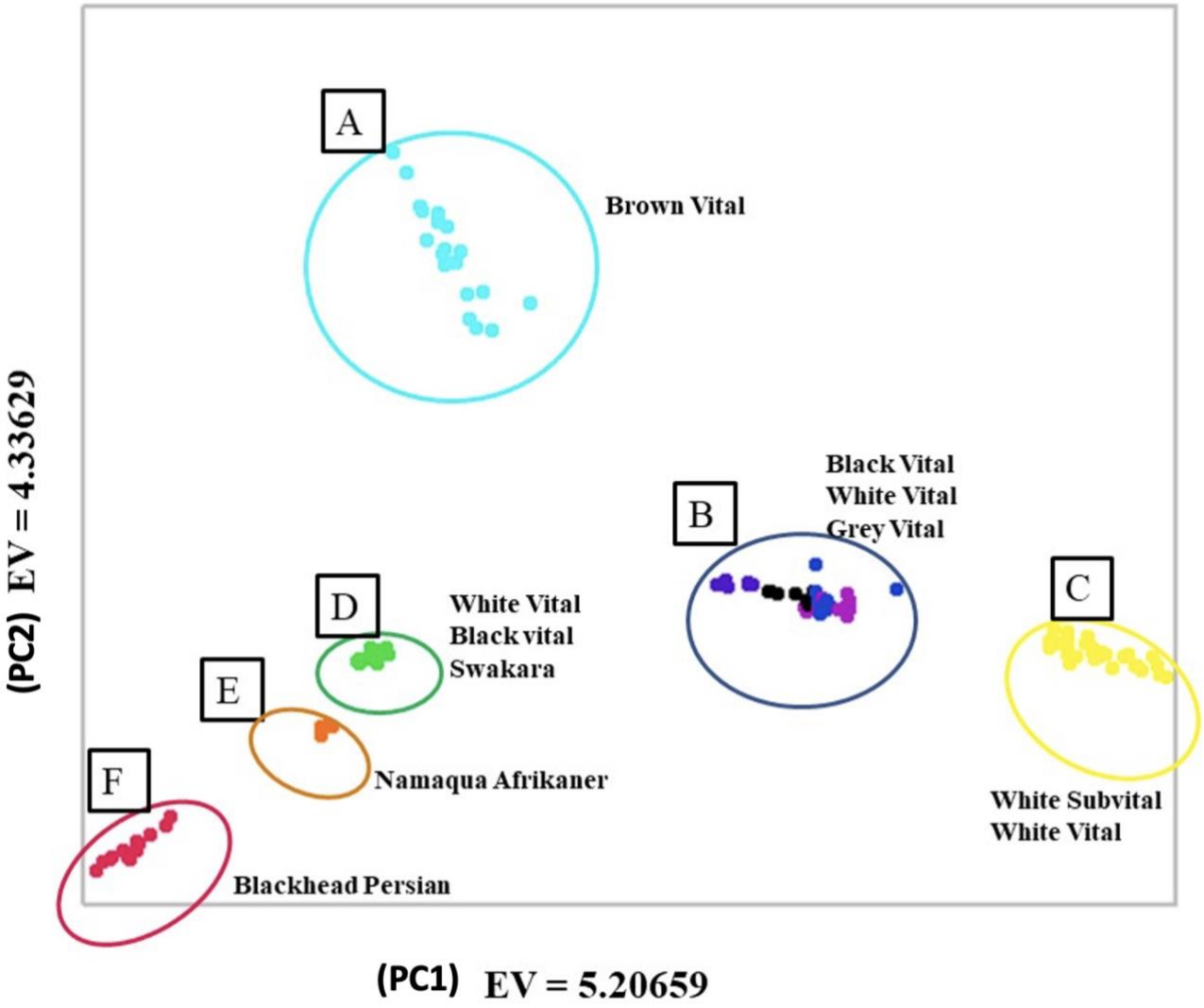


Figure 3.6 PCA clustering of individuals based on ADMIXTURE genetic proportions

3.4.6 Population Pairwise F_{ST}

Population pairwise F_{ST} values are shown in Table 3.4. The genetic differentiation between the different clusters ranged from F_{ST} = 0.102 between clusters B (Black, Grey, and White Vital) and C (White vital and Subvital) to F_{ST} = 0.356 between clusters E (Namaqua Afrikaner) and A (Brown Vital). Cluster E was more highly differentiated (F_{ST} > 0.25) from all other clusters with the highest differentiation being with cluster A as illustrated in Figure 3.7. Cluster B differed greatly from cluster E and F (Blackhead Persian) but showed the least differentiation with cluster C. Moderate diversity was also observed between cluster B and D which both consisted of the White Vital and Black Vital Swakara. High diversity (F_{ST} > 0.25) was also seen between cluster E (with only Namaqua Afrikaner) and all other clusters. Pairwise per marker comparison was plotted for clusters with F_{ST} > 0.25 as shown in Figure 3.7. These

included the (i) B and E (F_{ST} = 0.256), (ii) D and E (F_{ST} = 0.279, (iii) D and F (F_{ST} = 0.251), (iv) E and F (F_{ST} = 0.336), (v) A and E (F_{ST} = 0.356), and (vi) C and E (F_{ST} = 0.329), (vii) A and F (F_{ST} = 0.311), and (viii) C and F (F_{ST} = 0.289) clusters. Highly differentiating SNPs and their associated genes were detected.

Table 3.4 Pairwise genetic differentiation (F_{ST} Value) of sheep genetic clusters

	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E	Cluster F
Cluster A	0					
Cluster B	0.171					
Cluster C	0.237	0.102				
Cluster D	0.218	0.126	0.197			
Cluster E	0.356	0.256	0.329	0.279		
Cluster F	0.311	0.225	0.289	0.251	0.336	

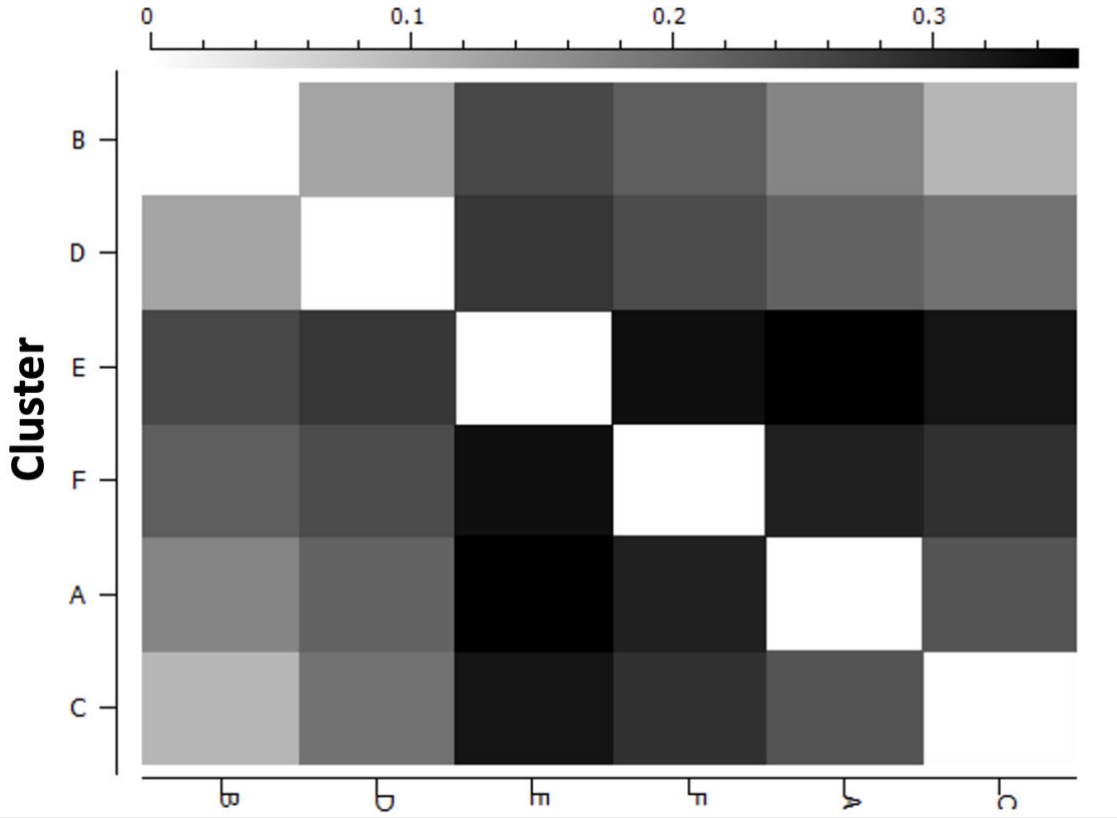
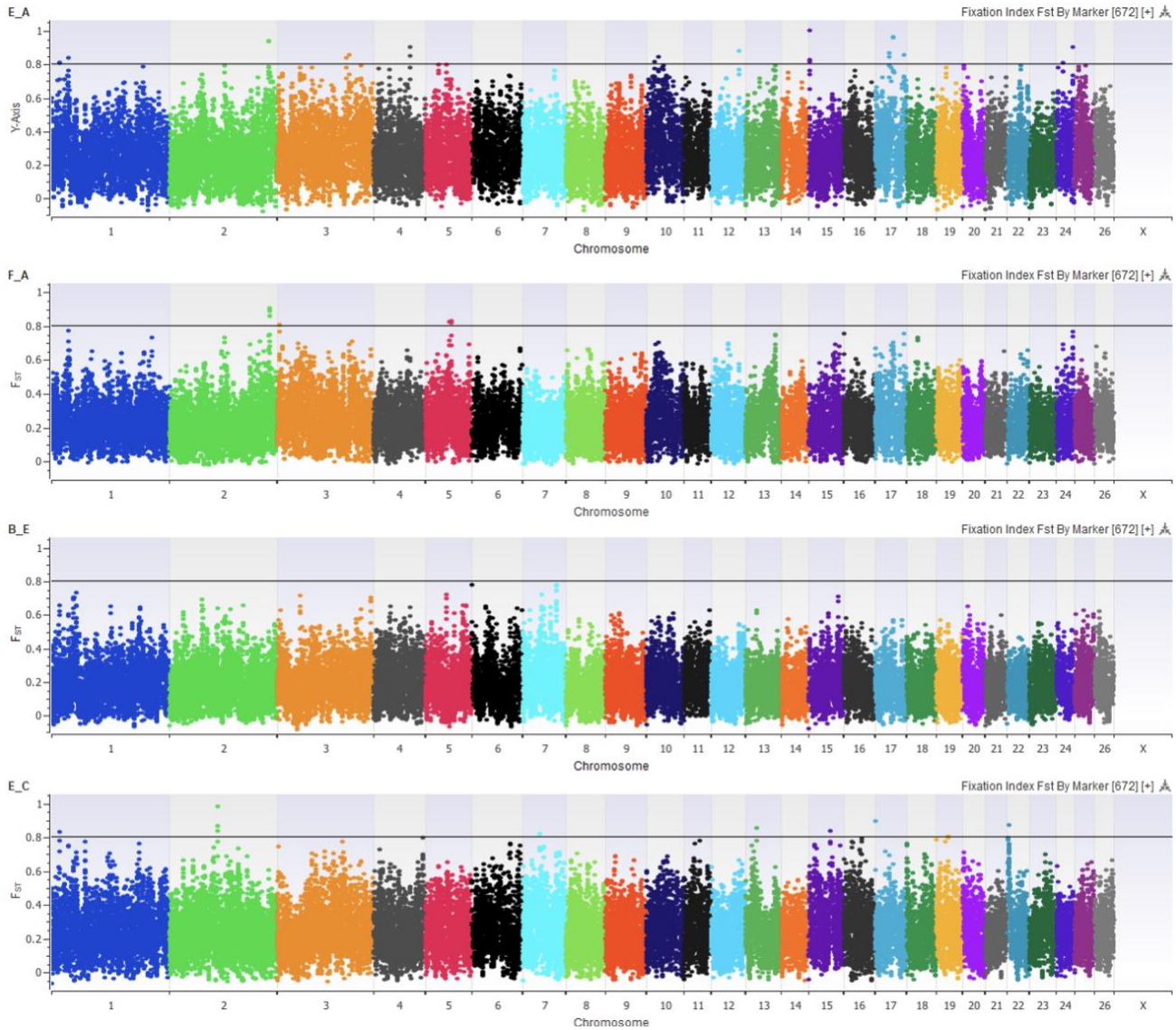


Figure 3.7 Genetic differentiation of sheep clusters

3.4.7 Per marker pairwise F_{ST}

Using a threshold of 0.8, the per marker F_{ST} analysis revealed a total of 114 SNPs ($F_{ST} \geq 0.80$) under selection that were annotated for genes within a 1Mb region (Figure 3.8). Pairwise comparison for all 6 clusters detected 48 differentiated SNPs with 36 candidate genes within the region of these SNPs (Table 3.5). Significant SNPs were mainly detected in chromosomes 1, 3, 4, 7, and 11 across all clusters. Chromosomes 5, 9, 10, 13, 14, 17 and 22 had the least number of differentiating SNPs. The loci with the greatest F_{ST} values (>0.95) were located on chromosome 11 between the D (Karakul) and E (Namaqua Afrikaner) cluster comparison with the *CARD14*, *SLC26A11*, and *RNF213* genes.



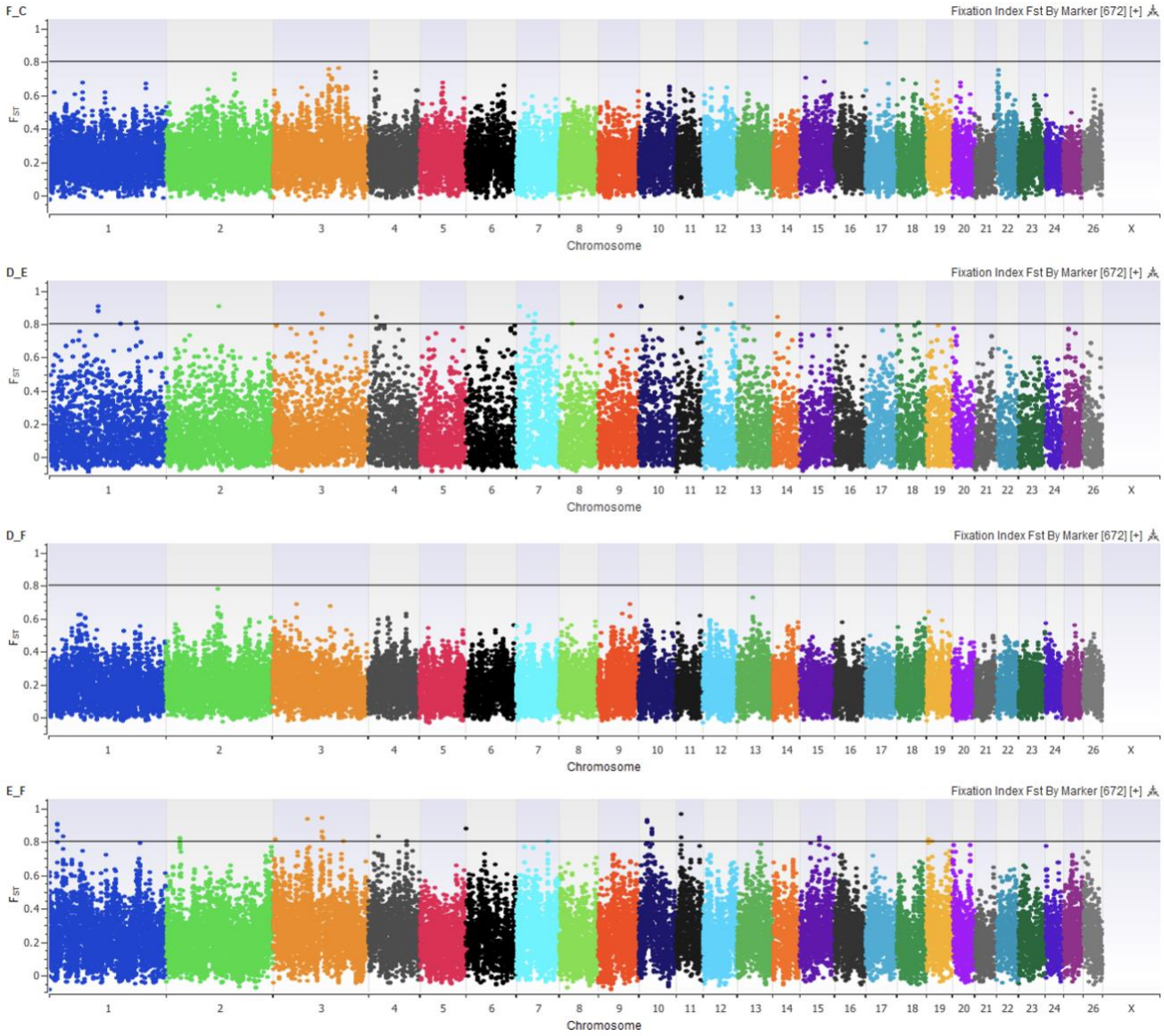


Figure 3.8 Distribution of per marker F_{ST} comparisons across all 26 autosomes of clusters A-F.

Table 3.5 Significant SNPs (F_{ST}≥0.80) with associated genes across six clusters

Clusters	CHR	SNP	F_{ST}	Gene	Function
A and E	3	OAR3_160874701.1	0.84	*RAB3IP*	Milk production (Saravanan *et al.*, 2021)
	1	OAR1_19750493.1	0.81	*RNF220*	Milk yield (Vanvanhossou *et al.*, 2021)
	4	OAR4_84603727.1	0.90	*CAMK2B*	Reproduction (Yao *et al.*, 2022)
	10	OAR10_27796789.1	0.84	*DCLK1*	Response to heat stress (Tolone *et al.*, 2022)
	2	OAR2_230317713.1 s07742.1	0.94 0.94	*SPAG16*	Sperm motility/reproductive system development (Roy and Matzuk, 2006)
	12	s50610.1	0.88	*ACBD6*	Growth and carcass traits (Roberts, 2018)
A and F	2	s66422.1 OAR2_233321958.1 s08098.1	0.90 0.89 0.86	*IGFBP2*	Hair follicle development (Tian *et al.*, 2022)
	5	OAR5_59984741.1	0.82	*POU4F3*	Fat yield (Kim *et al.*, 2013)
C and E	1	OAR1_19710996.1	0.83	*RNF220*	
	13	OAR13_25029451.1	0.85	*PTF1A*	Tail development (Han *et al.*, 2019)

	2	OAR2_110779829.1	0.84	*ELP3*	Stem cell maintenance/early development (Stafuzza *et al.*, 2019)
	22	OAR22_5540854.1	0.87	*PCDH15*	Immune response (Chen et al., 2021a)
	7	s11708.1	0.82	*UBR1*	Immune response (Serranito *et al.*, 2021)
C and F	17	OAR17_172595.1	0.91	*No gene*	Milk yield (Vanvanhossou *et al*., 2021)
D and E	1	OAR1_204344399.1	0.81	*IQCB1*	
	1	OAR1_204419440.1	0.81	*EAF2*	Disease resistance (Abied *et al*., 2020a)
	1	s65811.1	0.81	*No gene*	
	1	OAR1_114668260_X.1 OAR1_114647847.1	0.91 0.88	*KIRREL1*	
	11	s26679.1 s26637.1	0.96 0.96	*CARD14*	Inflammatory response *(Moisá et al.*, 2015)
	11	s22806.1	0.96	*SLC26A11*	Fat tail trait (Ahbara and Latairish)
	11	s18615.1	0.96	*RNF213*	Fecundity (Sánchez-Ramos *et al*., 2023)

Table 3.4 (Continued)

Cluster	CHR	SNP	F_{ST}	Gene	Function
D and E	12	OAR12_72007015.1	0.81	*HMCN1*	Teat number
	12	s55444.1 s50610.1 s67454.1	0.92 0.92 0.92	*ACBD6*	
	14	s21609.1	0.84	*CDH13*	Coat colour (Mastrangelo *et al.*, 2019)
	3	OAR3_114955879.1 OAR3_115060659.1	0.86 0.86	*TRHDE*	Postnatal growth trait (Zhang *et al.*, 2016)
	4	OAR4_18402375.1	0.84	*GLCCI1*	Sperm motility (Suchocki and Szyda, 2015)
	4	s03265.1	0.84	*ICA1*	Hair/Wool traits (Dzomba *et al.*, 2021)
	7	OAR7_28980510.1	0.85	*RYR3*	Tail shape (Xu *et al.*, 2021a)
	7	OAR7_44085704.1	0.86	*SOS2*	Meat and carcass
	7	OAR7_44184132.1	0.86	*CDKL1*	Gestation length (Purfield *et al.*, 2019)
	9	OAR9_49941341.1 OAR9_49953513.1 OAR9_50008246.1	0.91 0.91 0.91	*C8orf34*	Carcass weight (Lopez *et al.*, 2020)
E and F	1	OAR1_19710996.1 OAR1_19750493.1	0.91 0.90	*RNF220*	

11	s26637.1	0.82	*CARD14*	Regulation of fatty acids (Xie *et al*., 2021)
3	s33836.1	0.81	*MED27*	Regulate body size (Zheng *et al*., 2020)
3	OAR3_115060659.1	0.94	*TRHDE*	
3	OAR3_80722364.1	0.93	*FSHR*	Total lambs born
3	OAR3_165741097.1	0.80	*MSRB3*	Ear morphology (Cheng *et al*., 2022) (Posbergh and Huson, 2021, Paris *et al.*, 2020)

3.5 Discussion

This study sought to determine the genomic architecture of a largely understudied Swakara sheep breed with the main objective of investigating the genetic diversity between the different Swakara sub-populations as well as the level of differentiation from the ancestral breeds. Several studies have investigated population and breed diversity and illustrated how significant a deeper understanding of the genomic make-up can consequently provide more insight into breed associated traits. The release of the OvineSNP50 Beadchip single nucleotide polymorphisms has led to it becoming the most widely used type of genetic marker for extensive investigations of sheep diversity, genome-wide selection and overall population substructure (Kijas *et al.,* 2009). A wide range of global sheep breeds were utilised in the development of this SNP array including many indigenous African breeds such as the Namaqua Afrikaner, Red Masaai, and African Dorper. Several studies have used the OvineSNP50 Beadchip to investigate the genetic structure of different sheep breeds, determine significant loci under selection and dissect quantitative traits of economic importance (Molotsi *et al.,* 2017, Miller *et al.*, 2010, Usai *et al.*, 2019).

Understanding the genetic structure of sheep offers the chance of observing the influence of domestication, selection and subsequent evolution into discrete breeds and the various phenotypic characteristics observed in those breeds. Genome-wide studies can be essential in identifying genomic regions with associations to disease phenotypes. The origin and evolution of the Swakara breed from the various crossbreedings and selection have left its genetic construction to be poorly understood. Although previous studies have explored the genetic variation in Swakara, the present study sought to provide a larger overview of the genetic structure in relation to its ancestral breed in a large pool of samples with a greater geographic distribution.

The estimated heterozygosities demonstrate high inbreeding in Namaqua Afrikaner and White Vital Swakara (F_{IS} = 0.13±0.10 and 0.08±0.09, respectively) and low diversity levels in all the Swakara coat colour groups with high standard deviations suggesting a wider distribution and diversity in the dataset. Dzomba *et al.* (2021) reported similar low levels of genetic diversity for pelt producing Swakara sheep. (Buduram, 2004), observed lower levels of genetic diversity for the Namaqua Afrikaner which was relatively lower than the values reported for other indigenous breeds such as the Swazi and Pedi sheep breeds. The low genetic

diversity in Swakara can be attributed to the breed being kept as a closed population for many years and experiencing high selection pressures under non-random mating methods. This observation is supported by the significant clustering of the populations and admixture analysis. The number of animals for some breeds/sub-populations was significantly low which would have impacted the results that slightly differed from literature. Low MAF was observed for all Swakara sheep where the lowest value of 0.20±0.16 was reported in the Brown Vital Swakara with the highest being 0.23±0.15 in the Black and Grey sheep. There were no major differences observed in the MAF of these sub-populations, supporting the idea of geographically restricted groups observing low MAF (Kijas *et al.*, 2009; Dzomba *et al.*, 2021). Two of the sub-populations of Swakara (Grey and White) that also happen to experience the subvital effect and whose pelt is most popular had some of the lowest genetic diversity observed between all the sub-populations in this study. The relatively high levels of inbreeding and low genetic variability can be attributed to intense selection and breeding for economically important traits in conjunction with the absence of gene flow within these sub-populations.

Within the Swakara breed, effective populations size was the lowest in the Brown (40) and White Subvital (75) populations approximately twelve generations ago. A minimum Ne of 50 has previously been suggested by FAO for the maintenance of a positive genetic trend within a population (FAO, 2000). Other studies have recommended an effective population size of at least 100 animals in a generation (Meuwissen, 2009, Prieur *et al.*, 2017) whilst Franklin (1980) has suggested a size of 500 with the suggestion that an effective population of less than this makes the population at risk of extinction due to inbreeding effects. The low Ne observed in all these Swakara sub-populations is highlighted in the low genetic diversity which is propelled by strong selection effects for pelts and surviving in rural, underdeveloped systems. The highest Ne observed within the Swakara sub-populations was 133 for the Grey Swakara. Ne values varied between the sub-populations of Swakara and between their presumed ancestral breeds but were still similar within the same ranges. Expectedly, low Ne was also reported in all the ancestral breeds with numbers as low as 58 in the Namaqua Afrikaner. The Swakara and its presumed ancestors in this study are known to exist in fragmented populations and are considered to be endangered. Dzomba *et al.* (2021) has previously reported similarly low effective population size in Swakara sheep as well as in indigenous ancestral breeds of Karakul, Namaqua Afrikaner, and Blackhead Persian.

The results from the PCA and ADMIXTURE were aligned, grouping individuals according to their genetic composition and overall genetic similarities. Almost all Swakara coat colour

groups clustered together in cluster B with the exception being the Brown Swakara clustering separately in cluster A. This suggests the existence of a more pronounced genetic structure between the Brown and the rest of the Swakara coat colour groups. Out of all pelts colours the Brown sheep are definitively less popular thus are kept in relatively fewer Swakara herds (which are already smaller numbers) allowing very little gene flow between it and the other pelt colour sheep (Schoeman, 1998). With there being so many different breeds utilised in the development of the Swakara sheep there is very little known about the development of the brown coat colour suggesting that there possibly might be another ancestral breed used for crossbreeding that produced the Brown Swakara. Significantly, some White and Black sheep clustered with the Karakul demonstrating a weak genetic structure amongst these sub-populations. This was not surprising as these are the sub-populations heavily linked to the Karakul with the White having been involved in intense crossbreeding originally to produce white pelts. Additionally, the original Karakul was black, suggesting some genetic derivative and similarity with the Black Swakara that we see today. The first PCA demonstrated the effect of geographic origin with the sheep breeds sampled from Germany clustering together, those from Southern Africa clustering together and the Somalian-influenced breeds separating together from all the other clusters. The genetic division of breeds based on geographic history has similarly been reported by Kijas *et al.*, (2012) in a study of global sheep where European, Asian, American, and African sheep displayed phylogeographic patterns. Furthermore, individuals belonging to the same breed but sampled in different regions e.g., White vital clustering differently further supports the idea of geographic history having an influence on breed cluster.

Identifying, quantifying, and comparing genetic differentiation between and among populations is a fundamental objective in population genetics studies (Bird *et al.*, 2011). Arranging the individuals of the sub-populations into different clusters based on the admixture proportions allowed for better analysis of the clusters, and the different individuals in them so as to ascertain their genetic differentiation. Population pairwise F_{ST} revealed high genetic differentiation amongst six different cluster pairs. According to Machete *et al.* (2021), values ranging above 0.25 are considered to indicate very large genetic differentiation, whereas low genetic differentiation is generally considered to be between 0 and 0.05, and moderate differentiation is considered to be between 0.15 and 0.25. Genetic differentiation was relatively high between the different clusters with the lowest values recorded for two cluster pairs, clusters B (Grey/Black) and C (White) reporting a value of 0.102 and clusters A (Brown) and

B (0.171). Majority of the Swakara pairwise cluster comparisons fell within the moderate range of differentiation. The highest degree of differentiation exhibited between the Swakara, and the ancestral breeds was with the Brown Swakara and the Namaqua Afrikaner (0.356). With the ancestral breeds, lowest genetic differentiation was seen between the Karakul and White (0.197) and Grey/Black (0.126). The selection and breeding forces imposed on these breeds is further highlighted with the observed high genetic similarities. Genetically, the variation within these sub-populations was low therefore regardless of how they clustered individually the differentiation observed between them in the clusters will ultimately be low. The Swakara is a generally minor breed kept in small population sizes and poorly advanced breeding systems. The most highly differentiated ancestral cluster when compared against other Swakara clusters is cluster E with F_{ST} values of more than 0.25. Cluster E consisted of only the Namaqua Afrikaner population indicating that it is a highly homogenous and non-admixed population with less genetic correlation to the Swakara. The high F_{ST} levels observed in this cluster/breed against other indigenous Southern African sheep is consistent with the pairwise F_{ST} values of other studies (Dzomba *et al.*, 2020, Molotsi *et al.*, 2017). This may be attributed to genetic divergence as a result of natural selection pressures and the influence of founder effects in these sub-populations.

Per marker F_{ST} was estimated between pairs of highly differentiating clusters ($F_{ST} > 0.30$). Six cluster pairs were found to have high genetic variation, and this analysis was used to study candidate genes from SNPs having a threshold of $F_{ST} > 0.80$ and their associated traits. The Swakara sheep, is a fat-tailed breed that was first developed in Southern Namibia and North-Western Cape. These areas are known to be extremely dry and hot which has made the breed well adapted to harsh environments. High genetic variation was found between some Swakara sub-populations and the ancestral breeds with the most SNPs being identified between the Brown and Namaqua Afrikaner across 6 different chromosomes. Genes related to hair follicle development (*IGFBP2*), and heat tolerance (*DCLK1*) shows the key selection traits that were pivotal in the improvement of the Swakara artificially and naturally due to adaptation. The Swakara and its ancestral sheep breeds utilised in this study are known to have been able to adapt to the severe heat and dryness that characterises the climactic conditions in Southern Africa. Thus, the identification of these genes signalises their ability to have adapted to these conditions over a period of time. Hair and wool traits are often contradictory to pelt traits in Swakara, and genes related to hair follicles were identified as the major selection is on pelt quality. Notable genes between the indigenous breeds and the developed Swakara were found

to be related to immune response: *UBR1*, *PCDH15* (White and Namaqua Afrikaner), which is important for overall ovine health and survival, tail development: *PTF1A* (White and Namaqua Afrikaner), the fat-tail of the Swakara is a significant element to its survival, and other genes related to milk and meat production *RAB3IP* (milk production; Brown and Namaqua Afrikaner), *RNF220* (milk yield; Brown and Namaqua Afrikaner), *POU4F3* (fat yield; Brown and Blackhead Persian). Three genes were found to be related to growth: *TRHDE, ACBD6* and *CD86.*

Being one of the breeds used in the improvement of Karakul, the Blackhead Persian and Namaqua Afrikaner have had great influence in the genetic composition of Swakara sheep in Southern Africa today. The identification of a coat colour gene in this study (*CDH13*) was significant as it was an indication of selection targets in the Swakara sub-populations. The morphological consequence of selective breeding is clearly observed in the number of different coat-coloured sub-populations of the Swakara where the effects of selection have been detected and shown to have resulted as a response of a human-led breeding objective.

3.6 Conclusion

The overall objective of the study was achieved as it provided insight into the genomic structure of Swakara sheep sub-populations as well as the relationship with its presumed founding breeds. A strong genetic relationship between the Swakara sub-populations was seen with some differentiating characteristics. Some sub-populations such as the White and Brown presented differentiating traits related mostly to hair, milk production and immune response related traits. Hair/wool related traits were similarly seen in the Karakul and Namaqua Afrikaner with differentiating traits related to coat colour and tail traits observed. The genetic diversity investigation elucidated on the unknown effects of the selection and breeding practises that these sub-populations have been subjected to. Furthermore, the utilisation of the OvineSNP50K Beadchip and its ability to provide salient insight into breed characteristics and relations was demonstrated. Further analyses will provide more knowledge and understanding in discerning the influence of certain breeding practises on overall phenotypic traits and diseases. This study highlighted the lack of genetic diversity in indigenous smallholder sheep breeds and the necessity of implementing efficient breeding techniques that will maximise profitability and ensure preservation of important genetic resources.

.

4 DETECTION OF SELECTION SWEEPS IN SWAKARA SHEEP AND ITS ANCESTRAL BREEDS

4.1 Abstract

A native breed of Uzbekistan, the Karakul is known for being a multi-purpose breed used for milk and meat production, and mainly bred for its unique pelts. Since its introduction into Namibia and South Africa in the early 1900s, it has been transformed into an entirely new and different breed known as the Swakara (South-West African Karakul) characterised by unique and superior pelt. This has been achieved through crossbreeding with indigenous breeds such as the Namaqua Afrikaner and Blackhead Persian followed by intensive selection. Through the intense selection and resultant improved higher quality pelt, this breed is however, burdened with a lethal and subvital factor in the grey and white sheep respectively and a disorder in the brown, all of which are presumed to have emanated from the intensive selection practises. This study sought to investigate genomic regions that have been affected by selection pressures in the Swakara sub-populations as well as the believed ancestral breeds of the Namaqua Afrikaner and Blackhead Persian in an effort to make inferences on the observed phenotypes in Swakara subpopulations. Statistical methods were utilised to detect signatures of selection in 106 sheep belonging to six sub-populations of Brown (n=21), White (n=32), Grey/Black (n=24), Karakul (n=10), Blackhead Persian (n=14), Namaqua Afrikaner (n=5), followed by annotation of the genes associated with the identified genomic regions. Pairwise comparison of the Brown and White coat colour clusters yielded the most signatures with genes associated with meat colour and foreleg length (*GLRA3*), body height (*GALNTL6*), and quantity of internal fat (*TNS2, LIMA1, ASIC1, PRPF40B, KCNH3, CACNB3*). There was a commonality of five signatures among these subgroups. When comparing the Brown sub-populations to the Grey/Black sheep, QTLs for primary fibre diameter (*PTN, ADAMTS3*), face eczema susceptibility (*TEAD1, ANKRD17*), and milk protein yield (*DENNS2B*) were found. There were more discernible selection sweeps between the Swakara subpopulations than there were between the Swakara and the ancestral breeds. The Karakul and White Swakara pairwise comparison demonstrated selection sweeps associated with QTLs of testes weight, milk output, meat acid content, and worm count (*MTMR10, TRPM1, KLF13, TJP1, HMG20A*) that were limited to chromosome 8. Seven reoccurring genes were found in the Brown Swakara subpopulation, independent of the sheep cluster to which it was matched (*PTGER, PARVA, FBXO8, C1QL4, TROAP, NEIL2, TFCP2*). Almost 2000 regions were identified with the HapFLK analysis associated with over 700 genes. A small fraction of genes detected were shared amongst all three methods, with

only 6 between the XP-EHH and HapFLK methods (*CELF2, USPNL, ECHDC3, UPF2, CAMK1D, CHST11*), and only 3 between the iHS and XP-EHH analyses (*GALNT18, HMG20A, CSPG4*). QTLs essential in growth and health were detected in the Brown Swakara with genes such as *TMEM163* and *RPGRP1*. Detected signatures displayed the multi-purpose nature of Swakara sheep with traits related to meat production in Grey/Black cluster and milk related traits (*FBN2*) in the White Swakara identified. Overall, the Swakara sub-populations displayed some shared genetic similarities within and between the coat colour groups and their ancestral breeds.

Keywords: *iHS, selective sweeps, HapFLK, body weight, XP-EHH*

4.2 Introduction

The domestic sheep (*Ovis aries*) has long been the central component of many economies providing value through meat, wool, milk and other by products. Spanning over 8 million years, the genus Ovis has evolved throughout the years with eight main species constituting it today, mainly the urial *O. vignei*, argali *O. ammon*, thinhorn sheep *O. dalli*, Asiatic mouflon *O. orientalis*, the domestic sheep *O. aries*, European *mouflon O. musimon*, snow sheep *O. nivicola*, and bighorn sheep *O. canadensis* (Chen *et al.*, 2021b). The development of domestication syndromes which characterises/describes shared physiological and morphological traits among domestic species are believed to have resulted from selection to enhance tameness (Alberto *et al.*, 2018).

Swakara sheep are a specific breed of sheep found in Southern Africa and bred primarily for their luxurious and highly sought-after pelts. The objective of breeding Swakara sheep has been on producing pelts with particular characteristics, such as a dense, curly, and lustrous wool, as well as specific colour patterns. Swakara sheep have undergone selective breeding over generations to enhance desirable traits such as fur quality, growth rate, and resistance to diseases. These traits contribute to the high-quality and unique appearance of Swakara pelts, making them valuable in the fur industry. On the other hand, the subvital disorder experienced by some of the white and grey sheep that results in the early death of lambs as well as genetic disorders in the white and grey Swakara highlights the genetic imperfections that exist within this breed.

The intentional selection imposed on animals for the enhancement of desired phenotypes radically transformed the behavioural attributes from their wild counterparts and positive

selection is believed to have been triggered for many traits as a result (Naval-Sanchez *et al.*, 2018). The genetic signatures of selection in organisms refers to the specific genetic variations and patterns that have been favoured by selective breeding practices over time, leading to the distinctive traits observed in a particular population or breed (Nsoso and Madimabe, 1999). Genome screening studies and analyses have become the key in unlocking footprints of selection and their effects on breed performance. These studies involve examining the genomes of individual sheep within the breed and identifying regions of the genome that have experienced strong selective pressure (Dzomba *et al.*, 2021). Genetic signatures of selection refer to specific patterns of genetic variation that are indicative of natural or artificial selection acting on particular genomic regions. These patterns can include changes in the frequency or distribution of specific genetic variants (e.g., single nucleotide polymorphisms) or changes in the level of genetic diversity in a particular region (Schroeder *et al.*, 2019). One common approach to identify genetic signatures of selection is to analyse patterns of genetic variation across the genome. Selective breeding for specific traits can lead to changes in the frequency of certain genetic variants in the population. By studying these genetic signatures of selection, researchers can gain insights into the genetic basis of the desired traits. This knowledge can be used to further improve breeding programs, enhance the selection of desirable traits, and ensure the preservation and advancement of the breed's unique characteristics.

By comparing the genetic variation amongst the Swakara subpopulations and in relation to their ancestral breeds, research can help identify regions of the genome that show unusual patterns of genetic differentiation, indicating potential and differential selection amongst the associated breeds. The breed is plagued with diseases of subvital factor in the white and grey, and some lethal disorders in the brown Swakara. Furthermore, the harvesting of Swakara pelts within a day of the lambs' birth inadvertently enforces selection against postnatal growth and maturation resulting in a wider range of signatures which may elucidate on the prevailing selective sweeps and their importance in the overall appearance and health of the population. The improvement of pelt quality in Swakara relied on the crossbreeding of the original Karakul to the Namaqua Afrikaner and Blackhead Persian sheep. Where the Swakara is prioritised for its quality pelt production, the breeding goals of Namaqua Afrikaner and Blackhead Persian have always focused on meat production (Burger, 2015). Like the Swakara, the Namaqua Afrikaner and Blackhead Persian are known for being hardy with an ability to survive in unfavourable habitats, and lending to that survivability is the prominent feature of a fat-tail in

the breeds, an adaptive trait to counterbalance nutritional requirements in arid/unfavourable conditions.

Several studies have investigated the genetic signatures of selection in sheep (Kim *et al.*, 2016, Fariello *et al*., 2014, McRae *et al.,* 2014). A study by Schroder *et al.* (2019) analysed whole-genome sequence data of different sheep breeds and identified several genomic regions that showed evidence of selection. Identifying such genetic signatures in Swakara sheep may shed insight into the role of selection in shaping the genetic variation of Swakara sheep with respect to pelt quality and disease. Understanding these signatures can help breeders to develop more efficient breeding strategies to improve the quality and productivity of these animals. The aim of this study was to potentially identify signatures of selection in Swakara sheep and analyse how they differ from their ancestral breeds in order to understand the key selection pressure events and how they contributed to the genomic architecture of the different Swakara sub-populations whilst gaining an understanding of the genetic basis of the traits observed in this breed.

4.3 Materials and Methods

4.3.1 Data samples and quality control

Genotyped data using the OvineSNP50 Beadchip was obtained from a total of 244 animals and filtered as described in chapter 3, leaving a total of 106 animals. Sheep populations were grouped according to their admixture proportions in each PCA cluster as per Figure 3.7 in chapter 3. Individuals dominating the cluster remained and animals of other coat colours or ancestral breeds of fewer numbers were removed. After applying these filters, Cluster A consisted of the Brown Swakara (n=21), and Cluster B (n=24) was a mixture of Grey and Black Swakara as both groups were represented by equal numbers. Cluster C contained only White Swakara with a mixture of both vital and subvital animals. The Karakul was grouped in Cluster D (n=10) with a few Black Swakara which were retained with a justification of similar genetics seeing as the original Karakul was found in only black coat colour. The rest of the ancestral breeds of Namaqua Afrikaner and Blackhead Persian were in Cluster E (n=5) and Cluster F (n=14), respectively (Table 4.1).

Table 4.1 Breed grouping of Swakara sub-populations and ancestors

PCA Cluster	Sub-population	No.
A	Brown	21
B	Grey/Black	24
C	White	32
D	Karakul	10
E	Namaqua Afrikaner	5
F	Blackhead Persian	14

4.3.2 Detecting signatures of selection using *iHS*

The integrated haplotype score was used to detect selective sweeps within the clusters presented in Table 4.1. All chromosomes were recoded and .inp files were created for each chromosome. FastPHASE v1.4 was used to then create fastphase_hapguess_switch.outfiles. The *iHS* scores were computed for each locus using the rehh package (Gautier andVitalis, 2012). To infer the presence of selective sweeps in candidate genomic regions in within population calculations /analysis, a score of $iHS > 3.0$ was considered.

4.3.3 Detection of signatures of selection using XP-EHH

XP-EHH values were computed to determine regions under positive selection in one population referencing the other, utilising the haplotype data within the XP-EHH program (Sabeti *et al.,* 2007) and estimated using fastPHASE v1.4. This allowed for the comparison of integrated EHH profiles amongst two different populations at the same SNP location. Pairwise comparisons were conducted within the coat colour sub-populations (Clusters A-C) and between the coat colour groups and the ancestral breeds (Clusters D-F). Utilising the XP-EHH allowed for powerful detection of selection signatures regardless of small sample size (Eydivandi *et al.,* 2021). Significant XP-EHH scores were considered at a threshold of $(-\log_{10}[2\phi - |iHS|]) = 3$ ($P \leq 0.001$) and the associated SNPs were further analysed for associated candidate genes. The data was plotted using the qqman package in R.

4.3.4 Detection of signatures of selection using HapFLK

The HapFLK statistic is an extension of the FLK test where it accounts for the hierarchal structure of the populations through the utilisation of a multipoint linkage disequilibrium model (Fariello *et al.,* 2013). To determine the number of haplotype clusters the cross-validation procedure of fastPHASE was employed and an optimal number of 15 clusters was determined

which resulted in 10 EM runs to obtain a stable estimate of hapFLK for each chromosome (1-26). Moreover, the Imputeqc package in R studio was used to obtain the correct value of K. For this study, the Brown Vital Swakara was defined as the outgroup. P-values were computed for each SNP-specific value and a threshold of $p<0.05$ was applied to limit the false positives in the detection of significant sweeps. The manhattan plot was visually obtained using the qqman package in R.

4.3.5 Annotation of candidate genes

The genomic regions under possible selection using the three methods of signature detection were annotated for genes. The Ensemble gene database was used to annotate the genes within a 1Mb span of the SNP under selection. The OAR_Rambouillet_V 1.0 genome assembly was used to obtain the possible genes. Bedtools v2.31 (Quinlan and Hall, 2010) was utilised in determining the associated QTLs intersecting the sheep genome assembly/reference genome and the identified genomic regions.

4.4 Results

4.4.1 Selection signatures within populations utilising iHS approach.

Quality control yielded 38 116 SNPs for further analysis. A total of 73 selective sweep regions were identified across all six clusters/groups/sub-populations with the Karakul sub-population (Cluster D) containing the highest number of 22 selective sweeps with the least being identified in the Grey coat colour sub-population (Cluster B). Within the Swakara coat colour groups a total of 28 selective sweeps ($|iHS|>3.0$) were found across 10 chromosomes (OAR1, 2, 3, 5, 6, 7, 10, 12, 20, 22) ranging from iHS=3.1 to iHS=5.02 (Figure 4.1). A large number of the significant selective sweep regions ($|iHS|>3.0$) were identified in the ancestral groups across 16 different chromosomes (OAR, 1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 15, 17, 18, 21, 22, 23) with the most (10) being within the Blackhead Persian (Figure 4.2). Furthermore, within these regions were the *OSBPL9* (muscle weight in carcass) and *GRIK3* (lean meat yield percentage) genes on chromosome 1, *GOLGA3* (somatic cell score, bone density) on chromosome 17, and *ADGB* (internal fat amount) on chromosome 8. The highest iHS scores were found in the ancestral populations in the Namaqua Afrikaner on OAR2 ($|iHS|=4.39$; 213.18 Mbp), Blackhead Persian on OAR22 ($|iHS|=4.23$; 511.01 Mbp) and Karakul on OAR22 ($|iHS|=4.56$; 252.18 Mbp).

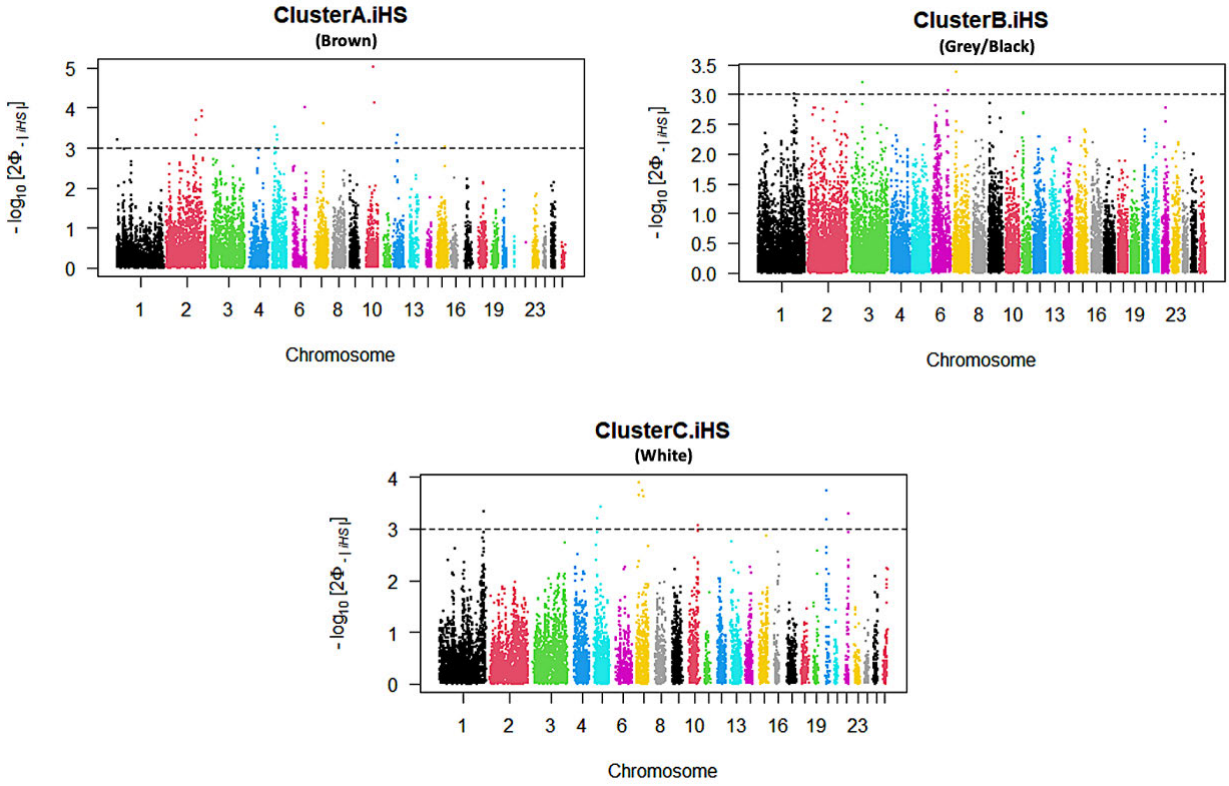


Figure 4.1 Genomic regions under selection with significant SNPs (|iHS|>3.0) in coat colour Swakara sub-populations

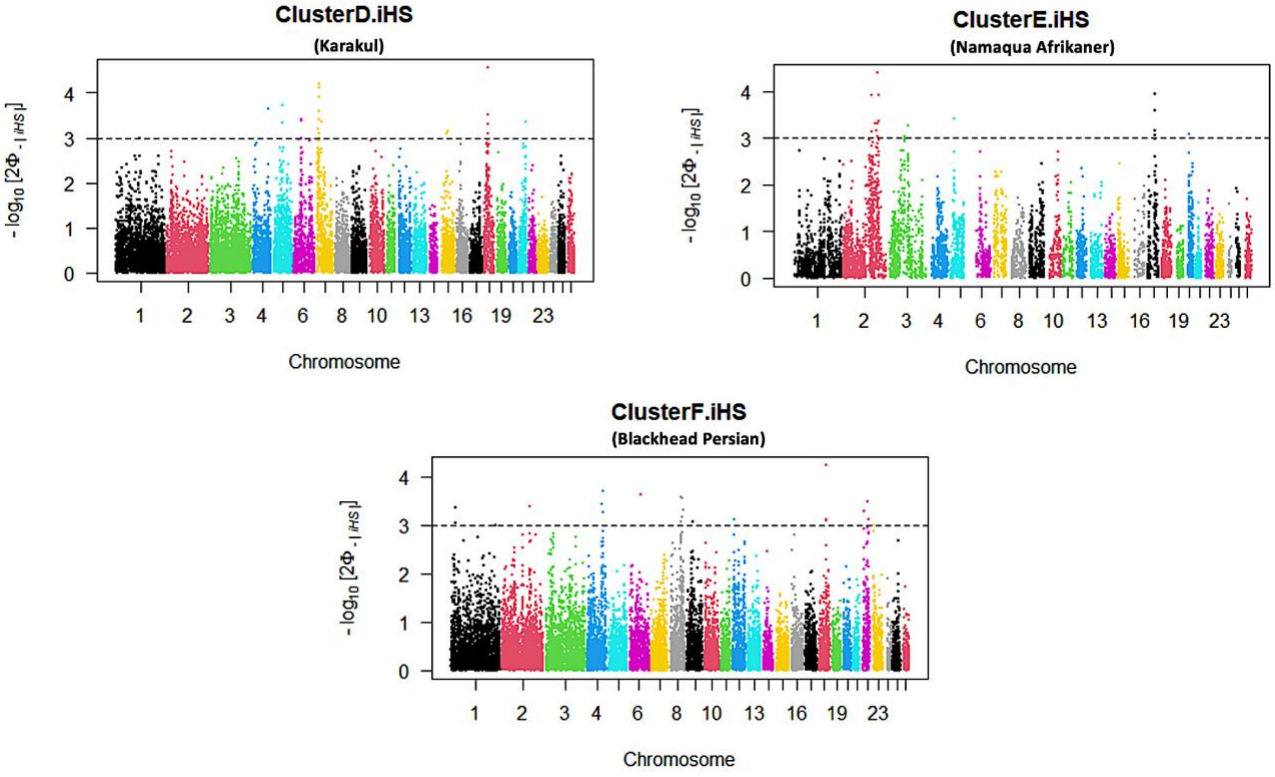


Figure 4.2 Genomic regions under selection with significant SNPs (|iHS|>3.0) in ancestral Swakara sub-population

The identified genomic regions in Table 4.2 demonstrate and provide clear evidence of selective forces in the chromosome wide scan of the different regions of the genome for the sub-populations (Figure 4.1 & Figure 4.2). 26 candidate genes overlapped with these regions including *LIPC*, *TMEM163*, *NAALADL2*, *FAM126B* and *PCMT1* which are associated with important traits. There were no genes found in common across the populations, however, common QTLs were detected within the Swakara sub-populations across different chromosomes. These QTLs were associated with body weight (*FBN2*, *CNTNAP5*, *ANKRD44*, *BOLL*, *FAM126B* and *RPGRIP1*) which were contained mainly within chromosomes 2 and 5 within the Brown and White Swakara and the Namaqua Afrikaner ancestral group. The milk fat percentage QTL was also found among genes such as *TMEM163*, *CNTNAP5*, *ANKRD44*, and *NDUFB3*.

Table 4.2 Signatures of selection identified within sub-populations and reported candidate genes

PCA Cluster	Population	Chr	Start length	End length	Gene name	QTLs
A	**Brown**	1	12017183	12258045	*GRIK3*	Muscle weight in carcass, lean meat yield percentage
		2	188501184	188780453	*TMEM163*	Meat arachidonic acid content, Milk fat percentage, Body weight, Hot carcass weight, Milk lactose yield
		5	44397076	44448204	*RPGRIP1*	Meat palmitoleic Acid content, Body weight, Foot angle
		12	31186419	31375067	No gene	
B	**Grey/Black**	1	229978373	231661320	*NAALADL2*	Meat acid content, Carcass fat percentage, Reproductive seasonality, Trichostrongylus colubriformis FEC, Teat number
C	**White**	1	277388474	277434989	*KY*	No QTL
		5	24160007	24388127	*FBN2*	Meat acid content, Body

						weight, Inherited ovine arthrogryposis, Milk fat yield
		7	29780864	29921369	No gene	
		7	52307605	52521797	*MY01E*	Staple length, Primary fiber diameter coefficient of variance, Haemonchus contortus FEC
		7	53011419	53184533	*LIPC*	Staple length, Haemonchus contortus FEC
D	**Karakul**					Fecal egg count, Fat weight in carcass, Muscle weight in carcass
		6	46933987	47058649	*GBA3*	Facial eczema susceptibility,
		15	42323672	42403144	*arntl*	Staple length
		15	44093851	44448218	*GALNT18*	Testes weight, Worm count, Salmonella abortusovis
		18	25259677	25271667	*MPHOSPH10*	susceptibility
		21	50465969	50516621	*NAP1L4*	Meat cis-vaccenic acid content

E	**Namaqua Afrikaner**	2	203610604	204839159	*CNTNAP5*	Meat docosapentaenoic acid content, Milk fat percentage, Hot carcass weight, Body weight, Fecal egg count, Somatic cell count
		2	213053605	213363940	*ANKRD44*	Meat acid content, Milk fat percentage, Body weight, Hot carcass weight, Change in haematocrit, Bone density, Fecal egg count
		2	213722490	213782576	*BOLL*	Worm count, Body weight
		2	217326043	217426829	*FAM126B*	Meat acid content, Body weight, Hot carcass weight, Bone density
		2	217389340	217482060	*NDUFB3*	Bond density, Change in haematocrit, Hot carcass weight, Milk fat percentage

		17	52311245	52348396	*GOLGA3*	Somatic cell score, Bone density
		17	52498439	52580788	No gene	
F	**Blackhead Persian**	1	27643177	27811776	*OSBPL9*	Muscle weight in carcass, Lean meat yield percentage, Carcass fat percentage
		6	72125952	72563683	*GABRB1*	Mean fiber diameter, Fat weight in carcass
		8	77809057	77995566		Internal fat amount, Trichostrongylus adult and larva count
		8	80706220	80746929	*PCMT1*	Fecal egg count, Milk lactose yield
		22	11732586	11826723	*LIPJ*	Somatic cell score

4.4.2 XP-EHH analysis to detect signatures of selection between sub-populations

The XP-EHH score was calculated to detect selection signatures between the coat colour sub-populations and between the coat colour sub-populations and their presumed ancestral breeds. Based on the analysis, 619 genomic regions harbouring putative selective sweeps were identified with majority being detected between the Brown (Cluster A) and Grey (Cluster B) sub-populations (181). The strongest signals were produced between the coat colour sub-populations with XP-EHH scores ranging from 5.27 to 3.93 (Figure 4.3). Further analysis detected signatures of selection related to 56 candidate genes, with some including *PTGER3*, *ADAMTS3*, *TROAP*, *NEIL2*, *FBXO8*, *BTBD10*, *KLF13*, *CCNT1*, *C1QL4*, *ASIC1*, *PARVA*, *LIMA1*, and *MTMR10*, that are essential in determining fat weight in carcass, milk protein percentage, hot carcass weight, testes weight as well as milk yield, indicating a diverse genomic selection from the human-led breeding/selection events/processes. Sub-populations compared with the Brown Swakara yielded the greatest diversity whereas the ancestral populations showed the least which was a pointedly common thread in the analysis (Figure 4.4).

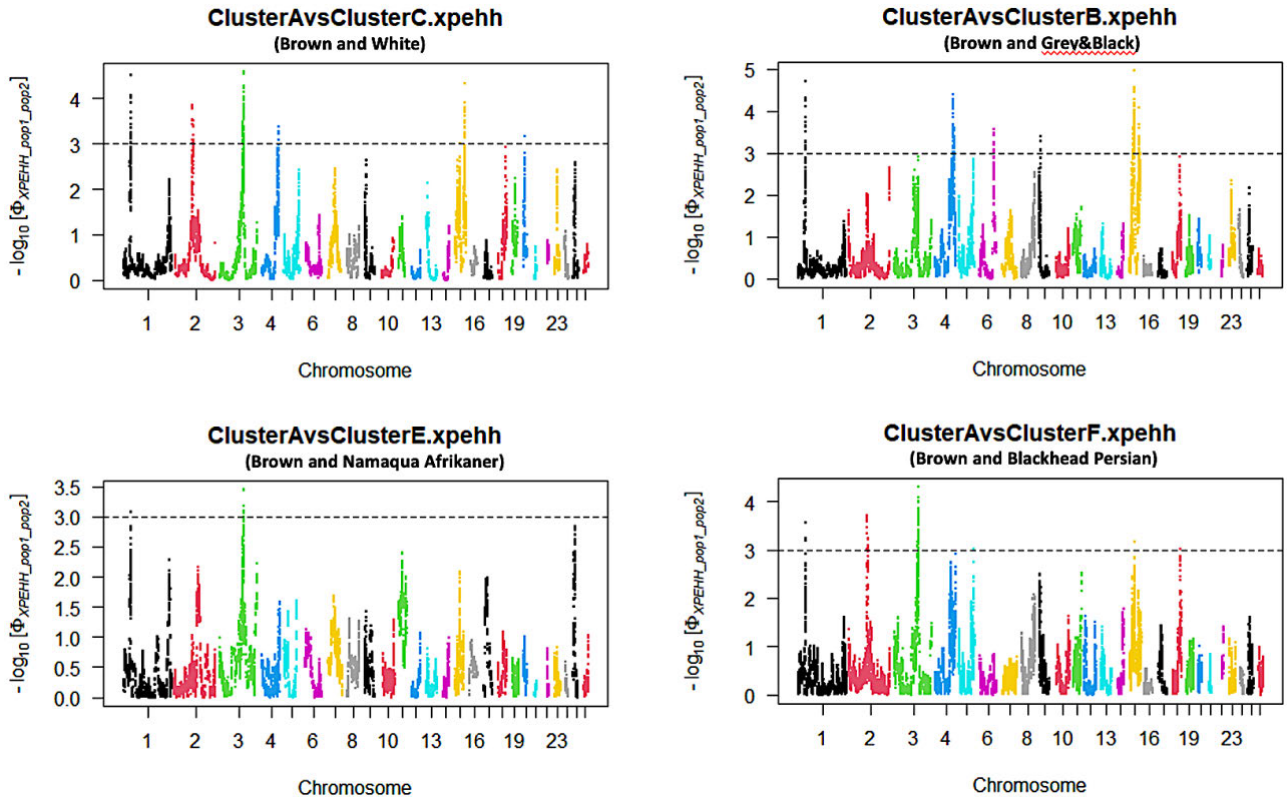


Figure 4.3 Genome-wide distribution of selection signatures detected by XP-EHH across 26 chromosomes between Brown Swakara and other coat colour sub-populations.

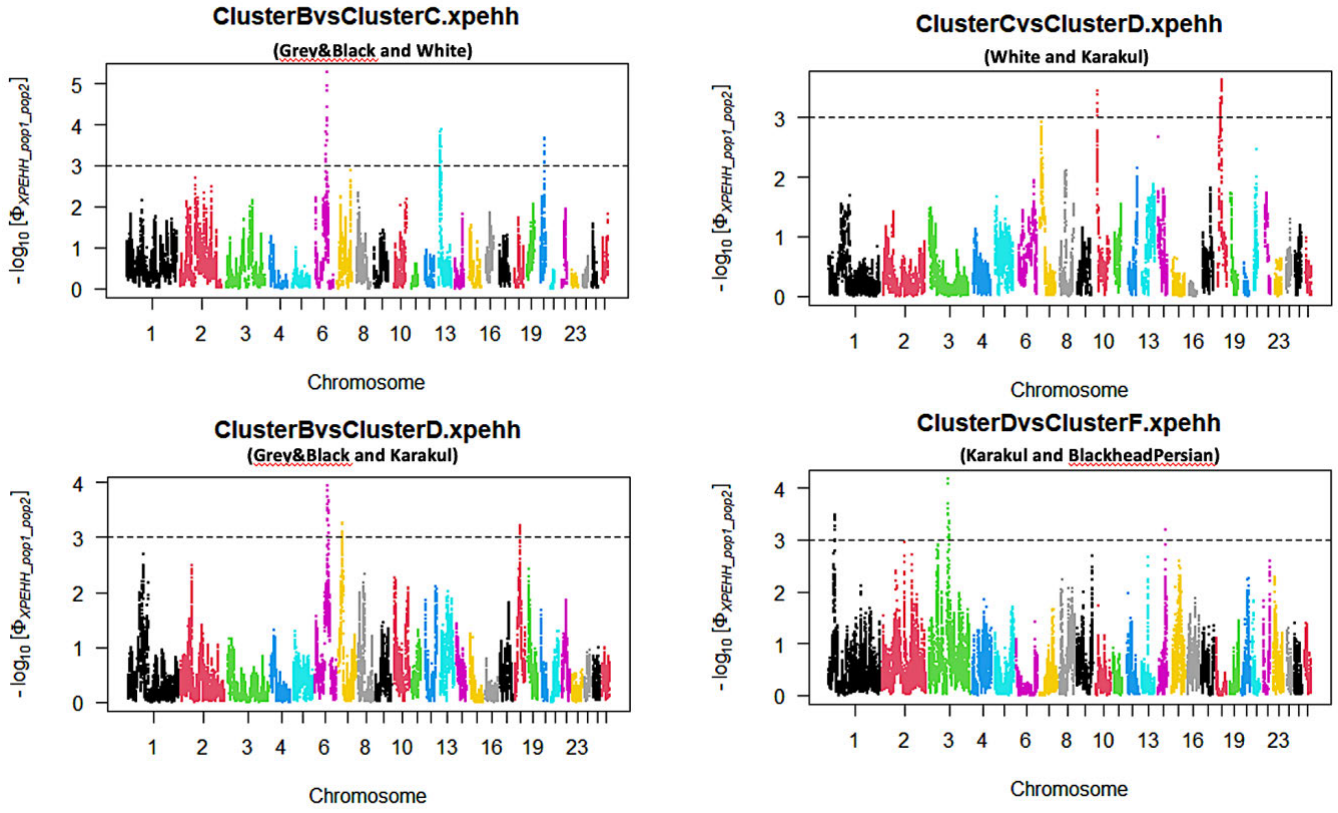


Figure 4.4 Genome-wide distribution of selection signatures detected by XP-EHH across 26 chromosomes between Swakara coat colour sub-populations and the ancestral breeds.

The observed selective sweeps with the highest signals in each cluster pair/comparison were maintained/conserved within only 5 chromosomes (OAR1, 3, 4, 6, 15, 18) with OAR3 harbouring the most (Table 4.3).

Table 4.3 Selective sweeps identified between Swakara sub-populations and ancestral breeds using XP-EHH

PCA Clusters	Population	Chr	Start length	End length	Gene name	QTLs
A vs C	**Brown_White**	1	49178631	49291401	*PTGER3*	Lean meat yield percentage, Muscle weight in carcass, Carcass bone percentage
		2	113785956	113834464	*GATA4*	
		2	113840190	113849697	*NEIL2*	Meat acid content, Meat colour, Ultimate pH, Shear force, Nematodirus FEC
		2	114307291	114762660	*GLRA3*	Meat colour, Ultimate pH, Shear force, Foreleg length, Horn type
		2	114950716	115082707	*FBXO8*	Hot carcass weight, Body weight, Meat colour, Milk fat percentage

	2	116185691	118014517	*GALNTL6*	Dressing percentage, Soft tissue depth, Leg yield, Backfat at third lumber, Body height
	3	142918112	142930441	*TNS2*	Staple length, Internal fat amount, Change in eosinophil number, Immunoglobin A level
	3	144589358	144636801	*TFCP2*	Meat conjugated linoleic acid, Body weight, Milk protein percentage
	3	146009413	146065629	*LIMA1*	Staple length, Internal fat amount, Change in eosinophil content, Milk protein percentage
	3	146148548	146170577	*ASIC1*	Internal fat amount, Meat conjugated linoleic acid content, Body weight, Milk

					protein percentage
	3	146540901	146560907	*PRPF40B*	Change in eosinophil number, Internal fat amount
	3	146610598	146629190	*KCNH3*	Meat conjugated linoleic acid content, Staple length, Internal fat amount, Body weight
	3	146820939	146823679	*C1QL4*	Internal fat amount, Meat conjugated linoleic acid, Change in eosinophil number, Body weight, Milk protein percentage
	3	146825079	146832287	*TROAP*	Body weight, Milk protein percentage

		3	147122772	147136176	*PRKAG1*	Staple length, Internal fat amount
		3	147268954	147281426	*CACNB3*	Change in eosinophil number, Meat conjugated linoleic acid
		3	147313219	147326258	*ADCY6*	Internal fat amount
		3	147371353	147407285	*CCNT1*	Body weight, Milk protein percentage
		4	110523021	110646648	*CREB3L2*	Primary fiber diameter coefficient
		4	110700794	110749383	*AKR1D1*	Primary fiber diameter coefficient
		15	43217761	43382308	*PARVA*	Facial eczema susceptibility, Staple length
A vs B	**Brown_Grey/Black**					
		1	49178631	49291401	*PTGER3*	Lean meat yield percentage, Carcass bone percentage, Muscle weight in carcass, Carcass bone percentage
		4	109812178	109920284	*PTN*	Primary fiber diameter coefficient

		4	109983753	110487852	*DGKI*	Pneumonia susceptibility, Primary fiber diameter coefficient
		6	96615346	96790243	*GC*	Useful yield content, Trichostrongylus colubriformis FEC
		6	97301328	97592253	*ADAMTS3*	Carcass fat percentage, Lean meat yield percentage
		6	98054239	98221458	*ANKRD17*	Mean fiber diameter, Fat weight in carcass,
		15	42245184	42321838	*BTBD10*	Facial eczema susceptibility, Staple length
		15	42803031	42976356	*TEAD1*	Milk yield, Milk fate yield, Milk protein yield
		15	43217761	43382308	*PARVA*	Facial eczema susceptibility, Staple length
		15	46895698	47017426	*DENND2B*	Milk protein yield
A vs E	**Brown_NamaquaAfrikaner**					Staple length, Internal fat amount
		3	146825079	146832287	*TROAP*	Milk protein percentage, Body
		3	146820939	146823679	*C1QL4*	

						weight, Change in eosinophil
A vs F	**Brown_BlackheadPersain**					
		1	49178631	49291401	*PTGER3*	Muscle weight in carcass
		2	113840190	113849697	*NEIL2*	Meat arachidonic acid, Milk fat percentage, Bod weight, Foreleg length
		2	114950716	115082707	*FBXO8*	Hot carcass weight, Trichostrongylus adult and larva count, Nematodirus FEC QTL, Horn type
		3	146820939	146823679	*C1QL4*	Staple length, Body weight
		3	150382842	150434300	*SLC38A1*	Teat number
		3	144589358	144636801	*TFCP2*	Immunoglobin A level
C vs D	**White_Karakul**	18	25300152	25341190	*MTMR10*	
		18	25349882	25450838	*TRPM1*	
		18	25552604	25589961	*KLF13*	Testes weight,
		18	26028635	26277057	*ENTREP2*	Staple length,
		18	26339648	26450903	*TJP1*	Milk yield, Meat
		18	26694775	26739580	*TARS3*	acid content,
		18	26892544	27087570	*PCSK6*	Salmonella abortusovis
		18	29389434	29456913	*TMEM266*	susceptibility,
		18	29598379	29975830	*SCAPER*	Worm count
		18	30213240	30373390	*PEAK1*	

		18	30540006	30615050	*HMG20A*	
		18	31031268	31065785	*CSPG4*	
		18	31253358	31296924	*SIN3A*	
		18	32139937	32239937	*SEMA7A*	
		18	32645619	32716452	*MCP-3*	
D vs E	**Karakul_NamaquaAfrikaner**					Hematocrit Reproductive seasonability, Nematodirus FEC, Subcutaneous fat thickness
		3	185866216	186128549	*CHST11*	
D vs F	**Karakul_BlackheadPersian**					Internal fat amount, Meat conjugated linoleic acid content, Trichostrongylus colubriformis FEC
		3	118144297	118379323	*KCNC2*	

4.4.3 Identifying signatures of selection between sub-populations utilising HapFLK

The HapFLK analysis revealed significant selection signals across many chromosomes with a total of 1931 significant regions identified spanning 21 chromosomes associated with 291 genes. Strong selection signals (P<0.05) were revealed on chromosome 3 (187.85Mbps; 187.87 Mbps; 188 Mbps) and 2 (159.57Mbps; 159.53Mbps; 159.47Mbp) (Figure 4.5). More than 40 of the strongest signals were identified within these two chromosomes with others being identified within chromosome 13, 5 and 10.

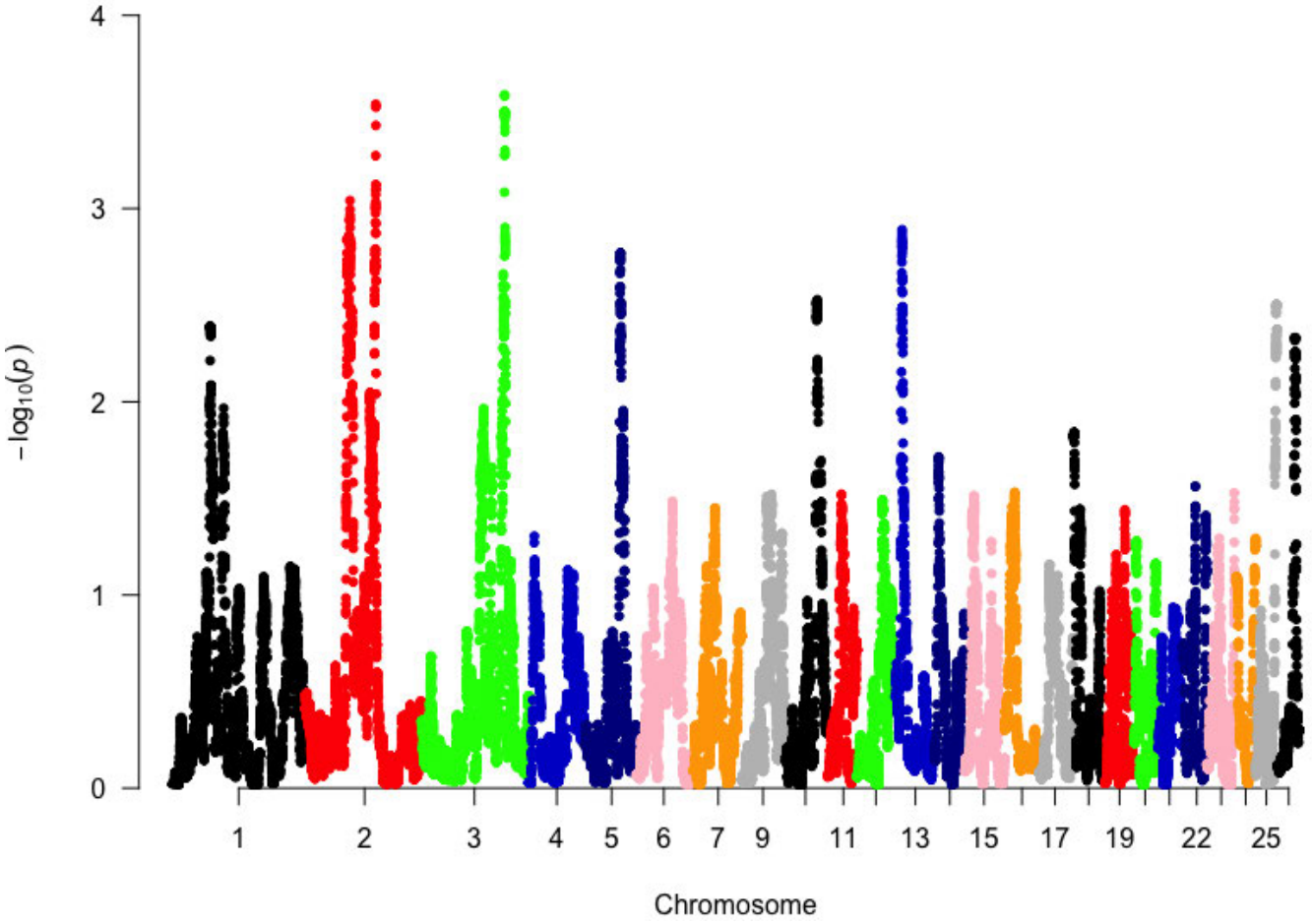


Figure 4.5 HapFLK Manhattan plot of detected signatures of selection

Genomic regions detected within chromosome 5 showed the highest number of annotated genes comprising of/aggregating to 62 genes with a few being found with associated QTLs. Majority of the genes were involved in traits such as determining body weight (*FAM172A*,

KIAA0825, *SLF1*, *MCTP1*), greasy fleece weight (*TENM2*), age at maximum daily gain (*LYSMD3*) and meat palmoteic acid content (*HMMR*) (Table 4.4). In addition, the analysis revealed genes associated with gastrointestinal parasites and overall health traits such as fecal egg count (*GPC6*, *KAZN*, *PLCG2*, *CWF19L2*, and *SFRP1*), pneumonia susceptibility (*ATF2*), facial eczema susceptibility (*KIRREL1*), haemonchus contortus resistance (*MCTP1*, *HS3ST3A1*) and trichostrongylus adult & larva count (*CNTLN*, *ZNF2864*). Multiple different genes were revealed but there were no annotated genes within the regions on chromosomes 4 and 6. Overall, many of the QTLs were found to overlap across the 19 chromosomes with relatively high selection signals

Table 4.4 Genomic regions under positive selection detected with HapFLK and their associated genes (*P-value* <0.05)

Chr	Start Length	End length	Gene name	QTLs
1	83687756	83907957	*OLFM3*	Lean meat yield percentage, Carcass fat percentage, Muscle depth at third lumbar, Backfat at third lumbar
1	84674722	84940130	*COL11A1*	Bone weight in carcass, Meat eicosapentaenoic acid, Meat polyunsaturated fatty acid content, Muscle depth at third lumbar, Milk yield
1	89000847	89366356	*NTNG1*	Muscle depth at third lumbar, Backfat at third lumbar, Meat polyunsaturated fatty acid content, Meat docosapentaenoic acid content, Bone weight in carcass
1	111080803	111243196	*KCNN3*	Bone density, Milk yield, Milk protein percentage, Body weight
1	111664770	111838346	*ASH1L*	Bone density
1	112060553	112076554	*SYT11*	Bone density
1	112134181	112135302	*RXFP4*	Bone density, Milk yield, Milk protein percentage, Body weight
1	114633082	114731957	*KIRREL1*	Bone density, Facial eczema susceptibility, Muscle weight in carcass

2	92371441	92798118	*CNTLN*	Meat arachidonic acid content
2	92809713	93031831	*SH3GL2*	Meat acid content, Milk fat percentage, Body weight, Hot carcass weight, Trichostrongylus adult and larva count, Meat colour a*, Ultimate pH, Meat colour L*, Meat colour b*, Somatic cell score
2	94142763	94733446	*ADAMTSL1*	Tail length
2	103064152	103159596	*IFT74*	Shear force
2	145538735	145616352	*ATF2*	Milk lactose yield, Milk yield
2	158691581	159310642	*KCNH7*	Pneumonia susceptibility, Mean corpuscular haemoglobin content, Milk yield,
3	156690838	156873558	*CNTN1*	
3	157522578	157736032	*LRRK2*	Staple length, Internal fat amount, Body weight, Strongyle FEC
3	157889821	158114054	*SLC2A13*	
3	158688253	158774272	*ABCD2*	
3	181747786	181786144	*SLC17A8*	
3	182075580	182154669	*NR1H4*	
3	182259739	182702403	*ANO4*	Hematocrit, Reproductive seasonality, Nematodirus FEC
3	182729278	182796898	*SLC5A8*	
3	182869000	182956422	*UTP20*	
3	133701643	133711212	*C3H12orf29*	Staple length, Internal fat amount, Meat conjugated linoleic acid content, Change in eosinophil number, Milk protein percentage
3	189818665	190430983	*LARGE1*	Subcutaneous fat thickness, Ear size
5	98927096	99331285	*FAM172A*	Body weight
5	99353536	99753270	*KIAA0825*	Body weight

5	99802184	99862080	*SLF1*	Body weight
5	99873658	100478089	*MCTP1*	Body weight, Haemonchus contortus resistance
5	78395242	78426595	*HMMR*	Meat palmitoleic acid content
5	82856718	83914578	*TENM2*	Greasy fleece weight
5	95068765	95074068	*LYSMD3*	Age at maximum daily gain
7	46775986	46884011	*CSNK1G1*	Staple length, Primary fiber diameter
7	48257899	48664379	*TLN2*	Longissimus muscle area, Milk protein percentage, Haemonchus contortus FEC
9	90605430	90660384	*DPY19L4*	Milk fat yield
9	90935433	91031336	*FSBP*	Milk fat yield
9	91634630	91677980	*TMEM67*	Milk fat yield
9	66765430	66769944	*AARD*	Muscle weight in carcass, Longissimus muscle area
9	68044013	68317285	*TRPS1*	Hot carcass weight, Muscle weight in carcass
9	57593102	57796089	*ZFHX4*	Longissimus muscle area
10	76114767	77457734	*GPC6*	Fecal egg count, Fat weight in carcass, Carcass fat percentage, Cheese yield, Teat number, Lean meat yield
11	30211661	30223888	*ZNF286A*	Internal fat amount, Trichostrongylus adult and larva count
11	31336940	31419236	*HS3ST3A1*	Body weight, Hot carcass weight
11	31845727	32247326	*ARHGAP44*	Milk yield, Milk yield persistency, Milk protein yield, Haemonchus contortus resistance
11	32264750	32347315	*MYOCD*	Milk polyunsaturated fatty acid content, Jaw length

11	32719069	32797017	*MAP2K4*	Milk lauric acid content, Milk capric acid content
12	56136003	57608267	*KAZN*	Fecal egg count, Carcass fat percentage
12	57674855	57787899	*PRDM2*	Carcass fat percentage
13	15705722	16254406	*CELF2*	Muscle weight in carcass, Salmonella abortusovis susceptibility
13	16934507	16980446	*CDC123*	Spleen weight
13	17059029	17447111	*CAMK1D*	Mean fiber diameter, Spleen weight
13	18717829	19363953	*PARD3*	Milk cis-10 heptadecenoic acid content, Soft tissue depth at GR site
14	8511740	9004226	*PLCG2*	Nematodirus FEC, Fat weight in carcass, Footrot susceptibility
14	9094606	9193777	*HSD17B2*	Total bone, Fecal egg count, Milk yield
14	9653072	10825799	*CDH13*	Dressing percentage
14	11012544	11053354	*SLC38A8*	Bone weight in carcass
14	11122689	11147342	*DNAAF1*	Fat weight in carcass
15	17583281	17775780	*CWF19L2*	Fecal egg count
15	20183591	20494608	*DDX10*	Horn type, Teat number
16	14486757	14766168	*ADAMTS6*	Lean meat yield percentage
16	14902797	15157209	*CWC27*	Body weight
16	15299540	15403238	*RGS7BP*	Subcutaneous fat area
16	15535182	16330361	*RNF180*	Dressing percentage
16	20316636	21581348	*PDE4D*	Staple length, Subcutaneous fat thickness
18	1599006	1662499	*UBE3A*	Hematocrit
18	2416150	2596685	*ATP10A*	Testes weight
18	16098671	16513865	*NTRK3*	Reproductive seasonality, Meat gadoleic acid content
18	16842099	16872862	*AEN*	Staple length, Milk yield
19	44921272	45007575	*PXK*	Dressing percentage
23	61898167	61981555	*TXNL1*	Milk yield

23	62029812	62425750	*WDR7*	Milk fat yield
25	45774525	45893920	*PARG*	Milk fat percentage, Testes weight
25	47579656	47622777	*COG2*	Testes weight
25	43407132	43421637	*MMRN2*	Staple length
25	43463346	43501564	*GLUD1*	Mean fiber diameter
25	44327860	44459369	*PTPN20*	Primary fiber diameter
26	41892170	42375690	*PSD3*	Stature
26	38911557	39198531	*ZMAT4*	Udder attachment
26	39585424	39625227	*SFRP1*	Fecal egg count

4.4.4 Common genes under selection between iHS, XP-EHH, and HapFLK

A total of 9 candidate genes were found to be commonly under selection using the three different statistical methods (Figure 4.6). Chromosome 13 harboured the most with 5 (*CELF2*, *USP6NL*, *ECHDC3*, *UPF2* and *CAMK1D*), and this was between the XP-EHH and HapFLK methods constituting a total of 1.5% of genes identified.). 3 candidate genes accounted for 0.8% of identified selection genes between iHS and XP-EHH (*GALNT18*, *HMG20A*, and *CSPG4*). These were harboured on chromosomes 15 and 18.

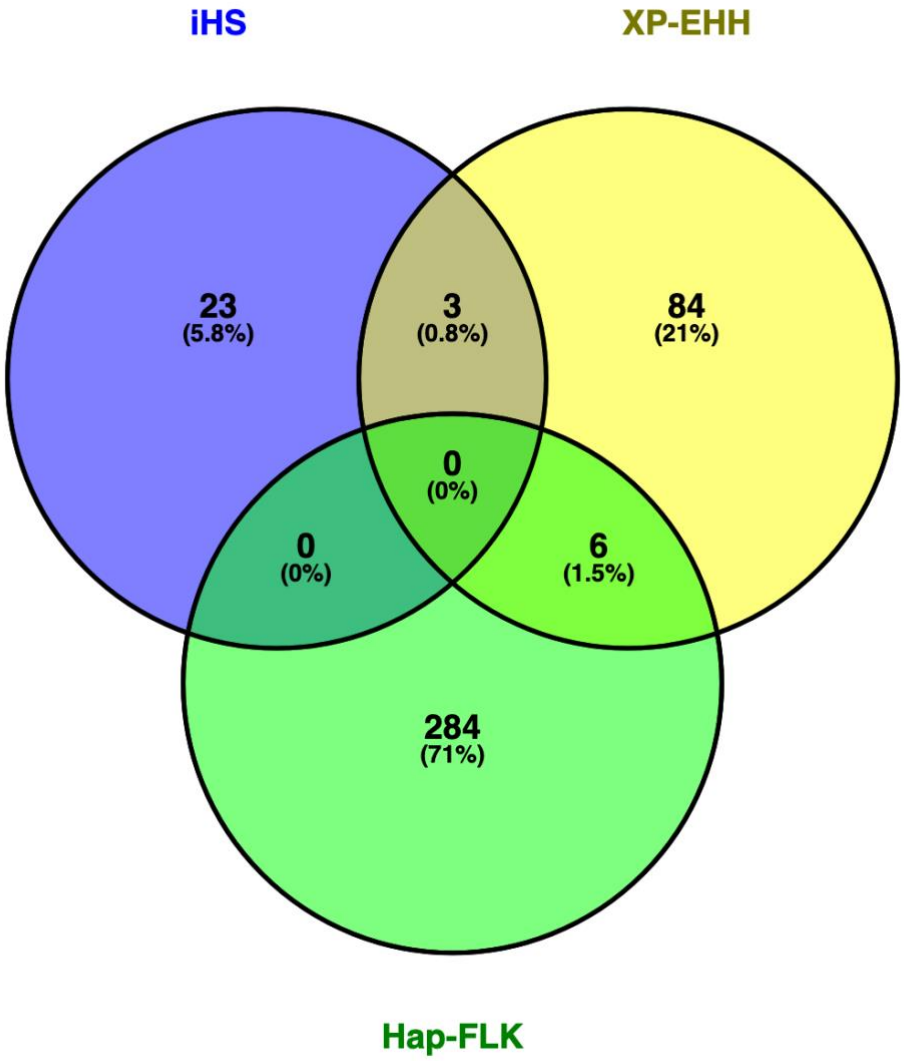


Figure 4.6 Common regions under selection and detected candidate genes using iHS, XP-EHH, and HapFLK

4.5 Discussion

Natural and artificial selection events can severely affect the genome and introduce patterns that alter allele frequencies in populations. These patterns, which are called signatures of selection, are essential in understanding regions that have undergone selective sweeps and the mechanisms of selection as they can give insights into the genes affecting phenotypic differences due to different selection pressures (McRae *et al*., 2014). To discover the selective sweeps, this study utilised three different statistical methods: (i) iHS, which was built to help hinder the influence that heterogenous recombination rates may have (Zhao *et al.*, 2015). This technique is an extension of the (ii) XP-EHH approach which was first developed by (Sabeti *et al.*, 2002) and compares a haplotype with extended homozygosity and high frequency with other haplotypes at a specific locus. The last method used was the (iii) HapFLK which takes into account a hierarchal structure to gather haplotype data and detect signatures (Onzima *et al.*, 2018). The variety of methods utilised enhances detection accuracy of signatures of selection whilst also removing unknown bias (Ma *et al.*, 2015). A major indicator of selection activity in a certain genomic area may be the discovery of signatures of selection through more than one tool. Regions detected as being sensitive to selection through one approach but not by another methodology does not, however, rule out the possibility of selection actually having taken place (Onzima *et al.*, 2002). Signature of selection analysis was done one Admixture based clusters of the sheep in this study that showed clear divergence of the Swakara sub-populations from their ancestral breeds as was shown in Figure 3.7 in Chapter 3. The clustering allowed the separation/grouping of individuals according to coat-colour by eliminating least dominant individuals in an admixed cluster. As was reported in Chapter 3, Figure 3.5, clustering of Swakara sub-populations together regardless of coat colour and geographical origin was witnessed, whilst the Karakul, Namaqua Afrikaner and Blackhead Persian formed separate clusters and provided clear evidence of complete divergence of the Swakara from their ancestors.

The iHS approach produced the least signatures with only 73 sweeps and 56 genes. Key genes such as *GRIK3* responsible for muscle weight in carcass, and *TMEM163* and *RPGRIP1* related to body weight were identified in the Brown Swakara. Body weight is widely known to be an important indication of growth and health (Jiang *et al.*, 2021), which is essential in meat production and although the Swakara are primarily bred for their pelts, meat and milk production are also prominent in the breed. Selection for meat traits was also apparent with the identification of meat acid content, and carcass fat percentage related signatures in the

Grey/Black Swakara cluster. The *FBN2* gene in the White Swakara was on the other hand observed to be related to milk fat yield. Signatures related to primary fiber diameter were detected in the White Swakara and provided evidence for the intense selection pressure in the production of pelts and improvement of pelt quality trait in the white pelt coloured subpopulation. Although there were no shared selection sweeps between the Swakara sub-populations and between themselves and their ancestral groups using iHS, common functions of different genes were observed with the most prominent being meat acid content observed within the Brown, Grey/Black and White Swakara groups, and carcass weight between the ancestral breeds and Swakara sub-populations. Genes reported in this study were also found to be candidate genes under selection in studies in other sheep populations (Alberto *et al.*, 2018, Fariello *et al.*, 2013), cattle (Zhao *et al.*, 2015), and buffaloes (Deng *et al.*, 2022). These included (*CNTNAP5*) which was discovered to be involved in milk production traits in cattle breeds (Stella *et al.*, 2010). The *NDUFB3* and *GBA3* genes were reported in bovine studies to be associated with QTLs of heat tolerance and weight (Flori *et al.*, 2019, Al-Mamun *et al.*, 2015). Interestingly, (Deng *et al.*, 2022) reported the *NAALADL2* gene to be involved in production traits in Chinese buffaloes.

Examining the variance in the frequencies of SNP alleles between the sub-populations and their ancestral populations has great potential in exposing the underlying genes responsible for phenotypic variation and determining the possible targets of previous selection practises. Detection of signatures of selection between the sub-populations utilising XP-EHH identified genomic regions with 56 candidate genes associated to various traits. Pairwise comparison of the Brown and White coat colour clusters produced the most signatures with genes related to meat colour and foreleg length (*GLRA3*), body height (*GALNTL6*) and internal fat amount (*TNS2*, *LIMA1*, *ASIC1*, *PRPF40B*, *KCNH3*, *CACNB3*). Five signatures were commonly shared between these sub-populations. The Brown sub-populations revealed QTLs of primary fiber diameter (*PTN, ADAMTS3*), facial eczema susceptibility (*TEAD1, ANKRD17*), and milk protein yield (*DENNS2B*) when compared to the Grey/Black sheep. Swakara sheep farming has had to adapt due to the ever-changing market value of the pelts that they are mainly bred for and follow a less traditional route by diversifying the products produced from this breed thereby becoming a multi-purpose breed with importance in milk, meat, wool, and pelt production (Nsoso and Madimabe, 1999). More selective sweeps were identified between the Swakara sub-populations than between the Swakara and the ancestral breeds as expected. The Karakul and White Swakara revealed selection sweeps contained only in chromosome 8 related

to QTLs of testes weight, milk yield, meat acid content, and worm count (*MTMR10, TRPM1, KLF13, TJP1, HMG20A*). Seven genes were observed following the Brown Swakara sub-population regardless of the sheep cluster it is compared to of which six were found to be in common between the Namaqua Afrikaner and Blackhead Persian ancestral breeds and the Brown Swakara sub-population. On chromosome 1 the *PTGER* was found in common between the Brown vs Grey/Black, and White, and Blackhead Persian. A common signature was again seen between Brown vs White, and Grey/Black on chromosome 15 with the *PARVA* gene. A signature unique to the Brown and the ancestral groups was found in chromosome 3 (*C1QL4*). A variety of phenotypes related to meat traits (lean meat yield, meat colour, carcass bone percentage, body weight, hot carcass weight), milk traits (milk protein percentage, milk fat percentage, milk fat yield, milk lactose yield and milk yield), wool/hair traits (staple length, primary fiber diameter, mean fiber diameter). The identification of meat related QTLs reveals intense selection for meat production. By far, the most common QTL shared between the Swakara sub-populations was related to body weight and included *FBXO8* on chromosome 2, *TFCP2, ASIC1, PRPF40B, KCNH3, C1QL4, ADCY6*, and *TROAP* on chromosome 3, and *PTGER* on chromosome 1. Strong sweep signals linked to this QTL for multiple genes further cements the acute selection for meat production where bigger sized bodies indicate higher profits. Similar results have been found in pig studies where QTLs for body size were revealed when analysing loci under strong selection in the course of pig domestication where fatness is viewed as an economically significant trait (Rubin *et al.*, 2012, Hlongwane, 2022). Furthermore, genetic correlation between lamb weight and pelt quality traits has been previously reported in the pelt producing Gotland sheep (Näsholm, 2008) which could explain the strong selection of this trait in the pelt producing Swakara. Genes associated to milk traits included *TFCP2, LIMA1, ASIC2, KCNH3, C1QL4, ADCY6, PARVA, TROAP, PTGER3*, and *BTBD10* which were detected within the Brown vs White and Brown vs Grey/Black Swakara cluster comparisons. Milk production being related to several genes suggests that it is a complex trait governed by several genes (Stella *et al.*, 2010). This hypothesis is further supported by identification of multiple candidate genes in the adaptation of goat and sheep breeds to hot and harsh environments in other studies (Onzima *et al.*, 2018, Kim *et al.*, 2016), and adaptation is known to be a very complex trait that has been under selection for several generations in livestock.

For the detection of significant SNPs investigated using the HapFLK method the threshold was set to $P<0.05$. Whilst several studies have used contrasting threshold values (Rostamzadeh-

Mahdabi *et al.*, 2021, MoioliPilla and Ciani, 2015), others have reported/utilised similar stringent threshold values (Maiorano *et al.*, 2022). HapFLK analysis revealed the highest number of putative selection sweeps out of all methods tested with well over 1900 regions under selection identified. Considering hierarchical structure is acknowledged in HapFLK the detection capacity of soft sweeps is greatly improved. This allows for various regions under selection and their significant genes to be detected that otherwise would not have been detected with other analyses such as iHS, and XP-EHH (Fariello *et al.*, 2013, Walugembe *et al.*, 2019). Several genes were found to be related to traits of importance. Chromosome 1 harboured the genes *ASH1L, SYT1*, and *RXFP4* which were related to bone density. Other QTLs shown were tail length (*ADAMTSL1*) on chromosome 2, greasy fleece weight (*TENM2*) on chromosome 5, teat number (*GPC6, DDX10*) on chromosome 15, horn type (*DDX10*), and udder attachment (*ZMAY4*) on chromosome 26. The *SP3* gene which is believed to play an essential role in fat tail development was detected in chromosome 2 (Yuan *et al.*, 2017). The characteristic of a thin tail is believed to be inherent to the wild ancestors of sheep leading to the belief that possession of the initial thin-tails was a unique characteristic to the first domesticated sheep and fat-tailed breeds were later developed through selection (Manzari *et al*., 2019). The identification of a signature related to tail development in this study focusing on significantly fat-tailed sub-populations of Swakara and their ancestors highlights the resultant development of advantageous traits for survivability in unfavourable environments. It is anticipated that to identify genes linked to economically relevant traits, it is assumed that the locations of signatures of selection must be connected to traits that been subjected to selection, natural or artificial, through evolutionary processes or human-led practises.

Common genes amongst the methods utilised in this study identified a total of 9 genes with a higher percentage of shared genes between HapFLK and XP-EHH. Chromosome 13 harboured majority of the genes (*CELF2, USP6NL, ECHDC3, UPF2, CAMK1D*) linked to traits of muscle weight in carcass, salmonella abortusovis susceptibility, and mean fiber diameter. *CHST11* located on chromosome 3 was related to staple length, hematocrit, reproductive seasonality, nematodirus FEC, and subcutaneous fat thickness. A key trait that has been the subject of selection and was impacted during domestication is the coat colour trait, and it has also been at the centre of selection and crossbreeding practises in the Swakara breed. No sweeps or loci overlapped with known QTLs related to this trait. Rubin *et al.* (2012) hypothesised that this could be as a result of a lack of genetic data revealing QTL locations that could potentially provide an explanation for the non-identification of the significant traits.

Numerous candidate genes with various biological functions were detected in this study from genomic regions that have possibly undergone selection. Over 2500 regions containing putative selective sweeps and 373 genes associated to those regions were detected across all the three statistical tests used. The number of selection signatures is higher than those reported in other studies of sheep breeds (Fariello *et al.*, 2014, de Simoni Gouveia *et al.*, 2017, Manzari *et al.*, 2019), and other pig genome studies (Rubin *et al.*, 2012). Moreover, majority of these studies used single methods of signature selection identification. This variance can be attributed to the use of different analytical approaches and each technique having varied parameters applied thus recording diverse patterns of selection in the genome.

Overall, across the 3 three methods was the consistent observation of muscle weight in carcass, mean fiber diameter, staple length, meat acid content and worm count to be under selection. Whilst the observation of genes associated with growth, meat production, and carcass weight were not expected to be observed in a breed predominantly bred for pelt production, their persistence across and amongst the methods might be a reflection of the traits that were selected against in Swakara sheep. Due to early harvesting of pelt, most Swakara sheep do not growth beyond 3 days which inadvertently results in selection against the traits that are important beyond postnatal such as growth, carcass quality etc. Traits related to wool/hair such as mean fiber diameter, staple length etc were most frequently identified in the coat colour sub-populations of Swakara, and within those the White and Black showed the most.

4.6 Conclusion

A wide range of genomic regions under selection were noted in this study. The associated candidate genes and functional pathways gave insight into the key traits that have shaped the phenotypes of these breeds. A large number of signatures detected were linked to meat production, milk production, reproduction, health, tail traits and wool traits. This is indicative of the main traits selected for and against in this breed. A high percentage of selection regions were detected in the Swakara sub-populations compared to their ancestral breeds under both iHS, and XP-EHH methods. Overall, the analyses highlighted the emphasis on meat, milk, and wool/pelt production during selection events as well as on overall health and adaptation to challenging conditions in the Swakara breed.

5 IDENTIFYING RUNS OF HOMOZYGOSITY IN SWAKARA AND FOUNDING BREEDS AS KEY TO UNRAVELLING PREVALENCE OF GENETIC DISORDERS

5.1 Abstract

Since its introduction into Southern Africa from Uzbekistan the Swakara sheep remain understudied, with a poor understanding on the population dynamics shaping its genomic architecture and production performance. The subvital factor plaguing the grey and pure white sub-population of Swakara sheep, is poorly understood. This presumed genetic disorder causes individuals to die within 48 hours, due to underdeveloped digestive organs thus hindering the animal from being able to feed Additionally, the brown Swakara is also known to suffer from some form of genetic disorder. Runs of homozygosity are defined as long stretches of homozygous haplotypes that occur because of parents with shared ancestors, transmitting identical DNA segments to offspring. Screening for these ROH allows for the detection of ROH islands which are hotspots indicative of selective sweeps across the genome. This study sought to utilise the OvineSNP50 beadchip data to investigate runs of homozygosity and infer the levels of inbreeding in the Swakara sub-populations and within its presumed ancestral breeds of Karakul, Blackhead Persian, and Namaqua Afrikaner to potentially unearth the causative agents in the occurrence of genetic disorders. 106 genotyped animals were screened for ROH using the detectRUNS package in R studio. 4 630 ROH were detected in the population with lengths ranging from 3.89 Mb to 53.42 Mb. ROH lengths were classified into 5 classes and the Brown, Grey/Black and White Swakara harboured the most ROH count in the $ROH_{5\text{-}10}$ class. Fewer ROH of more than 40 Mb in length were detected in the Swakara groups indicative of ancient inbreeding effects. The Grey/Black and Karakul groups displayed the lowest ROH mean length. Inbreeding based on runs of homozygosity (F_{ROH}) was investigated and ranged from $F_{ROH} = 0.012$ to $F_{ROH} = 0.309$ in the Swakara sub-populations. A broad span of high and low inbreeding was seen in these sub-populations. 28 ROH islands were detected across 18 autosomes which were linked to 722 genes with OAR7 having the most. These genes were related to metabolic pathways (*MOGAT1, ACSL3, SQLE, PIGH)* and disease related pathways (*RPS25, RPS23, DAPK2, ETV6*). Overall, a higher number of ROH, ROH islands, and inbreeding levels were observed in the ancestral populations compared to the Swakara.

Key words: *Runs of homozygosity, ROH island, inbreeding, OvineSNP50 beadchip, haplotype.*

5.2 Introduction

Sheep production has long been hailed as an essential economic activity in many countries, especially developing countries, where its value in subsistence farming systems through the provision of food has played a significant role (Fleming, 2009). This has become more evident in Southern African countries like Namibia, South Africa, and Lesotho. Although most sheep are bred for mainly meat and wool (Dzomba *et al.*, 2021), a few exceptions exist for breeds bred for their pelt like the Karakul (Kirsten, 1966). The Karakul is the only breed of fat-tailed sheep that has grown to be fairly wide-spread throughout Southern Africa. Developed from the Karakul breed through several generations of selection and crossbreeding, the Swakara sheep has become extraordinarily popular in the pelt industry with its high-quality unique pelts that come in four main colours (grey, black, brown, and white). The introduction and productivity characterization occurred in Namibia and South Africa in the early 1900's, but was also later on introduced in Botswana (Nsoso and Madimabe, 2003), where pelt industries based on the Karakul were previously developed.

The introduction of the Karakul into Southern Africa and subsequent crossbreeding with the indigenous Namaqua Afrikaner and Blackhead Persian resulted in a genetically distinct breed, compared to the original Karakul, which is now known and traded as the Swakara. The Swakara pelts are predominantly found in four coat colours, with five loci believed to be responsible for the expression of these colours (Schoeman, 1998). Due to the ability to be dyed to any desired hue to manufacture desirable items in the fashion industry, white pelts have become the most favoured pelt colour (Campbell, 2007). As a result, farmers have intensely selected for pure white pelts to maximise the profitability/remunerative power of these sheep. This study hypothesised that the intense selection geared towards the production of white pelts in Swakara sheep promoted inbreeding which may have inadvertently caused the development of the subvital disorder and other diseases. The subvital factor is characterised by distended rumens containing a fine, frothy substance with the walls of the abomasa and rumen being thin with very little muscle present. This results in the animal's inability to feed causing death within 48 hours. White Swakara are found in A-white and B-white with the pure sheep experiencing the subvital disorder considered A-white. B-white sheep do not experience the subvital disorder

but do not possess purely white coats. They can be found with white skin and small black patches. Although the pure white coats would be more desirable to farmers, A-white X A-white crossings are strongly avoided as they produce all subvital sheep since the disorder is homozygous recessive. Matings have mainly consisted of B-white X B-white (vital Swakara) resulting in 25% black sheep, 25% A-white pure sheep that display the subvital factor, and 50% B/C-white with small spots on the body (Muchadeyi *et al.*, 2014).

Similarly, grey x grey matings are avoided as 25% of the lambs carry the lethal subvital factor. The homozygous grey lambs can be identified by the ears, tongue and palate lacking pigmentation. Disorders have also been reported in the brown Swakara where matings of brown X grey have taken place where the lethal roan allele takes effect (Lundie, 2011). Understanding the genetics of these disorders can elucidate on their genetic origin and underlying mechanism, dissect the consequences of selection and inbreeding effects to maintain genetic diversity of this breed, assist breeding programmes and sustain the genetic gains emanating from improvement programs.

Populations experiencing high inbreeding may undergo purging/removal of deleterious alleles through natural selection. The effectiveness of the said selection is debatable in circumventing the consequences of inbreeding (appearance of disorders, extinction, etc) in relation to the rate of inbreeding (Reed *et al.*, 2003). With the Swakara being a breed kept primarily under smallholder farming systems with very small population sizes and experiencing constant selection pressures for pelt colour and quality this can lead to high inbreeding in the population. Investigating inbreeding in Swakara and its effects on populations can be important in the evolution of breeding strategies as well as in the elimination of deleterious variants.

Runs of homozygosity (ROH) serve as indicators of whole genome inbreeding levels, enabling the characterisation of the distribution of inbreeding depression on a phenotype and the identification of genes linked to economically significant traits (Mastrangelo *et al.*, 2018a). ROH are regions of the genome that have received the same alleles from both parents as a result of a shared ancestor. ROH analyses can shed light on genetic diversity, inbreeding, and potential health problems (Peripolli *et al.*, 2018). Longer ROH segments imply more recent inbreeding within a population, which can lower the degree of genetic variety (Meyermans *et al.*, 2020). Furthermore, long ROH can divulge the full detrimental effects of recessive

deleterious variants in ROH in inbred individuals (Ceballos *et al.*, 2018). On the other hand, short ROH show a population that has undergone a population bottleneck can be indicative of a deeper parental relatedness.

Screening the Swakara sheep for ROH can be helpful in determining the degree of genetic diversity. Investigating the frequency and distribution of ROH in Swakara sheep can assist in identifying ancestral qualities that are connected to alleles that may have been passed down from common ancestors and can further provide insights into human-driven selection. Information on the ROH can inform breeding and improvement of mating systems as well as conservation of genetic resources (Abied *et al.*, 2020b). The length and frequency of ROH are significant factors in determining the associated causes of changes in the genome throughout time (Nosrati *et al*., 2021). Several studies have been conducted investigating ROH in sheep breeds to quantify inbreeding (Selli *et al.*, 2021; Dzomba *et al*., 2021; Nosrati *et al*., 2021; Addo *et al.*, 2021).

ROH screening aid in the exploration/investigation of genomic regions in a population that are indicative of high inbreeding which are known as ROH islands (Gorssen *et al.*, 2021). These ROH hotspots carry additional information on the potential effects of artificial or natural selection from adaptation or breeding events (Liu *et al.*, 2021). Extensive research characterising autozygosity in sheep populations/breeds has proven/shown that selection is highly important in shaping the genome of sheep breeds and the contribution that ROH detection may have in identifying genomic regions responsible for determining traits of economic importance and the resultant effects (Mastrangelo *et al*., 2017, Li *et al*., 2022, Mastrangelo *et al.*, 2018a, Szpiech *et al.*, 2013).

Researchers often employ whole-genome sequencing or high-density SNP genotyping data to discover these ROH regions, examine their implications on a population's genetic parameters and phenotypes (Ferenčaković *et al.*, 2013; Meyermans *et al.*, 2020). The Illumina OvineSNP50 beadchip has become a comprehensive and widely used technology for producing high-throughput genotyping data to characterise population structure, genetic relatedness between individuals and detection of traits linked to harmful recessive alleles (Nosrati *et al.*, 2021). This study sought to employ the genomic proficiency/effectiveness of the OvineSNP50 beadchip to fully ascertain the presence and distribution of ROH, identify regions with notable

ROH islands and their association with phenotypic traits and infer on the causal genomic regions associated with the prevalence of the subvital factor and diseases in Swakara sheep

5.3 Materials and Methods

5.3.1 Sampling and data quality control

The 106 genotyped sheep from the six clusters (Brown, n=21; Grey/Black, n=24; White, n=32; Karakul, n=10; Namaqua Afrikaner, n=5; Blackhead Persian, n=14) as described in Chapter 3 were subjected to the MAF and LD pruning set at 0.01 and R^2>0.2 (using bins of 50 SNPs with a window of 5), respectively using PLINK v1.7 (Purcell *et al*., 2007).

5.3.2 ROH Detection

ROH analysis was computed for each individual using the detectRUNs package in r studio which makes use of two techniques: 1) Sliding-window and 2) Consecutive runs *(Biscarini et al.,* 2018), this study utilised the former. While consecutive runs is a window-free method of scanning the genome SNP by SNP, the sliding-window method is equivalent to PLINK (Purcell *et al.*, 2007). The threshold and parameters applied to succinctly/accurately define a ROH were (i) homozygous threshold window score for each SNP was set to 0,5; (ii) the minimum length of the ROH was extended to 1000kb as prevalence of short homozygous segments is common in the genome of sheep due to strong linkage disequilibrium (Mastrangelo *et al.,* 2017); (iii) the minimum number of consecutive SNPs in a ROH was 50; (iv) the density of one SNP per 50kb and (v) consecutive homozygous SNPs had a maximum gap of 1000kb between them. The minimum length of 1000kb defined the ROH to avoid shorter and more common ROH. The ROH were classified into five classes based on length: $ROH_{0\text{-}5\ Mb}$, $ROH_{5\text{-}10\ Mb}$, $ROH_{10\text{-}20\ Mb}$, $ROH_{20\text{-}40\ Mb}$, and $ROH_{>40\ Mb}$, respectively. The F_{ROH} (inbreeding coefficient) was estimated based on ROH for each individual as defined by (McQuillan *et al.*, 2008).

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{total}}$$

Where L_{ROH} is the total length of ROH above the minimum threshold length set, and Ltotal being the total length of the autosomes in the genome covered by markers.

5.3.3 ROH islands and their annotation

The topRuns argument in R was utilised to evaluate common regions/segments associated with ROHs across the sub-populations by calculating the proportion of SNPs in ROH and the SNPs with the highest incidence in all sub-populations which were then selected as an indication of ROH islands. A minimum threshold of 70% of ROH incidence was implemented for significant ROH detection. The obtained ROH islands were further investigated for shared commonality or uniqueness between the sub-populations.

The OAR_Rambouillet_V 1.0 genome model implemented in the ovine genome ENSEMBL was used to annotate for genes associated with each of the islands. KEGG (Kyoto encyclopedia of genes and genomes) pathways were identified with the obtained list of genes.

5.4 Results

5.4.1 ROH coverage and chromosome distribution

From the ovine 50k genotype data from 106 individuals representing 6 cluster populations, a total of 4 630 ROHs were found and distributed as follows: Cluster A = 944, B = 907, C = 1411, D = 434, E = 317, F = 617. The average ROH lengths ranged from 3.89Mb, which was detected in the Blackhead Persian population, to 53.42Mb in the Namaqua Afrikaner on chromosome 8. On average, chromosome 1 had the longest ROHs detected. ROH were classified by length from $ROH_{0\text{-}5}$, $ROH_{5\text{-}10}$, $ROH_{10\text{-}20}$, $ROH_{20\text{-}40}$, to $ROH_{>40}$. The average ROH length and average number of ROH per population are shown in Table 5.1 and Figure 5.2. Whilst the White coat colour group had the highest average number of ROH, the Grey/Black and Karakul sub-populations had the lowest (7.59) and (8.68). The Swakara sub-populations of Brown, Grey/Black, and White recorded the highest proportion of ROH in the short $ROH_{5\text{-}10}$ class suggesting more ancient inbreeding, conversely, the lack of recent inbreeding is evidenced by low average ROH counts (±0.5) in the $ROH_{>40}$ class in these sub-populations. All individuals had at least one ROH and almost all sub-populations had ROH in all length classes except for the $ROH_{>40}$ class. The Blackhead Persian group was the only one where no ROH were identified in the $ROH_{>40}$ class. The majority of the detected ROH were between 10 and 20Mb with an average proportion of 18% for all populations in that class. Namaqua Afrikaner had the highest proportion of ROH longer than 40Mb indicating some recent inbreeding in this breed. The Blackhead Persian had the greatest number of shortest ROH lengths with a ROH proportion of over 50% in the 0-5 class whilst having the fewest long ROH in the rest of the classes (38%, 11% and 0%). Overall, a higher number of ROH were found in

the Swakara sub-populations than the ancestral populations which displayed more shorter ROHs.

Table 5.1 Mean length of ROH in Mb across 26 autosomes in Swakara sub-populations

Chromosome	Sub-population					
	Brown	Grey/Black	White	Karakul	Namaqua Afrikaner	Blackhead Persian
1	9,79753988	8,68142779	11,2901021	8,61350308	12,0169307	6,25791654
2	12,9229954	7,44209598	9,42054129	8,40841474	12,1398821	6,39260217
3	13,0869633	9,18658213	9,85445083	8,69106955	15,3916735	6,80792097
4	11,0268371	8,26347229	9,20900258	6,40480476	8,96749393	6,23860755
5	10,5028089	8,00313509	6,87402769	9,28105189	12,926702	5,38350044
6	13,6784105	9,32031248	11,0416751	10,2730485	10,2777171	7,06212607
7	7,28356222	8,80911209	8,46960848	7,71613063	10,0213175	5,10890658
8	7,97621216	8,52395126	8,85022744	7,147961	25,6412467	6,38935092
9	10,1517418	6,06083179	8,21948138	7,44754033	5,58059775	5,62525303
10	11,3421713	5,89138041	8,77318112	6,13435795	9,09885938	5,8279667
11	11,2258687	11,5491494	9,23565667	6,58806244	12,8571604	5,3809832
12	11,8651504	8,82238209	9,71061352	7,4614219	8,9757065	5,66248347
13	12,1316752	8,80502827	7,54751461	9,510797	9,64751175	5,16887718
14	10,6554547	6,91429468	8,13521446	6,86531983	9,03095967	5,90979609
15	11,5141934	7,5512728	8,04794162	9,16083173	6,9703362	5,58401094
16	10,4666871	6,83168793	10,6967697	6,36296738	12,313043	5,77476176
17	12,3462314	7,517189	7,4644695	6,98982567	12,961962	7,367818
18	8,04839359	6,37746466	8,46129965	12,1583675	14,310123	6,74417179
19	6,62919742	8,12122473	9,35235074	13,0175895	12,0944977	5,11509386
20	5,73095988	5,36199127	8,4251833	6,03132338	9,2372472	4,63118864
21	8,90715509	9,11157306	7,38019212	8,1295805	10,414959	8,88358529
22	8,2539935	7,33628294	5,93562706	8,33909909	6,612233	6,25866279
23	9,41297176	10,8533351	8,04151352	4,91596333	9,8779184	6,5772001
24	8,54553518	10,2891853	9,6456828	7,0413832	6,7305785	7,24163133
25	6,26301346	7,37116706	8,6570485	5,22517329	9,02549227	3,63228
26	6,81439483	7,53584448	8,78772084	8,01167756	9,23536967	4,46242133

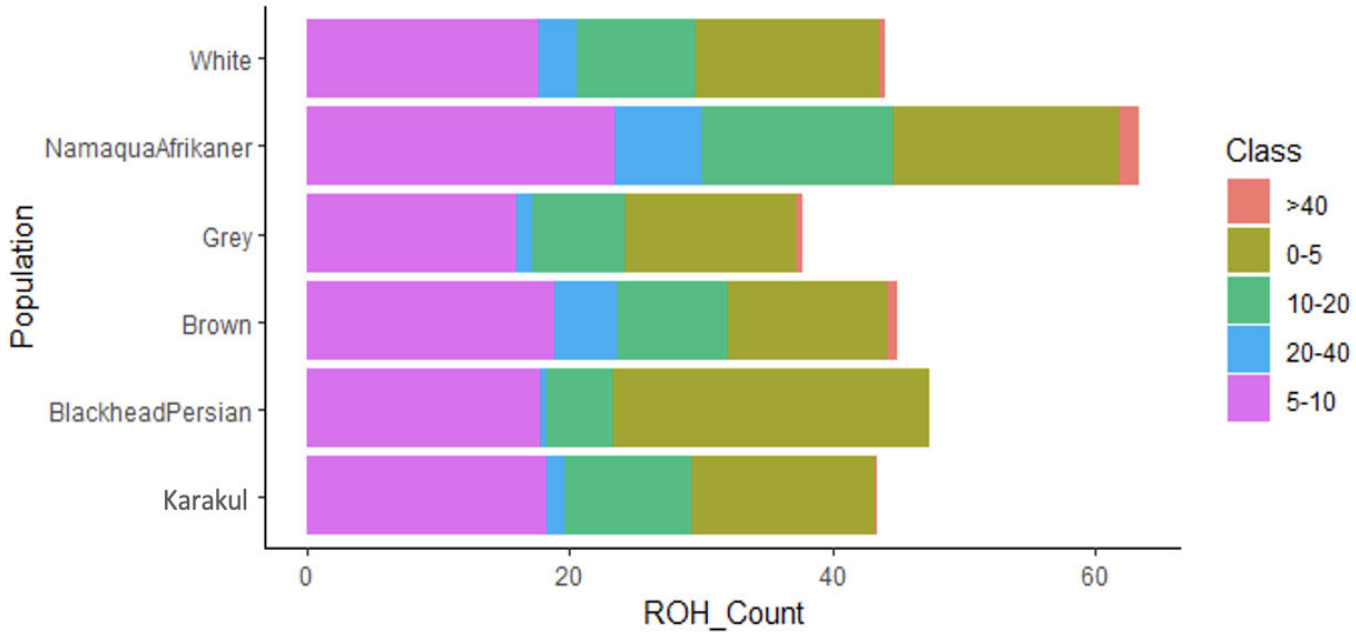


Figure 5.1 Average ROH number on populations divided into five classes with colour scale representing the length of runs in Mb

The proportion and distribution of ROH coverage was further investigated at a chromosomal level for all groups. Complementary to the preceding figures, the plotting of ROH per individual allowed the visualization of possible trends and patterns in the size and distribution of ROH in the genome for the populations. The chromosomes with the highest number of ROH and as a result the longest ROHs were OAR1, OAR2, OAR3, and OAR6. These scaffolds were magnified to identify clear patterns of quantity distribution (Figure 5.3). OAR1 had the highest number with 574 across all populations followed by OAR3 (456), OAR2 (449) and OAR6 (249). Long and short ROHs were detected at different proportions as seen in the Brown coat colour Swakara in OAR1 and OAR6 where longer clear runs are shown in chromosome 6 compared to shorter gapped runs in chromosome 1. This pattern was also detected for the Namaqua Afrikaner in OAR3 and OAR1.

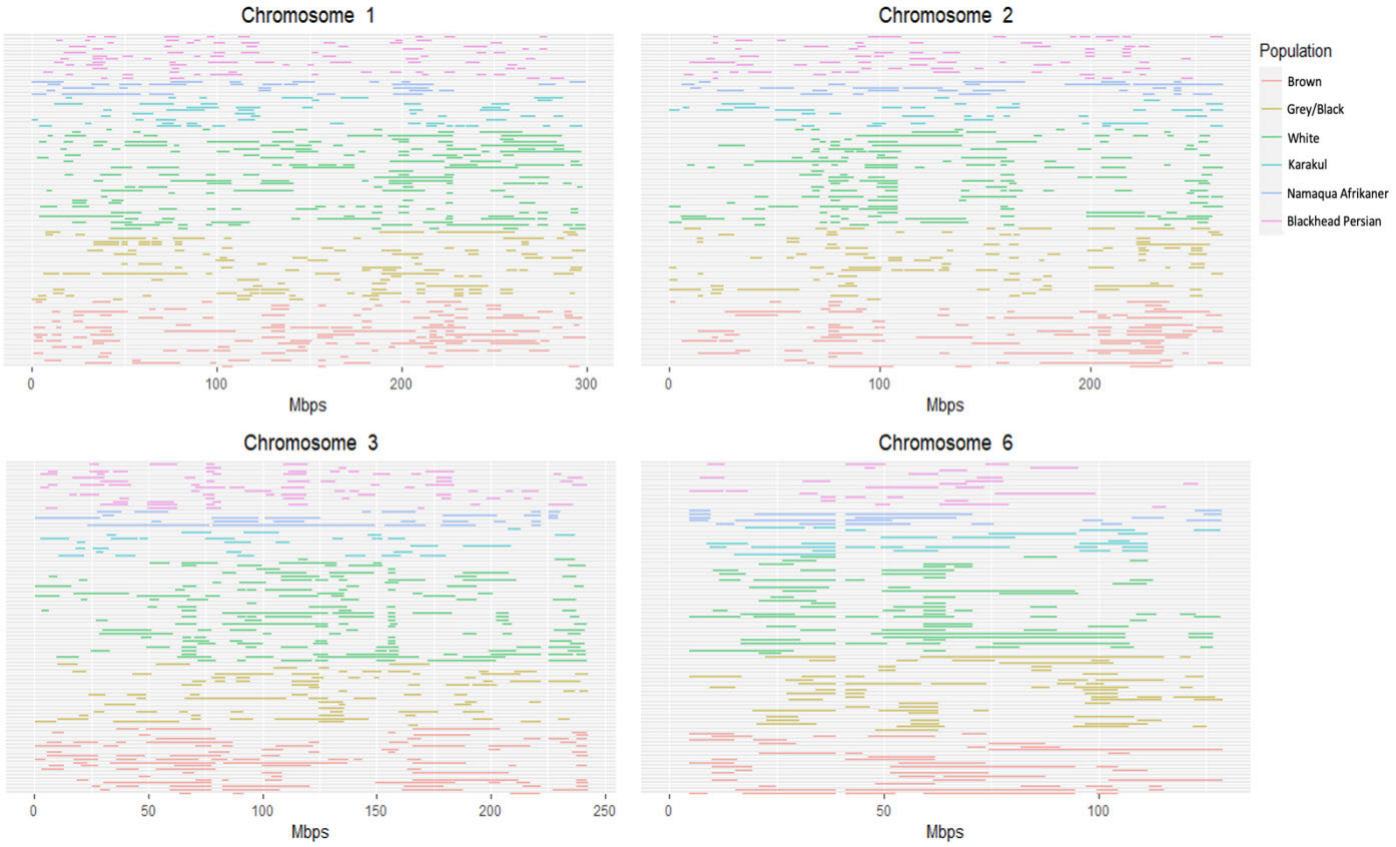


Figure 5.2 Distribution across chromosomes of four of the highest number of detected ROH

5.4.2 Analysis of inbreeding coefficient (F_{ROH})

Inbreeding values were computed and analysed for each individual in the population and varied among sub-populations and regions (Figure 5.3). Mean F_{ROH} values ranged from (F_{ROH} = 0.10 ± 0.03) for the Blackhead Persian to (F_{ROH} = 0.26 ± 0.05) for the Namaqua Afrikaner. The Brown and Grey/Black sub-populations had the highest inbreeding values amongst the coat colour Swakara with F_{ROH} = 0.296 and F_{ROH} = 0.309, respectively but also the lowest for some individuals. A greater range of inbreeding was witnessed in these sub-populations where individuals with the lowest inbreeding values of F_{ROH} = 0.017 (Brown) and F_{ROH} = 0.012 (Grey/Black) amongst the Swakara were also detected. Similar patterns were seen in the White Swakara with a range of animals exhibiting low inbreeding but those with high inbreeding still being present in the group. The highest inbreeding value of F_{ROH} = 0.254 was recorded with a low of F_{ROH} = 0.024. The higher range of inbreeding within the Swakara sub-populations compared to the Karakul, Namaqua Afrikaner and Blackhead Persian provided insight into the dynamics of the different sub-populations and their management. Overall, a higher level and

range of inbreeding was observed within the different Swakara coat colour groups compared to the ancestral groups.

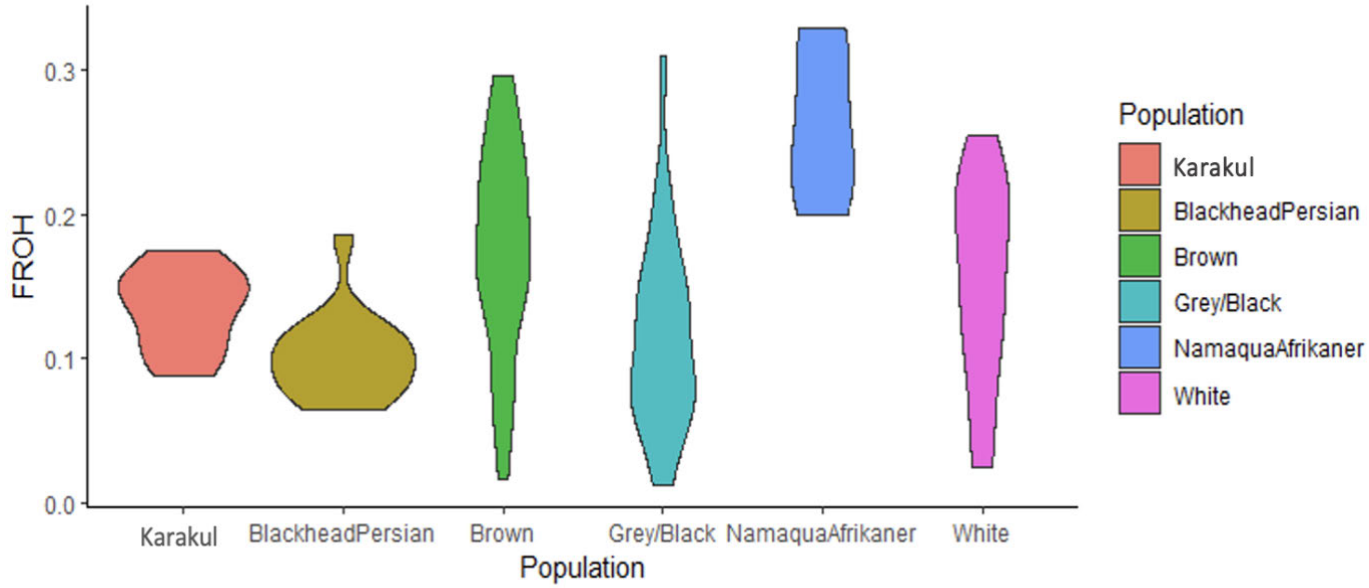


Figure 5.3 Violin plot showing inbreeding (F_{ROH}) in six sheep populations

5.4.3 Distribution and gene annotation of ROH islands

ROH islands were determined and a total of 28 ROH hotspots above the threshold of 70% (Figure 5.4) were identified across 18 autosomes (no islands were found on OAR14, 16, 18, 19, 21, 22, 23, 26).

722 genes were detected to be within these genomic hotspots with OAR7 harbouring the most (109) followed by OAR24 (106), OAR20 (99) and OAR1 (91). These regions were associated with genes involved in KEGG pathways such as metabolic pathways (*ACADL, LACC1, EXT2, MOGAT1, ACSL3, SQLE, PIGH and ARG2*), and coronavirus disease (*RPS23, RPS25*) (Table 5.2). OAR17 held the greatest number of ROH islands (4) harbouring the *FAT4, SCLT1, JADE1, LARP1B, ABHD18,* and *NEK2* genes. However, only one gene was found to be associated to a KEGG pathway which was for Basal transcription factors (*FAT4*

Table 5.2 ROH islands with identified genes and associated KEGG pathways

Rank	Chr	Genomic regions		Gene	KEGG Pathways
		Start	End		
1	13			*PSMA7*	Neurodegeneration (multiple diseases)
				TAF4	Basal transcription factors
		57402806	58349162	*CDH4*	Cell adhesion molecules
2	17	36649312	44373422	*FAT4*	Hippo signalling pathway
3	2	226002221	234264088	*ACADL*	Metabolic pathways
4	1			*HACD2*	Fatty acid elongation, Biosynthesis of unsaturated fatty acids
				MYLK	Calcium signalling pathway, Gastric acid secretion
		199950939	211907451		
5	10			*LACC1*	Purine metabolism, Metabolic pathways
		15802810	19965597		
6	11			*NLRP1*	NOD-like receptor signalling pathway
		37316663	38029087		
7	12			*RPS25*	Ribosome, Coronavirus disease
		72317972	80513310		
8	15			*EXT2*	Metabolic pathways, Glycosaminoglycan biosynthesis, heparan sulfate backbone
		78210546	81086980		
9	17	29629298	29666258	*No gene*	
10	17			*SCLT1*	No pathways
				JADE1	
				LARP1B	
		30432111	38223009	*ABHD18*	
11	17	39101959	44741554	*NEK2*	No pathways

12	2	144088913	147872409	*HOXD1*	Signalling pathways regulating pluripotency of stem cells
13	2	239828382	243254175	*MOGAT1*	Glycerolipid metabolism, Metabolic pathways
				ACSL3	Metabolic pathways, Ferroptosis, Adipocytokine signalling pathway
14	20	1040020	4480665	*SNRPC*	Spliceosome
15	20	6536394	13838496	*HCRTR2*	Neuroactive ligand-receptor interaction
16	24	1006379	5798772	*TELO2*	mTOR signalling pathway
17	25	42687885	48256402	*SLC18A3*	Synaptic vesicle cycle, Cholinergic synapse
18	3	217812222	222273968	*ETV6*	Transcriptional misregulation in cancer
				TAS2R42	Taste transduction
19	4	8389438	8687452	*ABCA13*	ABC transporters
20	5	85567966	88936480	*ATG10*	Autophagy-animal
				RPS23	Ribosome, Coronavirus disease
				XRCC4	Non-homologous end-joining
21	6	4772469	9213632	*MAD2L1*	Progesterone-mediated oocyte maturation, Cell cycle
22	6	43178901	52034048	*PPARGC1A*	AMPK signalling pathway, Longevity regulating pathway, Insulin resistance
23	7	3814182	8716513	*RAD51B*	Homologous recombination
				SNX1	Endocytosis
24	7	37125307	49826553	*DAPK2*	Pathways in cancer
				HERC1	Ubiquitin mediated proteolysis

25	7			*PIGH*	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis, Metabolic pathways
				ARG2	Metabolic pathways, Amoebiasis
		81734304	83427652		
26	8			*RFX6*	Maturity onset diabetes of the young
		22581673	27294787		
27	9			*WASHC5*	Endocytosis
				SQLE	Steroid biosynthesis, Metabolic pathways
		29234824	34085524		
28	3	78378522	79055518	*NRXN1*	Cell adhesion molecules

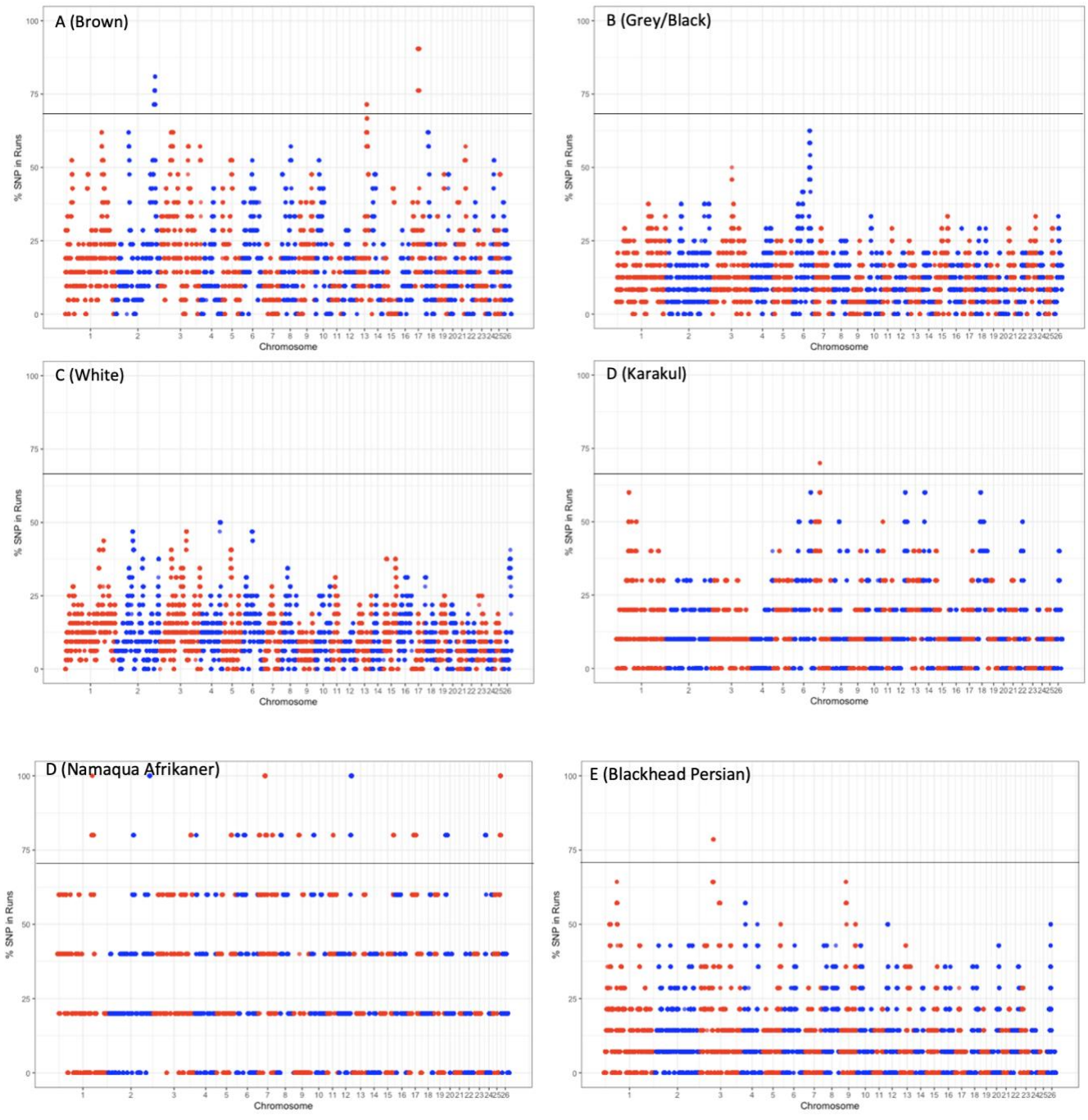


Figure 5.4 Distribution of ROH islands (>70%) in the whole population

5.5 Discussion

For almost a century, evolutionary biologists and agricultural researchers have prioritised the quantification of animal inbreeding with technological methods that are constantly improving. For the longest of times pedigree data has been traditionally used in estimating inbreeding. The convenience of high-density SNP markers has allowed for the exploration of other methods such as the utilisation of genome autozygosity. Normally, autozygosity materialises through the inheritance of similar DNA segments derived from a common ancestor in individuals that result in continuous homozygous regions known as runs of homozygosity (Pemberton *et al.*, 2012). ROH detection provides not only data on demographic history or population structure, but is also significant in determining ROH regions harbouring lethal recessive variants is highly important (Nosrati *et al.*, 2021). The concept of high inbreeding and increased autozygosity because of selection is one that is highly studied and clearly understood. However, lack of information on the consequences of genomic selection on ROH distribution and events behind appearance of lethal recessive alleles in the genome exists. In this work, we analysed the genomes of pelt colour variants of Swakara sheep and their presumed ancestral breeds to detect genomic inbreeding patterns and their implications on the appearance of subvital effects in the White and Grey coat colour Swakara and the occurrence of disorders in the Brown Swakara sheep.

This study utilised R studio to detect ROH in the populations. In other studies, the most common tools frequently used are Golden Helix SVS as well as PLINK (Forutan *et al.*, 2018, Purfield *et al.*, 2012, Xu *et al.*, 2021b). The aforementioned techniques all have the limitation of not taking into account the possible presence of heterozygous SNPs or those in close proximity to the ROH when calculating inbreeding coefficients (FerenčakovićSölkner and Curik, 2013). In such instances, it is more likely to be indicative of heterozygosity of the region. Additionally, minor allele frequency (MAF) and linkage disequilibrium pruning has been shown to affect the overall results (Ferenčaković *et al.*, 2013; Albrechtsen *et al.*, 2010). In this study, SNPs with a MAF lower than 1% were removed and LD pruning was applied at r^2>0.2 with window size of 50 and a shift of 5.

The average number of ROH present for each sub-population varied immensely in the study with an average of 43.68 ROH per animal. The Swakara coat colour groups produced the most ROH with 944 (Brown), 907 (Grey/Black), 1411 (White). Out of 4630 ROHs detected, the

ancestral populations accounted for 434 (Karakul), 316 (Namaqua Afrikaner), and 616 (Blackhead Persian) ROHs. High ROH counts in the White Swakara were similarly reported in a study by Dzomba *et al.* (2021) where 6155 ROHs were identified. The Swakara breed is known to be a small breed kept in traditional farming systems where mating is carried out on tribal lands with little or no flock boundaries such as fencing which culminates to uncontrolled breeding (Nsoso and Madimbe, 1999). Additionally, this breed can be found in smallholder sheep farms with low productive performance, lack of nutrition and proper breeding strategies. A small proportion of Swakara are found in commercial farms and the repeated use of fewer bucks in breeding cycles can be the reason behind the abundance of autozygous genomic sites witnessed in the Swakara coat-colour sub-populations (Liu *et al.*, 2022). Operating on a controlled open village breeding system may potentially decrease susceptibility to higher occurrences of consanguinity sites that lead to the accumulation of deleterious alleles and manifestation of disorders as observed in the White, Grey and Brown Swakara. Furthermore, with an active market in the pelt industry the constant need to maintain or improve genetic fitness within the Swakara coat colour breed coupled with the improper structure in these improvement programs, the higher occurrence of ROH is expected. During the 1970s, South West Africa had more 95% of Swakara out of 4.4 million sheep stock (Schoeman, 1998), but the collapse in the industry resulted in fewer stock with just over 250, 000 sheep being reported in Namibia in 2016 which constituted less than 10% of total breeding stock (Ministry of Industrialisation, Trade and SME Development, 2016). Population reduction is a big factor in the development of long homozygous sites along the genome and decreased effective population size in the Swakara may have also contributed to the observed effects of unfavourable inheritance of recessive alleles through population bottleneck effects (Purfield *et al.*, 2012; Zhang *et al.*, 2015).

Shorter ROHs were found frequently across the genome than the longer ROH with the Swakara coat colour sub-populations presenting more than 40% of their identified ROH between 5-10 Mb length. Higher percentages were also observed within the shortest class length (0-5 Mb) with more than 27% of ROH identified presented in each sub-population. Many studies have previously reported the prevalence of short ROH for various livestock breeds (Xu *et al.*, 2021b, Forutan *et al.*, 2018, Marras *et al.*, 2015, Dzomba *et al*., 2021). The classes representing short ROHs reported 654 ROHs in the Brown sub-population, 699 in Grey/Black, 1011 in White, 322 in Karakul, 203 in Namaqua Afrikaner and 546 in Blackhead Persian. Long ROHs are mostly indicative of relatively recent inbreeding. Longer ROHs were

also detected in the study although not at a higher degree. The appearance of both long and short ROH at various frequencies throughout the different sub-populations illustrates that both recent and ancient haplotypes have influenced the population with majority of the inbreeding events arising from past selection pressures. A difference in the genetic structures of the sub-populations, their history as well as exposure to various selection strategies may have resulted in the different ROH frequencies observed. Recent inbreeding and relatedness are characterised by long ROHs. An inverse correlation between recombination rate and ROH length has been reported (Hewett *et al.*, 2023). Recombination allows for separation of the identical regions, thus reducing their size relative to the size of the genome and their overall effects on population fitness. In a study by Hewett *et al.* (2023) forward genetic simulations were used to study the effects of population history, recombination and selection on the quantity and distribution of ROH in a wild red deer population. Population history was discovered to have had the most influence/impact on the distribution of ROH. The history of the Swakara, from its introduction and development from a previous breed and subsequent selection pressures, and divergence from the Karakul may play a key role in the ROH distribution. Recessive harmful mutations are often overrepresented in lower-diversity populations, which the Swakara is known to be, perhaps as a result of founding events that increased the frequency of otherwise uncommon alleles. If the same founding events that have amplified recessive harmful mutations are the cause of ROH, one may anticipate that such variants would be present in ROH areas resulting from such founder events, which are often short and medium length ROH (Szpiech *et al.*, 2013). Overabundance of lethal recessive variants in ROH sites may also be explained, to an extent, by an event where positive selection produces the observed long haplotype segments which in turn cause deleterious variants in the ROH neighbouring sites that are positively selected to hitchhike and appear in higher than normal frequencies (Pemberton *et al.*, 2012).

A diverse range of ROH frequencies was seen across the chromosomes in each of the sub-populations. The greatest number of ROH per chromosome was on OAR1 with 574 runs across 106 individuals followed by OAR3 with 456, then OAR2 with 449 and OAR6 with 249. A decrease in the number of ROH per chromosome was seen as chromosome length/size went down. Chromosome 24 reported the lowest number of ROHs in this study (56). These results are consistent with those reported by Islam *et al.* (2019) for different Chinese goat breeds. Patterns of ROH distribution were detected where different frequencies for some sub-populations varied in each chromosome. This demonstrated the non-random scattering/positioning of ROHs that occur as result of selection events. There were no ROHs

observed on chromosome 6 around the ±40Mb region (Figure 5.2). According to Dzomba *et al.* (2021) the presence of gaps in the coverage of the marker could be the result of such an occurrence.

The use of pedigree data to estimate inbreeding coefficient has been, for the longest time, the conventional method. However, limitations present themselves when studying cross-bred individuals with poorly established genealogical relationships in the paternal and maternal breeds, and even when established the pedigree data might be incorrect or incomplete (Xu *et al.*, 2021). Furthermore, an increased level of inbreeding is required for individuals involved in homozygous-mapping analyses since it does not enable smaller ROH detection which high density genotype data allows (Pemberton *et al.*, 2012). Therefore, an accurate representation of relatedness could not be confirmed with inbreeding coefficients estimated based on pedigree information. In this study, genomic data was employed in the estimation of inbreeding. The observed genomic inbreeding coefficient (F_{ROH}) in the whole population was relatively moderate. The Swakara coat colour groups displayed mean F_{ROH} of 0.17 (Brown), 0.12 (Grey/Black), and 0.15 (White). Whereas the Karakul, Blackhead Persian, and Namaqua Afrikaner recorded inbreeding coefficients of 0.13, 0.10, and 0.26, respectively. A wide range of inbreeding coefficients were observed in all Swakara sub-populations where animals with low and high inbreeding were recorded. This occurrence may be indicative of the existence of some individuals in the population who experience the disorders and thus are characterised by high inbreeding, and those who do not suffer from these disorders and exhibit lower inbreeding. Additionally, this can be heavily influenced by the management skills applied by farmers and the diversity in how the farmers manage animals within each sub-population to try and manage disorders whilst also trying to breed for a higher pelt quality. for Low inbreeding coefficients for the Swakara sub-populations have been shown by Muchadeyi *et al.* (2015) where F_{ROH} levels were as low as 0.011± 0.069 for the Grey Swakara and went up to 0.094 ± 0.075 for the Black Swakara with an overall F_{ROH} in all sub-populations of 0.089±0.063. The F_{ROH} levels reported are almost similar with slightly higher values reported in this study. The current crossbreeding of the White Swakara with the Black coat Swakara as means of mitigating and eliminating the presence of the lethal subvital factor is believed to impact the genetic variation in these populations (Muchadeyi *et al.*, 2015). Contrastingly, Dzomba *et al.* (2021) described high F_{ROH} for Swakara sub-populations. The high inbreeding values seen in some individuals of White, Brown and Grey/Black Swakara compared to their presumed ancestors can be attributed to the breed being kept in smallholder farming systems which are characterised by

small sizes, poor breeding management implying less genetic diversity, and with no natural gene flow or introduction of new genetic variation no opportunity is availed for heterozygous sites to disperse the ROHs. Furthermore, these sub-populations experience high selection pressures with the constant need for growth in the quality and quantity of pelts produced under small herd sizes. This level of inbreeding makes the sheep prone to inbreeding effects which can manifest in genetic disorders as seen in the White, Grey and Brown Swakara. The low F_{ROH} perceived in Karakul and Blackhead Persian has been witnessed with other fat-tailed sheep breeds (Edea *et al.*, 2017, Abdoli *et al.*, 2023) including the Blackhead Ogaden of Somalia. Although it was developed in South Africa, the Blackhead Persian is a direct forebear of the Somalian Ogaden, and it has become widespread throughout Africa in various production systems and even in the production of composite breeds (Wilson, 2011). The extensive availability of the Blackhead Persian may be the possible reason behind its low inbreeding levels based on ROH as the introduction of novel genetic variation has allowed for better recombination. In this work, the Namaqua Afrikaner had the highest inbreeding value of $F_{ROH} = 0.26 \pm 0.10$. These results differ from those of Retief (2020) of the breed exhibiting the lowest F_{ROH} in the study compared to other meat type and indigenous sheep populations.

A large proportion of the inbreeding identified can be mapped to chromosomes 1, 3, and 6 with average F_{ROH} of 0.17, 0.18, and 0.19, respectively across the 106 individuals studied which has also been reported by Addo *et al.* (2021) where chromosomes 1, 2 and 3 harboured the highest proportion of inbreeding. All 26 autosomes had inbreeding values of at least 0.10 with the lowest being seen in chromosomes 24 (0.10) and 20 (0.11). This can be attributed to larger chromosomes possibly containing more regions or genes that have been favourable in the process of selection in breeds.

The most homozygous regions that appear in more than 70% of the individuals in the population were identified as ROH islands in the genome. These are genomic regions with increased homozygosity surrounding positively selected targets undergoing strong selection pressure (Peripolli *et al.*, 2018, Xu *et al.*, 2021b, Mastrangelo *et al.*, 2018a). A strict threshold has been applied in ROH hotspot detection. Similar high thresholds have been observed in studies of dairy cattle (Mastrangelo *et al.*, 2018b), pigs (Di Gregorio *et al.*, 2023) and sheep (Selli *et al.*, 2021). A total 28 ROH islands were detected in 19 autosomes with over 400 genes found to be within these regions. A wide range of different biological processes, cellular components and molecular performance comprise the genes. Chromosome 17 was of

prominent importance detecting the largest number of islands. Only one pathway was notable in this region related to the *FAT4* gene (ROH2) which was the hippo signalling pathway. This pathway is known to be a highly-conserved developmental pathway playing a crucial role in controlling organ size and tissue generation (Juan and Hong, 2016) and has also been implicated in negatively affecting liver development and other internal organs in Hu sheep (Zhang *et al.*, 2017). The key characteristic of the subvital disorder is underdeveloped digestive organs which in turn hinders the lamb's inability to feed thus resulting in its death. The identification of this pathway could be associated to the appearance of the subvital disorder in Swakara sheep. Metabollic pathways were the most commonly identified in the ROH islands with genes such as *LACC1* (ROH6), *EXT2* (ROH8), *MOGAT1* (ROH13), *PIGH* (ROH25), *ARG2* (ROH25), and *SQLE* (ROH28). Metabolites, which are the end product of all metabolic reactions, are known to be sensitive indicators of issues within the genome. Metabolic status can be crucial in assessing damage of tissue, disorders of organ function and other physiological problems (Hernandez *et al.*, 2020) A highlighted pathway associated to environmental information processing was the neuroactive ligand-receptor interaction related to multiple diseases (*PSMA7, HCRTR2*) found to be a key component in the ATP-dependent proteolytic pathway in sheep (Filali *et al.*, 2014). The process of endocytosis is implicated in two ROH islands: ROH24 (*SNX1*), and ROH27 (*WASHC5*) was highlighted. The several disease related pathways such coronavirus disease (*RPS25, RPS23*), insulin resistance (*PPARGC1A*), pathways in cancer (*DAPK2*), maturity onset diabetes of the young (*RFX6*) and transcriptional misregulation in cancer (*ETV6*) were identified which were not expected but may be suggestive of some mechanisms that the genetic disorders seen in Swakara may be associated to. Many pathways related to metabolic function, diseases and resistance to diseases may also be indicative of possible selection pressure for these functions/traits. Mastrangelo *et al.* (2018a) has suggested that natural or artificial selection may result in some genomic regions being fixed in individuals for productivity or adaptability reasons. No ROH islands were observed on chromosome 6 in this study. Interestingly, Selli *et al.* (2021) found the region to be a common island for majority of the merino populations studied. Moreover, Abied *et al.* (2020b) and Gorssen *et al.* (2021) also identified candidate regions of ROH islands on chromosome 6 which were associated to key genes suggesting it to be crucial for multiple traits of economic significance in other livestock.

5.6 Conclusion

Measuring and characterising inbreeding in livestock has immense benefits. Increased selection can result in a higher number of ROH with some influencing negative effects on selected traits and harbouring deleterious alleles. The high number of short ROH accumulated in the Swakara sub-populations is evident of the multiple generations of inbreeding. The observed high and low inbreeding coefficients in the Brown, Grey/Black and White groups suggest the existence of a pool of individuals who may suffer from subvital disorders and diseases and those who are lowly inbred to be less likely to carry lethal variants. The accumulated ROH and their observed genes have illuminated on their possible association to the prevalent genetic disorders in Swakara. Understanding the underlying genetics of ROH distribution can be crucial in enforcing more effective selection as well as identification and possible elimination of causative mutations from consanguineous matings.

6 GENERAL DISCUSSION AND CONCLUSION

Sheep breeds are universally known for being multipurpose involved in the production of milk, wool, meat, and pelt. Although pelts are considered a by-product in sheep production, it has become the main breeding goal for the Swakara breed and has become the leading sheep breed in high-quality pelts. Smallholder farmers keep flocks of Swakara sheep in different areas of South Africa, Namibia, and Botswana. Found mainly in four coat colours: brown, white, grey, and black with the white pelt colour being the most superior due its high profit value in the fashion industry. Swakara sheep are well adapted to the arid, harsh environments and are resistant to most internal parasites (Ibragimov *et al.*, 2007). This breed is known for being non-selective grazers with strong fertility (Näsholm and Eythorsdottir, 2011). Farmers breed for high quality curl development, lustre, curl size and hair length but studies have shown that the heritability of pelt traits is very low (Martins and Peters, 1992b) highlighting the many generations subjected to intense selection in order to achieve the quality of pelt we see today in Swakara. The revered high quality of the pelt forces the lamb slaughtering at less than 3 days old to preserve the unique structure. For the longest time pelt production almost exclusively revolved around the black pelts (Martins and Peters, 1999), but the white pelt was put more into the spotlight. With increased interest in the white pelt came increased selection since the white coat colour is not inherent to the Karakul. Crossbreedings with other white-wooled indigenous breeds of the Namaqua Afrikaner and Blackhead Persian and further selection in the breed, is hypothesised to have caused the occurrence of subvital phenotype in the pure White and Grey Swakara sheep. This lethal phenomenon in Swakara sheep is poorly understood and largely remains understudied with farmers haphazardly resorting to the mating of black and white sheep in order to manage the subvital condition in pure white sheep which, in turn, results in the B and C White sheep.

The genomic architecture which is presumed to be heavily influenced by the result of crossbreeding and selection and its association to the subvital and other genetic disorders observed in the breed has not been adequately studied. A lack of information on the genomic structure of these breeds and the resulting implications of the preceding breed development practises is very evident. This study sought to investigate the genomic architecture of Swakara sub-populations and the relation to the Namaqua Afrikaner, Karakul and Blackhead Persian that are believed to be their ancestral breeds. The OvineSNP50 beadchip panel was used to analyse the population structure and genetic diversity in the population (Chapter 3), ascertain signatures of selection between and within the sub-populations (Chapter 4), and finally

investigate the presence and distribution of runs of homozygosity to infer inbreeding effects (Chapter 5). The OvineSNP50 beadchip was found to be highly informative in other studies that sought to understand the genetic diversity of sheep breeds and the implications on overall phenotype and genotype (Bedhiaf-Romdhani *et al.*, 2020, Molotsi *et al*., 2017, Yan *et al.*, 2017). The suitability/beneficial utility of the OvineSNP50 has been demonstrated in several South African breeds (Sandenbergh *et al.*, 2016). The high number of SNPs allows for greater coverage of the genome and is beneficial in studies investigating marker-trait associations. The different analyses were performed in order to determine genomic structure of the Swakara sub-populations and their genetic association to their founding breeds, to detect genomic regions under selection between the populations and assess their divergence, lastly to investigate runs of homozygosity and infer the cause of the prevalence of the genetic disorder with the ultimate goal of better understanding the Swakara sheep and make inferences on the genetics of the phenotypes prevalent in the breed with a focus on subvital and genetic disorders.

The population structure of the Swakara sheep was investigated by investigating genetic diversity within, between and amongst as the presumed sub-structures of different coat colour sub-populations that display contrasting phenotypic characteristics. The White Vital Swakara sub-populations reported high inbreeding (F_{IS} = 0.08±0.09) and low diversity demonstrating the high selection that the breed has been subjected to. The other coat colour sub-populations also demonstrated low diversity levels with Black, Grey, and White Subvital recording inbreeding (F_{IS}) values of 0.04±0.08, 0.03±0.10 and 0.01±0.1, respectively. The Swakara is commonly used in smallholder sheep systems which can result in the loss of genetic variation in the breed (Molotsi *et al.*, 2017). With that said, the White Swakara also showed the highest percentage of molecular variation (AMOVA) within the individuals at 98.65%. The low between population genetic variation of 6.82% observed amongst the Swakara sub-populations demonstrates the strong genetic ties regardless of differentiating phenotypic characteristics borne out of selection. Further investigation of the genetic structure in this chapter shed more light on genetic similarities and breed composition. The alignment of the PCA and ADMIXTURE confirmed the assumed effects of geographic origin on population structuring. F_{ST} pairwise comparison showed the highest differentiation to be between the Brown and White Swakara sub-populations. The Brown Swakara demonstrated the greatest genetic differentiation in the analysis when compared to the Namaqua Afrikaner (F_{ST} = 0.356). This differentiation is further seen between the Brown and other coat colour groups, supporting the results of the PCA which showed the clear separation of the Brown Swakara from all the other

sheep clusters. Pairwise per marker analysis showed the majority of the highly differentiating SNPs being detected on chromosomes 1, 3, 4, and 7 with genes linked to milk yield, reproduction, disease resistance, tail shape, response to heat stress etc. The Swakara breed are well known for their ability to adapt to hot and arid conditions which characterises most of Southern Africa, which accounts for the identification of genes related to response to the environment. Other genes related to immune response (*CYFIP1, UBR1, DOCK10, PCDH15*) highlighted their adaptive nature that has developed and evolved over a long period of time. Adaptation and innate immunity studies of sheep have generally focused on the *TLR* (Uematsu and Akira, 2006), *MHC* (Ballingall *et al.*, 2018), and *HSP-70* (Romero *et al.*, 2013) genes but novel genes have been discovered to have notable impact on regulating the immunity of different sheep breeds including the *CYFIP1* gene (Yang *et al.*, 2021) which was identified in this study. These results align with those of previous studies highlighting genes that are essential in adaptation and overall survival of the sheep breeds and honed characteristics that have become distinct to the breeds.

A great deal of change is seen under artificial selection events created as a response to the pressures and need to adapt. The population genetics theory states that distinctive patterns will emerge from functioning genes under selection as a result of selection preferences and these unique patterns are the signatures of selection. The effects of the intense selection that the Swakara has been subjected to was repeatedly demonstrated in Chapter 4 and 5 with high inbreeding, low genetic diversity, and clear patterns of selection pressures. Analysis of signatures of selection was key in understanding the genomic regions that have been under selection pressures in these populations. A large number of sweeps were recorded with all three methods of iHS, XP-EHH and HapFLK providing evidence of selective forces. 2.3% of the total genes identified were found to be in common amongst all three statistical methods. Three common genes were revealed between iHS and XP-EHH (*GALNT18, HMG20A and CSPG4*). These were linked to QTLs such as testes weight, staple length, meat acid content, worm count, facial eczema susceptibility and salmonella abortusovis susceptibility. Although no genes were found to be shared amongst all three methods, several identical QTLs were found to be expressed by different/multiple candidate genes. This pattern was also witnessed in chapter 3 and 5. Common QTLs detected were linked to body weight, milk production traits and meat production traits, metabolic and immune response related traits and pathways. These results highlight the vast of umbrella of complex traits that are of economic importance in sheep being selected for and controlled by multiple genes (Stella *et al.*, 2010, Kim *et al.*, 2016). The

identification of teat number and udder attachment QTLs in this study shows the significance of these traits as fundamental and crucial morphological and reproductive traits under selection.

In addition to investigating genomic regions that have been the subject of intense selection in Swakara and the indigenous breeds in this study, it was crucial to detect runs of homozygosity (ROH) regions to ascertain their role in harbouring potentially lethal variants. ROHs are generated when an individual inherits long segments that are identical by descent (IBD) from related parents. Longer ROHs are indicative of more recent ancestry, whilst short ROHs are derived from a more distant ancestor (Curik *et al.*, 2014). A higher frequency of shorter ROHs were detected for all sub-populations in this study. Low recombination in the breeds may be attributed to these results. This can be further exacerbated by poor breeding management and small population sizes. High inbreeding coefficients based on ROH (F_{ROH}) was reported for the White, Brown and Grey/Black Swakara. These results are supported by the low genetic diversity observed in these sub-populations that was reported in chapter 3. The identified ROH islands, which are genomic regions with high homozygosity near positively selected targets experiencing strong selection pressure, were found to be associated with various pathways. The most commonly shared in this population were related to metabolic pathways, and resistance to diseases. Parasite and parasite resistance related QTLs were habitually and frequently detected. Identification of parasite and parasite resistance genes in chapter 3 aligns with results of immune response genes found in chapter 5 which were linked to haemonchus contortus resistance, trichostrongylus adult and larva count, nematodirus FEC etc. A host response that prevents, lessens, or gets rid of parasitic infection is a major component of parasite resistance. Despite not fully rejecting the illness, a lower parasite burden is seen in resistant animals compared to susceptible individuals which is indicated by fewer parasite eggs in faeces. An animal's immunological capabilities is the determining factor of resistance when confronted with parasitoses (Alba-Hurtado and Muñoz-Guzmán, 2013). The breeds in this study, especially the Swakara have been established as having strong immune responses with resistance to most parasites (Ibragimov *et al.*, 2007). The overall results in chapter 3 and chapter 4 and 5 show that indigenous breeds that have adapted to their environment are characterised by poor production, and small effective population sizes, which was demonstrated in chapter 3, compared to high-performing breeds that lack disease-resistance qualities and this, in turn, results in higher ROH frequencies, low genetic diversity, but higher survivability rates.

The main characteristic distinction of the Swakara breed is the many coat colours it can be found in and one of the main focuses when it comes to selection and breeding. A high number of coat colour related genes were expected to be identified in this study, however only one gene (*CDH13*) was detected (chapter 3). The failure of methods/statistical protocols in identifying genes associated with coat colour across all analyses may be attributed to SNP ascertainment bias that comes with using the OvineSNP50 beadchip seeing as the Swakara was not a breed involved in its development. The PCA, admixture and F_{ST} statistics demonstrated poor population substructure amongst the different coat colour subpopulations coupled by clustering of more than one coat colour subpopulation in certain clusters. Such weak population structure between what was initially hypothesised as distinct populations could be suggestive of gene flow and continuous interaction amongst the subpopulations which will then diffuse the signatures of selection based on coat colour as was expected in this study. Observations has also shown that farmers and even research flocks often have multiple coat colour sheep raised on the same farm.

Conclusion

This study provided insight into the genomic architecture of Swakara sheep sub-populations, and the relationship shared with their presumed ancestors. Low levels of genetic diversity were observed in the overall population with the highest inbreeding levels being observed in the founding breeds of Namaqua Afrikaner, Blackhead Persian, and Karakul. Genetic variation in the Swakara varied, but the White and Black Swakara consistently displayed low variation amongst the sub-populations. Moderate genetic similarities between the founding breeds and Swakara were seen, but the obvious divergence witnessed was heavily influenced by selection pressures and progressive breeding goals. The varying levels of inbreeding observed in the White and Grey/Black groups points towards the existence of individuals experiencing the subvital disorder, additionally the observation of some long ROHs indicates recent inbreeding and all this may have led to the display of a recessive lethal disorder. The observed low genetic diversity, coupled with the high inbreeding levels and observation of pathways possibly linked to the occurrences of the disorder implies that intense selection and crossbreeding subjected to Swakara has resulted in the prevalence of disorders in the Brown, Grey, and White sheep.

7 STUDY CHALLENGES AND RECOMMENDATIONS

7.1.1 This study consisted of data from a previous study which consisted of a small sample size. This work intended to build on it with added sample numbers. However, the numbers utilised were still limited due to lack of availability of sampling sites. Furthermore, the lack of records and other essential information from the previous study prevented the continuation of other crucial analyses such as the screening of Copy Number Variants thus restricting the findings of this study that may have better elucidated on the study objectives. The Swakara breed is a largely understudied breed. Obtaining essential information in order to fully characterise the breed as well as make reasonable inferences on the obtained results was a challenge due to the lack of studies/information on this breed. Much of the information available is predominantly from the eras of the early and late 90's which may render it outdated based on the change in technologies of generating data.

7.1.2 It is recommended that improved management techniques are applied for the conservation of a diverse genetic pool in Swakara sheep as well as the Namaqua Afrikaner and Blackhead Persian. The low genetic diversity displayed in the study highlights the major risk of more increased inbreeding and complete loss of genetic variation in these breeds. This study had a limited animal genotype data which was largely owed to the generally low numbers of the Swakara breed. The implementation of breeding strategies will not only help farmers better breed for high quality pelts, but also increase population numbers. Disorders in Swakara sheep are still a major issue for farmers and it is recommended that more research be conducted on the direct mechanism of these disorders and how they can be eradicated completely from the population.

8 REFERENCES

ABDOLI, R., MIRHOSEINI, S. Z., GHAVI HOSSEIN-ZADEH, N., ZAMANI, P., MORADI, M. H., FERDOSI, M. H., SARGOLZAEI, M. & GONDRO, C. 2023. Runs of homozygosity and cross-generational inbreeding of Iranian fat-tailed sheep. *Heredity,* 130**,** 358-367.

ABIED, A., BAGADI, A., BORDBAR, F., PU, Y., AUGUSTINO, S. M., XUE, X., XING, F., GEBRESELASSIE, G., HAN, J.-L. & MWACHARO, J. M. 2020a. Genomic diversity, population structure, and signature of selection in five Chinese native sheep breeds adapted to extreme environments. *Genes,* 11**,** 494.

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. 2020b. Genome-Wide Analysis Revealed Homozygosity and Demographic History of Five Chinese Sheep Breeds Adapted to Different Environments. *Genes,* 11**,** 1480.

ADALSTEINSSON, S., LAUVERGNE, J., BOYAZOGLU, J. & RYDER, M. A possible genetic interpretation of the colour variants in the fleece of the Gotland and Goth sheep. Annales de Génétique et de Sélection animale, 1978. EDP Sciences, 329-342.

ADALSTEINSSONS, S. J. P. W. C. S. & BREED, B. C. 1980. Inheritance of colours in sheep.

ADDO, S., KLINGEL, S., THALLER, G. & HINRICHS, D. 2021. Genetic diversity and the application of runs of homozygosity-based methods for inbreeding estimation in German White-headed Mutton sheep. *PLOS ONE,* 16**,** e0250608.

AHBARA, A. & LATAIRISH, S. A genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation.

AL-MAMUN, H. A., KWAN, P., CLARK, S. A., FERDOSI, M. H., TELLAM, R. & GONDRO, C. 2015. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genetics Selection Evolution,* 47**,** 66.

ALBA-HURTADO, F. & MUÑOZ-GUZMÁN, M. A. 2013. Immune responses associated with resistance to haemonchosis in sheep. *BioMed research international,* 2013.

ALBERTO, F. J., BOYER, F., OROZCO-TERWENGEL, P., STREETER, I., SERVIN, B., DE VILLEMEREUIL, P., BENJELLOUN, B., LIBRADO, P., BISCARINI, F. & COLLI, L. J. N. C. 2018. Convergent genomic signatures of domestication in sheep and goats. 9**,** 813.

ALEXANDER, D. H. & LANGE, K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics,* 12**,** 246.

ALEXANDER, D. H., NOVEMBRE, J. & LANGE, K. J. G. R. 2009. Fast model-based estimation of ancestry in unrelated individuals. 19**,** 1655-1664.

AVILA, F., MICKELSON, J. R., SCHAEFER, R. J. & MCCUE, M. E. 2018. Genome-Wide Signatures of Selection Reveal Genes Associated With Performance in American Quarter Horse Subpopulations. *Frontiers in genetics,* 9**,** 249-249.

BALLINGALL, K. T., LANTIER, I., TODD, H., LANTIER, F. & ROCCHI, M. 2018. Structural and functional diversity arising from intra-and inter-haplotype combinations of duplicated DQA and B loci within the ovine MHC. *Immunogenetics,* 70**,** 257-269.

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**,** 109.

BEDHIAF-ROMDHANI, S., BAAZAOUI, I., CIANI, E., MASTRANGELO, S. & SASSI, M. B. 2020. Genetic structure of Tunisian sheep breeds as inferred from genome-wide SNP markers. *Small Ruminant Research,* 191**,** 106192.

BENNETT, D. C. & LAMOREUX, M. L. J. P. C. R. 2003. The color loci of mice–a genetic century. 16**,** 333-344.

BIÉNABE, E., KIRSTEN, J. & BRAMLEY, C. 2013. Collective action dynamics and product reputation. *Developing Geographical Indications in the South: The Southern African Experience***,** 51-72.

BIRD, C. E., KARL, S. A., SMOUSE, P. E., TOONEN, R. J. J. P. & CRUSTACEA, P. G. I. 2011. Detecting and measuring genetic differentiation. 19**,** 31-55.

BISCARINI, F., COZZI, P., GASPA, G. & MARRAS, G. 2018. detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes.

BOISSY, R. E., ZHAO, H., OETTING, W. S., AUSTIN, L. M., WILDENBERG, S. C., BOISSY, Y. L., ZHAO, Y., STURM, R. A., HEARING, V. J. & KING, R. A. J. A. J. O. H. G. 1996. Mutation in and lack of expression of tyrosinase-related protein-1 (TRP-1) in melanocytes from an individual with brown oculocutaneous albinism: a new subtype of albinism classified as" OCA3". 58**,** 1145.

BUDURAM, P. 2004. *Genetic characterization of Southern African sheep breeds using DNA markers.* University of the Free State.

BURGER, A. 2015. *Meat quality of indigenous fat-tail Namaqua Afrikaner, Dorper and the South African Mutton Merino breeds.* Stellenbosch: Stellenbosch University.

CAMPBELL, L. J. 2007. *Evaluation of two indigenous South African sheep breeds as pelt producers.* University of Pretoria.

CEBALLOS, F. C., JOSHI, P. K., CLARK, D. W., RAMSAY, M. & WILSON, J. F. J. N. R. G. 2018. Runs of homozygosity: windows into population history and trait architecture. 19**,** 220-234.

CHEN, Z.-H., XU, Y.-X., XIE, X.-L., WANG, D.-F., AGUILAR-GÓMEZ, D., LIU, G.-J., LI, X., ESMAILIZADEH, A., REZAEI, V. & KANTANEN, J. J. C. B. 2021a. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. 4**,** 1-15.

CHEN, Z.-H., XU, Y.-X., XIE, X.-L., WANG, D.-F., AGUILAR-GÓMEZ, D., LIU, G.-J., LI, X., ESMAILIZADEH, A., REZAEI, V. & KANTANEN, J. J. C. B. 2021b. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. 4**,** 1307.

CHENG, H., ZHANG, Z., WEN, J., LENSTRA, J. A., HELLER, R., CAI, Y., GUO, Y., LI, M., LI, R. & LI, W. J. B. 2022. Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep. 2022.05. 17.492311.

CIESLAK, M., REISSMANN, M., HOFREITER, M. & LUDWIG, A. 2011. Colours of domestication. 86**,** 885-899.

CLOP, A., VIDAL, O. & AMILLS, M. J. A. G. 2012. Copy number variation in the genomes of domestic animals. 43**,** 503-517.

CORTES-HERNÁNDEZ, J., GARCÍA-RUIZ, A., VÁSQUEZ-PELÁEZ, C. G. & RUIZ-LOPEZ, F. D. J. 2021. Correlation of Genomic and Pedigree Inbreeding Coefficients in Small Cattle Populations. *Animals,* 11**,** 3234.

DE SIMONI GOUVEIA, J. J., PAIVA, S. R., MCMANUS, C. M., CAETANO, A. R., KIJAS, J. W., FACÓ, O., AZEVEDO, H. C., DE ARAUJO, A. M., DE SOUZA, C. J. H., YAMAGISHI, M. E. B., CARNEIRO, P. L. S., BRAGA LÔBO, R. N., DE OLIVEIRA, S. M. P. & DA SILVA, M. V. G. B. 2017. Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. *Livestock Science,* 197**,** 36-45.

DENG, T. X., MA, X. Y., LU, X. R., DUAN, A. Q., SHOKROLLAHI, B. & SHANG, J. H. 2022. Signatures of selection reveal candidate genes involved in production traits in Chinese crossbred buffaloes. *Journal of Dairy Science,* 105**,** 1327-1337.

DI GREGORIO, P., PERNA, A., DI TRANA, A. & RANDO, A. 2023. Identification of ROH Islands Conserved through Generations in Pigs Belonging to the Nero Lucano Breed. *Genes,* 14**,** 1503.

DZOMBA, E., CHIMONYO, M., SNYMAN, M. & MUCHADEYI, F. J. A. G. 2020. The genomic architecture of South African mutton, pelt, dual-purpose and nondescript sheep breeds relative to global sheep populations. 51**,** 910-923.

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.

EDEA, Z., DESSIE, T., DADI, H., DO, K.-T. & KIM, K.-S. 2017. Genetic Diversity and Population Structure of Ethiopian Sheep Populations Revealed by High-Density SNP Markers. *Frontiers in Genetics,* 8.

ELHAIK, E. 2022. Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports,* 12**,** 14683.

ELHAIK, E., TATARINOVA, T., CHEBOTAREV, D., PIRAS, I. S., MARIA CALÒ, C., DE MONTIS, A., ATZORI, M., MARINI, M., TOFANELLI, S. & FRANCALACCI, P. 2014. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nature communications,* 5**,** 3513.

ENGELHARDT, B. E. & STEPHENS, M. 2010. Analysis of Population Structure: A Unifying Framework and Novel Methods Based on Sparse Factor Analysis. *PLOS Genetics,* 6**,** e1001117.

EXCOFFIER, L., LAVAL, G. & SCHNEIDER, S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics,* 1**,** 117693430500100003.

EYDIVANDI, S., ROUDBAR, M. A., ARDESTANI, S. S., MOMEN, M. & SAHANA, G. 2021. A selection signatures study among Middle Eastern and European sheep breeds. *Journal of Animal Breeding and Genetics,* 138**,** 574-588.

FARIELLO, M.-I., SERVIN, B., TOSSER-KLOPP, G., RUPP, R., MORENO, C., INTERNATIONAL SHEEP GENOMICS, C., CRISTOBAL, M. S. & BOITARD, S. 2014. Selection Signatures in Worldwide Sheep Populations. *PLOS ONE,* 9**,** e103813.

FARIELLO, M. I., BOITARD, S., NAYA, H., SANCRISTOBAL, M. & SERVIN, B. 2013. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. *Genetics,* 193**,** 929-941.

FERENČAKOVIĆ, M., SÖLKNER, J. & CURIK, I. 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution,* 45**,** 1-9.

FILALI, H., MARTÍN-BURRIEL, I., HARDERS, F., VARONA, L., HEDMAN, C., MEDIANO, D. R., MONZÓN, M., BOSSERS, A., BADIOLA, J. J. & BOLEA, R. 2014. Gene expression profiling of mesenteric lymph nodes from sheep with natural scrapie. *BMC genomics,* 15**,** 1-17.

FLEMING, E. 2009. 8. Economics of Sheep Production.

FLORI, L., MOAZAMI-GOUDARZI, K., ALARY, V., ARABA, A., BOUJENANE, I., BOUSHABA, N., CASABIANCA, F., CASU, S., CIAMPOLINI, R., COEUR D'ACIER, A., COQUELLE, C., DELGADO, J.-V., EL-BELTAGI, A., HADJIPAVLOU, G., JOUSSELIN, E., LANDI, V., LAUVIE, A., LECOMTE, P., LIGDA, C., MARINTHE, C., MARTINEZ, A., MASTRANGELO, S., MENNI, D., MOULIN, C.-H., OSMAN, M.-A., PINEAU, O., PORTOLANO, B., RODELLAR, C., SAÏDI-MEHTAR, N., SECHI, T., SEMPÉRÉ, G., THÉVENON, S., TSIOKOS, D., LALOË, D. & GAUTIER, M. 2019. A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology,* 28**,** 1009-1029.

FONTANESI, L., BERETTI, F., RIGGIO, V., GONZÁLEZ, E. G., DALL'OLIO, S., DAVOLI, R., RUSSO, V., PORTOLANO, B. J. C. & RESEARCH, G. 2009. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. 126**,** 333-347.

FONTANESI, L., DALL'OLIO, S., BERETTI, F., PORTOLANO, B. & RUSSO, V. 2011. Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) genes. *Animal,* 5**,** 8-17.

FORUTAN, M., ANSARI MAHYARI, S., BAES, C., MELZER, N., SCHENKEL, F. S. & SARGOLZAEI, M. 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics,* 19**,** 98.

GORSSEN, W., MEYERMANS, R., JANSSENS, S. & BUYS, N. 2021. A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genetics Selection Evolution,* 53**,** 1-10.

GRATTEN, J., BERALDI, D., LOWDER, B., MCRAE, A., VISSCHER, P., PEMBERTON, J. & SLATE, J. J. P. O. T. R. S. B. B. S. 2007. Compelling evidence that a single nucleotide substitution in TYRP1 is

responsible for coat-colour polymorphism in a free-living population of Soay sheep. 274**,** 619-626.
GREEFF, J., FAURE, A., MINNAAR, G. & SCHOEMAN, S. J. S. A. J. O. A. S. 1991. Non-genetic factors affecting pelt traits in Karakul sheep. 21**,** 173-178.
GREEFF, J., HOFMEYR, J., LOURENS, D. & VAN RENSBURG, R. J. Preliminary studies on colour inheritance and production traits in Karakul crossbred sheep. Proceedings of the 2nd World Congress on Sheep and Beef Cattle Breeding, 16-19 April 1984, Pretoria, South Africa., 1984. South African Stud Book and Livestock Improvement Association, 381-392.
HAN, J., YANG, M., GUO, T., NIU, C., LIU, J., YUE, Y., YUAN, C. & YANG, B. 2019. Two linked TBXT (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Animal Genetics,* 50**,** 772-777.
HAN, J.-L., YANG, M., GUO, T.-T., YUE, Y.-J., LIU, J.-B., NIU, C.-E., WANG, C.-F. & YANG, B.-H. 2015. Molecular characterization of two candidate genes associated with coat color in Tibetan sheep (Ovis arise). *Journal of Integrative Agriculture,* 14**,** 1390-1397.
HAN, Z., ZHOU, W., ZHANG, L., WANG, R., LIU, C., BAI, X. & LIU, S. 2022. Genetic Diversity and Runs of homozygosity analysis of Hetian Sheep Populations Revealed by Illumina OvineSNP50 BeadChip.
HEWETT, A. M., STOFFEL, M. A., PETERS, L., JOHNSTON, S. E. & PEMBERTON, J. M. 2023. Selection, recombination and population history effects on runs of homozygosity (ROH) in wild red deer (Cervus elaphus). *Heredity,* 130**,** 242-250.
HIRSCHHORN, J. N. & DALY, M. J. J. N. R. G. 2005. Genome-wide association studies for common diseases and complex traits. 6**,** 95-108.
HLONGWANE, N. L. 2022. *Genome-wide characterization of South African pig breeds.*
HOEKSTRA, H. E. & NACHMAN, M. W. J. M. E. 2003. Different genes underlie adaptive melanism in different populations of rock pocket mice. 12**,** 1185-1194.
IBRAGIMOV, Y., SVITOJUS, A., YUSUPOV, S. & BALTRENAITE, L. 2007. Management, use and conservation of Karakul sheep in traditional livestock farming systems in Uzbekistan. *People and animals, traditional livestock keepers: guardians of domestic animal diversity***,** 103-109.
ISLAM, R., LI, Y., LIU, X., BERIHULAY, H., ABIED, A., GEBRESELASSIE, G., MA, Q. & MA, Y. 2019. Genome-Wide Runs of Homozygosity, Effective Population Size, and Detection of Positive Selection Signatures in Six Chinese Goat Breeds. *Genes,* 10**,** 938.
ITENGE, T. & SHIPANDENI, M. 2015. Sale trends of Swakara pelt offered at the Copenhagen Fur Auction from 1994-2013. *Applied Animal Husbandry & Rural Development,* 8**,** 1-5.
JACKSON, P. J., DOUGLAS, N. R., CHAI, B., BINKLEY, J., SIDOW, A., BARSH, G. S. & MILLHAUSER, G. L. 2006. Structural and Molecular Evolutionary Analysis of Agouti and Agouti-Related Proteins. *Chemistry & Biology,* 13**,** 1297-1305.
JIANG, J., CAO, Y., SHAN, H., WU, J., SONG, X. & JIANG, Y. 2021. The GWAS Analysis of Body Size and Population Verification of Related SNPs in Hu Sheep. *Front Genet,* 12**,** 642552.
JOHNSTON, S. E., MCEWAN, J. C., PICKERING, N. K., KIJAS, J. W., BERALDI, D., PILKINGTON, J. G., PEMBERTON, J. M. & SLATE, J. J. M. E. 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. 20**,** 2555-2566.
JUAN, W. C. & HONG, W. 2016. Targeting the Hippo Signaling Pathway for Tissue Regeneration and Cancer Therapy. *Genes,* 7**,** 55.
KIJAS, J. W., LENSTRA, J. A., HAYES, B., BOITARD, S., PORTO NETO, L. R., SAN CRISTOBAL, M., SERVIN, B., MCCULLOCH, R., WHAN, V. & GIETZEN, K. J. P. B. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. 10**,** e1001258.
KIJAS, J. W., TOWNLEY, D., DALRYMPLE, B. P., HEATON, M. P., MADDOX, J. F., MCGRATH, A., WILSON, P., INGERSOLL, R. G., MCCULLOCH, R., MCWILLIAM, S., TANG, D., MCEWAN, J., COCKETT, N., ODDY, V. H., NICHOLAS, F. W., RAADSMA, H. & FOR THE INTERNATIONAL SHEEP GENOMICS,

C. 2009. A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLOS ONE,* 4**,** e4668.

KIM, E.-S., COLE, J. B., HUSON, H., WIGGANS, G. R., VAN TASSELL, C. P., CROOKER, B. A., LIU, G., DA, Y. & SONSTEGARD, T. S. 2013. Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PloS one,* 8**,** e80813.

KIM, E.-S., ELBELTAGY, A., ABOUL-NAGA, A., 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.

KIRIN, M., MCQUILLAN, R., FRANKLIN, C. S., CAMPBELL, H., MCKEIGUE, P. M. & WILSON, J. F. J. P. O. 2010. Genomic runs of homozygosity record population history and consanguinity. 5**,** e13996.

KIRSTEN, G. 1966. The Production and Marketing of Karakul Pelts. *Agrekon,* 5**,** 22-29.

KOSENIUK, A., ROPKA-MOLIK, K., RUBI'S, D. & SMOŁUCHA, G. J. A. A. B. 2018. Genetic background of coat colour in sheep. 61**,** 173-178.

LI, Y., CHEN, Z., FANG, Y., CAO, C., ZHANG, Z., PAN, Y. & WANG, Q. 2022. Runs of Homozygosity Revealed Reproductive Traits of Hu Sheep. *Genes,* 13**,** 1848.

LISCHER, H. E. & EXCOFFIER, L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics,* 28**,** 298-299.

LIU, J., SHI, L., LI, Y., CHEN, L., GARRICK, D., WANG, L. & ZHAO, F. 2021. Estimates of genomic inbreeding and identification of candidate regions that differ between Chinese indigenous sheep breeds. *Journal of Animal Science and Biotechnology,* 12**,** 95.

LIU, S.-H., MA, X.-Y., HASSAN, F.-U., GAO, T.-Y. & DENG, T.-X. 2022. Genome-wide analysis of runs of homozygosity in Italian Mediterranean buffalo. *Journal of Dairy Science,* 105**,** 4324-4334.

LOPEZ, B. I., LEE, S.-H., SHIN, D.-H., OH, J.-D., CHAI, H.-H., PARK, W., PARK, J.-E. & LIM, D. 2020. Accuracy of genomic evaluation using imputed high-density genotypes for carcass traits in commercial Hanwoo population. *Livestock Science,* 241**,** 104256.

LUNDIE, R. J. T. W. O. C. S. T. B. & COLOURED SHEEP BREEDER'S ASSOCIATION OF NEW ZEALAND, T. 2004. The genetics of colour in sheep-some basics. 111-122.

LUNDIE, R. S. 2011. The genetics of colour in fat-tailed sheep: a review. *Tropical Animal Health and Production,* 43**,** 1245-1265.

MA, Y., WEI, J., ZHANG, Q., CHEN, L., WANG, J., LIU, J. & DING, X. 2015. A genome scan for selection signatures in pigs. *PLoS One,* 10**,** e0116850.

MACHETE, J. B., KGWATALALA, P. M., NSOSO, S. J., HLONGWANE, N. L. & MOREKI, J. C. J. O. J. O. A. S. 2021. Genetic Diversity and Population Structure of Three Strains of Indigenous Tswana Chickens and Commercial Broiler Using Single Nucleotide Polymormophic (SNP) Markers. 11**,** 515-531.

MAIORANO, A. M., CARDOSO, D. F., CARVALHEIRO, R., JÚNIOR, G. A. F., DE ALBUQUERQUE, L. G. & DE OLIVEIRA, H. N. 2022. Signatures of selection in Nelore cattle revealed by whole-genome sequencing data. *Genomics,* 114**,** 110304.

MALESA, M. T. 2015. *Population genetics of Swakara sheep inferred using genome-wide SNP genotyping.*

MANOLIO, T. A. J. N. E. J. O. M. 2010. Genomewide association studies and assessment of the risk of disease. 363**,** 166-176.

MANZARI, Z., MEHRABANI-YEGANEH, H., NEJATI-JAVAREMI, A., MORADI, M. H. & GHOLIZADEH, M. 2019. Detecting selection signatures in three Iranian sheep breeds. *Animal Genetics,* 50**,** 298-302.

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. *Animal Genetics,* 46**,** 110-121.

MARTINS, C. & PETERS, K. J. 1992a. Alternative use of Karakul sheep for pelt and lamb production in Botswana. I. Reproduction and growth performance. *Small Ruminant Research,* 9**,** 1-10.

MARTINS, C. & PETERS, K. J. 1992b. Alternative use of Karakul sheep for pelt and lamb production in Botswana. II. Pelt production. *Small Ruminant Research,* 9**,** 11-19.

MASTRANGELO, S., CIANI, E., SARDINA, M. T., SOTTILE, G., PILLA, F., PORTOLANO, B. & CONSORTIUM, T. B. O. I. 2018a. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. 49**,** 71-81.

MASTRANGELO, S., CRISCIONE, A., SOTTILE, G., PORTOLANO, B., MARLETTA, D. & BORDONARO, S. 2019. Genome-wide analysis identifies potentially causative genes explaining the phenotypic variability in Pinzirita sheep. *Anim. Genet,* 50**,** 189-190.

MASTRANGELO, S., SARDINA, M. T., TOLONE, M., DI GERLANDO, R., SUTERA, A. M., FONTANESI, L. & PORTOLANO, B. 2018b. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *animal,* 12**,** 2480-2488.

MASTRANGELO, S., TOLONE, M., SARDINA, M. T., SOTTILE, G., SUTERA, A. M., DI GERLANDO, R. & PORTOLANO, B. 2017. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution,* 49**,** 84.

MCCARTHY, M. I., ABECASIS, G. R., CARDON, L. R., GOLDSTEIN, D. B., LITTLE, J., IOANNIDIS, J. P. & HIRSCHHORN, J. N. J. N. R. G. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. 9**,** 356-369.

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. 2008. Runs of homozygosity in European populations. *The American Journal of Human Genetics,* 83**,** 359-372.

MCRAE, K. M., MCEWAN, J. C., DODDS, K. G. & GEMMELL, N. J. 2014. Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genomics,* 15**,** 637.

MEUWISSEN, T. 2009. Genetic management of small populations: A review. *Acta Agriculturae Scand Section A,* 59**,** 71-79.

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**,** 94.

MILLER, J., POISSANT, J., KIJAS, J. & COLTMAN, D. 2010. A genome wide set of SNP detects population substructure and long range linkage disequilibrium in wild sheep.

MOIOLI, B., PILLA, F. & CIANI, E. 2015. Signatures of selection identify loci associated with fat tail in sheep. *J Anim Sci,* 93**,** 4660-9.

MOISÁ, S. J., SHIKE, D. W., SHOUP, L., RODRIGUEZ-ZAS, S. L. & LOOR, J. J. J. P. O. 2015. Maternal plane of nutrition during late gestation and weaning age alter Angus× Simmental offspring longissimus muscle transcriptome and intramuscular fat. 10**,** e0131478.

MOLOTSI, A. H., TAYLOR, J. F., CLOETE, S. W. P., MUCHADEYI, F., DECKER, J. E., WHITACRE, L. K., SANDENBERGH, L. & DZAMA, K. 2017. Genetic diversity and population structure of South African smallholder farmer sheep breeds determined using the OvineSNP50 beadchip. *Tropical Animal Health and Production,* 49**,** 1771-1777.

MUCHADEYI, F., MALESA, M., SOMA, P. & DZOMBA, E. Runs of homozygosity in Swakara pelt producing sheep: implications on sub-vital performance. Proc. Assoc. Advmt. Anim. Breed. Genet, 2015. 310-313.

MUHAGHEGH, D. M. & HABIBIZAD, J. 2014. Sequence characterization of promoter region at the melanocortin-1 receptor (MC1R) gene in karakul sheep breed.

MUIGAI, A. W. T. & HANOTTE, O. 2013. The Origin of African Sheep: Archaeological and Genetic Perspectives. *African Archaeological Review,* 30**,** 39-50.

MUSAVI, S. A. A., KHADIMIYAN, A. M. & AZIMI, A. M. 2022. Morphological Characterization of Karakul Sheep in North Part of Afghanistan. *Voice of the Publisher,* 8**,** 16-25.

NÄSHOLM, A. 2008. Genetic relationships between pelt quality, maternal ability, and lamb production in the Gotland sheep breed. *Livestock Science,* 117**,** 93-100.

NÄSHOLM, A. & EYTHORSDOTTIR, E. 2011. Characteristics and utilization of sheep pelts. *Small Ruminant Research,* 101**,** 182-187.

NAVAL-SANCHEZ, M., NGUYEN, Q., MCWILLIAM, S., PORTO-NETO, L. R., TELLAM, R., VUOCOLO, T., REVERTER, A., PEREZ-ENCISO, M., BRAUNING, R. & CLARKE, S. J. N. C. 2018. Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. 9**,** 859.

NORRIS, B. J. & WHAN, V. A. 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res,* 18**,** 1282-93.

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.

NSOSO, S. & MADIMABE, M. 1999. The sheep industry in Botswana: promoting the Karakul sheep industry. *South African Journal of Animal Science,* 29.

NSOSO, S. & MADIMABE, M. 2003. A survey of Karakul sheep farmers in Southern Kalahari, Botswana: management practices and constraints to improving production. *South African Journal of Animal Science,* 4**,** 23-27.

OLIVEIRA, J., PEREIRA, R., SANTOS, R. & SOUSA, M. Evaluating Runs of Homozygosity in Exome Sequencing Data - Utility in Disease Inheritance Model Selection and Variant Filtering. 2018 Cham. Springer International Publishing, 268-288.

ONZIMA, R. B., UPADHYAY, M. R., DOEKES, H. P., BRITO, L. F., BOSSE, M., KANIS, E., GROENEN, M. A. M. & CROOIJMANS, R. P. M. A. 2018. Genome-Wide Characterization of Selection Signatures and Runs of Homozygosity in Ugandan Goat Breeds. *Frontiers in Genetics,* 9.

PANIGRAHI, M., RAJAWAT, D., NAYAK, S. S., GHILDIYAL, K., SHARMA, A., JAIN, K., LEI, C., BHUSHAN, B., MISHRA, B. P. & DUTT, T. 2023. Landmarks in the history of selective sweeps. *Animal Genetics,* 54**,** 667-688.

PARIS, J. M., LETKO, A., HÄFLIGER, I. M., AMMANN, P. & DRÖGEMÜLLER, C. 2020. Ear type in sheep is associated with the MSRB3 locus. *Animal genetics,* 51**,** 968-972.

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. *The American Journal of Human Genetics,* 91**,** 275-292.

PERIPOLLI, E., STAFUZZA, N. B., MUNARI, D. P., LIMA, A. L. F., IRGANG, R., MACHADO, M. A., PANETTO, J. C. D. C., VENTURA, R. V., BALDI, F. & DA SILVA, M. V. G. B. 2018. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. *BMC genomics,* 19**,** 1-13.

POSBERGH, C. & HUSON, H. 2018. *Making Moorit: Mutations in TYRP1 are responsible for brown coat color in different United States sheep breeds*.

POSBERGH, C. & HUSON, H. J. A. G. 2021. All sheeps and sizes: a genetic investigation of mature body size across sheep breeds reveals a polygenic nature. 52**,** 99-107.

PRIEUR, V., CLARKE, S. M., BRITO, L. F., MCEWAN, J. C., LEE, M. A., BRAUNING, R., DODDS, K. G. & AUVRAY, B. 2017. Estimation of linkage disequilibrium and effective population size in New Zealand sheep using three different methods to create genetic maps. *BMC Genetics,* 18**,** 68.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. & DALY, M. J. J. T. A. J. O. H. G. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. 81**,** 559-575.

PURFIELD, D. C., BERRY, D. P., MCPARLAND, S. & BRADLEY, D. G. 2012. Runs of homozygosity and population history in cattle. *BMC Genetics,* 13**,** 70.

PURFIELD, D. C., EVANS, R. D., CARTHY, T. R. & BERRY, D. P. J. F. I. G. 2019. Genomic regions associated with gestation length detected using whole-genome sequence data differ between dairy and beef cattle. 10**,** 1068.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

REED, D. H., LOWE, E. H., BRISCOE, D. A. & FRANKHAM, R. 2003. Inbreeding and extinction: Effects of rate of inbreeding. *Conservation Genetics,* 4**,** 405-410.

RETIEF, A. 2020. *Whole genome investigation of the genetic structure of South African sheep breeds.* University of Pretoria.

ROBERTS, A. J. L. S. 2018. Genome-wide association study for carcass traits in a composite beef cattle breed. 213**,** 35-43.

ROBERTS, J. F. & JENKINS, T. J. T. W. J. O. A. 1926. A preliminary note on reversed badger-face pattern in sheep. 2**,** 70-73.

ROBERTS, J. F. J. J. O. G. 1924. Colour inheritance in sheep. 14**,** 367-374.

ROCHUS, C., WESTBERG SUNESSON, K., JONAS, E., MIKKO, S. & JOHANSSON, A. J. A. G. 2019. Mutations in ASIP and MC1R: dominant black and recessive black alleles segregate in native Swedish sheep populations. 50**,** 712-717.

ROMERO, R. D., MONTERO PARDO, A., MONTALDO, H. H., RODRÍGUEZ, A. D. & HERNÁNDEZ CERÓN, J. 2013. Differences in body temperature, cell viability, and HSP-70 concentrations between Pelibuey and Suffolk sheep under heat stress. *Tropical Animal Health and Production,* 45**,** 1691-1696.

ROSTAMZADEH-MAHDABI, E., ESMAILIZADEH, A., AYATOLLAHI MEHRGARDI, A. & ASADI FOZI, M. 2021. A genome-wide scan to identify signatures of selection in two Iranian indigenous chicken ecotypes. *Genetics Selection Evolution,* 53**,** 72.

ROY, A. & MATZUK, M. M. J. R. 2006. Deconstructing mammalian reproduction: using knockouts to define fertility pathways. 131**,** 207-219.

RUBIN, C.-J., MEGENS, H.-J., BARRIO, A. M., MAQBOOL, K., SAYYAB, S., SCHWOCHOW, D., WANG, C., CARLBORG, Ö., JERN, P., JØRGENSEN, C. B., ARCHIBALD, A. L., FREDHOLM, M., GROENEN, M. A. M. & ANDERSSON, L. 2012. Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences,* 109**,** 19529-19536.

SABETI, P. C., REICH, D. E., HIGGINS, J. M., LEVINE, H. Z., RICHTER, D. J., SCHAFFNER, S. F., GABRIEL, S. B., PLATKO, J. V., PATTERSON, N. J. & MCDONALD, G. J. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature,* 419**,** 832-837.

SABETI, P. C., VARILLY, P., FRY, B., LOHMUELLER, J., HOSTETTER, E., COTSAPAS, C., XIE, X., BYRNE, E. H., MCCARROLL, S. A., GAUDET, R., SCHAFFNER, S. F., LANDER, E. S., FRAZER, K. A., BALLINGER, D. G., COX, D. R., HINDS, D. A., STUVE, L. L., GIBBS, R. A., BELMONT, J. W., BOUDREAU, A., HARDENBOL, P., LEAL, S. M., PASTERNAK, S., WHEELER, D. A., WILLIS, T. D., YU, F., YANG, H., ZENG, C., GAO, Y., HU, H., HU, W., LI, C., LIN, W., LIU, S., PAN, H., TANG, X., WANG, J., WANG, W., YU, J., ZHANG, B., ZHANG, Q., ZHAO, H., ZHAO, H., ZHOU, J., GABRIEL, S. B., BARRY, R., BLUMENSTIEL, B., CAMARGO, A., DEFELICE, M., FAGGART, M., GOYETTE, M., GUPTA, S., MOORE, J., NGUYEN, H., ONOFRIO, R. C., PARKIN, M., ROY, J., STAHL, E., WINCHESTER, E., ZIAUGRA, L., ALTSHULER, D., SHEN, Y., YAO, Z., HUANG, W., CHU, X., HE, Y., JIN, L., LIU, Y., SHEN, Y., SUN, W., WANG, H., WANG, Y., WANG, Y., XIONG, X., XU, L., WAYE, M. M. Y., TSUI, S. K. W., XUE, H., TZE-FEI WONG, J., GALVER, L. M., FAN, J.-B., GUNDERSON, K., MURRAY, S. S., OLIPHANT, A. R., CHEE, M. S., MONTPETIT, A., CHAGNON, F., FERRETTI, V., LEBOEUF, M., OLIVIER, J.-F., PHILLIPS, M. S., ROUMY, S., SALLÉE, C., VERNER, A., HUDSON, T. J., KWOK, P.-Y., CAI, D., KOBOLDT, D. C., MILLER, R. D., PAWLIKOWSKA, L., et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature,* 449**,** 913-918.

SÁNCHEZ-RAMOS, R., TRUJANO-CHAVEZ, M. Z., GALLEGOS-SÁNCHEZ, J., BECERRIL-PÉREZ, C. M., CADENA-VILLEGAS, S. & CORTEZ-ROMERO, C. 2023. Detection of Candidate Genes Associated with Fecundity through Genome-Wide Selection Signatures of Katahdin Ewes. *Animals,* 13**,** 272.

SANDENBERGH, L., CLOETE, S., ROODT-WILDING, R., BESTER-VAN DER MERWE, A. & SNYMAN, M. 2016. Evaluation of the OvineSNP50 chip for use in four South African sheep breeds. *South African Journal of Animal Science,* 46**,** 89-93.

SARAVANAN, K. A., PANIGRAHI, M., KUMAR, H., PARIDA, S., BHUSHAN, B., GAUR, G. K., DUTT, T., MISHRA, B. P. & SINGH, R. K. 2021. Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. *Genomics,* 113**,** 955-963.

SCHOEMAN, S., GROENEVELD, H. & ALBERTYN, J. R. J. S. A. J. O. A. S. 1993. Factors influencing the quality of Karakul pelts, with emphasis on discrete characteristics. 23**,** 183-186.

SCHOEMAN, S. J. S. A. J. O. A. S. 1998. Genetic and environmental factors influencing the quality of pelt traits in Karakul sheep. 28.

SCHROEDER, A., SAMUELS, M. I., SWARTS, M., MORRIS, C., CUPIDO, C. F. & ENGELBRECHT, A. 2019. Diet selection and preference of small ruminants during drought conditions in a dryland pastoral system in South Africa. *Small Ruminant Research,* 176**,** 17-23.

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**,** 2696.

SERRANITO, B., TAURISSON-MOURET, D., HARKAT, S., LAOUN, A., OUCHENE-KHELIFI, N.-A., POMPANON, F., BENJELLOUN, B., CECCHI, G., THEVENON, S. & LENSTRA, J. A. J. F. I. G. 2021. Search for selection signatures related to trypanosomosis tolerance in African goats. 1305.

STAFUZZA, N. B., SILVA, R. M. D. O., FRAGOMENI, B. D. O., MASUDA, Y., HUANG, Y., GRAY, K. & LOURENCO, D. A. L. J. B. G. 2019. A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. 20**,** 1-11.

STELLA, A., AJMONE-MARSAN, P., LAZZARI, B. & BOETTCHER, P. 2010. Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. *Genetics,* 185**,** 1451-1461.

STRANGER, B. E., FORREST, M. S., DUNNING, M., INGLE, C. E., BEAZLEY, C., THORNE, N., REDON, R., BIRD, C. P., DE GRASSI, A. & LEE, C. J. S. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. 315**,** 848-853.

SUCHOCKI, T. & SZYDA, J. 2015. Genome-wide association study for semen production traits in Holstein-Friesian bulls. *Journal of Dairy Science,* 98**,** 5774-5780.

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.

TAM, V., PATEL, N., TURCOTTE, M., BOSSÉ, Y., PARÉ, G. & MEYRE, D. J. N. R. G. 2019. Benefits and limitations of genome-wide association studies. 20**,** 467-484.

TANG, H., CORAM, M., WANG, P., ZHU, X. & RISCH, N. J. T. A. J. O. H. G. 2006. Reconstructing genetic ancestry blocks in admixed individuals. 79**,** 1-12.

TIAN, Y., DU, J., YANG, X., ZENG, W., HE, J., ZHAO, B., FU, X., XU, X., WU, W., DI, J., HUANG, X. & TIAN, K. 2022. Sheep IGFBP2 and IGFBP4 promoter methylation regulates gene expression and hair follicle development. *Electronic Journal of Biotechnology,* 59**,** 46-54.

TOLONE, M., SARDINA, M. T., SENCZUK, G., CHESSARI, G., CRISCIONE, A., MOSCARELLI, A., RIGGIO, S., RIZZUTO, I., DI GERLANDO, R. & PORTOLANO, B. J. A. 2022. Genomic Tools for the Characterization of Local Animal Genetic Resources: Application in Mascaruna Goat. 12**,** 2840.

UEMATSU, S. & AKIRA, S. 2006. Toll-like receptors and innate immunity. *Journal of Molecular Medicine,* 84**,** 712-725.

USAI, M. G., CASU, S., SECHI, T., SALARIS, S. L., MIARI, S., SECHI, S., CARTA, P. & CARTA, A. J. G. S. E. 2019. Mapping genomic regions affecting milk traits in Sarda sheep by using the OvineSNP50 Beadchip and principal components to perform combined linkage and linkage disequilibrium analysis. 51**,** 1-19.

VANVANHOSSOU, S. F. U., YIN, T., SCHEPER, C., FRIES, R., DOSSA, L. H. & KÖNIG, S. J. F. I. G. 2021. Unraveling Admixture, Inbreeding, and Recent Selection Signatures in West African Indigenous Cattle Populations in Benin. 12.

VASIN, B. J. P. M. A. C. S. G. D. A. 1928. The genetics of the sheep. I. Inheritance of colour and pattern. 2.

WALUGEMBE, M., BERTOLINI, F., DEMATAWEWA, C. M. B., REIS, M. P., ELBELTAGY, A. R., SCHMIDT, C. J., LAMONT, S. J. & ROTHSCHILD, M. F. 2019. Detection of Selection Signatures Among Brazilian, Sri Lankan, and Egyptian Chicken Populations Under Different Environmental Conditions. *Frontiers in Genetics,* 9.

WANG, C., WANG, H., ZHANG, Y., TANG, Z., LI, K. & LIU, B. J. M. E. R. 2015. Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in C hinese domestic pigs. 15**,** 414-424.

WANG, J. 2022. Fast and accurate population admixture inference from genotype data from a few microsatellites to millions of SNPs. *Heredity,* 129**,** 79-92.

WILSON, R. 2011. Populations and production of fat-tailed and fat-rumped sheep in the Horn of Africa. *Tropical animal health and production,* 43**,** 1419-25.

WRIGHT, S. 1949. The genetical structure of populations. *Annals of eugenics,* 15**,** 323-354.

XIE, Y., LIU, Z., GUO, J., SU, X., ZHAO, C., ZHANG, C., QIN, Q., DAI, D., ZHAO, Y. & WANG, Z. J. F. I. G. 2021. MicroRNA-mRNA Regulatory Networking Fine-Tunes Polyunsaturated Fatty Acid Synthesis and Metabolism in the Inner Mongolia Cashmere Goat. 12**,** 649015.

XU, S.-S., GAO, L., SHEN, M. & LYU, F. 2021a. Whole-genome selective scans detect genes associated with important phenotypic traits in sheep (Ovis aries). *Frontiers in Genetics,* 12**,** 738879.

XU, Z., MEI, S., ZHOU, J., ZHANG, Y., QIAO, M., SUN, H., LI, Z., LI, L., DONG, B., OYELAMI, F. O., WU, J. & PENG, X. 2021b. Genome-Wide Assessment of Runs of Homozygosity and Estimates of Genomic Inbreeding in a Chinese Composite Pig Breed. *Frontiers in Genetics,* 12.

YAN, J., BLAIR, H. T., LIU, M., LI, W., HE, S., CHEN, L., DITTMER, K. E., GARRICK, D. J., BIGGS, P. J. & DUKKIPATI, V. S. R. 2017. Genome-wide detection of autosomal copy number variants in several sheep breeds using Illumina OvineSNP50 BeadChips. *Small Ruminant Research,* 155**,** 24-32.

YANG, H., YANG, Y.-L., LI, G.-Q., YU, Q. & YANG, J. 2021. Identifications of immune-responsive genes for adaptative traits by comparative transcriptome analysis of spleen tissue from Kazakh and Suffolk sheep. *Scientific Reports,* 11**,** 3157.

YAO, L., BAO, A., HONG, W., HOU, C., ZHANG, Z., LIANG, X. & ANIWASHI, J. J. P. 2019. Transcriptome profiling analysis reveals key genes of different coat color in sheep skin. 7**,** e8077.

YAO, X., YANG, F., EL-SAMAHY, M., LIU, B., ZHAO, B., GAO, X., ZHENG, J., FENG, X., FAN, Y. & WANG, F. J. G. 2022. Identification and characterization of unique and common lncRNAs and mRNAs in the pituitary, ovary, and uterus of Hu sheep with different prolificacy. 114**,** 110511.

YUAN, Z., LIU, E., LIU, Z., KIJAS, J. W., ZHU, C., HU, S., MA, X., ZHANG, L., DU, L., WANG, H. & WEI, C. 2017. Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Animal Genetics,* 48**,** 55-66.

ZHANG, H., WANG, Z., WANG, S. & LI, H. 2012. Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology,* 3**,** 26.

ZHANG, L., MA, X., XUAN, J., WANG, H., YUAN, Z., WU, M., LIU, R., ZHU, C., WEI, C. & ZHAO, F. J. P. O. 2016. Identification of MEF2B and TRHDE gene polymorphisms related to growth traits in a new Ujumqin sheep population. 11**,** e0159504.

ZHANG, Q., GULDBRANDTSEN, B., BOSSE, M., LUND, M. S. & SAHANA, G. J. B. G. 2015. Runs of homozygosity and distribution of functional variants in the cattle genome. 16**,** 1-16.

ZHANG, T.-T., ZHANG, G.-M., JIN, Y.-H., GUO, Y.-X., WANG, Z., FAN, Y.-X., EL-SAMAHY, M. A. & WANG, F. 2017. Energy restriction affect liver development in Hu sheep ram lambs through Hippo signaling pathway. *Tissue and Cell,* 49**,** 603-611.

ZHAO, F., MCPARLAND, S., KEARNEY, F., DU, L. & BERRY, D. P. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics Selection Evolution,* 47**,** 49.

ZHENG, X.-D., CHENG, J., QIN, W.-J., BALSAI, N., SHANG, X.-J., ZHANG, M.-T. & CHEN, H.-Q. J. F. I. G. 2020. Whole Transcriptome Analysis Identifies the Taxonomic Status of a New Chinese Native Cattle Breed and Reveals Genes Related to Body Size. 11**,** 562855.

ZHU, C., LI, N., CHENG, H. & MA, Y. 2021. Genome wide association study for the identification of genes associated with tail fat deposition in Chinese sheep breeds. *Biology Open,* 10.

9 APPENDIX

Appendix Table 1. Effective population size values of Swakara sub-populations, Namaqua Afrikaner and Blackhead Persian

Generations Ago	Sub-populations							
	Blackhead Persian	Brown Vital Swakara	Black Vital Swakara	Grey Vital Swakara	Karakul	Namaqua Afrikaner	White Subvital Swakara	White Vital Swakara
12	72	40	116	133	75	58	75	108
14	78	43	124	140	83	66	80	114
16	84	47	131	149	94	73	85	124
19	92	52	144	160	104	82	93	133
22	102	58	156	169	116	95	101	145
26	111	64	168	182	130	108	110	157
31	121	74	184	196	147	125	121	174
37	138	85	205	216	175	140	134	192
44	155	97	228	240	200	166	150	217
53	178	115	260	272	231	192	173	250
65	205	134	297	308	268	228	197	286
79	238	160	339	353	330	272	225	330
97	280	192	404	412	388	326	261	387
120	333	234	479	482	478	386	319	463
149	393	287	562	587	563	472	383	549
186	468	346	689	695	720	567	470	658
233	573	433	823	834	898	685	545	790
293	693	528	973	1001	1065	837	678	949
365	824	651	1191	1188	1343	998	808	1093
453	981	827	1359	1336	1621	1205	972	1285
552	1151	948	1599	1566	1951	1324	1114	1473
657	1278	1082	1727	1720	2286	1585	1275	1601
760	1506	1229	1839	1910	2520	1736	1373	1711
846	1588	1355	1946	2011	2719	1897	1529	1933
913	1682	1466	2200	2139	2862	2022	1557	1896
958	1772	1489	2502	2391	2861	1945	1818	2396

Appendix Table 2. Mean ROH length in Mb across 5 classes for 6 sheep clusters.

CLASS	CLUSTER POPULATION					
	A	**B**	**C**	**D**	**E**	**F**
0-5	3.950390	3.998075	3.964104	3.930538	4.250376	3.898288
5-10	7.239266	6.837794	6.802380	7.028818	7.067803	6.886443
10-20	13.957667	13.340692	14.083426	13.590193	13.943668	12.680624
20-40	26.898756	26.657960	26.331095	24.777376	25.461176	23.194656
>40	47.827118	45.657340	46.186055	41.030181	53.421566	N/A