Population genetics of Swakara sheep inferred using genome-wide SNP genotyping

by

MASOKO T. MALESA

Submitted in partial fulfilment of the academic requirements of

Master of Science

in Genetics

School of Life Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

August 2015

PREFACE

The research contained in this dissertation was completed by the candidate while based in the Discipline of Genetics, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg Campus, South Africa. The research was financially supported by NRF and ARC.

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

_____ _____

Signed: Mr EF Dzomba (Supervisor) Dr FC Muchadeyi (Co-supervisor)

Date: …… ……..

DECLARATION: PLAGIARISM

I, Masoko T. Malesa, declare that:

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

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

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

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

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

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

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

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

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

_____

Signed: Masoko T. Malesa

Date:

ABSTRACT

The pelts of Swakara sheep breed of Namibia are famous for their lustrous features and are unmatched in quality across the globe. These pelts occur in four subpopulations of white, grey, black and brown. The white pelt is the most preferred pelt colour. However, there is a challenge in producing white pelts due to the sub-vital factor that affects homozygous white lambs, thought to be a result of the negative impact of years of inbreeding and selection pressures. Sub-vital performance is a genetic disorder that causes digestive complications in lambs resulting in their death within days of birth. The ability to identify carriers or affected individuals of sub-vital performance at an early stage will save the Swakara industry production costs and reduce mortality rates in the white subpopulation. In this study, high-throughput SNP genotyping was used to perform a population genetics study, which investigates the genetic structure of Swakara with the aim of identifying structural differences between sub-vital individuals and the other coat colours, and by using GWAS to determine the variants contributing to sub-vitality in subtle effects and their associated genes in Swakara. This study will also look at selection pressures in Swakara, the genetic differentiation between subpopulations, level of inbreeding and the presence of ROH segments in Swakara and their associated genes that relate to sub-vital performance. Genetic statistical tools such as PLINK, ARLEQUIN and SVS were used to perform the analyses required for this population study. Ninety Swakara sheep were collected from Namibia and South Africa. These were sub-divided into four colour subpopulations of black (n = 16), grey (n = 22), vital white (n = 35) and sub-vital white (n = 17). The DNA from each of these was then genotyped using the OvineSNP50 beadchip. The genotyping success rate was > 93% across all the colour subpopulations. The quality control (QC) post genotyping was done by removing SNPs that departed from HWE > 0.0001, had low allele frequencies MAF < 0.01, SNP markers that had missing/low call rates (GENO) > 0.01 and individuals that had missing genotypes (MIND) > 0.1. This QC was done for every subpopulation described in this study prior to all other succeeding analyses. The inbreeding coefficient F was high in the black subpopulation (0.04798±0.069) and lowest in the grey (0.01074±0.079) Swakara sheep. The genetic diversity for the Swakara subpopulations showed a consensus in the results and showed the most diversity between the black and white subpopulations. The PCA results also

showed the genetically similar subpopulations, as identified by pairwise F_{ST} values, clustered together. Forty-two unique ROH were observed across 10 chromosomes in 33 (out of 90) individuals and spanned between 5198.93-7126.85 KB in length in the four colour subpopulations. The white sub-vital group had the highest number of ROHs (18). Seven overlapping/consensus regions of homozygosity were observed on chromosome 2, 7, 9, 10, and 20. There was no correlation between the frequency of ROHs in an individual and the level of inbreeding, however the black subpopulation had the highest level of inbreeding and had the highest average of ROH length among the other subpopulations but this was not a consistent trend with the other subpopulations. The GWAS revealed at least five SNPs associated to sub-vitality located on chromosome 3, 5, and 8. Chromosome 3 had three different SNPs associated to sub-vitality. The most prominent SNPs was located on chromosome 3, which is associated to a gene, *IGF1*, responsible for insulin-like growth factor and contributes to the development of foetal organs. The genes identified by GWAS and cROH pointed to the cause of sub-vitality due to the contributing effects of biological functions related to metabolic activities. A targeted gene sequencing study would be required to assess the differences in the sequences of the identified significant SNPs, the IGF1 genetic region, in order to examine the possible causal mutations of sub-vitality. A study focussed more on the white and sub-vital subpopulation, with an increased sample size and a targeted gene sequencing approach would assess the differences in the targeted sequences of both case and control individuals.

ACKNOWLEDGMENTS

I would like to thank my supervisors Mr E.F. Dzomba and Dr F.C. Muchadeyi for the time and effort that they have invested in me and for setting a standard of quality research, through their support and scientific guidance. The NRF and the ARC (Biotechnology Platform) for financial support and providing technical support, without which this research and the quality envisaged could not have been achieved. In particular, I want to thank Mr. K.S. Khanyile for his technical support in SNP genotyping. The Karakul Research Station and AGRA for seeking new ways of solving old problems and moving with technology advances in animal genetics and breeding, I thank them for providing the blood samples.

Lastly, I want to dedicate this dissertation to my family. I would like to appreciate the Malesa family for never failing to inspire, in particular mom and dad for always being there and providing encouraging words and strength, for believing in my dreams and always providing support when I needed it most.

Ke Mokhalaka wa Mmakhenyadi wa o nya dithaka bangwe ba khe nya dithokolo ra tšea ra haša! Ahee!

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1: INTRODUCTION

1.1 Rationale for the research

Karakul is a fat-tailed breed of sheep found across the world particularly in Russia, Iran, New Zealand, South Africa and Namibia among others (Campbell 2007) that is used in pelt production. Pelts are the harvested skins of lambs slaughtered within 24 hours of birth. Twelve Karakul rams of brown and grey coat colour were imported into Africa in the 1900s from Germany (Schoeman 1998). Around the 1960s, these Karakul were crossbred to three white haired breeds from the Namaqua, Afrikaner, and the black-headed Persian sheep breeds, resulting in over 20 000 crossbred sheep in Namibia in 1985. This initial crossbreeding resulted in a white coloured pelt characterized by hard and prickly fur. An improvement in the quality of the Karakul fur is the result of years of intense selection. The pelts currently harvested from Karakul sheep of Namibia are characterized by a luxurious pelt quality and are now referred to as South-West Africa Karakul (Swakara), which is the brand under which the pelt products are sold (Rothauge 2009). The term Swakara will be used to describe this improved Karakul breed throughout this study. The Karakul sheep breed found in Namibia has been renamed Swakara to distinguish it from the indigenous Asian/European breeds from which it was bred from (Schoeman 1998).

Swakara sheep are known for their high quality pelts that have different curling patterns compared to that of the original Karakul pelts (Campbell 2007). The Swakara sheep is farmed mainly in Namibia, South Africa, Botswana and other smaller parts of Africa. The breed is more adaptable to the drastic climatic conditions that are characteristic of arid lands (Nsoso & Madimabe 2003a). **S**wakara pelt can be in black, brown, grey, spotted and white colour. White pelts are the most preferred as they can be easily dyed into alternative colours for clothing and other pelt uses. These white pelts fetch higher prices on the market. The breeding of white Swakara sheep is however, hampered by the occurrence of sub-vital performance whereby the genetically pure white lambs have a short lifespan and die before reproductive age. The pure white animals have low stress tolerance and need special care which presents managerial constraints to breeders and farmers (Greeff & Faure 1991). Despite the many challenges

farmers face breeding white Swakara sheep, there is still a high demand for white pelts. The challenges include difficulties with replacing ewes for succeeding breeding flocks. As a practice the breeders mate *A-white* rams with black ewes to produce *B-white* and *C-white* progeny. Mating of *B-white* x *B-white* (*C-white*) produces 25% black; 50 % *B*/*C-white* and 25% *A-white*. Mortality among the *A-white* sheep is extremely high and results in major economic losses. In addition, only 25% of the F2 progeny is pure-white and as such further lowers the anticipated profit margins.

It is thought that sub-vitality performance that affects the production of white pelt from Swakara sheep in Namibia, South Africa and Botswana is caused by the presence of a recessive gene or alleles that causes death of some of the white individuals before they reach reproductive maturity. Genes that are detrimental to survival are termed lethal genes. In the case of Karakul, traits of "sub-lethality" and "sub-vitality" have been observed in the grey and white sheep (Nel 1966; Rothauge 2009). Homozygous white Karakul lambs die after reaching the weaning age which can be between 14 days and 6 months depending on the availability of feed (grain-based diet) (Groenewald 1993; Schoeman 1998). The symptom of sub-vital performance is that lambs are unable to digest feed and as such hampering their growth and development. The sub-vital trait prevalent in pure white pelt (*A-white*) sheep presents a challenge when lambs are selected for replacement flocks. Understanding the genetic basis of sub-vital performance in Swakara sheep is vital for future breeding strategies and optimum production of white pelt in the industry.

White Swakara pelts are greatly sought after in the fur industry because of their unique lustrous feel and quality. However, sub-vital performance negatively affects the production of white pelts and as such lowers the profit margins. Intensive management with special care is required for white Swakara if they are to be used as breeding animals. The Swakara industry is looking for ways of managing or eliminating sub-vital performance from their population so that they will be able to breed for white pelts without experiencing high mortality, low pelt yields and low survivability of replacements. Recently, an *A-white* population with none of the sub-vital characteristics was identified. The genetic differences between the

white sub-vital, white vital and other coloured Swakara sheep (black and grey Swakara) is not clear. It is important to identify genetic markers that can be used to select against sub-vital performance in white Swakara (Schlechter 2013). The ability to select against *A-white* animals susceptible to sub-vitality characteristics at an early stage will lower cost of managing sub-vital animals.

1.2 Justification

Sub-vitality has not been extensively studied in Swakara sheep. Understanding the genetic factors that contribute to the occurrence of sub-vital performance in Swakara sheep is important. Genetic regions differentiating *A-white* sub-vital Swakara sheep from the rest of *A-white* (normal vitality), grey and black Swakara sheep have not been identified. This study intends to use genome-wide SNP genotyping by using the ovineSNP50 beadchip to understand the genetics of sub-vital performance in Swakara sheep. The era of genomics has provided a great platform to explore the genetics of various livestock and other species. Large amounts of genomic data are made available through high-throughput genotyping platforms making it possible to screen for genomic variants that could be associated with sub-vital performance and other genetic disorders. The OvineSNP50 beadchip has been used in other genome-wide studies to understand the diversity of sheep breeds and associate traits of interest to SNPs and their related genes (Kijas *et al.* 2012). The application of the OvineSNP50 beadchip to the Swakara breed will help mine genetic information and help identify the mutations responsible for sub-vital performance in Swakara. An understanding of the genetic structure of the black, grey, white vital and sub-vital white individuals is important to unravel differences and similarities of the different Swakara subpopulations. Principal component analysis is one of the available methods used to determine population genetic structure of populations.

Years of intensive selection in Swakara sheep could have led to the accumulation of lethal recessive traits leading to expressions of genetic disorders (Verweij *et al.* 2014). The occurrence of sub-vital performance in white Swakara sheep is therefore hypothesized to be as a result of inbreeding as a product

of intensive selection in a small population. In this study, concepts pertaining to the genetic structure of Swakara and the SNPs responsible for sub-vital performance are employed to help understand the occurrence of sub-vitality at a genome-wide level. These concepts will include looking at admixture, which is used to describe a population admixed occurs as the result of two previously reproductive populations that are allowed to mate resulting in an admixed population (Tang *et al.* 2005). Admixture represents the mixture of alleles from different ancestries. Runs of homozygosity (ROHs) occur as a result of inheriting copies of a stretch of a genetic region from common ancestry. Extended regions of homozygosity are a result of inheriting a similar stretch of SNPs from both parents resulting in longer run of homozygous SNPs in an individual. This study therefore sought to assess the inbreeding levels in the different colour subpopulations as well as the occurrence of ROHs that could be associated with sub-vital performance in Swakara sheep. Genome-wide association studies (GWAS) has been used to determine causal mutations associated with traits of economic importance in a number of livestock species. A case/control GWAS study was done to determine genetic variants that may be responsible for the sub-vital performance in Swakara sheep. This study used Swakara sheep samples from Namibia where the sub-vital factor was first observed. Supplementary samples were acquired from South Africa (Northern Cape) for a broader analysis of genomic differences. The study utilized statistical methodology from three main software packages including (i) PLINK v.1.7 (Purcell *et al.* 2007), (ii) Golden Helix SVS (Bozeman 2014) and (iii) *R*-library package *qqman* (Turner 2014) for genome-wide genetic analyses and visualization of results.

1.3 Aims

The overall goal of the study was to investigate genomic regions that differentiated the different Swakara subpopulations and could be associated with sub-vital performance in white Swakara sheep.

1.4 Objectives

The overall aim was achieved using the following objectives:

1. the determination of the transferability of the Ovine SNP50K panel in studying Swakara sheep

2. the investigation of population structure and genetic differences between *A-white* (sub-vitality); *A-white* (normal vitality); grey and black Swakara sheep using whole genome SNP genotype data.
3. the determination and comparison of the levels of inbreeding and regions of extended homozygosity amongst the Swakara sheep subpopulations and assess their potential role in sub-vital performance.
4. the determination of genetic variants associated with sub-vital performance using GWAS analysis.

1.5 Outline of dissertation structure

Each chapter is mostly self-contained, containing a literature review, materials and methods, results and discussion, and conclusions.

Chapter 2 reviews the history of Swakara sheep and the occurrence of sub-vital performance and the different studies that have been done using the OvineSNP50 beadchip.

Chapter 3 assesses the transferability of the OvineSNP50 beadchip to the Swakara population.

Chapter 4 mainly determines the population structure of Swakara sheep and the presence of signals of selection using the fixation index.

Chapter 5 looks at the occurrence of runs of homozygosity and their putative gene functions that may have a contributing effect on sub-vital performance in Swakara.

Chapter 6 uses a genome wide association study to identify SNPs associated with the sub-vital performance

Finally, Chapter 7 provides conclusions on the research study, integrates the different findings, and offers research possibilities that can be focused on going forward.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Swakara is a sheep breed, known as the black diamond of Namibia and bred specifically for its pelt (Rothauge 2009). A pelt describes the skin of a lamb harvested within 24-hours of its birth. The Karakul industry was founded towards the end of the 19th century. Karakul breeding originated in Bukhara, Uzbekistan. The name Karakul is derived from a river in the same region of Bukhara. The practice of Karakul breeding for pelts spread to countries like Russia, Namibia, Afghanistan, and other Central Asian countries. Namibia was formerly known as South West Africa during the German colonial regime. The renaming of Karakul to Swakara comes from this as they were formerly traded under the name of South-West Africa Karakul. The unique characteristic of the locally produced pelts makes Swakara pelts easily distinguishable from Karakul pelts produced in Central Asia and Eastern Europe. The Karakul sheep (*Ovis aries.*) is believed to be one of the oldest breeds of domesticated sheep in the world. They are possibly one of the few animals that can survive the harsh, arid conditions of these regions while providing both a source of food and income to local people. Swakara is the most important breed for pelt production despite availability of other pelt breeds globally (Greeff & Faure 1993a; Rothauge 2009; Nasholm & Eythorsdottir 2011). Swakara is particularly preferred because of its lustrous pelts of different colours (Figure 2.1) and its tolerance of the harsh, arid climatic conditions of the Southern African regions (Sefidbakht & Farid 1977; Meadows *et al.* 2008). Swakara is considered a multipurpose breed that can also be bred for meat, wool and milk (Nsoso & Madimabe 2003b). In some sheep farming systems, pelts are considered a by-product, for example with the Swedish Gotland and Icelandic sheep breeds (Nasholm & Eythorsdottir 2011). Karakul sheep from which the Swakara sheep were developed are known to adapt well to arid regions of South Africa and Namibia. They are resistant to internal parasites and foot rot (Buduram 2004).

Figure 2.1: Pelt of a (i) black, (ii) white, (iii) brown and (iv) spotted white Swakara sheep

Swakara is largely produced in the arid parts of Namibia and other parts the Western Cape (Campbell 2007). Swakara sheep are not indigenous to Namibia or South Africa but originate from East Asia from where 750 Karakuls (the breed from which Swakara was developed) were imported over a period of two decades into Namibia and later distributed across the Northern Cape and surrounding areas of Southern Africa. Indigenous sheep such as the Blackhead Persian and Namaqua Afrikaner were later used for upgrading of the Karakul through cross-breeding (Rothauge 2009). In 1979, the Karakul sheep population in both South Africa and Namibia was estimated at 5 million animals. These numbers have, however declined due to the pelt industry losing popularity with its biggest consumer, the fashion industry, and changes in the global economy (Schoeman 1998).

When the Karakul breed was introduced into Africa in 1907, only brown and grey coat coloured sheep were part of the 12 rams imported into the country (Lundie 2011). Around the 1960s the Karakul was crossbred to the three white haired breeds of the Namaqua, Afrikaner, and the black-headed Persian (Lundie 2011). In 1985 over 20 000 crossbreeds were estimated in Namibia, 78 years after importation into the country. The white Karakul flock was formed in 1969 from a cross between black Karakul and white-wooled Persian breed (Steyn 1962). After years of selection, there was improved quality in the Karakul fur, particularly the pelts harvested from the Karakul sheep of Namibia which were then trademarked for their luxurious pelt quality and named Swakara, short for South West Africa Karakul (Lundie 2011; Nasholm & Eythorsdottir 2011).

Figure 2.2: The versatile uses of Swakara pelts (*image ref: http://swakara.net/new2015/design-studio/; http://swakara.net/new2015/wp-content/gallery/2009-collection/001.jpg*)

(Clockwise from top left: a shoe made from white Swakara pelt, the white Swakara trademark logo, a neckpiece made of metal and black Swakara belt, a bag made of black Swakara pelt, a white coat body made of white Swakara and other fur, a carpet made from different dyed Swakara wool.).

Swakara sheep farming is popularly known for the production of pelts (Nsoso & Madimabe 1999; Rothauge 2009). The industry is thriving, with the auction of pelts growing popular in the fashion industry (Rothauge 2009). Despite its growing success, the Swakara industry is faced with various challenges that are both managerial and genetic. One such challenged is that the white lambs experience sub-vital performance and require extra care and a specialized diet to help with their development, for instance, in the replacement ewes for breeding (Nsoso & Madimabe 2003a; Rothauge 2009). This review describes the Swakara industry i.e. the karakul breed's distinguishing characteristics and production challenges. The review also discusses the previous genetic studies on the breed using various genetic tools. Swakara breeders have managed to produce a lustrous sheen to their pelt and distinctive patterns

of swirl and curved curls earning these pelts the name 'the black diamond' of Namibia (Campbell 2007). Extensive work by researchers such as AD Thompson in the 1920s contributed to the popular pelt seen on international fur markets today, used to create different products (Figure 2.2). The flat curl type is still sought after and contributes to the higher prices obtained compared to that for other Karakul producing countries' average prices (Schoeman 1998).

Figure 2.3: Years of genetic selection have afforded the industry the sort-after watersilk Swakara pelt (Campbell 2007)

Improving the quality of Karakul pelts has been at the centre of the breeding objectives for centuries (Figure 2.3). Karakul from South Africa and Namibia have been marketed and traded as Swakara since the 20th century. Due to the marked differences between Swakara and Central Asian Karakul, the Namibian authorities envisaged to have a name change from Karakul to Swakara. Examining the difference at a genetic level will help support the idea that the Karakul and Swakara have diversified away from each other.

The Karakul breed is a unique breed with distinguished characteristics that include fur type, tightness and size of curl, lustre, pattern, texture and hair length and the desirability of the fur (Nasholm & Eythorsdottir 2011). Karakul sheep characteristically have a fat-tail, sloping rump (Figure 2.4) and respiratory appearance that refers to moderate angle ribs and lanky appearance that enable the animal to adapt to extreme climatic conditions.

Figure 2.4: A young light brown Karakul lamb still having the desired market pelt qualities alongside the herd of grown Karakul sheep and lambs *(image ref: http://www.plfkarakuls.com/hazel.jpg)*

Karakul has four basic curl types i.e. garlic, developed shallow, shallow and watered silk with intermediaries. The degree at which the hair curls determines the category of the pelt.

Table 2.1: The characteristics used to categorize Karakul pelts

Characteristic	Description
Curl type	degree at which the fibres swirl develops (curly to smooth)
Hair quality	texture and lustre, thickness of fibre; affects durability & pelt appearance
Pattern	design, extension and delineation of curls
Hair length	long or short fibres, affected by curl type, breeders prefer shorter curls

The pelt traits of Karakul are economically important and are usually assessed on a score system that determines the value and subsequent auction price of a pelt. The most important qualities of Karakul pelts include durability, silkiness and shape of curls (Table 2.1). These characteristics are lost with age hence it is imperative to slaughter lambs immediately after birth, to ensure a uniform distribution of curls as desired in fur trade. The quality of the curl on the pelt is determined by how soon the lamb was slaughtered after birth to preserve the curl.

2.2 Coat colour and sub-vital performance in Karakul sheep

The type of hair curl and colour of the pelt influence the price of a pelt. The basic karakul colours include black, brown, grey and white. Most Karakuls are black in colour and this is considered the dominant colour (Campbell 2007).

The white colour is not inherent to Karakul (Rothauge 2009). It was achieved by breeding the black Karakul to the white-wooled Persians at the Grootfontein College of Agriculture. An offspring of this crossing would typically display coloured features of black breakouts on the face and flecks on the legs, ears and the feet (Campbell 2007). Further intensive selection was conducted within these crosses to introduce the white colour that is seen in today's Swakara sheep.

Figure 2.5: A pelt showing breakouts of black in the white coat colour *(image ref: http://www.travelnewsnamibia.com/wp-content/uploads/flecke.jpg)*

The locus underlying the white coat colour is the W-locus that has been observed to have incomplete penetrance over the B and E loci, which both code for black colour (Schoeman 1998). Homozygous lambs (*A-white*) for this trait are completely white while the heterozygotes have breaks of black in their coat (Figure 2.5, Table 2.2).

Table 2.2: The different types of white Swakara reported

Type of white	Description
A-white	completely white with black pigmentation around eyes and other extremities i.e. ears & nose
B-white	black or brown patches only on the head, body or legs
C-white	clean white body with black or brown patches on the head, tail, legs, umbilicus and groin
D-white	similar to *C-white* same as *C-white*, but with black patches on the body
Spotted	similar to *D-white*, but with larger black/brown spots on the body

Heterozygotes have been observed to have better survival rate than the pure white homozygotes (Rothauge 2009). A study done on a breeding flock of Karakul showed that mortality was high in white Karakul and affected lambs from the ages of 1 to 48 months which being most prevalent between 4 and 9 months (Groenewald 1993). The sub-vital factor is characterized by complications of the digestive tract with gross distension of the abomasum (Schoeman 1998). A linkage between the sub-vital factor and the colour factor has however been suggested (Groenewald 1993; Schoeman 1998).

Digestive complications originating from the abomasum are the cause of death in homozygous grey Karakul lambs that carry the lethal factor (Nel & Louw 1953). Studies by Groenewald (1993) on the lethal factor that causes death in Karakul revealed large milk-filled rumen in the grey and white lambs, which was explained by the decreased number of myenteric ganglia & neurons in the wall. The grey and white Karakul populations are affected by a lethal factor; however, this is less severe in the white population hence the factor is regarded as a sub-vital performance in the white population. Replacement grey ewes are produced through mating of grey ewes to black rams to produce heterozygous grey

replacement ewes. Mating grey against another grey is known to produce more progeny that are homozygous for the recessive lethal factor and as a result this mating is avoided (Fishchenko 1978).

Genes of a lethal nature usually manifest themselves at the foetal stage of development (Greeff & Faure 1993b). The grey Karakul has a lethal roan factor that was identified in the Garole sheep breed. In past centuries, Russian Karakul breeders used to avoid grey to grey mating, which would result in individuals homozygous for the lethal gene (Campbell 2007). A comparison of black and lethal grey Garole sheep revealed unusual greyish or roan to reddish coat colour; different to what was observed in normal lambs (Banerjee 2010). A comparative post-mortem between lethal gene carriers and the normal Garole lambs showed that the lethal gene carriers had smaller sized organs when compared to the normal, black lambs. This is currently the only known documented sheep breed that shows similarities to sub-vitality in Swakara sheep.

Mating a homozygous *A-white* ram to another homozygous *A-white* ewe is generally expected to produce sub-vital lambs. The animal breeders at the House of Swakara have adopted a mating system that guarantees a certain percentage of the pelts produced from the mating are white while reducing the expected loss of lambs to the sub-vital factor. Farmers keep a breeding flock of ewes of the *B, C* and *D* phenotype that are then mated to an *A-white* ram. However, the progeny resulting from such a mating will be at least 50% *B, C* and *D*, spotted white phenotypes and sometimes, black rams would be obtained from these crosses and the hair quality will not be up to market standard. The current setback of mating spotted white ewes to *A-white* rams is the production of other spotted animals, which are unpopular in the market as it is costly removing the black patches from the pelt. Recently a pure white sheep population was identified that has not succumbed to the effects of the sub-vitality and has been living for over 7 years. This has stimulated interest into the genetics of the survival of the white sheep. Knowledge of the genetic of sub-vital and vital performance in white sheep will open possibilities of increasing production rates using genomic tools such as marker-assisted breeding.

2.3 Genes controlling coat colour in sheep

Sheep were domesticated for meat, milk and skin making them popular for clothing and played an important role in trade (McManus *et al.* 2010). Coat colour in sheep is an important breed characteristic and production trait. Classical genetics has been used to determine the genetic contribution to coat col

our variation. Using classical genetics, the alleles determining black coat colour were determined as a recessive black allele A^a of the agouti signalling protein (*ASIP*) gene. The agouti signalling protein is responsible for regulating pigmentation. The ovine *ASIP* gene was studied in order to assess its function and a 190 KB tandem duplication encompassing the ovine *ASIP* and its two neighbouring genes *AHCY* coding regions and the *ITCH* promoter region was discovered (Norris & Whan 2008). It is thought that this duplication is the cause of the white coat colour of the dominant white A^{Wt} agouti sheep (Norris & Whan 2008). The ovine *ASIP* gene has been characterized and located on chromosome 13 (Rochus *et al.* 2014).

The *ASIP* gene is hypothesized to be one of the coat colour loci for sheep and is a ligand for the melatonin receptor *MC1R* gene located on chromosome 14. Mutations on the *MC1R* gene have been identified in Swedish sheep breeds and determined to affect the coat colour variation of breeds (Rochus *et al.* 2014). Coat colour in mammals is due to the melanin pigment produced by melanocyte cells. The intensity of the colour depends on the ability to produce and migrate these pigments into the hair, skin and wool (Cockett *et al.* 2001). The binding of alpha-melanocyte-stimulating hormone (*MSH*) to receptor (*MC1R*) on melanocyte cell surfaces initiates the production of eumelanin; in its absence phaeomelanin is produced (Cockett *et al.* 2001). It has been suggested that the white colour is caused by a number of factors including the absence of melanocytes or the decreased effectiveness of melanin production (Cockett 1999).

2.4 Molecular markers and genomic tools used to study population genetics

Molecular markers are polymorphisms at the DNA level (Vignal *et al.* 2002) that may be associated with given traits. There are three main variation types at the DNA level *i.e.* single nucleotide changes

known as SNPs (single nucleotide polymorphisms), insertions or deletions (INDELS) of various lengths from one to several hundred base pairs and VNTR (variable number of tandem repeats) (Vignal *et al.* 2002). SSR markers describe tandem repeats of short nucleotide motifs ranging in size from 1 – 6 bp (Varshney *et al.* 2005; Pashley *et al.* 2006) that are used for genetic mapping, population and evolutionary studies, fingerprinting and pedigree analysis in different species (Tang *et al.* 2002).

A SNP is a single base change in a DNA sequence with two possible nucleotide alternatives at a given position (Vignal *et al.* 2002). In order for such a base change to be considered a SNP the least frequent allele should have a frequency of 1% or greater in a population (Vignal *et al.* 2002). The information content of a single SNP is limited compared to other markers that are multi-allelic. However this is overcome by the relative abundance of SNP loci (Vignal *et al.* 2002). SNP abundance is highly variable in eukaryotic genomes and is affected by natural, domestication, and breeding history, mating systems, mutation, migration, genomic rearrangements, recombination and other factors (Kolkman *et al.* 2007).

Genomics is the study of structure, function and interrelationships of individual genes and the complete genome of a species of interest (Fadiel *et al.* 2005). The availability of sequencing and genome-wide SNP panels has opened up avenues to investigate population genetic structure as well as causal mutations affecting traits of economic importance and address various research questions. Genome-wide SNPs panels are ideal QTL fine mapping, marker-assisted breeding and genetic association studies (Dekkers 2012). SNPs are known to occur frequently near or in coding sequences and are suspected to be the cause in functional differences between alleles (Lai *et al.* 2005).

2.4.1 OvineSNP50 beadchip panel.

The OvineSNP50 beadchip is the most comprehensive genome-wide genotyping tool developed for sheep and has approximately 54 241 SNP loci that are widely distributed across the ovine genome. The chip has an average density of about one SNP per 50.9 KB. The developers of the OvineSNP50 strategically selected informative SNP markers across the ovine genome. The SNPs were selected under

the criteria of minor allele frequency (MAF), allele count, Infinium Assay quality scores and chromosomal spacing. The beadchip panel is robust with a high density and validated in over 3 000 *Ovis aries* breeds. This makes it highly transferable to a variety of breeds as it is comprehensive and robust. Globally, a number of studies have been done on sheep breeds using the OvineSNP50 beadchip. It has been used to assess genomic selection and improvement of economically important traits in the New Zealand Sheep industry (Auvray *et al.* 2011), the genetic relationships and divergence time based on haplotype sharing between global breeds (Kijas *et al.* 2012), the fine mapping of QTLs and GWAS studies for milk production traits in dairy sheep, paratuberculosis and chondrodysplasia in various sheep breeds (Suarez-Vega *et al.* 2013) and copy number variation studies (Liu *et al.* 2013).

2.4.2 Linkage disequilibrium

Linkage disequilibrium (LD) is the non-random association of alleles at two or more loci, which is largely influenced by population history and its evolution. Linkage disequilibrium localizes genes affecting quantitative traits allowing the detection of associations between a quantitative trait and marker of interest (Nakamoto *et al.* 2006). It is also reflective of population history, breeding systems, and patterns of geographic subdivision (Farnir *et al.* 2000; Sabatti & Risch 2002; Du *et al.* 2007).

Various factors affect the extent of LD, including population history and structure, sample size, the density and distribution of markers and allele frequency distribution (Sabatti & Risch 2002). Regions associated with low recombination levels will have higher LD, for example, the centromere and at the sex chromosomes (Zavattari *et al.* 2000). Inversely, recombination hotspots have a smaller extent of LD. Statistical equations derived for LD include *D'* and r^2 that measure the LD between bi-allelic markers (Sabatti & Risch 2002; Stumpf 2002; McRae *et al.* 2005). *D'* has measures that range between -1 and 1 for any set of allele frequencies for any pair of polymorphic bi-allelic markers and relies heavily on sample size becoming biased if the sample size is small (Lin 2005; VanLiere & Rosenberg 2008). The parameter r^2 is the correlation coefficient between SNP pairs value of 0 would indicate no equilibrium while a value of 1 reflects absolute disequilibrium. The advantage of using the r^2 estimate

is that it takes allele frequencies into consideration and measures the statistical association between two markers (Lin 2005). Minor allele frequency distribution also effects r^2. Extensive LD has been observed in domestic animals compared to humans most likely due to the small effective population size and the stronger selection pressures in livestock species.

2.4.3 Haplotype analysis

Characterizing patterns of LD and haplotype structure helps identify mutations related to complex disease traits (Tishkoff & Verrelli 2003). Linkage disequilibrium blocks are observed in humans (Zhai *et al.* 2004) and livestock species (McRae *et al.* 2005; Meadows *et al.* 2008). SNPs that are in close proximity to each other are usually inherited as a unit or a block called haplotype blocks. These haplotype blocks can be used to find associations relating to quantitative traits. SNPs that are informative for association mapping are usually found in high LD block-like structures that are separated by regions of high recombination (Zhai *et al.* 2004). Haplotype blocks can be used in identifying common variants for complex traits using the hypothesis that a common disease trait occurs as a result of a common variant between populations (Tishkoff & Verrelli 2003; Zhai *et al.* 2004; Nakamoto *et al.* 2006). Testing for association using this approach has increased the chance of finding causal SNPs for traits of interest (Al-Mamun *et al.* 2013). Haplotype blocks are more informative as they are usually inherited together and have been found to harbour quantitative trait regions that are economically important (Khanyile *et al.* 2014).

2.4.4 Runs of homozygosity (ROHs)

The availability of high-density SNP arrays has allowed the examination of runs of homozygosity (Zhuang *et al.* 2012). These are contiguous homozygous regions of a DNA sequence where two haplotypes inherited from parents are identical. Runs of homozygosity are usually formed in the genome of offspring of parents that share a recent common ancestor as a result of inbreeding, with segments of DNA that are identical-by-descent rather than identity-by-state are thus created. ROHs can have differing lengths that give a glimpse into a population's genetic history, with longer segments indicating

recent inbreeding between individuals belonging to a recent ancestor and shorter segments representing ancient inbreeding (Purfield *et al.* 2012). The length and frequency of ROHs informs an animal's ancestry and the history of its population. Because ROHs inform on the inbreeding levels of an individual, these stretches are a common tool for determining autozygosity derived from recent or distant ancestors (Howrigan *et al.* 2011). Runs of homozygosity have also been used in identifying genetic variants responsible for recessive genetic disorders in livestock using genome wide SNP data. The discovery of a 31-bp deletion on exon 36 on the *RELN* gene that is responsible for lissencephaly (smooth brain) in sheep (Suarez-Vega *et al.* 2013) is one good example.

2.4.5 Signatures of selection

Signatures of selection are regions in the genome that have been preferentially increased in frequency and fixed in a population due to their functional importance (O'Brien *et al.* 2014). Selection sweeps cause a fixation in haplotypes therefore reduce haplotype diversity as the haplotypes that surround favourable alleles in a population are inherited along with these alleles when they are in LD (O'Brien *et al.* 2014).In cattle, years of adaptation and interbreeding have focused on the development of individual breeds that have different milk yields, meat quality and tolerance to diseases (Bovine Genome Sequencing and Analysis Consortium *et al.* 2009). The domestication and selection of favourable traits affects the genetic composition of animals. Coat colour has been selected for over the years in livestock as it serves as a form of identification and production trait. Selection is an important aspect of evolution and the population history of every breed as it shapes the pattern of genetic variation (Akey *et al.* 2002).

Selective sweeps can lead to an excess of homozygosity (Lobo *et al.* 2014). Alleles showing increased differences in allele frequencies between populations indicate differences in sweeps. Selective sweeps are used to determine signatures of selection. Assessing signatures of selection by looking at regions belonging to economically important traits helps understand the forces of selection over the regions and the strength of their selection over years (Akey *et al.* 2002; Lobo *et al.* 2014). High density SNP panels such as the OvineSNP50 beadchip provide an opportunity of great statistical power to interrogate the

genome for signatures of selection (Akey *et al.* 2002). Two targets of selection were identified in Brazilian local adapted sheep that are putative for age at lambing *NRG3* (neuregulin 3) and *PTPN20B* (Lobo *et al.* 2014).

Genetic differentiation (F_{ST}) assess allele frequency of loci and their distribution across the genome, chromosome and individual genes (Akey *et al.* 2002) and is one way of examining the variation in SNP allele frequencies between populations and can be used to detect signals of selection (Akey *et al.* 2002; Kijas *et al.* 2012; Lobo *et al.* 2014). Coat colour and pelt characteristics in Swakara sheep are important traits of great economic value and we would expect in our population to observe strong signatures of selection pertaining to these two characteristics as years of extensive breeding would have focussed on improving these two traits to increase the value of the pelts.

2.4.6 Genome-wide association studies (GWAS)

Genomics has ushered in an era that allows for selective breeding of lowly heritable disease traits (Teneva 2009; Anderson *et al.* 2010). The use of molecular data in selection programmes has the power to increase productivity of farm animals. Genome-wide association study (GWAS) is a genomic approach used to compare highly variable SNP markers between individuals that show a specific trait, usually traits of economic importance such as disease susceptibility. GWAS are popularly used to find SNPs that are responsible for disease traits (usually in medical genetics) or traits of economic importance in livestock. GWAS studies are favourable for studying complex traits that have both a genetic and environmental component determining the susceptibility risk (Gorbach *et al.* 2010; Bolormaa *et al.* 2011; Demars *et al.* 2013; Anderson *et al.* 2014). The prominent strength of GWAS studies is that they are able to detect novel associations to disease while also being able to confirm candidate genes if these have been identified in previous studies (Lewis & Knight 2012). High density SNPs offer a cost effective method for analysing genomes on a large scale (Ramos *et al.* 2009; Pareek *et al.* 2011). Through GWAS, it is possible to detect causal variants with modest effects and identifying regions harbouring causal variants for complex traits. The availability of genomic sequences for domestic animals such as dog,

horse, sheep, pig and cattle has meant progressive use of GWAS in livestock for causal gene identification. Genome-wide SNP beadchips are now available for a number of domestic animals (cattle, pig, sheep, horse, and chicken) a result of the availability of complete reference genomes. These high-density beadchips range size from 1K – 600 K.

Traits of economic importance in different animals have been looked at using genome wide association studies (Table 2.3). The Illumina OvineSNP50 beadchip has been used to identify candidate genes that affect growth and meat production traits in 329 purebred sheep (Zhang *et al.* 2013). Studies Genes also identified more disease association and economically important trait in sheep such as the ones listed in Table 2.3. Application of the OvineSNP50 beadchip has also been used the identification of quantitative trait loci, genome wide association studies, characterization of genetic variability among breeds and determining population genetic history among breeds (Grasso *et al.* 2014). It was developed using SNPs used in assembling the sequence data of the ovine genome and it is representative of the ovine genome as it covers all 26 autosomes and the *X* chromosome. Each chromosome has at least a few hundred SNPs that are evenly spaced along the chromosome. The beadchip was also developed using the data of globally economically important breeds.

Table 2.3: Genes and the traits in livestock they are responsible for, as discovered by GWAS

Animal	Associated SNP/gene and trait
Cattle	*DGAT1* (milk yield), 11 genes (milk quality), *SLC6A6* (Tuberculosis in dairy cattle)
Pig	37 SNPs on chromosome 1 & 6 (synthesis and metabolism of androgens)
	Knobbed acromosome defect (KAD-associated gene chromosome 15)
	MC4R (back fat), *TYRP1* (coat colour, chromosome 1)
Horse	*MSTN* (optimum racing distance traits: lethal inherited disease, lavender
	Foal syndrome) located close to gene implicated in dwarfism in humans
	Single based a deletion on exon 30
Sheep	
Chondrodysplasia	Malformed legs, associated with group of consecutive SNP markers
Paratuberculosis	Johne's disease, *Mycobacterium avium* subspecies
Rickets	Corriedale sheep, nonsense mutation (*R145X*) on exon 6 of *DMP1 (Zhao et al. 2011)*
Horns	Soay sheep, *RXFP2* (determines sexual characteristics in humans)
Lissencephaly	31-bp deletion in exon36 of *RELN* gene (Suarez-Vega *et al.* 2013)
Microphthalmia	Texel sheep, missense mutation on *PITX3* involved in vertebrae lens forming (Becker *et al.* 2010)

2.5 Conclusion

Owing to domestication of sheep, pelt production has been in existence for centuries across the globe, and through time, breeders have worked towards improving the quality of pelts for use in the fashion industry. Swakara pelt production is an important industry in Namibia that contributes to the agricultural GDP of the country. The production of white Swakara pelts is being hampered by a sub-vital factor that affects some white individuals and has metabolic implications i.e. failure of lambs to digest feed and underdevelopment of the intestines. This translates to huge production losses in the production of white pelts. The sub-vital factor also creates managerial challenges as these affected individuals have low stress tolerance and therefore require great care. Limited studies have been documented about sub-vital performance in sheep.

The OvineSNP50 beadchip has had remarkable success in identifying causal mutations in different sheep breeds. Here we suggest the use of the OvineSNP50 beadchip to identify the underlying genetic causes of sub-vitality in white Swakara. We examine that applicability of the OvineSNP50 beadchip to the Swakara pelt-producing sheep. Inbreeding levels and ROHs were estimated for the different sheep breeds as well as population structure and genetic differentiation (F_{ST}) between the colour populations. Ultimately, a GWAS between the white and other colour subpopulations was conducted.

CHAPTER 3: PERFORMANCE OF THE OVINESNP50 BEADCHIP IN SWAKARA PELT PRODUCING SHEEP

3.1 Abstract

The emergence of high-end genotyping platforms has made it possible for researchers to study complex diseases and the genetic structure of different livestock using genetic variants such as Single Nucleotide Polymorphisms (SNPs). The development of the Illumina OvineSNP50 beadchip has revolutionized sheep genomics studies and has helped in the discovery of underlying selection pressures and also in understanding of the genetic structure of different sheep breeds across the globe. This study determined the applicability of the OvineSNP50 genotyping beadchip in Swakara sheep that are a pelt-producing breed of Namibia, also farmed in parts of South Africa. Ninety-six Swakara sheep from Namibia, South Africa and Germany belonging to four colour subpopulations; black (n = 16), grey (n = 22), vital white (n = 41) and sub-vital white (n = 17), were genotyped using the OvineSNP50 beadchip. Less than 491 markers failed the HWE test ($P > 0.0001$).The combined population had 80.84% polymorphic markers and an average MAF of 0.26. More than 40% of markers had MAF between 0.4-0.5 indicative of high polymorphism and informative SNPs in the Swakara sheep population. The OvineSNP50 has shown to be a resourceful tool for genomic studies in Swakara sheep.

Keywords: OvineSNP50 beadchip, minor allele frequency, beadchip application, Hardy-Weinberg equilibrium

3.2 Introduction

Swakara is a pelt producing domestic sheep breed farmed mainly in Namibia and is one of the three important breeds for pelt production in the world, in conjunction with Zandi and Gray-Shiraz (Nanekarani *et al.* 2011). The Swakara sheep were developed through years of selection by breeding for pelt characteristics such as lustre, curl type and pelt colour that are now unique to the breed (Schoeman 1998). The current focus for the industry is increasing the production in order to meet the growing demand for pelts globally (Rothauge 2009). The major challenge faced by the Swakara industry is the sub-vital factor that results in the premature death of affected white lambs. The factor is thought to be a

manifestation of mutations in metabolic pathways or in the genes responsible for pelt/coat colour in sheep (Schoeman 1998; Louwrens *et al.* 2004; Coertzen 2009; Rothauge 2009). Very little research has been done concerning the genetic factors that affect the production of pelts in Swakara and to date the causes of sub-vitality in this breed have not been resolved.

The OvineSNP50 beadchip has approximately 54 241 SNP loci that are widely distributed across the ovine genome and has been used in various studies including assessing genomic selection and improvement of economically important traits. The Sheep Genome Consortium is constantly working towards improving the genomic tools available for sheep breeds and have recently produced a 600k ovine beadchip. Few studies have been conducted on African sheep breeds and none to date on African pelt-producing sheep such as the Swakara. Although the OvineSNP50 has been previously used in the many sheep breeds of the world, it has not been used in Swakara sheep. The main goal of this study was to determine the utility of the OvineSNP50 beadchip for studying pelt producing Swakara sheep.

3.3 Materials and methods

3.3.1 Animal Populations

Approximately 4 ml blood was collected from 96 Swakara sheep from Gellap-ost (GO) which is a Swakara sheep research station in Namibia (n = 36), two private Swakara farms from Namibia (PF) (n = 24), Karakul Research Station of Northern Cape (NC), South Africa (n = 30) and later imported and transferred for experimental work at the Biotechnology Platform, Onderstepoort, South Africa. In addition, hair samples were obtained from Karakul sheep from Halle (HG) in Germany (n = 6), this was done in an effort to determine if there are any genetic similarities between populations from the three countries (South Africa, Namibia and Germany). The German flock was included particularly as after years of developing the Swakara breed through selection and that the Karakul breed of Germany is the founding population of Swakara of Namibia. The animals in this study population were randomly

selected to represent the four subpopulations categorized according to pelt colour and sub-vitality as vital white (n = 41), grey (n = 22), black (n = 16) and sub-vital white (n = 17).

3.3.2 Purification of DNA

Two commercial extraction kits of Qiagen DNeasy Blood & Tissue Kit and Macherey-Nagel Nucleo-Spin® Blood kit were used in isolating genomic DNA from the sheep blood samples. The genomic DNA from the hair samples was isolated using a commercial NucleoSpin Tissue XS kit designed specifically for isolating DNA from tissues such as hair and meat. The commercial kits are silica, spin column-based offering a simple approach for DNA extractions with high yields of quality DNA requiring lesser time and labour than conventional DNA extraction protocols (Figure 3.1).

Figure 3.1: A summary of a typical isolation procedure using silica-based column tissue kit for genomic DNA extraction

3.3.3 DNA quantification and quality check

The concentrations were measured using 2 µl of DNA and Qubit®, a next generation flourometer designed to work with highly specific and sensitive DNA quantification assays including RNA and protein assays.

3.3.4 SNP genotyping workflow

Individual DNA from the 96 animals was genotyped using the OvineSNP50 beadchip at the Biotechnology Platform, Onderstepoort Veterinary Institution in South Africa. Each DNA concentration

was adjusted to 40ng/µl according to the specifications of the genotyping process. Genotyping followed the Infinium assay, which is a three-day workflow consisting of whole genome amplification on day one, fragmentation and hybridization on day two and beadchip staining and scanning on day three. The beadchips were imaged on the Illumina HiScan scanner, which detects the intensity of fluorescence signals in the beads to read the SNPs on each locus. Illumina's GenomeStudioTM v1.0 genotyping module was used to analyse collected data. The SNP genotypes called using GenomeStudio were then converted into PLINK formats, .map and .ped, that could be used for downstream analysis.

3.3.5 SNP quality control and analysis

PLINK v.1.7 (Purcell, 2009), was used to prune SNPs using thresholds set for MIND> *0.1,*GENO > *0.01*, MAF< *0.01* and SNPs deviating from HWE ≤ *0.0001* were excluded from subsequent analysis. The SNPs are pre-ascertained and as a result may introduce sampling bias that affects the outcome of an analysis. PLINK takes this into account through the HWE function as SNPs that are fail the HWE are likely the ones to introduce ascertainment bias. Quality control was imposed on the genotyped individuals separately for each subpopulation *i.e.* black, grey, vital white and white sub-vital as well as for the combined population to ensure stringency in the statistical analysis. Missing genotypes and genotype calling parameters inform on the success and integrity of the genotyping.

The minor allele frequency within each colour subpopulations and entire population was calculated using the *--freq* function in PLINK using SNP data that had already been treated for quality control parameters. The distribution of the SNPs was grouped according to SNP MAF value in the form of six bin groups. PLINK used the observed and expected number of homozygous genotypes to calculate the inbreeding coefficients of individuals per population.

3.4 Results

3.4.1 Genotyping and SNP frequencies pruning

PLINK was able to distinguish the 96 individuals according to their sex, and reported 25 males and 71 females which correlated with the gender labels recorded during sample collection. Plink excludes markers or individuals that have more than the set threshold of markers missing. Two females were excluded from further analysis due to having a low genotyping rate, MIND > *0.1,* meaning more than 10% of the genotyped markers were missing from the individuals, leaving 94 individuals. The HWE cut-off value was set to 0.0001, 491 markers failed the set threshold and thus were excluded from further analysis. Fewer SNPs were lost as a result of a deviation from the expected HWE equilibrium. The vital-white group lost the most SNPs, 219, in comparison to the other colour subpopulations whereas the sub-vital white only lost 28 markers due to failing the HWE threshold. A total of 6 614 SNPs failed the missingness test while 6 794 SNPs fell below the MAF cut-off value. Most markers were lost due to low MAF in the alleles in which the sub-vital white group lost 11 567 markers, followed by the black subpopulation that lost 10 991 markers (Table 3.1).

Table 3.1: The number of SNPs lost after QC parameters and the number of individuals lost due to low genotyping success (MIND > 0.1)

Subpopulation	n	MIND > 0.1	GENO > 0.01	HWE < 0.0001	MAF < 0.01
Black	16	0	5 494	35	10 991
Grey	22	0	4 088	47	7 594
Vital white	41	2	4 292	219	7 501
Sub-vital white	17	1	3 500	28	11 567
Combined	96	2	6 614	491	6 794

Overall, 43 849 (80.84%) SNPs markers in 94 individuals were left after quality control (QC) (Table 3.2). The grey subpopulation had the highest rate, 84.37%, of genotyped markers that passed the QC parameters while the black subpopulation had the lowest genotyped markers at 75.98% (Table 3.2).

3.4.2 SNP polymorphisms and MAF

Overall, the proportion of polymorphic markers was above 75% (Table 3.2). The average MAF in each group was above 0.2 with the highest MAF average of 0.257 in the grey population and the lowest MAF of 0.219 in the black population (Figure 3.2). The median MAFs were within the same range as the average MAFs in each of the groups showing that the combined allele frequencies were different (Figure 3.2).

Table 3.2: The number of polymorphic loci in the colour subpopulations of Swakara sheep after QC parameters

Populations	Samples	Polymorphic loci	Genotyping rate %	Mean MAF	Median MAF
Black	16	41 216 (75.98%)	93.85	0.219	0.20
Grey	22	45 765 (84.37%)	93.99	0.257	0.25
Vital white	39	45 517 (83.91%)	94.07	0.252	0.26
Sub-vital white	16	42 361 (78.09%)	94.10	0.255	0.25
Combined	94	43 849 (80.84%)	94.02	0.254	0.26

3.4.3 MAF frequency distribution

The allele frequency of each group was calculated in PLINK and the frequencies were binned in order to determine the proportion of SNPs in each allele frequency category for each colour subpopulation. The average MAF was calculated in each group as shown in Table 3.3.

Table 3.3: The production of alleles within each colour subpopulation and their minor allele frequencies

	Black		Grey		Vital white		Sub-vital white		Combined	
MAF Bins	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD
0.0-0.05	0.0282	0.0120	0.0338	0.0113	0.0249	0.0106	0.0314	0.0008	0.0214	0.0123
0.05-0.1	0.0672	0.0029	0.0790	0.0115	0.0704	0.0142	0.0779	0.0155	0.0705	0.0137
0.1-0.2	0.1336	0.0275	0.1473	0.0253	0.1474	0.0295	0.1559	0.0257	0.1488	0.0290
0.2-0.3	0.2336	0.0272	0.2503	0.0321	0.2495	0.0292	0.2497	0.0256	0.2492	0.0251
0.3-0.4	0.3342	0.0272	0.3523	0.0255	0.3519	0.0294	0.3449	0.0256	0.3495	0.0294
0.4-0.5	0.4435	0.0346	0.4491	0.0299	0.4515	0.0278	0.4457	0.0322	0.4495	0.0284

The proportion MAF per bin was similar across the four colour subpopulations. Minor allele frequencies in the 0.0-0.05 bins, ranged between 0.0249 in the white-vital and 0.0338 in grey subpopulation with standard deviations of 0.0106 and 0.0113, respectively. This bin represented a proportion of SNPs/markers that were fixed or nearly fixed within a population. The bin category that represented informative SNPs was 0.4-0.5, and had a high proportion of SNPs ranging between 0.4435 in the black and 0.4515 in the vital-white subpopulations. The graph in Figure 3.2 showed a similar trend across their distribution of MAFs.

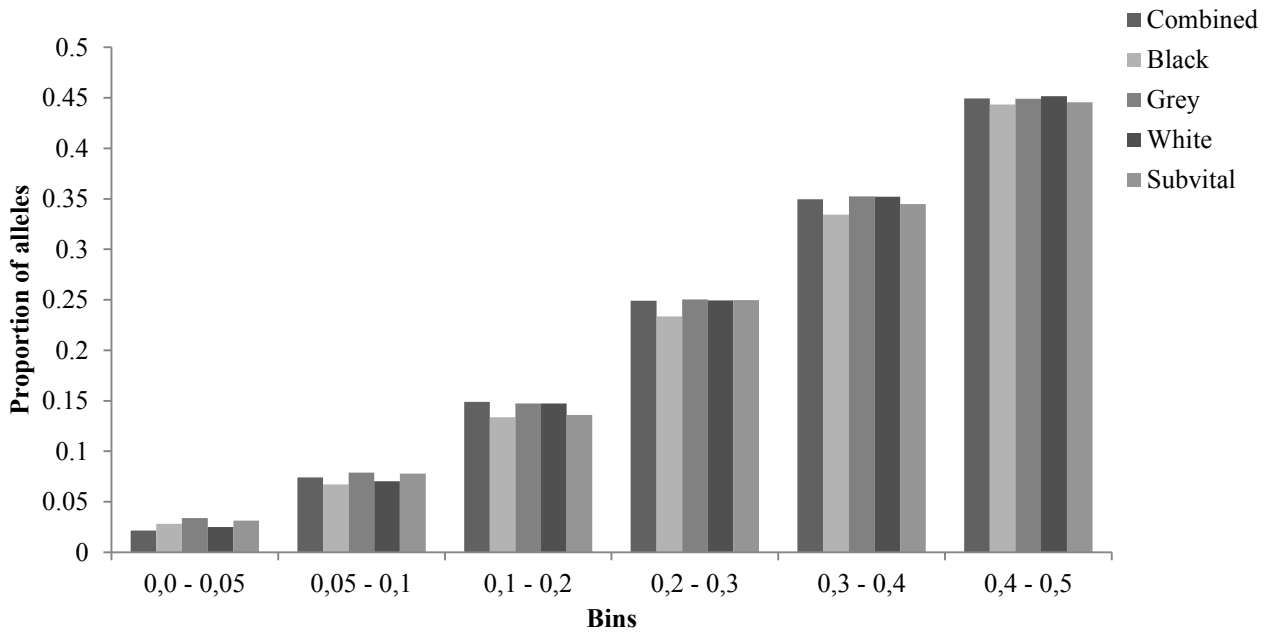

Figure 3.2: Minor allele frequency distribution across combined Swakara population and the individual colour subpopulations

Using the MAF < 0.01 threshold resulted in a larger proportion of SNPs that were common and more informative thus ensuring statistical integrity for downstream analysis. Rare SNPs indicate unique alleles to a population and could potentially harbour association with traits of economic importance.

3.5 Discussion

Single nucleotide polymorphisms refers to mutations at single base positions in the genome of living organisms (Twyman 2003). The move towards high-density SNP genotyping for genomics analysis has been reported recently in many livestock such as horse, pig, sheep and cattle (Ramos *et al.* 2009; Anderson *et al.* 2014). Illumina in collaboration with the International Sheep Genome Consortium has designed a genotyping array (OvineSNP50) for domestic sheep (International Sheep Genomics Consortium *et al.* 2010). The OvineSNP50beadchip has 54 241 evenly spaced probes that are distributed across the 26 chromosomes, including the sex chromosome. The SNP markers are distributed at a density of least one marker per 46 kb. In constructing the OvineSNP50 beadchip, the minor allele frequency, allele count, chromosomal spacing and location of the SNP markers were taken into account

to include SNPs that were relatively informative in populations. Autosomal, mitochondrial and sex linked SNPs are well represented in the sheep beadchip content. The OvineSNP50beadchip has been used in various genomic studies including association studies for various traits of interest, signatures of selection, genetic diversity within breeds and the history of domestication of sheep of the world (Kijas *et al.* 2012; Demars *et al.* 2013; Suarez-Vega *et al.* 2013). The current study investigated the utility of the OvineSNP50 beadchip as a genomic tool for studying Swakara sheep. Transferability of a beadchip to a breed requires that the percentage of loci scored per animal, the proportion of animals scored per locus/genotype and any deviations from HWE be determined such that only informative and randomly segregating SNPs are used (Dodds *et al.* 2009).

The quality control procedure was done for each of the subpopulations belonging to the different pelt colours and susceptibility to sub-vitality. Quality control was also done for the combined population. The call rates for both SNPs and individuals are used as a general measure for SNP data quality and screening tool for genomic studies and evaluation (Cooper *et al.* 2013). The call rate is defined as the ratio of number of genotypes exceeding the threshold value to the total number of genotypes (Demars *et al.* 2013). In this study, it was SNPs that had a call rate lower than 90% and MAF less than 0.01. The QC parameters also included departures from HWE at $P < 0.0001$. SNPs that failed the QC parameters were considered not suitable for downstream analysis and would therefore be not used in further analysis of the Swakara population.

Generally, the genotyping of the Swakara population using the OvineSNP50 beadchip was successful with genotyping rates of above 93% for each of the colour subpopulations. Two individuals (one black and one white sub-vital) were removed from the study due to low genotyping success (missing genotypes >10%). The general practice, in genotyping studies is to exclude individuals with call rates <0.95. Other studies have been observed to set thresholds that are more stringent. In a GWAS to identify SNPs associated with hyperprolificacy, the call rate was set to exclude individuals with < 0.98 (Demars *et al.* 2013). In another study for lissencephaly with cerebellar hypoplasia in sheep, individuals were

excluded if the call rate was < 0.95 (Suarez-Vega *et al.* 2013). Generally the acceptable threshold for call rates is between 85% and 95% (Wiggans *et al.* 2010). The call rate and therefore success of genotyping was high in this study and comparable to other studies (Dodds *et al.* 2009).

Under HWE, SNPs segregate randomly in the population. PLINK considers the expected and observed genotyping frequencies when calculating HWE using the Pearson chi-square test to assess the goodness of fit. Deviations from HWE can reduce the statistical power of an association study as deviations may be caused by genotyping error, assortative mating, selection and population stratification (Lewis & Knight 2012). The deviations from HWE can also be compounded by small sample sizes, which increase the chances of introducing ascertainment bias in the results, which in turn affects allele frequency, giving biased results. However, in our study population, few SNPs were lost due to deviations from HWE, in total only 329 SNPs were lost across the four subpopulations while in the combined analysis, 491 SNPs were lost. Having few SNPs that deviate from HWE points to a sample that is likely homogenous and in random mating thus satisfying the conditions of HWE. This also means increased statistical power.

Minor allele frequency of a population considers the proportion at which the least common alleles occur in that population. In order to eliminate spurious results, only informative SNPs markers (MAF > 0.05) should be used in genomics studies. The OvineSNP50 beadchip has SNPs that have been validated in over 75 economically important and localized breeds such as the Awassi, Namaqua Afrikaner, Churra, Merino and Romney (Illumina 2010). The average MAF of 0.28 has been reported in the validation of this chip. In the current study, MAF averaged between 0.21-0.25 showing the robustness of this genomic tool in vast number of sheep breeds including the Swakara. The combined population had 80.84% polymorphic SNPs with an average MAF of 0.25. The high proportion of polymorphic SNPs implies that the chip is informative for use for Swakara sheep genomic studies. More than 50% of the reported allele frequencies for the genotyped loci were binned in 0.04-0.05, showing that most markers were polymorphic and are useful to study Swakara sheep population (Figure 3.2).

The OvineSNP50 has been successfully used in a number of diverse sheep breeds (Illumina 2010; Auvray *et al.* 2011) that led to the discovery of quantitative trait loci, identification of associated alleles to disease traits of interest and the genetic diversity that exists among breeds (Kijas *et al.* 2012; Suarez-Vega *et al.* 2013). A proportion of 80.72% of the OvineSNP50 SNPs could be used after imposing QC parameters in the construction of haplotype structures and determining the decay of LD in Spanish Churra sheep (Garcia-Gamez *et al.* 2012). The study by (Kijas *et al.* 2012) used 90.4% of the SNPs on this beadchip to investigate the level of historic admixture in sheep breeds of the world as well as the role of selection. The average MAF per subpopulation in the current study was between 0.219 and 0.257. In this study, the proportion of alleles based on MAF showed a similar trend across the four colour subpopulations. In the 0.0 - 0.05 bins, the range across the subpopulations ranged between 0.0249 (vital-white) and 0.0338 (grey) with standard deviations of 0.0106 and 0.0113 respectively. This represents that a proportion of SNPs/markers that are fixed within a population. Informative SNPs were also in similar proportions across the subpopulations as represented in the 0.4-0.5 category and had a high proportion of SNPs represented ranging between 0.4435 (black) and 0.4515 (vital white). With more than 44.35% of SNPs within each colour subpopulation having MAF between 0.4-0.5, it can be said that the OvineSNP50 beadchip can be utilized for further genetic studies such as population structure and GWAS within the Swakara sheep breed.

3.6 Conclusions

The results from the study indicate that the OvineSNP50 beadchip could be a resourceful genomic tool that can be applied successfully to study Swakara pelt sheep genomics. The results showed acceptable proportions of call rates in individuals as well as the SNPs deviating from HWE. The MAF distribution indicated the panel is polymorphic and informative in the four colour subpopulations of Swakara.

CHAPTER 4: Determining the population genetic substructure of Swakara sheep using genome wide SNP data

4.1 Abstract

Selection and domestication contribute greatly to the genetic structure of populations. The advent of high-density SNP panels has made it possible to determine genetic diversity, population stratification and population history in a number of livestock species including sheep. In this study, the OvineSNP50 beadchip was used to interrogate 96 individuals of the Swakara collected from four locations i.e. Halle, in Germany, Northern Cape, South Africa and Gellap-ost and two private farms in Namibia. The individuals belonged to four different coat colour subgroups. Population genetic structure was determined using principal component analysis (PCA) as implemented in Golden Helix's SVS genome suite, the Weir & Cockerham F statistics implemented in ARLEQUIN and ADMIXTURE. The fixation index (F_{ST}) was used to determine the level of population divergence between the four coat colour subpopulations. The grey subpopulation was the least inbred (F_{IS} = 0.0092) and most diverse (H_E = 0.3417±0.1434) whilst the white subpopulation was the most inbred (F_{IS} = 0.0856), and the least genetically diverse (H_E = 0.3301±0.1473). The level of population differentiation (F_{ST}) was greatest between the white and black population and the SNP with the highest F_{ST} = 0.6846 was on chromosome 5. Lowest F_{ST} of 0.00938 was observed between the vital white and grey subpopulation. The overall population F_{ST} was 0.2905 for a SNP on chromosome 13. The same SNP had a high F_{ST} of 0.6164 between the white and black. PCA and admixture analyses revealed the white and black subpopulations have similar genetic ancestry as individuals from both subpopulations clustered together in Principal Component Analyses plots. The genetic structure of Swakara sheep is the result of years of selective breeding for coat colour and pelt characteristics, the result of which are subpopulations whose genetic diversity has been altered. The black and sub-vital white subpopulations had the least diversity.

Keywords: admixture, principal component analysis, genetic diversity, population structure

4.2 Introduction

Sheep are among the first livestock to be domesticated. The movement of sheep across the globe to different geographical regions has led to localized trait specialization in the different breeds (Kijas *et al.* 2009) for traits such as wool, meat, milk and production environments. Selection and domestication have a consequence on the genetic structure of populations as they alter allele frequencies and therefore patterns of genetic variation (Groenewald *et al.* 2010). The Swakara breed has been localized and is specialized for pelt production. Swakara breed development and selection pressures over the years has mainly focused on improving pelt characteristics such as coat colour, curl type and curl length (Schoeman *et al.* 2010). Years of intensive breeding in Swakara has resulted in the unique pelt characteristics that are seen today. The white Swakara were specifically produced through intense crossbreeding to other fat-tailed sheep breeds. Imprints of domestication and breed development can be seen in the genetic structure of populations and some detectable signatures of selection (Bovine Genome Sequencing and Analysis Consortium *et al.* 2009). The hair texture and colour characteristics were the main focus for improvement especially the white pelt that has been improved extensively to meet the production demands. There are at present, four main coat colour subpopulations in Swakara namely the black, grey, white and brown (Campbell 2007). The two popular pelts are the black and white subpopulations and there is a preference for the white pelt as it can be altered into different colour forms used in making the different products of Swakara such as coats, dresses, shoes as well as bags (Coertzen 2009).

White pelt producing Swakara, however, experience sub-vital performance (Schoeman 1998), a trait that causes premature death in lambs. It is important to understand the underlying genetic mechanisms that may be responsible for this sub-vital performance trait observed in Swakara that only affects some individuals in the white population. Understanding the population structure and genetic diversity of Swakara is necessary for the eventual mapping of the causal mutations that are responsible for sub-vital performance. The consequence of focused improvement of pelt features is a divergence between subpopulations and the accumulation of specific alleles in some subpopulations more than others that

may not be desirable (Zechner *et al.* 2002). A possible accumulation of unique alleles in the white subpopulation may have been the cause of the sub-vital performance observed today. It is important to determine the within and between population diversity and extent of genetic sub-structuring of the Swakara sheep as a prerequisite to successful determination of causal mutations for sub-vitality. Fixation indices (*F statistics*) methods have been used to determine within and between population diversity and the level of population sub-structuring in a number of species (Edea *et al.* 2013) using a number of molecular markers including SNPs. Fixation indices measure the variation in gene frequency among sub-structured populations (Nei & Chesser 1983) and also give insights into deviations from HWE due to small population sizes, selection, inbreeding and other genetic factors that influence the genetic makeup of a population (Guries & Ledig 1982). F_{ST} measures the divergence of a subpopulation from the ancestral population. Population structure analysis interrogates the allele frequency differences between subpopulations that may have risen as a result of systematic ancestry differences. A popular way of determining population sub-structuring is through principal component analysis (PCA) which clusters individuals on an independent of any population genetic model (Patterson *et al.* 2006; Jombart *et al.* 2010). Other structuring methods include ADMIXTURE, a Bayesian clustering based approach (Jombart *et al.* 2010).

The purpose of this study was to determine genetic diversity and population substructure of Swakara sheep. Understanding the genetic diversity of Swakara will provide baseline information for sustainable management of subpopulations and in identifying genomic regions that differ between subpopulations and could be associated with sub-vital performance. Specifically this study sought to:

(i) determine within and between pelt colour subpopulation diversity using Wright's fixation indices

(ii) determine level of population divergence using genome wide and pairwise per marker F_{ST} analysis. In this analysis genome wide and per marker F_{ST} were computed for (a) vital-white against white sub-vital; (b) vital-white against black (c) vital-white against

grey (d) grey against black (e) grey against white sub-vital and (f) black against white sub-vital

(iii) determine the level of population substructure using principal component analysis and Bayesian-based clustering

4.3 Materials and methods

4.3.1 Animal population

The 96 Swakara sheep used in this study were described in Chapter 3 and consisted of the black (n = 16), grey (n = 22), white-vital (n = 41) and sub-vital (n = 17). The two white subcategories were treated as different populations in all the analyses.

4.3.2 Quality control (QC)

The OvineSNP50 beadchip developed by Illumina was used to genotype the 96 animals with over 54 000 SNPs per individual. PLINK v.1.7 (Purcell, 2009) was used to prune SNPs based on the thresholds for missing genotypes (GENO) > *0.01*, genotype calling (MIND) > *0.1*, minor allele frequency (MAF) < *0.01* and SNPs deviating from Hardy Weinberg Equilibrium (HWE > *0.0001*). Quality control was imposed on the genotyped individuals to ensure high quality data is used in the downstream statistical analysis. The number of SNPs pruned out using the above-mentioned parameters is shown in Table 4.1.

Table 4.1: Number of markers and individuals that lost to quality control parameters

Parameter	*P-value*	Number of markers failing parameter
MIND	0.1	2 individuals (low genotyping success)
GENO	0.01	6614
MAF	0.01	6794
HWE	0.0001	491

After QC, 43 849 markers and 94 individuals remained for further analysis. The average genotyping rate of the remaining individuals in this study was 94.02%. Of the 94 animals, 88 were used for within and between population diversity and population structure analysis after removing the six individuals collected from Halle, Germany. The Halle group was included to show how much the Karakul and Swakara populations have since diverged from each other as the Swakara breed was developed from a Karakul flock from Germany many years ago. The Halle group showed to be an outgroup that clustered differently from the rest of the populations collected from the other locations (South Africa and Namibia) that would otherwise have created population stratification and a misrepresentation of Southern African Swakara population structure and was consequently removed from other analyses. This clustering can be seen in the PCA analysis in Figure 4.4.

4.3.3 Expected and observed heterozygosities and inbreeding estimates

PLINK v.1.7 was used to estimate the frequency of each marker in the population and the observed (H_O) and expected heterozygosity (H_E). The inbreeding coefficient (F_{IS}) was estimated as the difference between the expected and observed heterozygosity divided by the expected heterozygosity *[*($H_E - H_O)/H_E$]. The inbreeding coefficient was also calculated with PLINK v.1.7 using the *--het* function. This analysis does not consider SNPs that are haploid which are found of the *X* -chromosomes.

4.3.4 Wright's fixation indices for determination of population substructure

The Wright's fixation indices as implemented in ARLEQUIN (Excoffier & Lischer 2010) were used to infer on population substructure. In this analysis variance in allele frequencies was computed for within (F_{IS}), between (F_{ST}) and overall population (F_{IT}) for the (i) vital group consisting of black, grey and white-vital sheep (ii) white consisting of white-vital and white sub-vital sheep and (ii) overall population consisting of all 88 Swakara sheep. The population differentiation of the Swakara pelt/coat colour subpopulations was also assessed using the pairwise fixation index (F_{ST}) implemented in Golden Helix SNP Variation Suite v.8.1. The fixation index measures the genetic divergence between pairs of

subpopulations and values range between 0 interpreted as "no genetic divergence" between subpopulations and 1 which shows complete isolation of the subpopulations from each other respectively (Excoffier & Lischer 2010). The significance of the fixation indices was determined by 1000 permutation tests. Pairwise fixation indices were estimated genome-wide based on (i) subpopulation and (ii) per marker. A HeatMap was plotted in Golden Helix SVS to illustrate the population pairwise differences while a Manhattan plot was used to illustrate per marker per subpopulation comparisons.

4.3.5 Population clustering using Principal component analysis

Principal component analysis (PCA) implemented in SVS was used to analyse population structure. Plots were made using PC 1 and PC 2 illustrating the clustering of the coat colour subpopulations in relation to each other. PCA was carried out to determine breed relationships directly on allele frequencies by using a multivariate method which groups the information from large number of alleles and loci into readable variables as principal components (Jombart *et al.* 2010).

4.3.6 Bayesian-based clustering using ADMIXTURE

ADMIXTURE v1.23 software (Alexander *et al.* 2009) was used to infer the most probable number of ancestral populations for the Swakara sheep based on the SNP genotype data from which the population sub-structuring was determined. Admixture analyses structure populations by using a Bayesian model-based method that assumes there are *K* populations/clusters that contribute to the genotype of each individual and characterizes each individual with a set of allele frequencies at each marker locus. Cross validation (CV) is used to determine the best *K* value, where *K* is the number of underlying ancestral populations. CV does this by predicting the systematically withheld data points (Alexander & Lange 2011). The best *K* value is one with the lowest CV value (Alexander & Lange 2011). Admixture was run from $K = 2$ to $K = 5$, a number of populations and the optimal number of clusters (*K* value) were determined as one with the lowest cross validation error (CV-error).

4.4 Results

4.4.1 Genetic differentiation

PLINK v.1.7 was used to estimate the frequency of each marker in the population from which the observed (H_O) and expected heterozygosity (H_E) were calculated in Table 4.2. The F_{IS} statistic, an estimate of inbreeding, was estimated as the difference between the expected and observed heterozygosity divided by the expected heterozygosity. The expected (H_E), and observed (H_O) heterozygosity and the resultant F_{IS} for each subpopulation and for the overall population are presented in Table 4.2. The grey and white sub-vital subpopulations had the highest genetic diversities H_E = 0.3417±0.1434 and H_E = 0.3396±0.1432 respectively (Table 4.2). The lowest level inbreeding was observed in the grey subpopulation F_{IS} = 0.0092.

The average observed heterozygosity was H_O = 0.3109±0.1615 and the expected heterozygosity was H_E = 0.3266±0.1554. The F_{IS} statistic for the overall Swakara population was 0.0480 (Table 4.2). Genetic diversity was highest in grey (H_E = 0.3417±0.1434) while the lowest was in white (H_E = 0.3301±0.1473). Grey had the lowest inbreeding coefficient (F_{IS} = 0.0092), making it the least inbred subpopulation while the black subgroup had the highest inbreeding value (F_{IS}=0.0496). Sub-vital (H_E = 0.3396±0.1432) had an inbreeding statistic of 0.0197 and the vital-white group had an F_{IS} = 0.0351 making it more inbred than the grey and sub-vital white populations.

Weir and Cockerham's fixation indices indicated that there was little genetic differentiation within the populations (F_{ST}) and greater between populations as shown in Table 4.3.

Table 4.2: Expected and observed heterozygosity and F_{IS} in four colour subpopulations of Swakara and the fixation indices (F_{IS})

Population	Samples	$H_O \pm SD$	$H_E \pm SD$	F_{IS}
Black	16	0.3221±0.1699	0.3389±0.1430	0.0496
White vital	35	0.3185±0.1514	0.3301±0.1473	0.0351
Sub-vital white	17	0.3329±0.1718	0.3396±0.1432	0.0197
Grey	22	0.3385±0.1637	0.3417±0.1434	0.0092
Combined	90	0.3109±0.1615	0.3266±0.1554	0.0480

Table 4.3: Weir and Cockerham's Wright's fixation indices

Population	F_{IS}	F_{ST}	F_{IT}
All Pop	0.069	0.025	0.092
WV vs WSV	0.085	0.026	0.109
BV vs WV	0.092	0.037	0.126
BV vs GV	0.045	0.031	0.074

A calculation of the genome-wide fixation indices between pairs of the four colour subpopulations revealed the highest genetic diversity was between the black and the white subpopulations with an F_{ST} of 0.03759 (Table 4.4). The least genetic diversity was between the grey and white subpopulations (F_{ST} = 0.009383), indicative of genetic similarity between the two subpopulations. The white and sub-vital group had an F_{ST} value of 0.02689, lower than that of sub-vital white and grey (F_{ST} = 0.03215) as illustrated in Figure 4.1.

Figure 4.1 illustrates the extent of genetic differences between the four colour subpopulations using colour shades as a code where white shade represent genetically similar populations (min F_{ST} = 0) and darker shade representing more distinct populations (max F_{ST} = 0.38). The grey and white-vital subpopulations were genetically very similar to each other with F_{ST} = 0.009. The most differentiation was between the white and black subpopulations with F_{ST} = 0.03759. The sub-vital white and grey were also more genetically diverse with a F_{ST} value of 0.03215. The black and grey showed some genetic differentiation with an F_{ST} value of 0.03084 (Table 4.4; Figure 4.1).

Table 4.4: Pairwise F_{ST} values between colour subpopulations

Coat colour	White	Grey	Black
White			
Grey	0.00938		
Black	0.03759	0.03084	
Sub-vital	0.02689	0.03215	0.02719

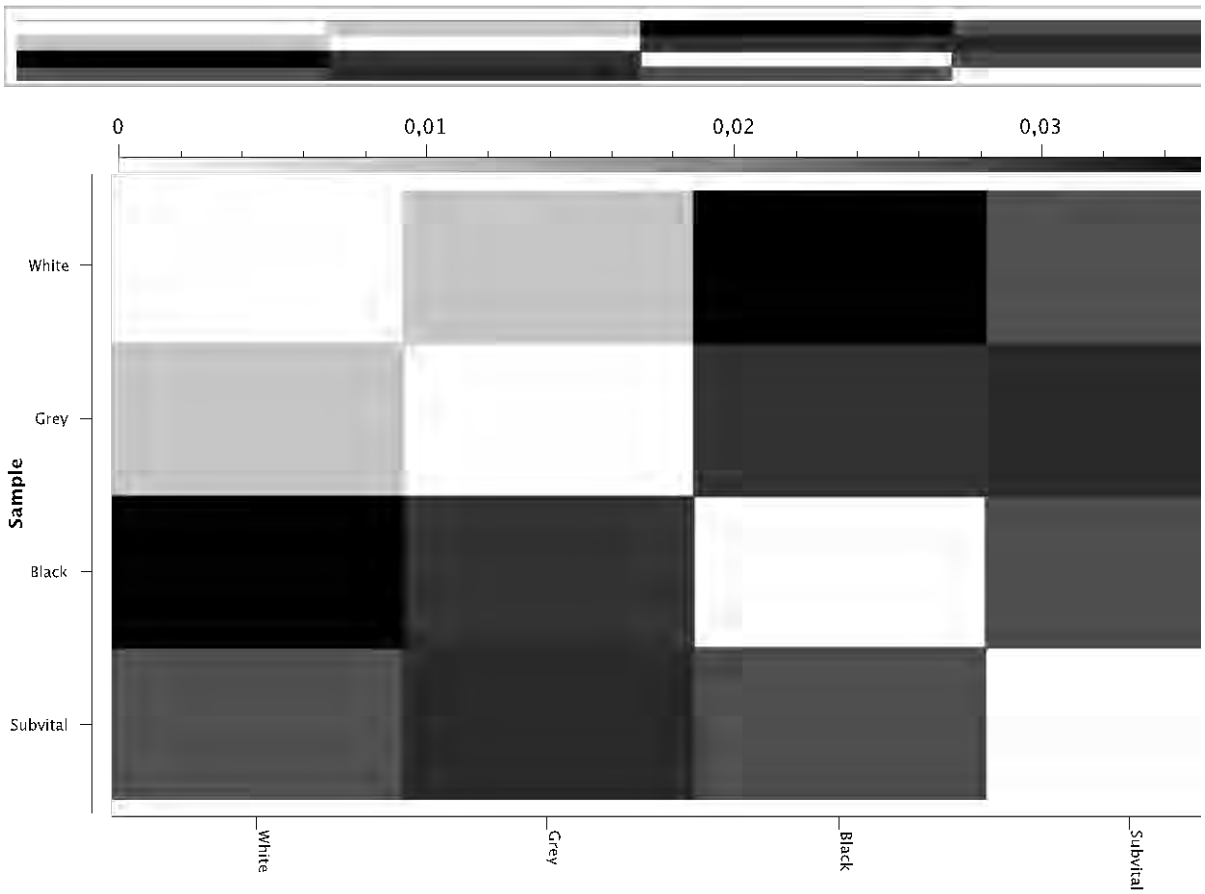

Figure 4.1: The genetic differentiation across the four colour subpopulations

4.4.2 Fixation indices based on markers per subpopulation

The genetic differentiation between the different colour subpopulations was also determined using SVS per marker fixation index to determine the most differentiated SNPs or markers between subpopulations. The SNPs with the highest F_{ST} values within each comparison were identified and the associated SNP/genes functions were determined using *ENSEMBL* ovine genome browser and results presented in Table 4.5. The genetic differentiation by markers between the white and the sub-vital white group revealed a SNP on chromosome 10 with an F_{ST} = 0.4908 (Figure 4.2). A SNP on chromosome 7 had the highest F_{ST} of 0.3772 for pairwise comparison between white and grey populations (Table 4.5, Figure 4.3). The population with the highest F_{ST} values was the grey vs sub-vital genetic differentiation with a F_{ST} value of 0.6846 for a SNP on chromosome 5. Chromosome 15 was also identified between the white vs black subpopulations where it had a F_{ST} of 0.6164.

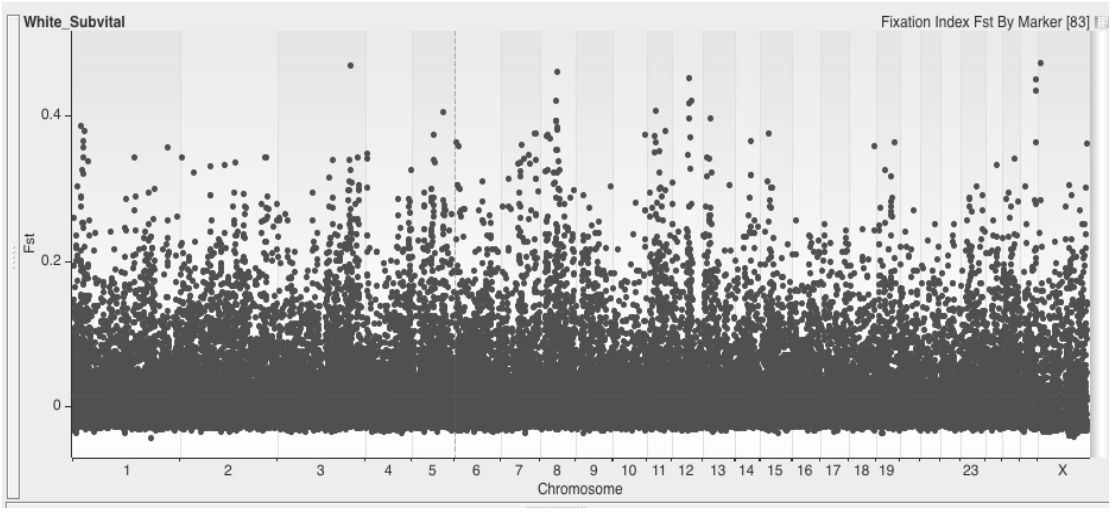

Figure 4.2: F_{ST} by markers between the white and sub-vital white
The comparison between the sub-vital white and the white subpopulations showed a number of SNPS however the chromosome with the F_{ST} value of 0.4908 was found on chromosome 10.

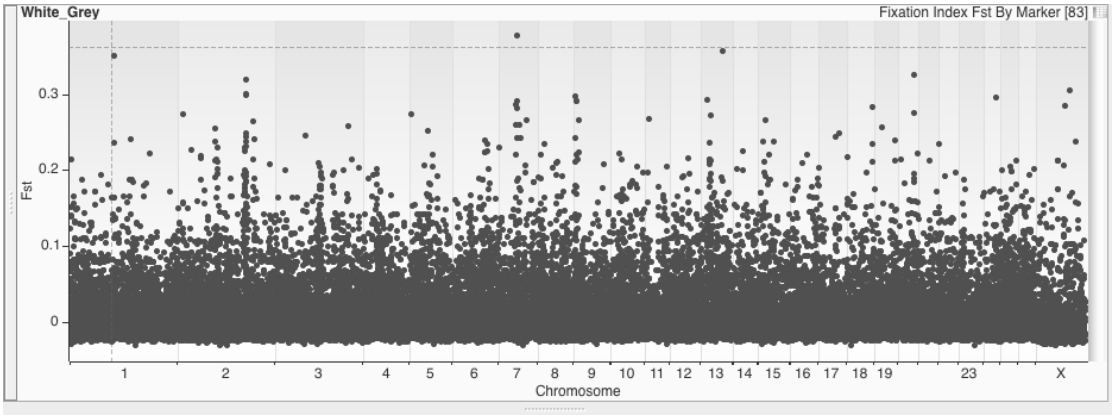

Figure 4.3: F_{ST} by markers between the white and grey subpopulations
The comparison between the white and grey subpopulations using SNP markers to determine the F_{ST} values reveal a SNP on chromosome 7 (Figure 4.3).

Table 4.5: SNPs with highest F_{ST} by marker statistic between colour subpopulations and the associated *ENSEMBL* genes

Subpopulation	F_{ST}	SNP ID	CHR	Related gene
White_Grey	0.3772	s53466.1	7	no gene
White_Black	0.6164	OAR13_16487810.1	13	no gene
White_Subvital	0.4908	OAR10_87219588.1	10	No gene
Grey_Black	0.5209	s38022.1	10	No gene
Grey_Subvital	0.6846	OAR5_55263158.1	5	No gene
Black_Subvital	0.5727	OAR6_72030005.1	6	*GABRB1*

The highest per marker F_{ST} was on chromosome 5 (F_{ST} = 0.6846) for the grey vs sub-vital white comparison. The associated SNPs on these signals of selection had no genes associated to them except for a significant SNP on Chromosome 6 that differentiated between Black vs Sub-vital white sheep was associated to the *GABRB1* (gamma-aminobutyric acid A receptor, beta 1) gene.

4.4.3 Population structure within the Swakara population

Principal component analysis revealed the presence of population sub-structuring in the study population based on both location and pelt colour. The six individuals sampled from Halle clustered separately from the other individuals collected at different locations in South Africa and Namibia (Figure 4.4, D), making this subpopulation an outgroup and therefore having a different genetic structure from the rest of the individuals. The Northern Cape sheep formed a cluster (B) separate from the two other clusters consisting of individuals collected from private farms (C) and the Gellap-ost (A) Research Institute in Namibia. Within the Northern Cape cluster, there were however individuals collected from private farms and Gellap-ost indicating some level of genetic admixture within sheep from Namibia and South Africa.

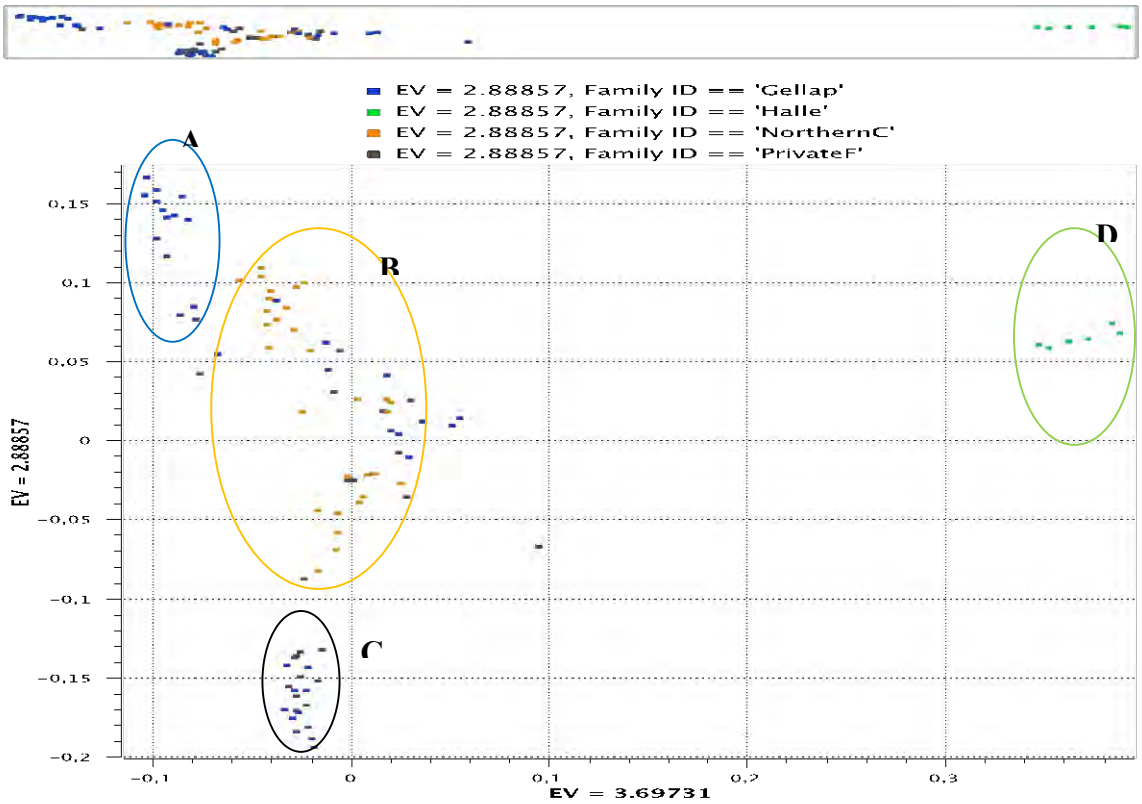

Figure 4.4: PCA-based clustering of Swakara subpopulations from South Africa and Namibia and a Karakul breed from Halle, Germany

Due to the individuals collected from Halle creating a subgroup, they were consequently removed from further analysis. A second PCA was performed using only the pelt colour based subpopulations from Namibia and South Africa (Figure 4.5). The first cluster (A) revealed genetic similarity between the white and sub-vital group made up of 12 individuals. The sub-vital and black sheep grouped together in the other cluster (C). The grey subpopulation formed a larger cluster and revealed genetic similarity mostly to vital white sheep (B).

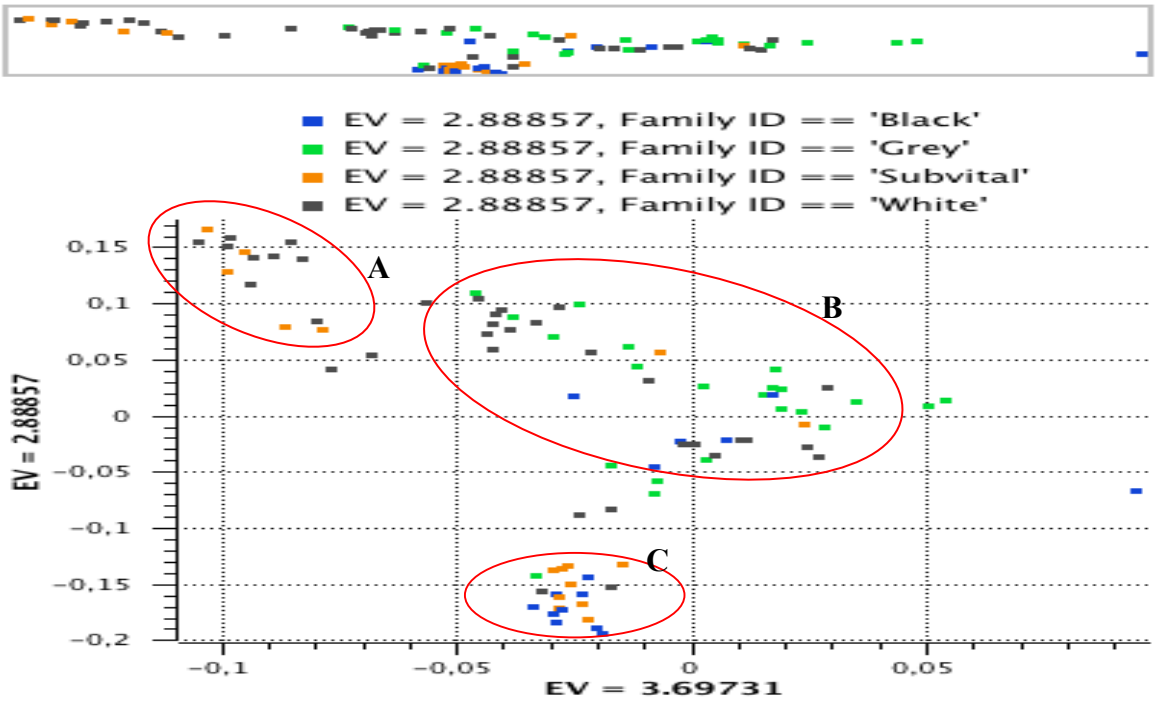

Figure 4.5: PCA-based clustering of Swakara subpopulations from South Africa and Namibia

4.4.4 Bayesian-based clustering

ADMIXTURE was used to estimate an individual's ancestry using maximum likelihood estimates in a parametric model from SNP data. Admixture was run from $K = 2$ to $K = 5$. The lowest cross validation error rate ($CV = 0.669$) was for $K = 4$ and this was determined as the most probable number of ancestral population of the Swakara sheep understudy (Figure 4.6). This suggests that there are four ancestral populations that contributed to the genetic structure of this animal population.

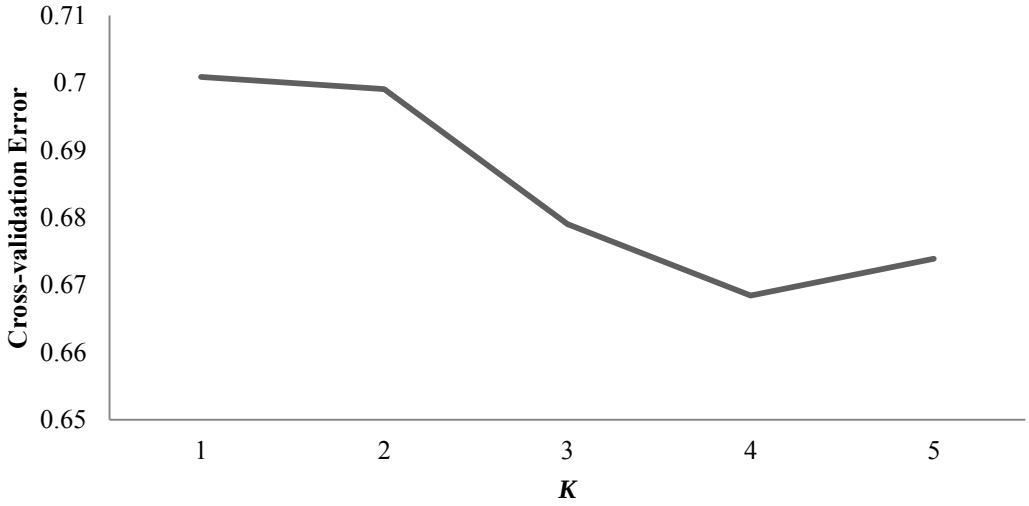

Figure 4.6: The CV error rate for different *K* values

K = 2 suggested a unique sub-structure consisting of individuals from the white vital sheep whilst the rest of the sheep belonged to one cluster (Figure 4.7). Within the *K = 2* CV error, the subpopulation with the most admixture was the white-vital subpopulation. However, there was uniformity in the structure of the population across the individuals of all four colour groups.

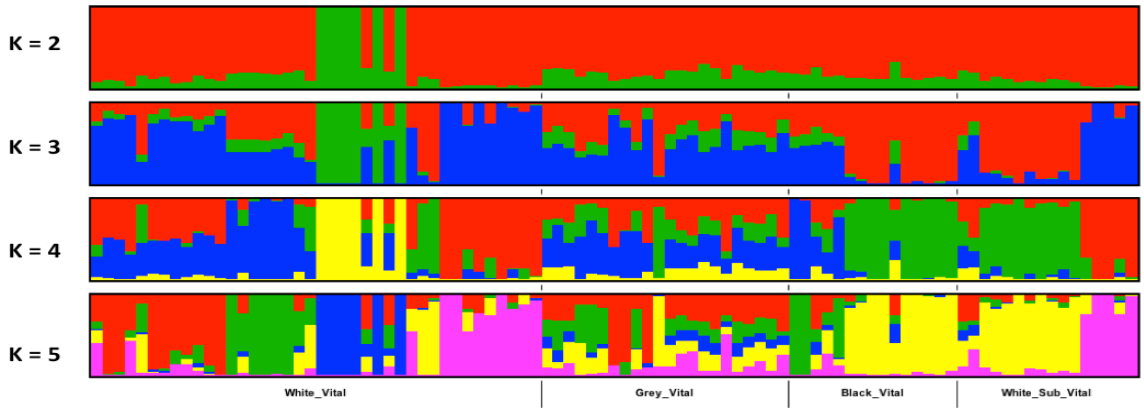

Figure 4.7: ADMIXTURE results of four Swakara subpopulations based on *K* = 2 to *K* = 5

K=3 resulted in 2 genetic clusters from individuals of the vital-white and sub-vital white while the rest of the individuals shared genomic components. Results for this *K* value also indicated individuals that were clustered the same in the white-vital and white sub-vital subpopulations.

4.5 Discussion

Studies of genetic relationships within populations reveal useful information within and between populations that can be used for breed improvement programs (Kantanen *et al.* 2000). Problems of sub-vital performance are seen in some of the white Swakara sheep and it is hypothesized to be as the result of intensive selection for white pelt characteristics in the original Swakara population. The Swakara sheep of Namibia were introduced in 1907 from Germany and developed through years of selection and crossbreeding with breeds such as the Blackhead Persian and Namaqua. The Swakara sheep has continued to show superiority in terms of producing pelts with unique and distinctive characteristics that characterize the breed (Schoeman 1998). Years of intensive breeding can lead to inbreeding depression, a phenomenon where recessive genes with undesired phenotypes are expressed in the population and it is possible the sub-vitality observed in Swakara is a culmination of these factors.

The study population consisted of Swakara sheep collected from Namibia and South Africa and also six individuals from Germany. Assessing the genetic diversity between Swakara and Karakul would highlight how much the Swakara has been developed such that it is different genetically from the Karakul of Germany. The preliminary results indicated differentiation of the Halle population from the other colour subpopulations, which were therefore excluded for further analysis. The Swakara sheep belonged to different pelt colour subpopulations of white, grey and black. The white sheep were either of vital white or sub-vital white performance and were treated as two subpopulations whilst the grey and back were of vital performance. Using the OvineSNP50 beadchip with 54 241 evenly spaced SNP markers, genetic diversity within, the presence of population structure and the genetic differences amongst Swakara subpopulations was investigated.

The inbreeding coefficient (F_{IS}), population differentiation and F_{ST} statistics were used to investigate the relationships amongst the colour subpopulations. The inbreeding coefficient measures the probability that an individual has two alleles at a locus that are identical by descent by comparing the frequency of heterozygotes in a population relative to the frequency of expected heterozygotes under random mating

(H_E). *F* statistics in general measures the degree by which heterozygosity (genetic diversity) is reduced in a population under HWE causing populations to diverge. Genetic diversity and population structure are important; characterizing these two aspects in Swakara sheep will help with an understanding of the genetics of disorders such as sub-vital performance and this could lead to possible breed improvements at the genetic level.

The most gene diversity was in the grey population (H_E = 0.3417) which had the lowest inbreeding coefficient (F_{IS}) of 0.0092. The population with the lowest genetic diversity was the black population with H_E of 0.3389 and F_{IS} of 0.0496 followed by the vital-white population with H_E of 0.3301. In the Karakul/Swakara pelt industry, coat colour is an important trait that has been selected for over years. The black is the most abundant and readily available, while the white is preferred because the pelts can be altered to different colours and made into different garments. Selective breeding in most instances compromises the diversity of a population (Wiggans *et al.* 2010), which could have been the case with the white and black populations. Loss of genetic diversity can affect a population's evolution and reduces individual fitness. Higher F_{ST} values were observed between the grey and white-sub-vital that had a genetic differentiation F_{ST} value of 0.6846 on chromosome 13. These results indicate that the SNP on chromosome 13 is highly differentially selected between the sub-vital white and grey. Overall, the genetic differentiations revealed observed on chromosomes 5, 6, 7, 10 and 13. Chromosome 10 had two SNPs with high F_{ST} values between the white-vital and sub-vital Swakara sheep (F_{ST} =0.4908) and between the grey and black populations (F_{ST} = 0.5209). These SNPs were checked for genetic functions or related genes and none were found. The other SNPs which showed high differentiation were also investigated for genomic functions and only the SNP on chromosome 6 with an F_{ST}= 0.5727 between the black and white-sub-vital sheep was associated to the *GABRB1* (gamma-aminobutyric acid A receptor, beta 1) gene (Table 4.5), which is responsible for protein encoding in the central nervous system. A study on signals of selection was conducted on sheep to a SNP with a high F_{ST} in the global sheep study population was located close (29.5 Mb) to gene responsible for horns (*RXPI2*). Over the years breeders have been selecting against horns and this supported by the history of sheep breeding as

a strong signal of selection (Kijas *et al.* 2012). Only F_{ST} values of over 0.5 are considered for signals of selection in this population study and the SNPs that qualified showed no gene related any function.

Genetic structure was also determined using principal component analysis. The high eigenvalues of the first (36.89%) and the second (28.8%) principal component indicate the presence of non-random population structure (Figure 4.4). The principal components showed a clustering of the six individuals collected from Halle, Germany (Figure 4.4 D) separate from the rest of Southern Africa populations thus becoming an outgroup and over the years the Karakul and the Swakara have come to differ genetically. The individuals collected from Gellap-ost, Namibia (Figure 4.4 A) clustered in two groups with another group congregating individuals collected from private farms in Namibia (Figure 4.4 C). This extent of genetic relatedness may be due to the fact that the Swakara industry in Namibia is consolidated (Louwrens *et al.* 2004) and therefore the same genetic pool is used to create different flocks. The population collected from the Northern Cape also formed a distinct cluster (Figure 4.4 B) that contained individuals from Gellap-ost, showing some genetic similarities among some individuals of the two geographical populations. The Swakara populations would likely have had less genetic gain and as such flocks are likely to have some genetic similarity (Groenewald *et al.* 2010). The breeding history of a flock also contributes to its population structure and the divergence between individuals from different stations could be due to isolated breeding and the difference in breeding management between stations. The second PCA separated populations according to pelt colour. The white-vital and some of the white sub-vital sheep clustered together forming an isolated cluster around 0.15 and -0.1 coordinates (Figure 4.5 A). The other white sub-vital individuals also formed a second cluster together with some of the black individuals (Figure 4.5 C). The grey individuals were dispersed over a wider admixed cluster with traces of black and sub-vital individuals in that cluster (Figure 4.5 B).

F_{ST} and PCA-based results revealed that the vital-white subpopulation was more genetically similar to the grey population and more distinct from the black subpopulation. By comparison the white subpopulation shared a selection signal with the sub-vital white population at chromosome 10 that had

an F_{ST} value of 0.4908 but the SNP (OAR10_87219588) had no associated gene function to it. The white and grey subpopulations shared an F_{ST} value of 0.3772 for a SNP (s53466.1) on chromosome 7 but had no function or gene associated to it. The highest fixation index (F_{ST} = 0.6164) between the subpopulations that shared the least genetic similarity was on SNP OAR13_16487810.1 on chromosome 13 that had no gene function associated to it.

This supports the within subpopulation genetic diversity analysis that revealed low genetic diversity in the white and black subpopulations. Highly inbred or least diverse populations will be more diverged from other populations because of the reduced allelic variation within them. The South African Karakul industry was developed from flocks taken from Namibia (Campbell 2007) and therefore would be expected to be more genetically similar than the Halle group. The analysis based on location showed a more hierarchical structure and admixed clusters were observed in which the Gellap-ost, private farms and Northern Cape individuals were grouped together. In a study of global sheep populations, PCA analysis grouped individuals of same breed and geographic location such as Africa, East Asia and North American breeds together (Kijas *et al.* 2009). In our study population we found that the Halle group diverged significantly from the Namibia and South Africa populations. A principal component analysis for the African populations that excludes the Halle group would be important in order to look at how much the South Africa and Namibia populations differ at a geographic location level.

Bayesian-based analysis showed some overlapping genomic regions between subpopulations particularly the vital-white and sub-vital white populations as well as the black and white sub-vital sheep. ADMIXTURE is a model-based clustering method that partitions the genome of each individual into predefined number of components. Population sub-structuring is examined using a clustering algorithm. The analysis for the population structure in Swakara sheep using ADMIXTURE was similar to that of PCA that revealed how closely related the subpopulations were to each other. The black and the sub-vital subpopulations shared some genomic regions that were different from the rest of the subpopulations. The Swakara population is more genetically similar amongst the subpopulations. The

extensive breeding on each subpopulation may have resulted in the admixture seen in Swakara population. The grey subpopulation also experiences the lethal factor that affects growth and development and as such it is usually cross-bred to eliminate the factor (Schoeman 1998; Rothauge 2009) which might explain the genetic diversity in this subpopulation due to the breeding practices i.e. avoiding grey x grey mating.

4.6 Conclusions

The population structure of Swakara has little genetic differentiation when considering the entire population. The Swakara colour subpopulations showed little divergence in population structure and between the individuals collected from different geographical regions. The similarity between the South African and the Namibian populations shows they are the product of similar ancestry even showing more divergence to the Karakul population from Germany. This supports historical accounts of how the breed was developed and was shared across the two countries while also having undergone years of selection and cross-breeding, which supports the account that the Swakara breed was developed from the Karakul.

CHAPTER 5: Inbreeding levels and distribution of Runs of homozygosity in Swakara sheep

5.1 Abstract

Swakara sheep experience a sub-vital trait in their white subpopulation that affects their production performance. The sub-vital performance is thought to be due to the negative impacts of years of inbreeding. Runs of homozygosity (ROH) are long stretches of contiguous DNA fragments that are homozygous, occurring as a result of parents transmitting similar haplotypes to their offspring due to shared ancestry. This study aimed to investigate the levels of inbreeding and also screen for extended ROHs in the different Swakara subpopulations. Additionally the study investigated genes found within the observed ROHs and made inferences on associations with sub-vital performance. Genomic DNA from 90 Swakara individuals that belonged to the grey (n = 22), black (n = 16) & white (n = 35) and white sub-vital (n = 17) were genotyped using the OvineSNP50 beadchip. PLINK v.1.07 was used to assess the quality of SNPs and prune those with a missing genotyping rate > 0.01, SNPs and individuals with low genotype calls of > 0.1, MAF < 0.01 and those that deviated from HWE < 0.0001. A total of 43 849 from 54 241 SNPs after pruning was used for further analysis. The inbreeding levels and ROHs were estimated using the PLINK genome wide analysis tool. *Ovis aries* genome browser v.3.1 and *ENSEMBL* were used to determine the putative biological functions of SNPs located within ROH segments. The inbreeding coefficient F_{IS} was high in the black (0.04798±0.069) subpopulation and lowest in the grey (0.01074±0.079) Swakara sheep. Forty-two unique ROH regions were observed on 10 chromosomes in 33 individuals that spanned between 5198.93-7126.85 KB. The white sub-vital group had the highest number of ROHs at 18. Seven overlapping regions of homozygosity were identified on chromosomes 2, 7, 9, 10 and 20. Region cROH1 on chromosome 9 was the most common overlap found in 10 individuals, and the SNPs on this region were associated with *TG* (thyroglobulin) and *SLA* proteins. There was no correlation between the frequency of ROHs in an individual and the level of inbreeding. The black subpopulation however, had the highest level of inbreeding as well as the highest average length of ROHs among the other subpopulations. Larger sample sizes and more particularly of the sub-vital would give a more conclusive association of ROH to sub-vitality.

Keywords: **sub-vital, runs of homozygosity, inbreeding, pelt sheep**

5.2 Introduction

The sub-vital factor causes some of the white Swakara lambs to succumb to death in less than 48 hours of birth. The failure to survive past the weaning stage is due to metabolic complications where lambs are unable to digest feed (Greeff & Faure 1991; Groenewald *et al.* 2010). In practice, farmers avoid breeding pure white x pure white Swakara, as the offspring of such a breeding is likely to express sub-vital performance. Hence, a breeding flock would mainly consist of black rams and white ewes as the mating between these reduces the occurrence of sub-vital performance (Rothauge 2009). One of t possible reasons for the appearance of the sub-vital factor in white Swakara could be due the accumulation of a homozygous recessive allele. The production of white Swakara sheep is the product of generations of intensive selection and likely extensive inbreeding, resulting in the accumulation of homozygous recessive alleles. The hypothesis is that sub-vital performance is the result of inbreeding in the Swakara population that has led to the accumulation of lethal homozygous alleles in the white population and not in the other pelt colour populations. There is no information on the inbreeding levels of Swakara available in the public domain. An understanding of the genomic structure and diversity amongst the vital-white ,sub-vital white and two other Swakara colour variants is important in determining the genetics of sub-vital performance in order to aid in the breed improvement programmes of Swakara.

Populations that have largely homozygous genomes are known to harbour variants linked to genetic disorders and other diseases (Purfield *et al.* 2012). Recessive, autosomal homozygous alleles expressed in highly inbred animals are largely the reason for decreased production performance (VanRaden *et al.* 2011). Inbreeding depression describes reduced production fitness and the risk of inheriting genetic defects as a detrimental consequence of mating closely related individuals (Howrigan *et al.* 2011; VanRaden *et al.* 2011). Runs of homozygosity are defined as long and continuous stretches of homozygous genotypes (Ku *et al.* 2011) that result from parents transmitting identical haplotypes to

their offspring. The extent and frequency of ROHs informs the ancestry of individuals and of a population, for example long stretches of ROH allude to recent inbreeding events in the population as a result of identity by descent (Ku *et al.* 2011; Purfield *et al.* 2012). Recent studies have found associations of regions of extended homozygosity to complex and simple gene disorders (Yang *et al.* 2010; Ku *et al.* 2011; Kijas *et al.* 2012; Suarez-Vega *et al.* 2013). Through the ROH association approach, it has become possible to identify loci harbouring recessive variants of complex diseases and traits (Ku *et al.* 2011).

Due to the small population sizes of Swakara breeding flocks and intense selection, high levels of inbreeding and longer extended ROH could be occurring at a higher frequency and also contributing to the expression of the sub-vital factor (Coertzen 2009). The breeding community of Swakara is relatively small and it is possible that mostly animals of superior economically traits have been used in reproducing subsequent generations thus increasing chances of having homozygous or closely related individuals. In this study levels of inbreeding were determined and the genome of Swakara sheep was screened for the occurrence of ROHs. The aims of the study were therefore to:

(i) investigate levels of inbreeding in the different Swakara sheep colour subpopulations,

(ii) screen for ROHs in the Swakara sheep population using the Ovine SNP50K data and

(iii) investigate associations between ROH and the coat colour patterns and sub-vital performance in Swakara sheep.

5.3 Materials and methods

5.3.1 Animals

Blood was collected from 90 Swakara sheep sampled from research stations and farms in Namibia (n = 60) and a research station in South Africa (n = 30) i.e. the 6 individuals from Halle, Germany were removed from further analyses. The 90 animals consisted of four colour subpopulations *i.e.* black (n = 16), grey (n = 22), vital-white (n = 35) and sub-vital white (n = 17). DNA isolations were done using a commercial Blood and Tissue kit from Macherey-Nagel. The DNA quantification was done using Qubit

flourometer that measured, DNA concentration ranged between 37.5 to 141.1ng/µl. The ($A_{260/280}$) ratio was between1.8-2.0. All samples were recorded as having good DNA quality and concentration. The DNA concentrations were adjusted to 40ng/µl for SNP genotyping in order to normalize for the call of genotypes across the study population.

5.3.2 SNP genotyping

The OvineSNP50 beadchip was used for genotyping of the 90 Swakara sheep. The chip has 54 241 SNPs. The call rate was above 98% for all individuals. SNP quality control parameters imposed on the genotyped individuals included deviations from Hardy-Weinberg Equilibrium > 0.001, minor allele frequency < 0.01, missing genotypes > 0.01 and individuals that failed genotyping > 0.1. After QC parameters, at least 80.35 % of the SNPs were left for further analysis. Two individuals (one white sub-vital and one black) were excluded from downstream analysis leaving 88 animals for further analysis.

5.3.3 Runs of homozygosity

Runs of homozygosity (ROH) were determined using PLINK v.1.7 (Purcell *et al.* 2007), following an algorithm that considers a window of *x* SNPs and slides it across the ovine genome. At each window, the extent of homozygosity of the region is determined. PLINK allows for screening of ROH within an individual by defining a run as the number of homozygous SNPs encompassing a certain distance in kb. It follows an algorithm that takes approximately 50 SNPs and slides the window across the genome at given SNP intervals. With every sliding window position, the homozygosity of that position is determined by comparing the signal intensity of that allele. The proportions of homozygous windows that overlap the window position at a particular SNP are then calculated.

In this analysis, the ROHs were defined as homozygous stretches along a 1000kb sliding window with a minimum of 100 SNPs allowed in the tract. This window allowed at least 2 missing SNPs and 1 heterozygote. The gap allowed between 2 consecutive SNPs was 50kb. These parameters were set to avoid redundancy and underestimation of ROHs that could otherwise result from genotyping errors or

missing genotypes. The parameters defined for the algorithm should allow for missing calls or heterozygotes that may otherwise break up a true ROH, rendering it undetectable.

Pools of overlapping and potentially matching segments were also estimated using PLINK. An overlapping region of homozygosity was determined as the consensus region between individuals i.e. a common region found in different individuals on the same chromosome. A region of overlap was confirmed if at least 95% of jointly non-missing, homozygous segments were identical. The PLINK – *group* function generated a pool of individuals that share a consensus ROH (cROH) on the same chromosome and within an overlapping region as well as its distribution in the sub-vital and vital sheep.

5.3.4 Identifying potential trait association of SNPs in ROH segments

The SNPs located within the consensus ROH segments were individually investigated for the putative genomic functions using the Ovine assembly genome browser v.3.1 (www.livestockgenomics.csiro.au/sheep/oar3.1.php) and *ENSEMBL* (www.ensembl.org/sheep). The genomic base location of each SNP was used to pinpoint the SNPs and identify their associated biological function or absence thereof.

5.4 Results

5.4.1 SNP genotyping quality control

Quality control parameters removed two individuals due to low call rates and 7399 markers that had MAF below 0.01. In total, 10 657 SNPs were removed from further analysis after failing to meet the QC parameters, leaving 80.35% markers for use in heterozygosity and inbreeding coefficient analysis as well as determination of ROHs in the Swakara population.

5.4.2 Inbreeding levels within individuals of the different subpopulations

In the four subpopulations, the grey subpopulation had the lowest F_{IS} value (0.01074±0.069) (Figure 5.1). The black subpopulation had the highest F_{IS} value = 0.04798±0.079, making it the most inbred

population relative to the rest in the group. Inbreeding levels in the vital white and sub-vital white were 0.03501±0.0618 and 0.02152±0.1080, respectively.

The ten white vital individuals from Northern Cape (NC) had the least inbreeding (Figure A1, Appendix A), within this location group WV24 had the highest F value of 0.1317, while WV10 had the lowest of F = -0.06064. Sheep WV10 from the Northern Cape subgroup had the lowest (F_{IS} = 0.002) values. WV85 from Gellap-ost (GO) location had the highest F_{IS} value (F_{IS} = 0.1567) within the white vital population. The sub-vital white individuals were only collected from GO and Private Farm (PF) in Namibia. The Gellap-ost sub-vital population had the highest inbreeding coefficient on individual WSV82 (0.3636) and the lowest F_{IS} values were by two sub-vital individuals (Figure A2, Appendix A). Two sub-vital white sheep, WSV38 and WSV42, sampled from the private farms had low F_{IS} values of -0.044 and -0.048 respectively. The individuals from the private farms had inbreeding coefficients below 0.15 while the 5 individuals from GO had higher F_{IS} values.

Inbreeding was high within the black Swakara sheep (Figure A3, Appendix A). The Northern Cape subpopulation (NC), had the highest F_{IS} value of 0.28 for individual BV14 while the other individuals within the NC population had lower F_{IS} values with one individual, BV21 reporting a very low F_{IS} value of -0.0116. Average inbreeding levels of the Gellap-ost subpopulation were below 0.25. Figure A4 (Appendix A) illustrates the inbreeding levels in the grey Swakara from the Northern Cape and Gellap-ost. The Gellap-ost individuals (G62-G72) all except one (G61) were observed to have very low levels of inbreeding. G61 had an F_{IS} value of 0.0917 while other individuals between G62 and G72 had values ranging from -0.01075 to -0.09483. Comparatively, the individuals from NC had high levels of inbreeding. The NC population had high F_{IS} values with the highest observed being individual GV32 (0.1418). Two individuals (GV13 and GV34) had low inbreeding values in the negatives.

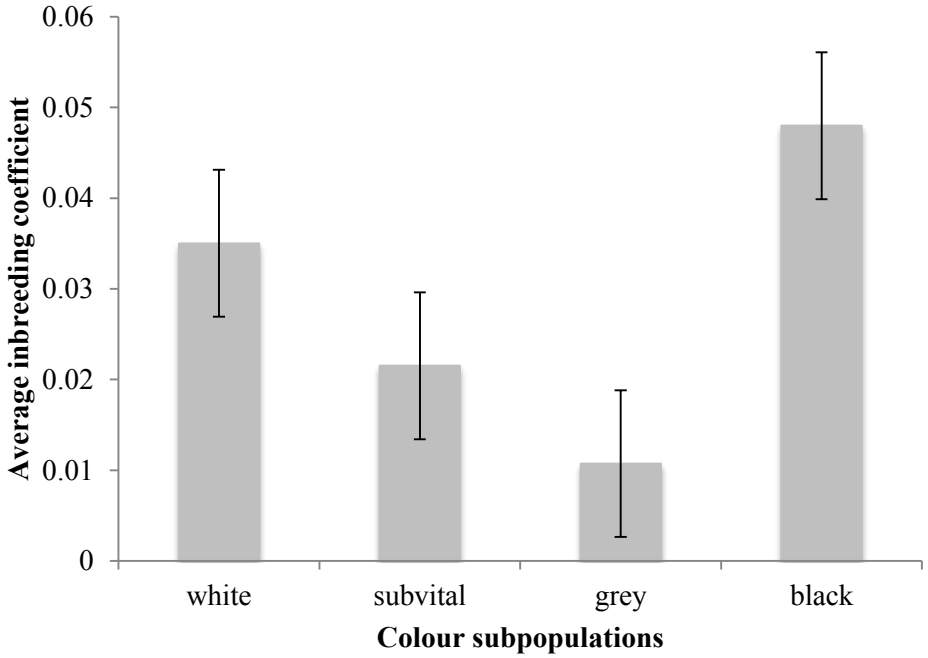

Figure 5.1: The average inbreeding coefficient *F* in all four Swakara subpopulations
**The inbreeding coefficients of each individual are located in the appendix.*

5.4.3 ROH prevalence and patterns in Swakara sheep

A total of 42 unique ROH were observed in 33 Swakara sheep. The number of ROH per individual sheep ranged between 1 and 3 (Table 5.1). Eight ROHs were observed in 7 white sub-vital sheep. Six and five ROH were observed in the grey and black subpopulations, respectively. The vital white subpopulation had 18 ROHs in 15 individuals. The individual with the most ROHs reported is BV14 from the black subpopulation that had three ROHs. The average length of ROH of sheep BV14 was 5606.72±607.47 KB (Table A1, Appendix A). Individuals with the highest frequency of ROH (Table 5.2) segments (≥2) were BV14 (average length = 5606.72±607.47 KB), BV96 (average length = 6104.55±125.28 KB), GV17 (average length = 4881.61±45.39 KB), GV72 (average length = 5081.75±165.72), WSV84 (average length = 7936.85±2344.42 KB), WV22 (average length = 6293.63±503.50 KB), WV3 (average length = 5077.93±338.34 KB) and WV46 (average length = 6782±2141.96 KB). Table A1 illustrates the length of ROH in the 33 individuals reported for ROH.

Table 5.1: The number of ROHs observed in the different subpopulations

Subpopulation	No. of ROH	No of individuals	Ave. length of ROH (KB)
Black (B)	8	5	9552.50±5040.15
Grey (G)	8	6	7569.46±2393.70
White vital (WV)	18	15	7066.84±2835.09
White Sub-vital(WSV)	8	7	7253.35±3842.94

5.4.4 Chromosomal coverage of ROH in Swakara population

Table 5.2 illustrates the number of ROHs per chromosome and their average lengths. Only eleven chromosomes (1, 2, 3, 5, 6, 7, 8, 9, 10, 12, and 20) out of 26 (excluding the *X* chromosome) had ROH. The ROH length averaged between 5198.93 KB and 7126.85±1516.26 KB. The ROHs in this study covered at least 0.2282% of the sheep genome.

Chromosome 8 only had one ROH observed in individual GV72. This ROH was the shortest observed across chromosomes and individuals. Chromosome 7 had 4 ROHs that averaged 7126.85±1516.26 KB, which were the highest average ROH length observed on a chromosome. Chromosome 9 had the highest number of 14 unique ROHs that ranged from 4849.51KB (GV17) to 9290.12 KB (GV35). Two black Swakara sheep (BV14, BV83), 4 grey (GV17, GV35, GV62, GV72), 1 white sub-vital (WSV51) and 7 white-vital (WV24, WV3, WV33, WV46, WV60, WV85, WV92) individuals had ROHs on this chromosome. The average ROH length on chromosome 9 was 5994.42±1411.95 KB. The distribution of ROHs in the individuals is summarized in Table A1.

Table 5.2: The average ROH length per chromosome and the occurrence of ROH on each chromosome

CHR	Ave. length (KB)	SD	No. of ROHs	CHR length (Mbp)	% of chromosome
1	5547.67	*	1	299.628	0.0018
2	5465.64	508.71	6	263.099	0.0124
3	5388.14	777.04	2	242.728	0.0044
5	5467.78	282.02	2	116.288	0.0094
6	5235.90	*	1	129.02	0.0041
7	7126.85	1516.26	4	108.869	0.0262
8	5198.93	*	1	97.681	0.0053
9	5994.42	1411.95	14	100.623	0.0834
10	5967.47	1520.13	8	94.082	0.0507
12	6649.65	*	1	85.913	0.0077
20	6324.17	435.87	2	55.380	0.0228
					Total: 0.2282%

**indicates no standard deviation as only one segment is identified in the chromosome and therefore the average cannot be calculated.*

5.4.5 Consensus (overlapping) ROHs observed in Swakara population

The 42 unique ROHs grouped into 7 main consensus ROHs (cROH). Some of these regions only occurred in sub-vital sheep while others occurred in both vital and sub-vital sheep. Other segments were shared among individuals of sub-vital and vital (Table 5.3). Chromosome 9 had the most consensus ROHs (n = 3). cROH1 (on chromosome 9) was the most abundant cROH with 10 individuals making it the most common ROH, while chromosome 20 (cROH5) had the least individuals. cROH3 was shared

between black and white (vital) sheep and cROH6 only a white sheep was identified. cROH4 and cROH7consensus regions were shared between white only individuals (vital and sub-vital).

Table 5.3: The consensus regions shared between individuals colour subpopulations and their respective chromosomes

Consensus Region	CHR	No. of Individuals	Vital/Sub-vital	Colour subpopulation
cROH1	9	10	Vital	BV, GV, WV
cROH2	10	8	Both	BV, GV, WSV
cROH3	7	4	Vital	BV, WV
cROH4	2	4	Both	WV, WSV
cROH5	20	2	Both	BV, WSV
cROH6	9	2	Vital	WV
cROH7	9	2	Both	WV, WSV

The longest consensus segment of 6434.68 KB belonged to cROH7 (Figure 5.2) and had 2 white individuals (WV46 and WSV51) that were both sampled from private farms. This region is located on chromosome 9. cROH2 had the shortest stretch of homozygosity accounting only 1020.32 KB in length. This region had 8 individuals sharing the consensus region on chromosome 10, shared between BV14 (NC), GV15 (NC), WSV53 (PF), WSV84 (GO), WV48 (PF), WV90 (GO), WV88 (GO) and GV17 (NC) i.e. three Northern Cape individuals, two private farms and three from Gellap-ost. The segmentcROH1 was the most prevalent ROH segment that was shared between the most individuals consisting of 10 Swakara animals. This consensus region was located on chromosome 9 and was 3190.87 KB in length.

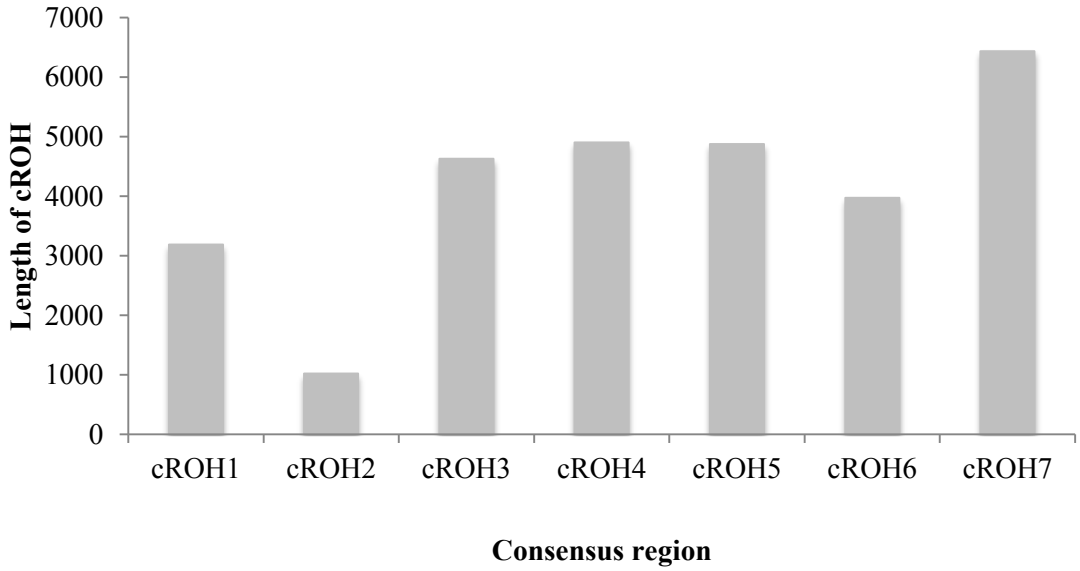

Figure 5.2: The length of consensus runs of homozygosity found between subpopulations of Swakara

5.4.6 Putative function/association of observed ROH

Table 5.4: SNPs marking homozygosity blocks and their biological associated functions

cRegion	First SNP	GENE	Second SNP	GENE
cROH1	OAR9_22183589.1	*TG and SLA*	OAR9_25374455.1	N/A
cROH2	OAR10_83473815.1	*PCCA*	OAR10_84494139.1	N/A
cROH3	OAR7_181698.1	N/A	s63040.1	*HMGCR*
cROH4	s33265.1	N/A	OAR2_4905166_X.1	N/A
cROH5	OAR20_40664275.1	N/A	OAR20_45542235.1	N/A
cROH6	OAR9_80693962.1	N/A	s16612.1	*CPQ*

**N/A - no biological function association was identified*

The first SNP on the consensus cROH1 on chromosome 9 was linked to Thyroglobulin *(TG)* and Src-like-adapter 1 (*SLA1*) genes (Table 5.4). On consensus cROH2, found on chromosome 10 and the first SNP was associated to *PCCA*, a propionyl CoA carboxylase gene while the last SNP had no associated

gene or function. The last SNP of cROH3 found on chromosome 7 was associated with *HMGCR* (3-hydroxy-3-methylglutaryl)-CoA reductase), last SNP on cROH6 was associated with *CPQ* (carboxypeptidase Q). The remaining segments cROH4, cROH5, and ROH7 had no associated functions on either their first or last SNPs.

5.5 Discussion

The advent of high-density SNP genotyping arrays has facilitated contemporary methods of estimating inbreeding such as measuring the loss of heterozygosity and inbreeding levels which have been estimated using high-density genotyping arrays in a number of studies (Keller & Waller 2002; Fan *et al.* 2010). Molecular inbreeding coefficient defines the probability of two alleles on a locus of an individual are identical by descent (Keller & Waller 2002). Runs of homozygosity are expected to be more frequent and longer in populations that have experienced recent inbreeding events (Kirin *et al.* 2010). Previous studies have associated ROH to economically important traits and in this study ROHs were analysed in the context of inbreeding in Swakara subpopulations while the patterns of ROH and putative functions were determined.

The size of a population also affects the allele frequencies which in turn reduce heterozygosity (Allendorf 1986). The loss of genetic variation when certain alleles are fixed results in reduced fitness of a population that is likely to result in the accumulation of lethal recessive allele known as inbreeding depression (Vrijenhoek 1994). The Swakara breeding flocks of Namibia and Northern Cape, South Africa are both small and close together, with intensive selection could have resulted in high levels of inbreeding and accumulation of homozygous loci.

Previously ROHs have been used in genome wide association studies to identify genes/genetic variants that could be associated with a disease of interest (Karimi 2013). In this study, we set out to find genes or other genetic variants within a screened ROH that may be associated to the sub-vital performance in the Swakara sheep using the ovine genome browser and *ENSEMBL* genome-browsing platform. ROHs

were observed in each colour subpopulation and the white vital subpopulation had the most ROH (n = 18) in 15 individuals. The rest of the colour subpopulations each had 8 ROHs reported at an average of 1 ROH per individual. The black population had the highest average length of ROH blocks that were 9552.5±5040.15 KB long while the white individuals had the lowest average length of 7066.84±2835.09 KB.

Along with having the most ROH segments, BV14 also had the highest F_{IS} (0.2805) value in the black population showing some correlation between ROH and inbreeding. Individual GV72, however, had 2 ROH segments and an F_{IS} of -0.0209. The black subpopulation had the highest inbreeding coefficient of 0.0496 as well as largest number of ROHs. It was suggested in a study on the effects of population history on runs of homozygosity that ROHs in outbred populations could be attributed to unusual mutations, linkage disequilibrium or low recombination rates (Kirin *et al.* 2010). The inheritance of haplotype blocks within an individual due to strong linkage disequilibrium could be the reason ROH were observed in outbred sheep (those with negative inbreeding coefficients). There was no clear correlation between ROH and inbreeding coefficient. Population bottlenecks, selection pressures and breeding management all affect the pattern and length of ROHs (Curtis *et al.* 2008)..

There were seven overlapping/consensus ROH (cROH) in the Swakara population. The length of cROH1 was 3190.87 KB. The first SNP on the consensus cROH1 on chromosome 9 was linked to Thyroglobulin (*TG*) and Src-like-adapter 1 (*SLA1*) genes. The thyroglobulin gene is responsible for the thyroid gland regulation of hormones and mutations on this gland upset the normal chemical reactions of metabolic activities causing hypothyroidism in sheep (Hetzel & Mano 1989). Consensus cROH2 is found on chromosome 10 and the first SNP was associated to *PCCA* (propionyl CoA carboxylase) gene while the last SNP had no associated gene or function. The second SNP of cROH3 found on chromosome 7 was associated with *HMGCR* (3-hydroxy-3-methylglutaryl)-CoA reductase), SNP2 on cROH6 was associated with *CPQ* (carboxypeptidase Q). The associated gene was *HMGCR*, which codes for 3-hydroxy-3-methylglutaryl-CoA reductase that is an enzyme responsible for the biosynthesis of

cholesterol (Do *et al.* 2014). In pigs, a mutation on the *HMGCR* gene was found to be associated with intramuscular fat content along with cholesterol biosynthesis (Hetzel & Mano 1989). The remaining segments cROH4, cROH5, and cROH7 had no associated functions on either their first or last SNPs. cROH7 was shared between two sub-vital white and vital white individuals and was the longest cROH segment, 6434.68 KB and cROH2 was the shortest consensus region of 1020.32 KB. In this study there were no ROHs that were unique to the sub-vital white subpopulation and therefore no gene could be directly associated to the sub-vital trait conclusively as in order for an association to be declared, a cROH must be of considerate length and occur in higher frequency in the cases than in the controls. No such cROH was identified in this Swakara population.

5.6 Conclusions

The level of inbreeding in the Swakara population was low across the subpopulations with the black population showing the highest level of inbreeding. This can be expected as the genetic diversity in the subpopulations was moderate. The occurrence of ROHs is often attributed to high levels of inbreeding or a recent mating of closely related individuals. However, Swakara individuals showed a low occurrence of ROHs that only accounted for 0.2282% of the genome. There was no consistent correlation between the frequency of ROHs in an individual and the level of inbreeding, however the black subpopulation had the highest level of inbreeding and also had the highest average of ROH length among the other subpopulations. Overlapping runs of homozygosity in Swakara were identified; however no putative function could be associated to sub-vitality. A more focused approach is necessary to identify the SNPs that are in close proximity to ROHs or the genes that lie within the tract of ROH as well as an increase in sample size particularly of sub-vital sheep.

CHAPTER 6: GENOME-WIDE ASSOCIATION STUDY OF SUB-VITAL PERFORMANCE IN SWAKARA SHEEP USING THE OVINESNP50 GENOTYPING ARRAY

6.1 Abstract

The sub-vital factor that causes premature death in white Swakara lambs is a major production constraint for breeders as it reduces profit margins and the number of replacement ewes in breeding flocks. In this study, the Illumina OvineSNP50 array was used to conduct a genome-wide association study for sub-vital performance in white Swakara sheep. The GWAS revealed at least five significant SNPs on chromosome 3, 5, and 8 when the sub-vital white sheep were compared against all other Swakara sheep from the black, grey and vital white subpopulation. Chromosome 3 had three SNPs associated to sub-vitality. The most promising SNP was located on chromosome 3, which is associated to a gene, *IGF1*, responsible for insulin-like growth factor and contributes to the development of foetal organs. Another GWAS involving only the vital white and sub-vital white revealed no significant SNPs associated to sub-vitality.

Keywords: GWAS, sub-vitality, OvineSNP50, Manhattan Plot

6.2 Introduction

A GWAS is a genetic study design that attempts to identify genetic variants commonly occurring and contributing to a given phenotype such as susceptibility or resistance to disease (Anderson *et al.* 2010; Bush & Moore 2012). Genetic association studies have served as platforms for correlation between disease status and genetic variants in identifying candidate genes or genomic regions that contribute to specific diseases (Lewis & Knight 2012). Genome wide association studies are the preferred modern tool when there is a need to investigate the genetic basis of complex diseases or traits. The advantage of using GWAS over traditional methods is the power to detect causal variants that have minimal effect on

the trait of interest while defining narrower genomic regions that carry causal variants and also providing a glimpse of the genetic mechanisms that govern complex traits (Zhang *et al.* 2012).

The availability of genome-wide SNP data from highly automated genotyping platforms makes it possible to undertake genetic association studies for genetic disorders (Fan *et al.* 2010; Stranger *et al.* 2011; Zhang *et al.* 2012). The OvineSNP50 beadchip is a comprehensive, genome-wide SNP array that was developed for *Ovine aries* (sheep) and other related species (*Ovis musimon, Ovis vignei* and *Ovis dalli* etc.) for genome wide population genomic studies. The OvineSNP50 beadchip has been used in assessing copy number variations, fine mapping of QTLs, the global divergence of sheep breeds (Kijas *et al.* 2012) and genomic selection in the New Zealand sheep industry (Auvray *et al.* 2011), and GWAS looking at traits such as milk production traits, paratuberculosis and chondrodysplasia (Suarez-Vega *et al.* 2013).

The causal variant of sub-vitality in Swakara is unknown. The process of determining the susceptibility to sub-vitality is time consuming and costly. The ability to determine susceptibility of white sheep to sub-vitality is important to the successful breeding and production of white pelt at limited costs. Determination of the causal variant of sub-vital performance in Swakara will shed light on its genetic causes and in finding markers that may be used in selection against sub-vital performance when breeding for production. The main aim of this study was to conduct a GWAS identifying SNP(s) that are associated with sub-vitality performance affecting the white pelt Swakara sheep. This was based on the hypothesis that sub-vitality only affects some individuals of the white pelt colour subpopulation and therefore the genetic variant(s) will occur in high frequencies in the sub-vital population.

6.3 Materials and methods

6.3.1 Animal population, SNP genotyping and quality control

The Swakara sheep used and the genotyping and SNP quality controls for this study were described in previous chapters. The sub-vital trait is observed in white lambs alone however it does not affect all the individuals of this colour population. It was therefore worth noting the differences between the normal white individuals and the sub-vital white individuals in a GWAS study. The 90 animals were divided into two categories (Table 6.1) and a GWAS was conducted using data sets from each category. The first group call *AA* (All Animals) consisted of all 90 animals divided into 17 cases that were the white sub-vital and 73 controls belonging to the grey, black and white vital subpopulations. The second group named *WO* (White Only) consisted of only the white sheep and subdivided into 17 cases of the white sub-vital sheep and 35 controls of vital white individuals. Table 6.1 describes the sample size within each group and in either cases or controls and also the number of SNPs and individuals that were pruned out during quality control. The imposed QC parameters included (i) Hardy-Weinberg Equilibrium (*HWE* > 0.0001), (ii) minor allele frequency (*MAF* < 0.01), (iii) individuals with low genotyping (*MIND* > 0.1) and (iv) SNPs failing the missingness test (*GENO* > 0.01).

Table 6.1: Number of SNPs and individuals removed after QC parameters in each group

Group	(n)	Cases	Controls	SNPs after QC	MIND	GENO	HWE	MAF
(AA)	90	17	72	43 584 (80.35%)	2	6527	250	7399
(WO)	52	17	35	43 624 (80.42%)	1	4281	229	9413

In the *AA* group, two individuals were removed (one sub-vital white and one black individual) after quality control, whilst one sub-vital individual was pruned due to low genotyping success rate in the *WO* group. The *AA* had 80.35% of the SNPs (n = 43 584) remaining and the *WO* group had 80.42% of SNPs (n = 43 624) remaining after QC.

6.3.2 Genome wide association analyses

A basic test of association involving comparing of allele frequencies between cases and controls across the 26 autosomes and the *X* chromosome on the OvineSNP50 array was conducted using algorithms implemented in PLINK v.1.07. This basic association test was performed with permutations to provide empirical *p*-values to determine the significance of the variance in allele frequency.

A standard case/control association analysis was done for each of the *AA* and *WO* population categories. The phenotype being tested for association was the sub-vitality trait and this was done by making the white sub-vital the cases in all tests. The *--assoc* command generated an output reporting results on the SNP name, chromosomal and base position of the SNP and the associated *p*-value. An additional *--ci* was included to determine the 95% confidence intervals for odds ratios. The output included allele counts for both cases and controls for all SNPs used in the analysis.

6.3.3 Visualization GWAS data

The *p*-values generated in PLINK were first converted to the $-\log_{10}$ (*p*-value) which were then used to construct Manhattan plots for the visualization of results. The Manhattan plots were plotted for each $-\log_{10}$ (*p-value*) of each SNP interrogated on the bead chip (Y-axis) versus the SNP chromosomal position (X-axis) using the *qqman* package function implemented in *R* library. Thresholds association were as set at $-\log_{10}$ ($5e^{-8}$) for significant association and $-\log_{10}$ ($1e^{-05}$) for suggestive association.

6.3.4 Putative functions of SNPs identified

The biological functions related to the significant SNPs discovered were investigated by using the SNP IDs to search for genes in *ENSEMBL Genome Browser* (Flicek *et al.* 2014), using the sheep genome browser and the sheep reference genome. *ENSEMBL* allows the user to view the SNPs identified in a genomic context by showing the SNP in the context of other genes that lie along the same track on the chromosome and also by providing various tabs and links to other databases such as NCBI and the reference genome of the species being studied.

6.4 Results

6.4.1 The distribution of genotyped SNPs using QQ plots

QQ plots were generated for each study group to look at the distribution of SNPs across the ovine genome in order to test the null hypothesis of no "association" in the AA group (Figure 6.1) and the *WO* group (Figure 6.2). The SNPs in the *AA* group (Figure 6.1) did not follow a normal distribution and deviated from the line of expected distribution (red diagonal).

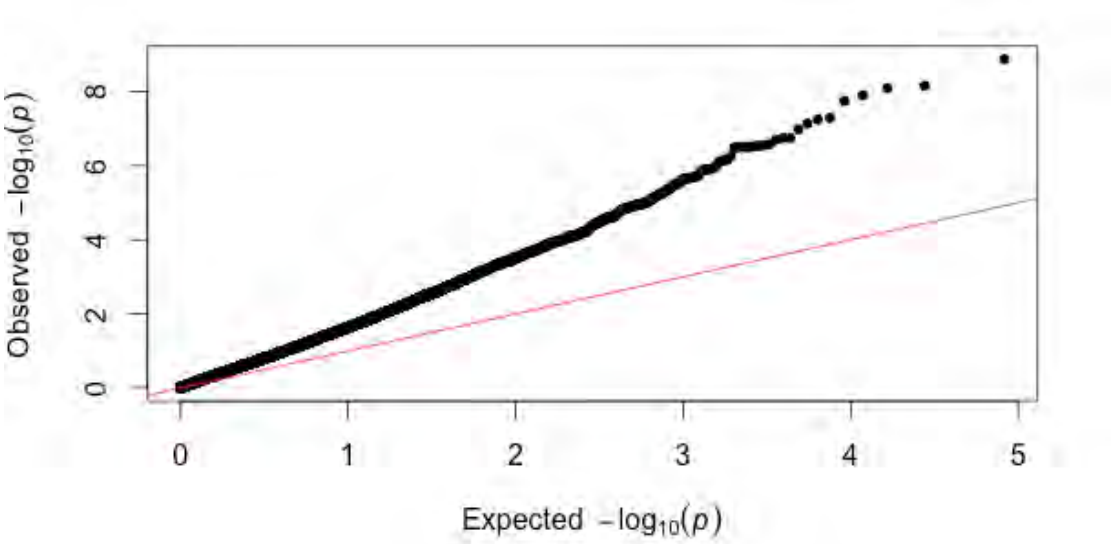

Figure 6.1: QQ plot showing the distribution of SNPs in the *AA* Swakara study population

The SNPs in the *WO* group also did not follow a normal distribution across the expected and observed comparisons (Figure 6.2).

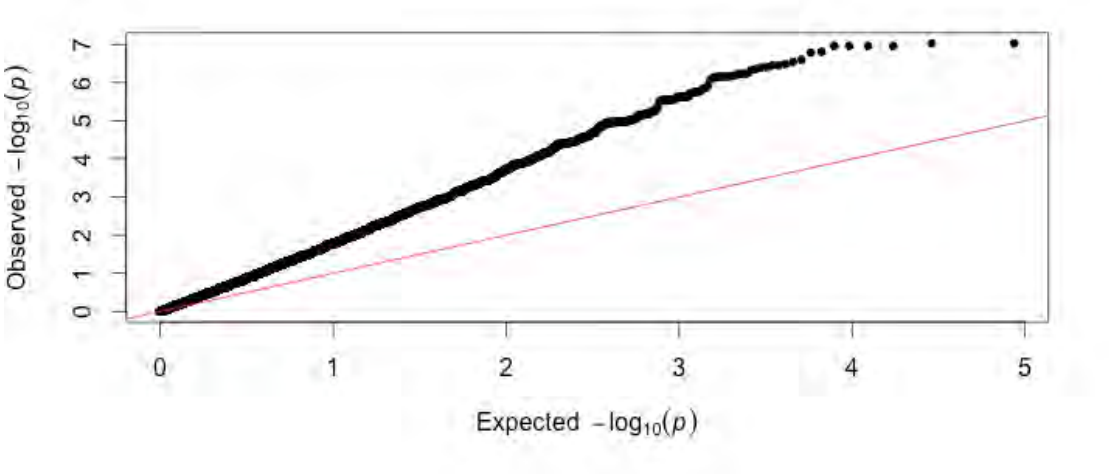

Figure 6.2: QQ plot showing the distribution of SNPs in the *WO* Swakara study population
When SNPs do not follow a normal distribution it suggests that there is some level of association that can be expected from the SNPs that deviated the most from the distribution.

6.4.2 Genome-wide association analysis

The AA group had five SNPs that showed a high level of significant association. Two significant SNPs were found on chromosome 3, one significant SNP on chromosome 5 and two on chromosome 8 that were observed with SNPs having *p*-values over the set threshold value of significant association. The SNPs identified and the group they were found in is shown in Table 6.2. The *WO* group had no SNPs identified that could be associated to sub-vitality.

Table 6.2: Associated SNP comparison across two study groups

CHR	SNP ID	*p*-value	Cases	Controls	AA	WO
3	OAR3_184018243.1	1.35 e^{-09}	0.3125	0.01389	Y	N
	s24957.1	8.11e^{-09}	0.4062	0.04861	Y	N
5	OAR5_55263158.1	1.28 $e^{-08€}$	0.9375	0.3819	Y	N
8	OAR8_42787515.1	7.03 e^{-09}	0.3438	0.02778	Y	N
	OAR8_44102022.1	1.81 e^{-08}	0.375	0.04167	Y	N

The association results were visualized using Manhattan plots that displayed the distribution of SNPs across the chromosomes of *Ovis aries* and projecting the –log10*p* values of the *p*-values determined by the association function in PLINK v.1.7. The Manhattan plot shows significant SNPs in of the *AA* population category (Figure 6.3) and *WO* group (Figure 6.4).These are SNPs that have values above the $-\log_{10}(5e^{-8})$ threshold which is represented by the significant association line (red line). The blue line on the Manhattan plot is the line of suggestive association with a threshold of $-\log_{10}$ $(1e^{-05})$. These thresholds are used to determine the genome-wide association of SNPs to the traits of interest in this study being sub-vitality.

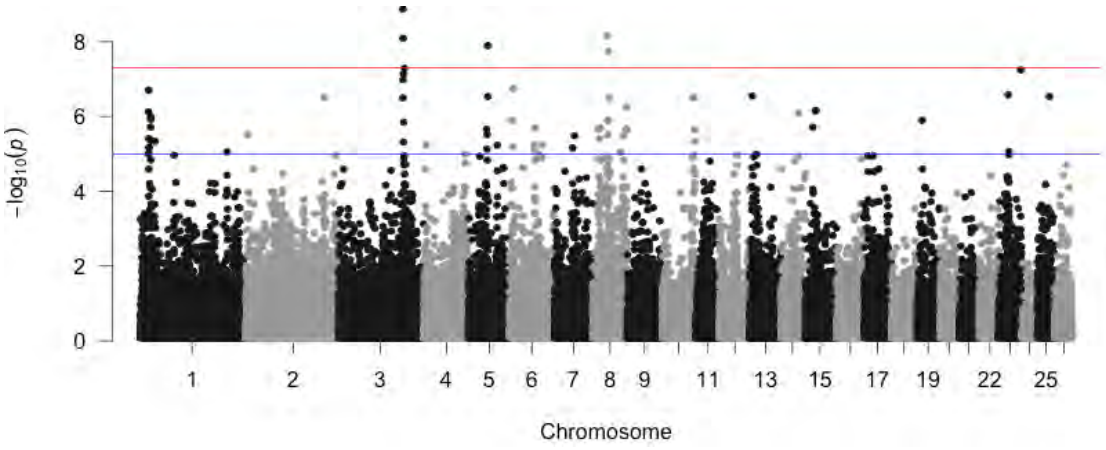

Figure 6.3: Genome-wide association analysis of white sub-vital versus the black, grey and white-vital subpopulations (*AA* group)

The sub-vital white versus vital white GWA was tested and the results were visualized using Manhattan plot (Figure 6.4). There were no SNPs detected above the genome association threshold, however, a number of SNPs went above the suggestive line of $-log_{10}5$. The SNP markers with highest *p*-values are located on chromosomes 8 and 10.

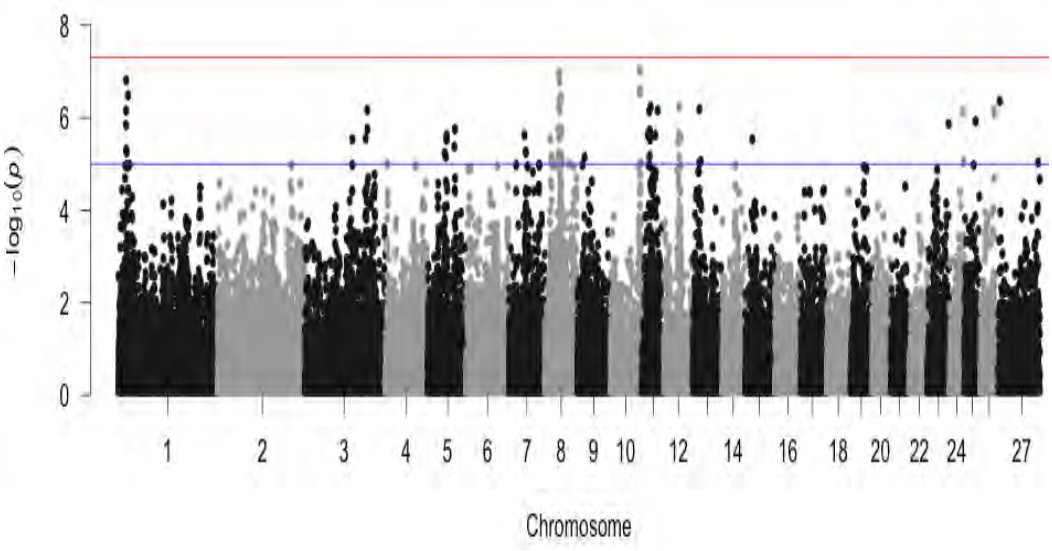

Figure 6.4: Genome-wide association analysis of sub-vital white versus vital-white subpopulations (*WO* group)

6.4.3 Putative biological functions of SNPs identified

The significant SNPs were analysed for possible gene functions. The related genes to the SNPs discovered in this study were identified through *ENSEMBL Genome Browser* (www.ensembl.org/Ovis_aries/Info/Annotation) and the sheep genome browser (www.livestock**genomics**.csiro.au/**sheep**/oar3.1.php). In addition, the biological implications of these SNPs and their associated SNPs or genes that were in close proximity were also investigated (Table 6.3). Few of the SNPs that had a high association level had an associated biological function. The genes reported here are either directly associated to the SNP of interest or are found close to the SNP identified to have strong association. In total six genes were identified and their biological functions were investigated. On chromosome 3, we identified *IGF1* (insulin growth factor 1) for both SNPs that had high association *p-value* i.e. OAR3_184018243.1 ($1.35e^{-09}$) and s24957.1 ($8.11e^{-09}$) located on position 171 415 798 and 171 452 281 bp, respectively. The two SNPs identified were 36 492 base pairs apart (36.492 kbp) and the *IGF1* gene was the most closely located gene for both these SNP s.

Table 6.3: Identified ovine SNPs and their biological functions

SNP ID	Base position	Distance (bp)	Gene (Biological implication)
OAR3_184018243.1	171415798	88046	*IGF**(Insulin Growth Factor)
s24957.1	171452281	124529	*IGF1**(Insulin Growth Factor-1)
OAR5_55263158.1	50972897	76504	*FGF1**(Fibroblast Growth Factor-1)
		412567	*NDFIP1**(Nedd4 family interacting protein)
OAR8_42787515.1	39687564	40983	*KLHL32/MMS22L*(Kelch-like 32/MMS22-like)
OAR8_44102022.1	41015484	256830	*FUT9**(α-(1,3)-fucosyltransferase 9)
		385247	*MANEA**(Mannosidase, endo-alpha)

**Implies genes that were in close proximity to the identified SNP/genes flanking the SNP, distance indicated in bp*

6.5 Discussion

Production of white pelt from Swakara lambs is hindered by the presence of the sub-vital factor that greatly impacts the white pelt market due to the loss of white lambs and reduced production. Limited studies have been done on sub-vitality. The OvineSNP50 genotyping array has facilitated the identification of QTL, candidate genes and markers through GWAS and has been used in characterization of genetic variability among breeds and determining population genetic history among breeds (Grasso *et al.* 2014). Genome wide association studies are designed to aid in the discovery of genetic variants for economically important traits (Collins *et al.* 2003). Genome wide association studies have achieved success in livestock with notable discoveries in bovine, porcine, and sheep (Anderson *et*

al. 2014). Genome wide association studies in sheep have revealed important SNPs in different breeds for economically important traits such as (i) rickets in Corriedale sheep, (ii) horns in Soay sheep, (iii) lissencephaly (smooth brain) caused by 31-bp deletion on *RELN* gene and on (iv) microphthalmia in Texel sheep that associated with a missense mutation on *PITX3* (Becker *et al.* 2010; Zhao *et al.* 2011; Suarez-Vega *et al.* 2013). The current study used the OvineSNP50 data from Swakara sheep to perform a GWAS to determine the causal variants for sub-vitality.

The distribution of SNPs of Swakara over the ovine genome was observed using QQ plots. These plots assess the behaviour of SNPs by comparing the expected versus the observed $-log_{10}(p)$ values. Under the null hypothesis, the *p-values* are expected to adhere to the normal distribution resulting in a straight diagonal line. This null hypothesis assumes that there in there is no association of the SNP loci being interrogated. QQ plots align the genotyped SNPs from the smallest to the highest $-\log_{10}(p)$ value and deviations from the distribution line are considered to be as a result of a true association to the trait under study (Pearson & Manolio 2008). In this study, none of the $-\log_{10}(p)$ followed the expected normal distribution in both the *WO* and *AA* study groups. This suggested SNPs that were highly significant and associated to sub-vitality.

Tools are now available for the visualization of data sets generated from genome-wide association studies for variant-disease/trait associations. Within the *R* environment there exists a number of packages such as *qqman, ggplot2* and other scripts that help biologists visualize their data in graphs known as Manhattan plots (Turner 2014). Manhattan plots for *AA* population category revealed significant SNPs on chromosome 3, 5 and 8 and. the SNPs with the highest *p-values* were of $8.11e^{-09}$ and $1.35e^{-09}$ respectively (Figures 6.3 and 6.4). SNP s24957.1 on chromosome 3 had a frequency of 0.4062 in the sub-vital sheep and 0.04861 in the controls whilst SNP OAR3_184018243.1 had a frequency of 0.3125 in sub-vital sheep and 0.01389 in the vital individuals. SNPs OAR5_55263158.1 on Chromosomes 5 and OAR8_42787515.1 and OAR8_44102022.1 on chromosome 8 were also above

the GWAS threshold. These three SNPs also occurred in high frequencies in the sub-vital white cases compared to the controls (Table 6.2).

SNP s24957.1 with the highest p-value was found on chromosome 3 and is in close proximity to the ovine insulin-growth like factor 1 (*IGF1*) gene which is an anabolic hormone in postnatal life that has been found to be responsible for endocrine regulator of foetal growth (Harding *et al.* 1994). The expression of several components of the *IGF* system are regulated by nutrition (Wathes *et al.* 1998). It has been noted that in sub-vital Swakara the digestive system tends to be compromised and feed management is very important such that animals kept for breeding are fed specialized diet to increase their ability to digest feed, thus increasing their chances of survival. The *IGF-1* promotes the growth of major foetal organs, endocrine glands, and skeletal maturation (Lok *et al.* 1996) and variations in this gene could therefore be responsible for sub-vitality in Swakara sheep. The high frequencies of the two SNPs (OAR3_s24957.1 and OAR3_184018243.1) on chromosome 3 in the sub-vital sheep make them more likely to have an influence on sub-vital performance in white Swakara. The linkage disequilibrium (r^2) between these two SNPs was = 0.571 indicating strong linkage disequilibrium, which is a measure of the chances that two SNPs are inherited together. These two SNPs are therefore most likely inherited together and based on their frequency could be associated with sub-vital performance.

A GWAS done on checkered giant rabbit revealed that the *KIT* gene is associated with the English Spotting coat colour locus and congenital megacolon in these rabbits. The study also showed that sheep that were homozygous recessive at the English spotting locus (white) were sub-vital individuals (Fontanesi *et al.* 2014). The *KIT* gene is responsible for the migration of melanocytes that are produced by the *MC1R* gene to hair follicles. This potentially supports the idea that sub-vitality in homozygous white Swakara is caused by pigmentation factors that also have metabolic implications. The sub-vitality performance GWAS in our study, however, did not reveal any possible association to the *KIT* gene, which is located on chromosome 13. The symptoms of sub-vital performance in the English spotting locus of the checkered giant rabbit are similar to the symptoms in the Swakara population. Both these

lethal genes affect white, homozygous individuals. The results yielded in Swakara reveal that there could be multiple SNPs associated to sub-vitality and therefore sub-vital performance occurs as a result of a combined effect from different genetic variants that are responsible for different developmental functions.

Other significant SNP was associated with *alpha*-mannosidases gene on chromosome 8 that has been implicated in enzymes that facilitate newly formed *N*-glycans and degrade mature glycoproteins. Mannosidases located in the endoplasmic reticulum have been suggested to have a role in catabolic activities. The *Nedd4* family interacting protein, *Kelch-like 32*, and *MMS22-like* genes did not have putative functions that related to sub-vitality as far as we know.

None of the SNPs were significant in the white only (*WO*) population category. It is possible that the samples size in the *WO* population category was small i.e. fewer controls to cases, therefore negatively impacting on significant thresholds. A software such as EMMAX (efficient mixed-model association eXpedited) can be used to validate the associations identified through GWAS in PLINK. This software takes into account the sample structure and size in determining association.

6.6 Conclusion

The GWAS study identified SNPs associated to sub-vital performance when analysis was conducted using all colour subpopulations. The SNPs identified in this study point more to a metabolic consequence of sub-vitality based on the genes associated and sub-vitality could therefore be a culmination of the small effects of different SNPs. GWAS alone was however not conclusive enough as no significant SNPs were identified between the white sub-vital and white vital sheep.

CHAPTER 7: DISCUSSION AND CONCLUSIONS

The Swakara population study set out to investigate the population structure, genetic diversity and genomic differences that exist amongst the pelt colour subpopulations using the OvineSNP50 beadchip. This was done in an effort to identify the underlying causes of sub-vitality and subsequently a GWAS was conducted in search of SNPs that are associated to sub-vital performance in white Swakara.

The first analysis looked at the transferability of the OvineSNP50 beadchip to Swakara sheep. The transferability of the OvineSNP50 beadchip was a success as each subpopulation had genotyping success rates of above 93% with only 2/90 individuals having low genotyping success. Another measure of successful genotyping and transferability is the deviations from HWE, which is important for eliminating markers that are not informative and only 329 SNPs were found to deviate from HWE < *0.0001*. The quality control parameters excluded SNPs with a call rate lower than 95% and markers that had a MAF < 0.01. Minor allele frequency (MAF) represents the frequency of the least common allele in a given population. MAF allows us to identify informative SNPs that can be used in down-stream analyses. The MAF average was between 0.21 – 0.25. The overall population showed 80.84% polymorphic markers. The presence of a high proportion of polymorphic SNPs makes the beadchip informative in the Swakara population. Such results support the use of OvineSNP50 beadchip as it is robust and applicable to diverse sheep breeds.

The OvineSNP50 genotyping platform has been successfully used in other genetic studies in sheep of different breeds (Becker *et al.* 2010; Michelizzi *et al.* 2010; Kijas *et al.* 2012; Al-Mamun *et al.* 2013; Do *et al.* 2014). The information generated from using high-density and high-throughput genotyping platforms is extensive and can be used in a variety of studies depending on the research questions. The OvineSNP50 beadchip has been applied successfully in other population genetics studies on a number of diverse sheep breeds. This has led to the discovery of QTL, identification of associated alleles to disease traits of interest and the genetic diversity that exists among breeds (Kijas *et al.* 2012; Suarez-

Vega *et al.* 2013) in breeds such as the Spanish Churra (80.72% polymorphic SNPs) (Garcia-Gamez *et al.* 2012) and in a global sheep study for historic admixture and selection, 90.4% polymorphic SNPs (Kijas *et al.* 2012). The genotyping success of the OvineSNP50 beadchip in the Swakara breed shows good transferability of the beadchip to the breed and that it is usable in the study of population genetics in different breeds of interest.

The second analysis looked at inbreeding levels, genetic diversity and runs of homozygosity in Swakara subpopulations. Extensive inbreeding and selective breeding reduce genetic diversity, which results in the accumulation of lethal recessive alleles in populations (Verweij *et al.* 2014). The observed heterozygosities were all below the expected heterozygosities and thus subpopulations were inbred. Comparatively, the combined population (includes vital white, sub-vital white, black and grey individuals) had an inbreeding coefficient value of 0.048 with an observed heterozygosity of 0.3019±0.1615 and high level of expected heterozygosity 0.3266±0.1554. The highest level of genetic diversity was in the grey population (H_E = 0.3417) and it was also the least inbred population (F_{IS} = 0.0092). The most inbred between the subpopulations was the black population with an F_{IS} = 0.0496 which is closely similar to the overall population inbreeding values showing high levels of inbreeding in the population. However, the population with the lowest level of genetic diversity was the white population (H_E = 0.3301) and (F_{IS} =0.0351). The most highly selected SNP between the grey and sub-vital white that had a genetic differentiation F_{ST} value of 0.6846 and was located on chromosome 13. The other SNPs with high F_{ST} values were also investigated for genomic functions and only the SNP on chromosome 6 with an F_{ST}= 0.5727 between the black and sub-vital white sheep was associated to a gene, *GABRB1* (gamma-aminobutyric acid A receptor, beta 1).

The pairwise $F_{ST,}$ PCA and Weir & Cockerham F_{ST} showed supporting results. The black vs white subpopulations were the most genetically different based on these platforms with values of F_{ST} = 0.037. The analyses showed different results for the most genetically similar subpopulations. The white vs grey subpopulations were identified by pairwise F_{ST} with a value of 0.00938 while the Weir and Cockerham

identified the white vs sub-vital subpopulations as more genetically similar. The PCA based on pelt colour results showed the white and sub-vital group to be more genetically related, further supporting the Weir & Cockerham F_{ST}. The PCA analysis also supported the result from Pairwise F_{ST} as the two subpopulations were grouped into the same cluster.

A total of 42 unique ROHs were observed in 33 (37.5%) of the individuals. ROH were found in each of the colour subpopulations and the vital-white population had the most ROH (n = 18) from 15 individuals. The rest of the colour subpopulations each had 8 ROHs reported. The black population had the highest average length of ROH blocks of 9552.5±5040.15 KB while the white individuals had the lowest average of 7066.84±2835.09 KB. It was suggested in a study on the effects of population history on runs of homozygosity that ROHs in outbred populations could be attributed to unusual mutations, linkage disequilibrium or low recombination rates (Kirin *et al.* 2010). There was no clear correlation between ROH and inbreeding coefficient in this study. There were seven overlapping/consensus ROH (cROH) in the Swakara population. In this study there were no ROHs that were unique to the sub-vital white subpopulation.

A few significant SNPs were identified using the GWAS but only when comparing the white sub-vital against all the other pelt colour subpopulations. SNP s24957.1 on chromosome 3 had the highest *p*-value and was in close proximity to the ovine *IGF1* (insulin-growth like factor 1) gene. This SNP was located 36.483 kbp away from the second SNP that also occurred in high frequency in sub-vital Swakara sheep. The insulin-growth like factor 1 is an anabolic hormone in postnatal life and may be responsible for endocrine regulator of foetal growth (Harding *et al.* 1994). The *IGF-1* promotes the growth of major foetal organs, endocrine glands, and skeletal maturation (Lok *et al.* 1996) and variations of genes for *IGF* could be responsible for sub-vitality in Swakara sheep. The frequency of the two SNPs on chromosome 3 that had highly significant frequencies in the sub-vital individuals, presents them as likely candidates for sub-vitality performance in white Swakara. Contrary to expectations was the failure to observe significant SNPs in the white only (*WO*) group. The study contained a lot of individuals from

Namibia and South Africa for the GWAS test. The South African population is however, not severely affected by sub-vitality and so a study that focuses on individuals from a flock known to be affected by sub-vitality could yield stronger associative SNPs through a GWAS.

Although the SNPs identified and interrogated for putative functions were not similar, the genes identified for gene functions using cROH and GWAS were more related to growth and metabolic activities i.e. insulin growth factor and thyroglobulin which both produce hormones for growth and development. Results showed that within colour subpopulations there was genetic diversity, however not so much between subpopulations. Admixture and PCA showed the genetic similarities between subpopulations, these showed that some colour subpopulations are more genetically similar than others, which could be as a result of similar breeding practices. The inbreeding levels were relatively low when looking at colour subpopulations. GWAS results gave candidate genes that can potentially be implicated as the contributing factors of sub-vital performance. The attempt to identify associated SNPs and their putative functions using cROHs and GWAS did not yield conclusive results. The SNPs identified using both methods were found on different chromosomes. It is possible that the sub-vital performance is the result of the combined effect of various SNPs, each contributing a small genetic effect. Increasing the sample size of sub-vital white subpopulation would increase the power of finding an associated SNP or finding other candidate SNPs to investigate further.

REFERENCES

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

Al-Mamun H.A., Kwan P., Tellam R.L., Kijas J.W. & Gondro C. (2013) A study on effects of family and haplotype blocks on conservation of gene expression traits in half sib sheep families. *Association for the Advancement of Animal Breeding and Genetics* **20**, 266-9.

Alexander D.H. & Lange K. (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**.

Alexander D.H., Novembre J. & Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655 - 64.

Allendorf F.W. (1986) Genetic drift and loss of alleles versus heterozygosity. *Zoo Biology* **5**, 181-90.

Anderson C.A., Pettersson F.H., Clarke G.M., Cardon L.R., Morris A.P. & Zondervan K.T. (2010) Data quality control in genetic case-control association studies. *Nature Protocol* **5**, 1564-73.

Anderson R., McEwan J., Brauning R., Pickering N., Kijas J.W., Dalrymple B.P., Daetwyler H.D., Worley H., Heaton M.P., van Stijn T., Clarke S., Baird H. & Khan A. (2014) Development of a high density (600K) Illumina ovine SNP chip and its use to fine map the yellow fat locus.

Auvray B., Dodds K.G. & McEwan J.C. (2011) Brief Communication: Genomic selection in the New Zealand sheep industry using ovine SNP50 beadchip.

Banerjee S. (2010) A lethal genetic factor in Garole lambs - a case study. *World Applied Sciences Journal* **10**, 397-401.

Becker D., Tetens J., Brunner A., Burstel D., Ganter M. & Kijas J.W. (2010) Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. *PLos One* **5**.

Bolormaa S., Hayes B.J., Savin K., Hawken R., Barendse W., Arthur B.P.F., Herd R.M. & Goddard M.E. (2011) Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science* **89**, 1684-97.

Bovine Genome Sequencing and Analysis Consortium, Elsik C.G., Tellam R.L., Worley K.C. & Gibbs R.A. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522 - 8.

Bozeman M.T. (2014) SNP & Variation Suite (SVS). Golden Helix, Inc.

Buduram P. (2004) Genetic characterization of Southern African sheep breeds using DNA markers. In: *Department of Animal, Wildlife and Grassland Sciences*. University of the Free State.

Bush W.S. & Moore J.H. (2012) Chapter 11: Genome-wide assoociation studies. *PLoS Collections*.

Campbell L.J. (2007) Evaluation of two indigenous South African sheep breeds as pelt producers. In: *Animal and Wildlife Sciences*. University of Pretoria, Pretoria.

Cockett N.E. (1999) Genomics of Sheep. *AgBiotechNet* **1**.

Cockett N.E., Shay T.L. & Smit M. (2001) Analysis of the sheep genome. *Physiological Genomics* **7**, 69 -78.

Coertzen A. (2009) Increasing the efficiency of Karakul pelt production by accelerated lambing. In: *RING*. AGRA, Windhoek, Namibia.

Collins F.S., Green E.D., Guttmacher A.E. & Guyer M.S. (2003) A vision for the future of genomics research. *Nature* **22**.

Cooper T.A., Wiggans G.R. & VanRaden P.M. (2013) Short communication: relationship of call rate and accuracy of single nucleotide polymorphism genotypes in diary cattle. *Journal of Dairy Science* **96**, 3336-9.

Curtis D., Vine A.E. & Knight J. (2008) Study regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. *Annals of Human Genetics* **72**, 261-78.

Dekkers J.C.M. (2012) Application of genomics tools to animal breeding. *Current Genomics* **13**, 207 - 12.

Demars J., Fabre S., Sarry J., Rossetti R., Gilbert H., Persani l., Tosser-Klopp G., Mulsant P., Nowak Z., Drobik W., Martyniuk E. & Bodin L. (2013) Genome-wide association studies identify two novel BMP15 mutations responsible for an atypical hyperprolificacy phenotype in sheep. *Public Library of Science Genetics* **9**.

Do D.N., Ostersen T., Strathe A.B., Mark T., Jensen J. & Kadarmideen H.N. (2014) Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat, and weight gain in pigs. *BMC Genomics* **15**.

Dodds K.G., Auvray B., Pickering N. & McEwan J.C. (2009) Quality control for ovine SNP50 beadchip genotypes. *Association for the Advancement of Animal Breeding and Genetics* **18**, 296 - 9.

Du F., Clutter A.C. & Lohuis M.M. (2007) Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences* **3**, 166-78.

Edea Z., Dadi H., Kim S., Dessie T., Lee T., Kim H., Kim J. & Kim K. (2013) Genetic diversity, population structure and relationship in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Frontiers in Genetics* **4**.

Excoffier L. & Lischer H.E.L. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-7.

Fadiel A., Anidi I. & Eichenbaum K.D. (2005) Farm animal genomics and informatics: an update. *Nucleic Acids Research* **33**, 6308-18.

Fan B., Du Z., Gorbach D.M. & Rothschild M.F. (2010) Development and application of high-density SNP arrays in genomic studies of domestic animals. *Asian-Australasian Journal of Animal Science* **23**, 833-47.

Farnir F., Coppieters W. & Arranz J.J. (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**, 220-7.

Fishchenko O.P. (1978) Defects of Karakul sheep skins and ways of eliminating them. *Ovtsevodstvo* **5**, 26-7.

Flicek P., Amode M.R., Barrell D., Beal M., Billis K., Brent S., Carvalho-Silva D., Clapham P., Coates G., Fitzgerald S., Gil L., Giron C.G., Gordon L., Hourlier T., Hunt S., Johnson N., Juettemann T. & Kahari A. (2014) Ensembl 2014. *Nucleic Acids Research* **42**.

Fontanesi L., Viargiolu M., Scotti E., Latorre R., Faussone P., Mazzoni M., Asti M., Chiocchetti R., Romeo G., Clavenzani P. & de Giorgio R. (2014) The *KIT* gene is associated with the *English Spotting* coat color locus and congenital megacolon in checkered giant rabbits (*Oryctolagus cuniculus*). *PLos One* **9**.

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

Gorbach D.M., Fan B., Onteru S.K., Zhao X., Du X., Garrick D.J., Dekkers J.C.M. & Rothschild M.F. (2010) Genome-wide association studies for important economic traits

in domestic animals using high density SNP genotyping. *Animal Industry Report: AS 656, ASL R2540*.
Grasso A.N., Goldberg V., Navajas E.A., Iriarte W., Gimeno D., Anguilar I., Medrano J.F., Rincon G. & Ciappesoni G. (2014) Genomic variation and population structure detected by single nucleotide polymorphism arrays in Corriedale, Merino and Creole sheep. *Genetics and Molecular Biology* **37**.
Greeff J.C. & Faure A.S. (1991) Genetic and phenotypic parameters of pelt traits in a karacul control flock. *South African Journal of Animal Science* **21**.
Greeff J.C. & Faure A.S. (1993a) Genetic trends of selection for pelt traits in karakul sheep: I. direct responses. *23* **5/6**.
Greeff J.C. & Faure A.S. (1993b) Genetic trends of selection for pelt traits in karakul sheep: ii. correlated responses. *South African Journal of Animal Science* **23**.
Groenewald H.B. (1993) Ultrastructure of the myenteric ganglia in the rumen, reticulum, omasum and abomasum of grey, white, and black karakul lambs. *Onderstepoort Journal of Veterinary Research* **60**, 189-95.
Groenewald L.F., Lenstra J.A., Eding H., Toro M.A., Scherf B., Pilling D., Negrini R., Finlay E.K., Jianlin H., Groenewald E., Weigend S. & Consrtium T.G. (2010) Genetic diversity in farm animals - a review. *Animal Genetics - Immunogenetics, Molecular Genetics and Functional Genomics* **41**, 6-31.
Guries R.P. & Ledig F.T. (1982) Genetic diversity and population structure in pitch pine (*Pinus rigida Mill.*). *Evolution* **36**, 387-402.
Harding J.E., Liu L., Evans P.C. & Gluckman P.D. (1994) Insulin-like growth factor 1 alters feto-placental protein and carbohydrate metabolism in fetal sheep. *Endocrinology* **134**, 1509-14.
Hetzel B.S. & Mano M.T. (1989) A review of experimental studies of iodine deficiency during lamb development. *The Journal of Nutrition* **119**, 145-51.
Howrigan D.P., Simonson M.A. & Keller M.C. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* **12**.
Illumina I. (2010) OvineSNP50 genotyping beadchip. URL www.illumina.com/products/ovinesnp50_dna_analysis_kit.html.
International Sheep Genomics Consortium, Archibald A.L., Cockett N.E., Dalrymple B.P., Faraut T., Kijas J.W., Maddox J.F., McEwan J.C., Oddy V.H., Raadsma H.W., Wade C., Wang J., Wang W. & Xun X. (2010) The sheep genome reference sequence: a work in progress. *Animal Genetics: Immunogenetics, molecular genetics and functional genomics* **41**, 449-53.
Jombart T., Devilliard S. & Balloux F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structures populations.
Kantanen J., Olsaker I., Holm J.E., Lien S., Vilkki J., Brusgaard K., Eythorsdottir E., Danell B. & Adalsteinsson S. (2000) Genetic diversity and population structure of 20 North European cattle breeds. *Journal of Heredity* **91**, 446 - 57.
Karimi Z. (2013) Runs of homozygosity patterns in taurine and indicine cattle breeds. *European Master in Animal Breeding and Genetics*.
Keller L.F. & Waller D.M. (2002) Inbreeding coefficients in wild populations. *Trends in Ecology and Evolution* **17**.
Khanyile K.S., Dzomba E.F. & Muchadeyi F.C. (2014) Haplo-block structure of Southern African village chicken populations. *10th World Congress of Genetics Applied to Livestock Production*.

Kijas J.W., Lenstra J.A., Hayes B.J., Boitard S., Neto L.R.P., Cristobal M.S., Servin B., McCulloch R., Whan V.A., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J. & Dalrymple B.P. (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *Public Library of Science Biology* **10**.

Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F., McGrath A., Wilson P., Cockett N.E., Oddy H.V., Nicholas F.W. & Raadsma H. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *Public Library of Science One* **4**.

Kirin M., McQuillan R., Franklin C.S., Campbell H., McKeigue P.M. & Wilson J.F. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLos One* **5**.

Kolkman J.M., Berry S.T., Leon A.J., Slabaugh M.B., Tang S., Gao W., Shintani D.K., Burke J.M. & Knapp S.J. (2007) Single nucleotide polymorphisms and linkage disequilibrium in Sunflower. *The Genetics Society of America* **177**, 457-68.

Ku C.S., Naidoo N., Teo S.M. & Pawitan Y. (2011) Regions of homozygosity and their impact on complex diseases and traits. *Human Genetics* **129**, 1-15.

Lai Z., Livingstone K., Zou Y., Church S.A., Knapp S.J., Andrews J. & Rieseberg L.H. (2005) Identification and mapping of SNPs from ESTs in Sunflower. *Theor Appl Genetics* **111**, 1532-44.

Lewis C.M. & Knight J. (2012) Introduction to genetic association studies. *Cold Spring Harbor Protocols*.

Lin C. (2005) Linkage disequilibrium measures. In: *Department of Statistics*, p. 59. University of Carlifonia, Los Angeles.

Liu J., Zhang L., Xu L., Ren H., Lu J., Zhang X., Zhang S., Zhou X., Wei C., Zhao F. & Du L. (2013) Analysis of copy number variations in the sheep genome using 50K SNP beachip array. *BMC Genomics* **14**.

Lobo A.M.B.O., Paiva S.R. & Lobo R.N.B. (2014) Signatures of selection for age at lambing in Brazilian local adapted sheep. *10th World Congress of Genetics Applied to Livestock Production*.

Lok F., Owens J.A., Mundy L., Robinson J.S. & Owens P.C. (1996) Insulin-like growth factor promotes growth selectively in fetal sheep in late gestation. *American Journal of Physiology* **270**.

Louwrens A., van der Merwe J.A. & Coetzee S.F. (2004) Progress of the Karakul studs at Gellap-Ost research station over the period of 1994 to 2003. (ed. by AGRICOLA).

Lundie R.S. (2011) The genetics of colour in fat-tailed sheep: a review. *Tropical Animal Health Production* **43**, 1245 - 65.

McManus C., Paiva S.R. & de Araujo R.O. (2010) Genetics and breeding of sheep in Brazil. *Revista Brasileira de Zootecnia* **39**.

McRae A.F., Pemberton J.M. & Visscher P.M. (2005) Modeling linkage disequilibrium in natural populations: the example of the Soay sheep population of St. Kilda, Scotland. *Genetics* **171**, 251-8.

Meadows J.R.S., Chan E.K.F. & Kijas J.W. (2008) Linkage disequilibrium compared between five populations of domestic sheep. *BioMed Central* **9**.

Michelizzi V.N., Wu X., Dodson M.V., Michal J.J., Zambrano-Varon J., McLean D.J. & Jiang Z. (2010) A global view of 54 001 single nucleotide polymorphisms (SNPs) on the Illumina BovineSNP50 beadchip and their transferability to Water Buffalo. *International Journal of Biological Sciences* **7**, 18-27.

Nakamoto K., Wang S., Jenison R.D., Guo G.L., Klaassen C.D., Wan Y.Y. & Zhong X. (2006) Linkage disequilibrium blocks, haplotype structure, and htSNPs of human CYP7AI gene. *BMC Genetics* **7**.

Nanekarani S., Amirinia C. & Amirmozafari N. (2011) Genetic analysis of Karakul sheep breed using microsatellite markers. *African Journalk of Microbiology Research* **5**.

Nasholm A. & Eythorsdottir E. (2011) Characteristics and utilization of sheep pelts. *Small Ruminant Research* **101**, 182 - 7.

Nei M. & Chesser R.K. (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics* **47**, 253-9.

Nel J.A. (1966) Genetics studies in Karakul sheep. *University of Stellebosch*, D.Sc. (Agric).

Nel J.A. & Louw D.J. (1953) The lethal factor in grey Karakul. *Farming in South Africa* **5**, 169 - 72.

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 Research* **18**, 1282-93.

Nsoso S.J. & Madimabe M.J. (1999) The sheep industry in Botswana: promoting the karakul sheep industry. *South African Journal of Animal Science* **29**.

Nsoso S.J. & Madimabe M.J. (2003a) A survey of Karakul sheep farmers in Southern Kalahari, Botswana: management practices and constraints to improving production. *South African Society of Animal Science* **4**.

Nsoso S.J. & Madimabe M.J. (2003b) A survey of karakul sheep farmers in southern Kalahari, Botwana: management practices and constraints to improving production. *South African Society of Animal Science* **4**, 23-7.

O'Brien A.M.P., Utsunomiya Y.T., Meszaros G., Bickhart D.M., Liu G.E., Van Tassell C.P., Sonstegard T.S., Da Silva M.V.B., Garcia J.F. & Solkner J. (2014) Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genetics Selection Evolution* **46**.

Pareek C.S., Smoczynski R. & Tretyn A. (2011) Sequencing technologies and genome sequencing. *Journal of Applied Genetics* **52**, 413-35.

Pashley C.H., Ellis J.R., McCauley D.E. & Burke J.M. (2006) EST databases as a source of molecular markers: lessons from *Helianthus*. *Journal of Heredity* **97**, 381-8.

Patterson N., Price A.L. & Reich D. (2006) Population structure and eigenalysis. *Plos Genetics* **2**.

Pearson T.J. & Manolio T.A. (2008) How to interpret a genome-wide association study. *JAMA* **299**, 1335-44.

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. & Sham P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**.

Purfield D.C., Berry D.P., McParland S. & Bradley D.G. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* **12**.

Ramos A.M., Crooijmans R.P.M.A., Affara N.A., Amaral A.J., Archibald A.L., Beever J.E., Bendixen C., Churcher C., Clark R., Dehais P., Hansen M.S., Hedegaard J., Hu Z., Kerstens H.H., Law A.S., Megens H., Milan D., Nonneman D.J., Rohrer G.A., Rothschild M.F., Smith T.P.L., Schnabel R.D., Van Tassell C.P., Taylor J.F., Wiedmann R.T., Schook L.B. & Groenen M.A.M. (2009) Design of a high density assay in the pig using SNPs identified and characterized by next generation sequencing technology. *Public Library of Science One* **4**.

Rochus C.M., Mikko S., Nasholm A. & Johansson A.M. (2014) ASIP and MC1R muations causing black coat colour in five Swedish sheep breeds. In: *10th World Congress of Genetics Applied to Livestock Production*.

Rothauge A. (2009) Increasing the efficiency of Karakul pelt production by accelarated lambing. In: *AGRA Co-operative LTD Member's Newsletter*. RING, Namibia.

Sabatti C. & Risch N. (2002) Homozygosity and Linkage disequilibrium. *Genetics* **160**, 1707-19.

Schlechter D. (2013) Namibia: Genetic mapping to protect local species. In: *New Era*. AllAfrica Global Media, Windhoek, Namibia.

Schoeman S.J. (1998) Genetic and environmental factors influencing the quality of pelt traits in Karakul sheep. *South African Journal of Animal Science* **28**, 125-39.

Schoeman S.J., Cloete S.W.P. & Olivier J.J. (2010) Returns on investment in sheep and breeding in South Africa. *Livestock Science*.

Sefidbakht N. & Farid A. (1977) Effect of early weaning and hormone treatments on induction of estrus, conception, and lambing of fall-lambing karakul. *Journal of Animal Science* **45**, 311-9.

Steyn M.G. (1962) Teling van volbloed Karakoelskape.

Stranger B.E., Stahl E.A. & Raj T. (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367-83.

Stumpf M.P.H. (2002) Haplotype diversity and the block structure of linkage disequilibrium. *Trends in Genetics* **18**.

Suarez-Vega A., Gutierrez-Gill B., Cuchillo-Ibanez I., Saez-Valero J., Perez V., Garcia-Gamez E., Benavides J. & Arranz J.J. (2013) Identification of a 31-bp deletion in the RELN gene causing lissencephaly with cerebellar hypoplasia in sheep. *Public Library of Science One* **8**.

Tang H., Peng J., Wang P. & Risch N. (2005) Estimation of individual admixture: analytical and study design considerations. *Genetic Epidemiology* **28**, 289 - 301.

Tang S., Yu J., Slabaugh M.B., Shintani D.K. & Knapp S.J. (2002) Simple sequence repeat map of sunflower genome. *Theor Appl Genetics* **105**, 1124-36.

Teneva A. (2009) Molecular markers in animal genome analysis. *Biotechnology in Animal Husbandy* **25**.

Tishkoff S.A. & Verrelli B.C. (2003) Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. *Current Opinion in Genetics and Development* **13**, 569-75.

Turner S.D. (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRXiv*.

Twyman R. (2003) Techniques patents for SNP genotyping. *Pharmacogenomics* **4**, 67-9.

VanLiere J.M. & Rosenberg N.A. (2008) Mathematical properties of the r^2 measure of linkage disequilibrium. *Theoretical Population Biology* **74**, 130 - 7.

VanRaden P.M., Olson K.M., Null D.J. & Hutchinson J.L. (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science* **94**, 6153-61.

Varshney R.K., Graner A. & Sorrells M.E. (2005) Genomics-assisted breeding for crop improvement. *Trends in Plant Science* **10**.

Verweij K.J.H., Abdellaoui A., Veijola J., Sebert S., Koiranen M., Keller M.C., Jarvelin M. & Zietsch B.P. (2014) The association of genotype-based inbreeding coefficient with a range of physical and psychological human traits. *PLos One* **9**.

Vignal A., Milan D., SanCristobal M. & Eggen A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **34**, 275-305.

Vrijenhoek R.C. (1994) Genetic diversity and fitness in small populations. *Conservation Genetics* **68**, 37-53.

Wathes D.C., Reynolds T.S., Robinson R.S. & Stevenson K.R. (1998) Role of the insulin-like growth factor system in uterine function and placental development in ruminants. *Journal of Diary Science* **81**, 1778-89.

Wiggans G.R., VanRaden P.M., Bacheller L.R., Tooker M.E., Hutchinson J.L., Cooper T.A. & Sonstegard T.S. (2010) Selection and management of DNA markers for use in genomic evaluation. *Journal of Dairy Science* **94**, 3202-11.

Yang T., Guo Y., Zhang L., Tian Q., Yan H., Papasian C.J., Recker R.R. & Deng H. (2010) Runs of homozygosity identify a recessive locus 12q21.31 for human adult high. *Journal of Clinical Endocrinol Research* **95**.

Zavattari P., Deidda E., Whalen M., Lampis R., Mulargia A., Loddo M., Eaves I., Mastio G., Todd J.A. & Cucca F. (2000) Major factors influencing linkage disequilibrium by analysis of dfferent chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Human Molecular Genetics* **9**, 2947-57.

Zechner P., Solkner J., Bodo I., Druml T., Baumung R., Achmann R., Marti E., Habe F. & Brem G. (2002) Analysis of diversity and population structure in the Lipizzan horse bred based on pedigree information. *Livestock Production Science* **77**, 137-46.

Zhai W., Todd M.J. & Nielsen R. (2004) Is haplotype block identification useful for association mapping studies? *Genetic Epidemiology* **27**, 80-3.

Zhang H., Wang Z., Wang S. & Li H. (2012) Progress of genome wide association study in domestic animals. *Animal Science and Biotechnology* **3**.

Zhang L., Mousel M.R., Wu X., Michal J.J., Zhou X., Ding B., Dodson M.V., El-Halawany N.K., Lewis G.S. & Jiang Z. (2013) Genome-wide genetic diversity and differentially selected regions among Suffolk, Rambouillet, Columbia, Polyplay, and Targhee Sheep. *PLos One* **8**.

Zhao X., Dittmer K.E., Blair H.T., Thompson K.G., Rothschild M.F. & Garrick D.J. (2011) A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in Corriadale sheep. *PLos One* **6**.

Zhuang Z., Gusev A., Cho J. & Pe'er I. (2012) Detecting identity by descent and homozygosity mapping in whole-exome sequencing data. *Public Library of Science One* **7**.

APPENDIX A: INBREEDING CO-EFFICIENT OF THE INDIVIDUALS IN THE FOUR COLOUR SWAKARA SUBPOPULATIONS.

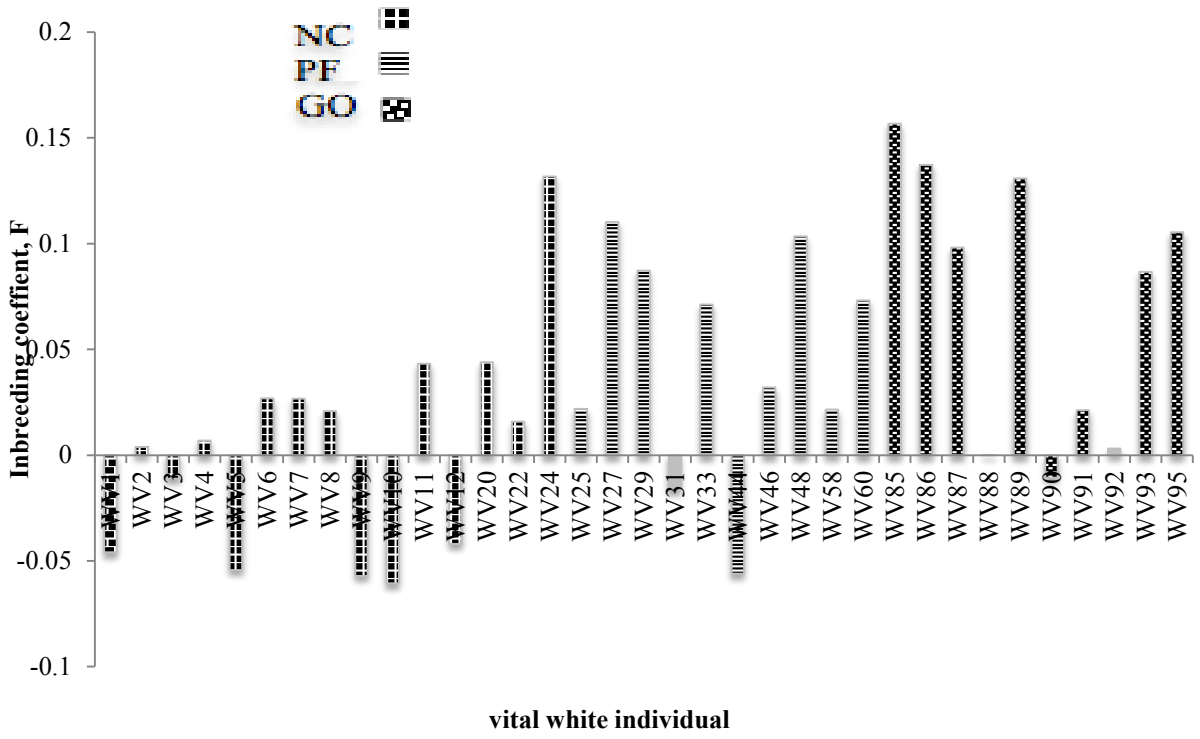

Figure A1: The inbreeding coefficient of the vital white individuals

**NC – Northern Cape, PF – Private farms and GO – Gellap-ost*

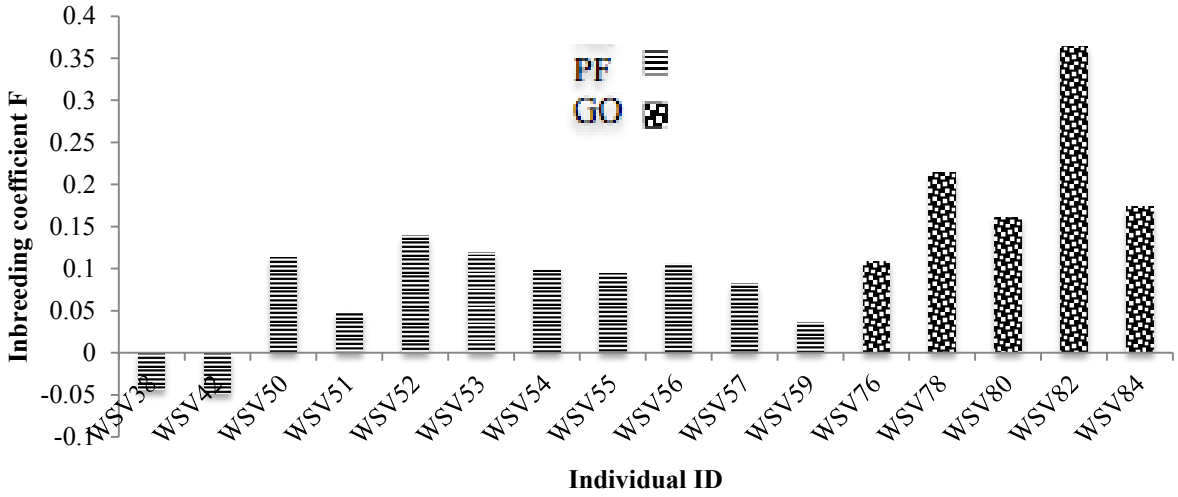

Figure A2: Inbreeding coefficients of sub-vital white subpopulation

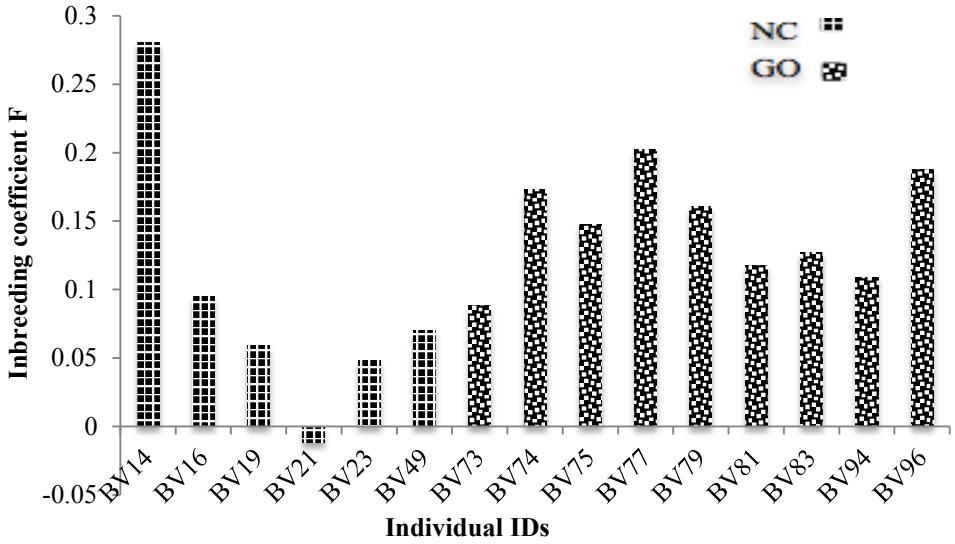

Figure A3: Inbreeding coefficient in black individuals

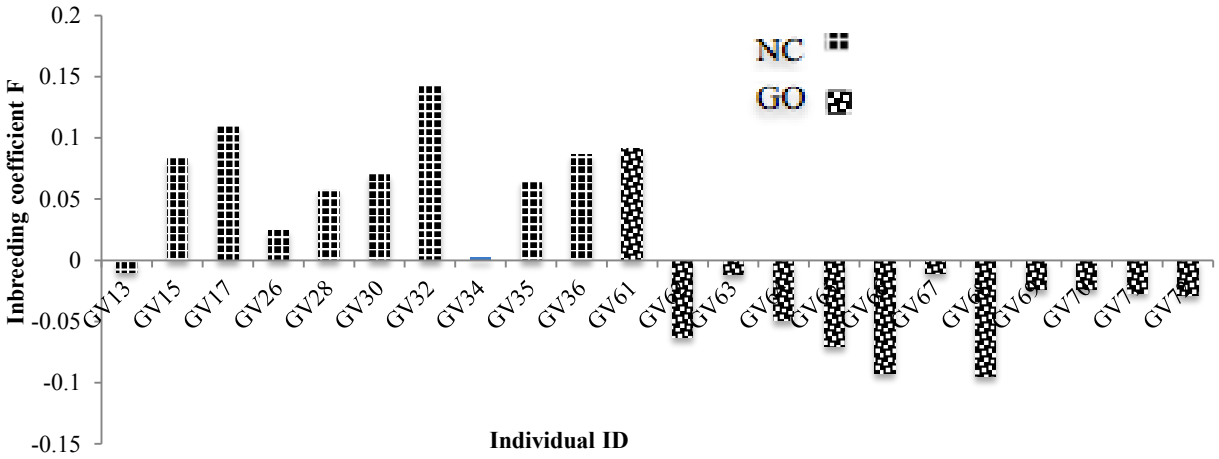

Figure A4: Inbreeding coefficient in grey subpopulation

Table A1: The number of ROHs in each individual including their total length (KB) sum and the average (KBAVG)

Individual	Vitality status	Number of ROHs	Total KB	KBAVG
WV3	Vital	2	10155.9	5077.93
WV7	Vital	1	5508.67	5508.67
WV8	Vital	1	4902.71	4902.71
BV14	Vital	3	16820.2	5606.72
GV15	Vital	1	5349.04	5349.04
GV17	Vital	2	9763.22	4881.61
WV22	Vital	2	12587.3	6293.63
WV24	Vital	1	5058.57	5058.57
WV27	Vital	1	8128.31	8128.31
WV33	Vital	1	5000.88	5000.88
GV35	Vital	1	9290.12	9290.12
WV46	Vital	2	13565.9	6782.96
WV48	Vital	1	5265.89	5265.89
WSV51	Sub-vital	1	6434.68	6434.68
WSV53	Sub-vital	1	5902.85	5902.85
WSV55	Sub-Vital	1	6632.39	6632.39
WSV56	Sub-Vital	1	5235.9	5235.9
WSV59	Sub-Vital	1	5146.26	5146.26
WV60	Vital	1	5531.16	5531.16
GV62	Vital	1	5318.13	5318.13
GV71	Vital	1	5532.76	5532.76
GV72	Vital	2	10163.5	5081.75
BV73	Vital	1	8677.28	8677.28
BV75	Vital	1	5143.36	5143.36
WSV80	Sub-Vital	1	5547.67	5547.67
BV83	Vital	1	4912.56	4912.56
WSV84	Sub-Vital	2	15873.7	7936.85
WV85	Vital	1	7008.53	7008.53
WV88	Vital	1	5265.89	5265.89
WV90	Vital	1	5265.89	5265.89
WV92	Vital	1	6967.32	6967.32
WV95	Vital	1	5789.65	5789.65
BV96	Vital	2	12209.1	6104.55