# Genetic characterisation of the Tankwa Goat (*Capra hircus*) using genome-wide SNP data and the identification of genes associated with climate adaptation

by

Anna Sophia Kropff

Submitted in fulfilment of the requirements for the degree

MSc(Agric): Animal Breeding and Genetics

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

2022

Supervisor: Prof C. Visser

Co-supervisor: Prof A. Kotzé

# Declaration

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

Signature:……  ……

Date:……30 November 2022…….

# Acknowledgements

Firstly I would like thank to my parents for their continued physical and emotional support throughout my studies. Without them I would never have become the person I am today. Thank you for always being available for me to run home to when everything else fails.

To my significant other, Dylan Meyers, thank you for your love and support and for putting up with all my stress related mood swings and weekends working. You are the light in my life and I could not have done this without you. You have been there for me through all my ups and downs and have been my greatest pillar of support. Your delicious snacks also helped a lot.

Thank you Prof Carina Visser and Prof Antoinette Kotzé for this wonderful opportunity to complete my MSc under you. You were amazing and always very supportive even when I was not so efficient. Thank you for not losing patience with me and always helping wherever I needed assistance.

Thank you to SANBI for funding my MSc and providing me with the equipment needed to complete this study. And thank you to Northern Cape DALRD for funding the genotyping component and allowing me to use the data for this study.

To my colleagues in the Genetics Services Unit, and especially my deputy-director Karin Behr, thank you for driving me to complete this and always motivating me when I needed it.

And lastly my brothers Johan and Derek Kropff, thank you for your support and listening to me rant while showing me another viewpoint for my problems.

# Abstract

The feral Tankwa goat is a landrace breed of domesticated goat (*Capra Hircus*) originating from the Tankwa Karoo National Park. This landrace breed is being managed and conserved by the Northern Cape Department of Agriculture, Environmental Affairs, Rural Development and Land Reform (DAERL) on a farm in Carnarvon. This study aimed to genetically characterise the Tankwa goat, identify signatures of selection for adaptability and propose recommendations for the successful management of this landrace breed. A total of 360 Tankwa goats was genotyped on the Illumina GoatSNP50 BeadChip at the Biotechnology Platform of the Agricultural Research Council. In addition, data on 48 Angora and 40 dairy genotypes were provided by the University of Pretoria. The two commercial goat breeds (Angora and dairy) was used as reference for comparison with the Tankwa goats. Sample and marker based quality control were completed using PLINK. The Tankwa population was genetically characterised by calculating genetic diversity parameters that included observed ($H_O$) and expected ($H_E$) heterozygosity, minor allele frequency (MAF), runs of homozygosity (ROH) and linkage disequilibrium (LD), inbreeding using PLINK; effective population size ($N_e$) using SNeP; population structure using ADMIXURE and relatedness using GCTA. Signatures of selection were determined using the $F_{ST}$ method in PLINK. The Tankwa goat had a MAF of 0.249, Expected and Observed Heterozygosity of 0.368 and 0.367 respectively and LD using $r^2$ of 0.469. Inbreeding was calculated using the $F_{IS}$ and $F_{ROH}$ statistics, with low levels reported (Average $F_{IS}$ = -0.107 and Average $F_{ROH}$ = 0.006). $N_e$ was estimated at 60 individuals 12 generations ago using SNeP. The Tankwa goat population showed no sub-structure within the population but had a clear separation from the other breeds confirming their genetic uniqueness. This result was supported by the principal component analysis (PCA). Signatures of selection identified 50 SNP's under selection, the top 0.1% was considered significant, 49 genes were identified as possibly significant for adaptation. Of these *GJB2*, *GJB6* and *GJA3* on Chromosome 12 were previously associated with adaptation to heat and drought in other breeds. Genes *GJB2* and *GJB6* are involved in the sensory perception of sound, while *GJA3* and *OPA3* is involved in visual perception. These genes could be important for foraging or escape from predators.

# Table of Contents

# List of Abbreviations

| | |
|---|---|
| ARC | Agricultural Research Council |
| CfCS | Centre for Conservation Science |
| CGIAR | Consultative Group on International Agricultural Research |
| CHI | Chromosome |
| CV | Cross-validation |
| DAERL | Northern Cape Department of Agriculture, Environmental Affairs, Rural Development and Land Reform |
| DALRD | National Department of Agriculture and Land Reform |
| DNA | Deoxyribonucleic Acid |
| EDTA | Ethylenediaminetetraacetic Acid |
| FAO | Food and Agriculture Organization of the United Nations |
| $F_{IS}$ | Inbreeding Coefficient |
| $F_{PED}$ | Inbreeding calculated with phenotypic data |
| $F_{ROH}$ | Inbreeding calculated using Runs of Homozygosity |
| $F_{ST}$ | Fixation Index |
| $H_E$ | Expected Heterozygosity |
| $H_O$ | Observed Heterozygosity |
| HSP | Heat Shock Protein |
| HWE | Hardy-Weinberg Equilibrium |
| IBD | Identity by Descent |
| LD | Linkage Disequilibrium |
| MAF | Minor Allele Frequency |
| *ma*$F_{ST}$ | Moving Average $F_{ST}$ |
| MB | Megabase |
| N | Number |

| | |
|---|---|
| $N_e$ | Effective Population Size |
| PCA | Principal Component Analysis |
| ROH | Runs of Homozygosity |
| SANBI | South African National Biodiversity Institute |
| SANParks | South African National Parks |
| SFS | Site Frequency Spectrum |
| SNP | Single Nucleotide Polymorphism |
| UP | University of Pretoria |

# List of Tables

# List of Figures

# Chapter 1: Introduction

## 1.1 Introduction

Goats were one of the first domesticated ungulates, this happened about 10 000 years ago in the Fertile Crescent (Benjelloun *et al.*, 2015; Amills *et al.*, 2017). It is commonly accepted that domesticated goats originated from the bezoar (*Capra aegagrus*) which still occurs in the wild today (Naderi *et al.*, 2008; Masseti, 2009; Bertolini *et al.*, 2018). After domestication, goats dispersed from the Fertile Crescent to surrounding areas following human migration routes, reaching Greece between 8 500 and 9 000 years ago, North Africa 7 000 years ago and the British Isles 4 000 years ago (Amills *et al.*, 2017; Onzima *et al.*, 2018b).

As humans migrated, goats adapted well to different environments and husbandry practices. Goat are considered one of the most adaptable domestic species and are found in desert areas, mountains and the tropics (Amills *et al.*, 2017; El-Halawany *et al.*, 2017). There are more than 500 recognised goat breeds (FAO, 2015), and they differ from each other in terms of phenotypic characteristics, reproductive performance, production performance and environmental adaptation (Brito *et al.*, 2017; Bertolini *et al.*, 2018). Goats are an essential livestock species, especially in developing countries, where they have important cultural, nutritional and socio-economic significance (Onzima *et al.*, 2018b).

Heat stress reportedly affects growth and production in small ruminants, it also affects meat quality by changing the physiochemical and sensory attributes of the animals (Gaughan *et al.*, 2019). Goats are generally considered the domesticated animal with the best adaptation to harsh, warm climates. They possess unique characteristics such as water conservation capability, high sweating rate, lower metabolic rate, higher respiratory rate, higher surface temperature and relatively constant heart rate and cardiac output. These physical traits provide the ability of goats to survive and perform in these high heat environments (Gaughan *et al.*, 2019; Sejian *et al.*, 2019).

The genetic characterisation of goats can be used to study the variation between breeds. This can also be used to investigate breed history and guide management decisions. Genetic studies can also be used to identify regions of the genome associated with the adaptability of a specific goat breed to its environment (Ajmone-Marsan *et al.*, 2014; Bertolini *et al.*, 2018).

Small populations are at risk of loss of diversity due to inbreeding, cross-breeding and natural factors such as disease. The conservation of such populations requires continuous genetic monitoring over time to ensure the management decisions made will not adversely impact the

population by increasing inbreeding or reducing population size below the recommended effective population size (FAO, 2011; Meek *et al.*, 2015; Allendorf, 2017; Monau *et al.*, 2020a). Genome-wide SNP data is a tool that can be used to investigate the whole genome and thus monitor the population (Colli *et al.*, 2018).

The Tankwa feral goat is a unique genetic resource in South Africa (Kotzé *et al.*, 2014). The Tankwa goat is successfully established at the Carnarvon Research Station (Northern Cape, South Africa) where they are free roaming. Carnarvon is semi-desert Karoo with annual temperatures ranging from 7 to 34 °C and little rainfall. The goats are only minimally managed, with health checks and data and sample collection happening twice a year. The original herd comprised of 60 animals and has grown to about 300 individuals (Personal communication with Deon Kriel - DKriel@ncpg.gov.za).

The need to assess genetic variability within the landrace breed and to measure adaptation to the harsh environment, with high temperatures and little food and water was identified by the Northern Cape Department of Agriculture, Environmental Affairs, Rural Development and Land Reform (DAERL). This study will include the comprehensive genetic characterisation using SNP's of the landrace to inform future conservation actions. Identifying any signatures of selection and genes associated with adaptation to the harsh environment where they occur can contribute to optimal management of the population.

### 1.2 Aims and Objectives

The aim of this project is to genetically characterise the Tankwa goat population and to identify signatures of selection with a specific focus on adaptation. This was done through attaining the following objectives:

1. Estimate population parameters of the Tankwa population using the commercial goat SNP chip in terms of level of linkage disequilibrium (LD), average minor allele frequency (MAF), heterozygosity ($H_E$ and $H_O$), effective population size ($N_e$) and inbreeding levels.
2. Study the genetic structuring within and between the Tankwa goat population and two commercial breeds.
3. Identify signatures of selection associated with adaptation by comparing the Tankwa goat population with two commercial goat breeds.
4. Propose management recommendations for the conservation of the Tankwa goat based on the genetic results of the analysis.

# Chapter 2: Literature Review

## 2.1 Introduction

The world domestic goat (*Capra hircus*) population is currently estimated to be more than one billion animals (FAOStat, accessed 12 September 2022). The majority of goats are kept in Asia (51.36%) and Africa (43.35%) with the remainder in America (3.47%), Europe (1.44%) and Oceania (0.38%) (FAOStat, accessed 12 September 2022; Table 2.1). Since the year 2000, goats have experienced the largest population growth of all the major livestock species, most likely due to an increased demand for subsistence farming rather than the development of the species (Dubeuf & Boyazoglu, 2009). Worldwide, more than 500 goat breeds are recognised that generally occur in rural areas with basic management and poor nutrition (FAO, 2015). Goats are considered one of the most adaptable domesticated species in the world (Amills *et al.*, 2018; Bertolini *et al.*, 2018; Onzima *et al.*, 2018b). These animals are considered easy to keep, with low maintenance cost, which makes them a fundamental part of many households in rural communities (Bertolini *et al.*, 2018).

## 2.2 Origin and domestication of goats

Goats were one of the first ungulates to be domesticated about 10 000 years ago in the Fertile Crescent which spans from Eastern Anatolia to the Zagros mountains in Iran (Benjelloun *et al.*, 2015; Amills *et al.*, 2017). It is widely accepted that goats were domesticated from a unique wild ancestor, the bezoar (*Capra aegagrus*). The bezoar still occurs in the wild today, and is found in the South Western parts of Asia (Naderi *et al.*, 2008; Masseti, 2009; Bertolini *et al.*, 2018). After domestication, goats dispersed to surrounding areas following human migration routes, reaching Greece between 8 500 and 9 000 years ago, North Africa 7 000 years ago and the British Isles 4 000 years ago (Amills *et al.*, 2017; Onzima *et al.*, 2018c). Disease in central Africa delayed the dispersal of goats to southern Africa where they arrived approximately 2 000 years ago (Amills *et al.*, 2017; Colli *et al.*, 2018). Domestic goats migrated downwards into southern Africa through Botswana, most probably following the migration routes of the Khoisan (Amills *et al.*, 2017; Visser, 2019). Goats are now an important part of the South African livestock industry making up 13% of all livestock in South Africa (FAOStat, accessed 12 September 2022).

With the movement of human populations across the globe, goats adapted well to a wide range of environmental conditions and husbandry practices. Goats are commonly considered as a very adaptable domestic animal and are found in desert areas, mountains and the tropics (Amills *et al.*, 2017; El-Halawany *et al.*, 2017). Many goat populations gradually adapted to their local environment due to natural selection, leading to the development of diverse landraces and populations (Benjelloun *et al.*, 2015; Marsoner *et al.*, 2018). These populations were known to interbreed, with limited selection for specific traits and thus maintaining high phenotypic diversity (Benjelloun *et al.*, 2015; Marsoner *et al.*, 2018). Artificial selection however, gave rise to specialized commercial breeds (Henkel *et al.*, 2019) with specific production purposes (e.g. meat, milk or fibre).

Goats are differentiated into more than 500 breeds (FAO, 2015) that differ from each other in phenotypic characteristics, reproductive and production performance and environmental adaptation (Brito *et al.*, 2017; Bertolini *et al.*, 2018). Goats are an important livestock species on all the continents (Table 2.1; FAOStat, accessed 12 September 2022) and especially in developing countries, where they have important cultural, nutritional and socio-economic significance (Onzima *et al.*, 2018c).

Table 2.1 Population sizes of goats by continent and globally (FAOStat, accessed 12 September 2022)

| Continent | Goat population size | Percentage of global population |
|---|---|---|
| Africa | 489 021 886 | 43.35% |
| Asia | 579 347 344 | 51.36% |
| Australia and Oceania | 4 301 278 | 0.38% |
| Europe | 16 241 452 | 1.44% |
| North America | 2 685 122 | 0.24% |
| South America | 36 509 154 | 3.23% |
| **Global Total** | **1 128 106 236** | 100% |

### 2.3 Indigenous goats

Indigenous breeds are described as locally adapted goats that underwent no or limited artificial selection, with natural selection playing a significant role in their development (Onzima *et al.*, 2018b). They are considered a valuable genetic resource due to their adaptation to a diverse range of harsh environments, including specific environmental conditions such as high temperatures, high incidences of disease and a lack of water availability (Webb & Mamabolo, 2004; Ajmone-Marsan *et al.*, 2014; Benjelloun *et al.*, 2015; Onzima *et al.*, 2018a). Harsh

environments allowed the formation of breeds with unique adaptive traits to survive and reproduce in areas where they are kept or farmed (Mwai *et al.*, 2015; Monau *et al.*, 2020a). Indigenous breeds play an important role in maintaining across-breed genetic diversity as a reservoir for rare genetic material (Biscarini *et al.*, 2015).

Indigenous breeds are predominantly part of small-holder farming systems (Mhlanga *et al.*, 2018) and contribute to the livelihoods of many resource poor communities, playing varying roles across different religious and cultural groups (Mataveia *et al.*, 2018). Goats are important for the production of meat, milk, manure, mohair, cashmere and skins; are an easily accessible source of income for rural farmers; and are also used in traditional ceremonies (Mataveia *et al.*, 2018; Mhlanga *et al.*, 2018). Goats are opportunistic herbivores (Aldezabal & Garin, 2000), and are usually browsers rather than grazers (Badenhorst, 2006) and can utilize land that is not suitable for growing crops (Mhlanga *et al.*, 2018). Indigenous goats are found in a wide range of environments in Southern Africa where they are mostly used to produce meat, milk and skins and to trade for other items (Onzima *et al.*, 2018a; Monau *et al.*, 2020b). Indigenous goats make up 95% of the goat populations in Africa (Visser, 2019). The indigenous goat breeds are commonly identified by the region or tribe they are associated with, and are characterised based on morphology and coat variation. However, indigenous populations commonly share phenotypic descriptions and are often only separated by geographical origin (Mdladla *et al.*, 2017; Onzima *et al.*, 2018c; Visser, 2019).

Phenotypic characterisation is one of the first ways to classify a breed or landrace, however this does not accurately reflect the underlying genetic similarities or differences and thus genetic characterisation is also needed (Selolo *et al.*, 2015; Visser, 2019). Limited studies on phenotypic characterisation have been performed in Southern Africa, usually such characterisations form part of a larger study focussing on a specific breed (Baleseng *et al.*, 2016; Monau *et al.*, 2018; Visser, 2019). Baleseng *et al.* (2016) studied performance in the Tswana goats in Botswana, characterising them at the same time.

Cosmopolitan breeds, such as the Angora, Boer and Saanen goats, are described as goats that underwent artificial selection for specific traits, usually phenotypic and production traits (Brito *et al.*, 2017; Henkel *et al.*, 2019). Farming with cosmopolitan breeds have become more widespread due to short term economic benefits and this resulted in the substitution of indigenous / locally adapted breeds (Martínez *et al.*, 2012; Benjelloun *et al.*, 2015). In 2015, the Food and Agriculture Organization of the United Nations (FAO) estimated that 18% of global indigenous goat breeds were threatened or extinct (FAO, 2015). These breeds contribute to locally adapted genetic resources and face various threats such as high levels of inbreeding (Ajmone-Marsan *et al.*, 2014; Onzima *et al.*, 2018a), as well as indiscriminate cross

breeding with commercial (cosmopolitan) breeds, leading to the loss of unique genetic variation and adaptability. Inbreeding caused by a reduction in population size due to farmers replacing the indigenous animals with more commercial breeds, is considered a more serious threat than cross breeding (Ajmone-Marsan *et al.*, 2014; Onzima *et al.*, 2018a). Thus, it is crucial to assess the indigenous goat genetic resources worldwide through genetic characterisation in order to sustain and conserve unique populations and breeds that can contribute to food security, especially in areas affected by climate change (Ajmone-Marsan *et al.*, 2014; Benjelloun *et al.*, 2015; Monau *et al.*, 2018).

### 2.4 Adaptation and climate change

It is well known that animals that are well adapted to a specific environment, are more likely to survive and pass on their genes to next generations (Aleena *et al.*, 2018; Gaughan *et al.*, 2019). These populations will have a higher overall proportion of alleles that favour survival in the specific environment, especially if the animals are not intensively managed, as may be the case with indigenous goats (Sejian *et al.*, 2019).

Climate change is considered a major threat to species survival (Benjelloun *et al.*, 2015; Mhlanga *et al.*, 2018). Animals occupy habitats by maintaining a balance between foraging for food and water, and sheltering from extreme temperature and rainfall pressures (Dunbar & Shi, 2013). Climate change will affect this balance making it harder for animals to survive in their current habitats (Dunbar & Shi, 2013). The specific elements affected by climate change include: the quality and availability of food and water sources; an increased rate of new pests and disease outbreaks; longer periods of drought; and an increase in the frequency of extreme weather events such as heat waves (Sejian *et al.*, 2019). The effects of climate change may be direct or indirect, animals may succumb in below zero temperatures or a drought might limit food availability, either in quantity or quality (Dunbar & Shi, 2013).

With climate change placing pressure on natural resources, efficient livestock production will have to increase drastically to supply food for the human population, while using less land and water (Van Marle-Köster & Visser, 2018). There are limited global studies on the effects of climate change on livestock systems (Rojas-downing *et al.*, 2017; Gaughan *et al.*, 2019; Sejian *et al.*, 2019; Giger-reverdin *et al.*, 2020). Singh *et al.* (2012) studied the effect of climate change on livestock in the Western Himalayan and Middle Gangetic Plain region of India where more than 50% of farmers experienced a decrease in productivity and 60% reported an increase in disease incidences. The Consultative Group on International Agricultural Research (CGIAR) published a paper on the impact of climate change in Africa (Thornton *et al.,* 2015). They concluded that a wide range of adaptive options exist to cope with climate

change, however no single method can be used universally. They suggested methods such as selective breeding for more robust animals, the management of genetic resources to ensure higher diversity in livestock and governments taking responsibility for generating markets to ensure income. When faced with the effects of climate change, farmers will need to implement intensive and expensive management strategies for the survival of commercial breeds. Conversely, if they farm with indigenous breeds there should be no need to extensively manage the population for their survival (Singh *et al.*, 2012).

One of the most crucial factors affecting the productive and reproductive ability of small ruminants, such as goats, appears to be heat stress (Sejian *et al.*, 2019). Heat stress is usually associated with lack of water availability and intake, and leads to a decrease in foraging (Giger-Reverdin *et al.*, 2020). The effect of heat stress on small ruminant production systems include reduced feeding and increased water consumption with a reduction in the quality and quantity of products such as meat and milk. These effects of climate change lead to an overall increase in the financial burden placed on rural farmers (Gaughan *et al.*, 2019; Sejian *et al.*, 2019).

Animals respond to changing environments by changes in their phenotype and genotype over many generations. An animal's adaptability is evaluated based on their ability to survive as well as reproduce in harsh environments. This adaptability is controlled by many different factors, including morphology and behaviour (Sejian *et al.*, 2019). Knowledge of the adaptive responses to heat stress and the genetic basis for these responses may allow for the identification of breeds with a high heat tolerance (Gaughan *et al.*, 2019). Thus genetic strategies to combat climate change in the livestock industry should include the identification of breeds that are adapted to climatic stress, the identification of genes associated with heat resistance and cross-breeding of heat resistant and higher productivity breeds (Gaughan *et al.*, 2019). Some studies have been done to identify genes associated with adaptation in indigenous goats (Aleena *et al.*, 2018; Onzima *et al.*, 2018b; Gaughan *et al.*, 2019). Aleena *et al.* (2018) studied three indigenous goat populations in India with a focus on heat shock protein 70 (HSP70). To this end, indigenous goats (such as the Black Bedouin goat in the Middle East) are well adapted to survive in harsh climates and breeds from arid areas can cope with heat stress better than those from milder areas (Giger-Reverdin *et al.*, 2020).

An example of adaptation to heat stress can be seen in the Black Bedouin goat, which shows adaptability to dehydration by losing water from the rumen to ensure a normal water balance in the blood and tissues. They can lose up to 40% of their body weight when faced with prolonged dehydration and can drink between 20% and 40% of their body weight at a time when a water source is found (Giger-Reverdin *et al.*, 2020).

Shaji *et al.* (2016) studied heat shock proteins (HSP) in Osmanabadi goats. These proteins can be used for evaluating thermal adaptability in animals. They reported that the heat shock protein HSP70 was expressed when the animals were exposed to both heat and nutritional stress. Dangi *et al.* (2015) performed a similar study that investigated Barbari goats. They found an increase in the expression of HSP60, HSP70, HSP90 and HSP105/110 under heat stress. This study was supported by a study on HSP70s in Indian goats (Aleena *et al.*, 2018)

A total of 119 candidate genes associated with adaptation to a hot humid environment were identified in Egyptian Barki sheep and goats (Kim *et al.*, 2016). These genes were involved in energy metabolism, the nervous and endocrine system, body development, testicular embryogenesis, immune and inflammatory responses, and coat colour.

### 2.5 The feral Tankwa goat

The feral Tankwa goat was declared a landrace by National Department of Agriculture and Land Reform (DALRD). These goats have been free-roaming in the Tankwa Karoo National Park for more than 80 years with little human interference (Kotzé *et al.*, 2014; Ahmed *et al.*, 2018). Now known as the Tankwa goat, a small population still occurs in the Tankwa Karoo National Park where they were originally discovered. With the establishment of a new national park, South African National Parks (SANParks) initiated the removal of most of these goats in 2007 as they were considered a threat to the natural and unique succulent Karoo plant diversity (Chynoweth *et al.*, 2013).

The Department of Agriculture, Environmental Affairs, Rural Development and Land Reform (DAERL), Northern Cape moved 60 animals to the research station in Carnarvon in 2009 (Mdladla *et al.*, 2018). The Carnarvon population is currently being maintained at approximately 300 individuals (Ahmed *et al.*, 2018).

There is a high degree of phenotypic variation with regards to the general characteristics such as coat colour, coat length, horn shape, nose shape and general morphology in the Tankwa goat (Figure 2.1). The goats are all horned, the average adult female weighs between 20 and 40kg and the average adult male weighs between 25 and 55kg.

Figure 2.1 Phenotypic variation observed in the feral goats from the Tankwa Karoo National Park (Photos by Thinus Jonker)

The Tankwa goats have been under intense natural selection pressure due to the harsh environment they lived in, namely the semi-desert Karoo. Over the last 10 years, the Tankwa Karoo National Park had temperatures ranging from an annual minimum of 4°C to a maximum of 28°C. The average annual rainfall is 137 mm per year, which only occur on an average of 58 days a year (World Weather Online, Accessed 9 June 2021). Over the same period, temperatures in Carnarvon ranged from a yearly minimum of 7°C to a maximum of 34°C, the average annual rainfall is 78 mm per year and it only rains on an average 40 days a year (World Weather Online, Accessed 9 June 2021). The Tankwa goats currently survive in this hot and dry environment, where they roam freely with minimal management intervention.

### 2.6 Genetic characterisation of indigenous goats

Characterization of breeds is considered a strategic priority for the development of a national plan for the management of animal genetic resources (FAO, 2011; Ajmone-Marsan *et al.*, 2014). In addition, it supports conservation as well as the effective utilization of the indigenous genetic resources (Visser, 2019). Optimal utilization of such resources is key for the effective

conservation, possible breed improvement and the management of small populations (Monau *et al.*, 2020a). Genetic characterisation is required to determine estimates of genetic diversity and breed differentiation (Visser, 2019), however studies of genetic characterisation on goats are still lacking (Brito *et al.*, 2017; Monau *et al.*, 2020a). Investigating the genetic diversity of populations by using various molecular markers, can also expand our insight into breed history (Martínez *et al.*, 2012; Ajmone-Marsan *et al.*, 2014). Studies on genetic diversity, population structure and genetic relatedness can guide decisions on breed development (Monau *et al.*, 2020a).

Genetic characterization can be performed using genetic markers. The two most common examples of such markers are microsatellites and Single Nucleotide Polymorphisms (SNP's). Microsatellites are fragments of the genome where short genetic sequences are repeated, and the number of these repeats differ between individuals. One limitation of microsatellites is that they are only found in the non-coding regions of the genome and can therefore not be used in studies of gene expression and selection (Lenstra *et al.*, 2012; Miller *et al.*, 2014). A benefit of these markers are their high level of polymorphism, thus increasing the amount of data that can be obtained per marker. SNP's on the other hand are single base pair changes found in the genome, spread throughout the genome and thus more useful for studies on genes, expression and selection. SNP markers are bi-allelic, decreasing the amount of data obtainable per marker and thus more markers are needed to obtain useful information (Lenstra *et al.*, 2012; Miller *et al.*, 2014).

### *2.6.1 Microsatellite markers*

The FAO recommends a panel of microsatellite markers for the genetic characterisation of each of the main livestock species (FAO, 2011) with 30 markers suggested for goats. The use of the same panel of markers across different projects would allow researchers to combine different datasets for improved analysis. However, this was only partially successful as most studies only used a small subset of the suggested marker panel (Ajmone-Marsan *et al.*, 2014; Benjelloun *et al.*, 2015; Brito *et al.*, 2017). In addition, although microsatellite markers are highly polymorphic and therefore informative for diversity studies (Ajmone-Marsan *et al.*, 2014; Van Marle-Köster & Visser, 2018), genotyping and scoring may be labour intensive and is difficult to standardise between laboratories and instruments, if a large number of shared samples are not available (Ajmone-Marsan *et al.*, 2014). Some examples of previous genetic studies on South African goat populations using microsatellite markers are listed in Table 2.2.

Table 2.2 A non-comprehensive list of studies using microsatellite markers to study South African goat populations

| Breed/s | Type of Study | n Animals | n Markers | Reference |
|---|---|---|---|---|
| Tankwa<br>Boer<br>Angora<br>Saanen | Genetic Structure | 224 | 8 | Kotzé *et al.*, 2014 |
| Saanen<br>Toggenberg<br>British Alpine | Genetic Diversity | 240 | 25 | Bosman *et al.*, 2015 |
| Tswana | Genetic Variation | 87 | 12 | Maletsanake *et al.*, 2013 |
| Angora | Parentage | 381 | 12 | Garritsen *et al.*, 2015 |
| Angora | Parentage | 192 | 14 | Visser *et al.*, 2011b |
| Angora | Quantitative Trait Loci Identification | 695 | 134 | Visser *et al.*, 2011a, 2013 |
| Boer<br>Kalahari Red<br>Savannah | Genetic Characterization | 177 | 18 | Pieters *et al.*, 2009 |
| Indigenous<br>Toggenburg | Genetic Characterization | 595 | 19 | Chenyambuga *et al.*, 2004 |

Microsatellite studies on southern African goats showed that indigenous populations generally have high genetic diversity (based on heterozygosity parameters) and low levels of inbreeding (Pieters *et al.*, 2009; Maletsanake *et al.*, 2013),

### *2.6.2 Single Nucleotide Polymorphisms (SNP)*

The development of species-specific tools for genomics studies (such as SNP arrays) allows for the study of genomes at a high resolution. Current genomic technologies available for goats includes the Illumina CaprineSNP50 BeadChip (Tosser-Klopp *et al.*, 2014) which was developed using ten geographically and biologically different breeds and a goat reference

genome (Dong *et al.*, 2013; Bickhart *et al.*, 2017). This SNP array contains 53,347 SNP's distributed across the goat genome covering 97.3% of the total genome length (Ajmone-Marsan *et al.*, 2014; Tosser-Klopp *et al.*, 2014). The number of SNP's range from 855 on CHI25 (Chromosome 25) to 3256 on CHI1, with between 90.6% (CHI18) and 99.6% (CHI17) of the genome covered by SNP's. An updated version of the SNP array was developed in 2021 with 59,727 SNP's (Goat Genome). The ARS1 reference goat genome published in 2017 (Bickhart *et al.*, 2017) has improved gene annotation and allowed for the update of the SNP positions on the CaprineSNP50 BeadChip (Bertolini *et al.*, 2018). Even though goats are of high economic importance, few genome-wide studies for this species have been reported (Bertolini *et al.*, 2018, Table 2.3).

SNP are the more abundant and well-distributed sequence variants in the genome (Sobrino *et al.*, 2005). An advantage of SNP's is that they have low mutation rates. One disadvantage of SNP based studies is ascertainment bias (Willing *et al.*, 2012; Ajmone-Marsan *et al.*, 2014). As only 10 different breeds were used to develop the Illumina CaprineSNP50 BeadChip, detection of variants common in those breeds are more likely than rare or unique variants, which may skew the Minor Allele Frequency (MAF) (Willing *et al.*, 2012). The 50K SNP chip technology is considered robust, has a relatively lower cost than older technologies such as microsatellites considering the amount of data generated and has automatic allele calling reducing the labour intensity of the genotyping (Ajmone-Marsan *et al.*, 2014). Large-scale collaborative projects such as the AdaptMap Project (Stella *et al.*, 2018) is possible when using the 50K SNP chip to generate data across several different projects.

Table 2.3 A non-comprehensive list of studies on South African goats using the CaprineSNP50 BeadChip

| Breed | Type of study | n Animals | Authors |
|---|---|---|---|
| Angora<br>Saanen<br>British Alpine<br>Toggenburg | Genetic Diversity | 88 | Lashmar *et al.*, 2016 |
| Angora | Genetic Structure and Signatures of Selection | 43 | Paim *et al.*, 2019 |
| Angora | Genetic Structure | 40 | Visser *et al.*, 2016 |
| Boer<br>Kalahari Red<br>Savanna<br>Tankwa<br>Indigenous breeds | Genetic Structure | 239 | Mdladla *et al.*, 2016 |
| Indigenous Breeds | Genetic Structure and Relatedness | 217 | Chokoe *et al.*, 2022 |
| Tswana<br>Boer<br>Kalahari Red<br>Swazi | Genetic Structure | 219 | Monau *et al.*, 2020b |
| Boer<br>Indigenous Breeds | Genetic Structure and Growth | 72 | Ncube *et al.*, 2020 |
| Boer<br>Indigenous Breeds | GWAS | 72 | Ncube *et al.*, 2022 |
| Indigenous Breeds | Genetic Diversity | 117 | Chokoe *et al.*, 2020 |

## 2.7 Estimating genetic diversity and population structure

Genetic diversity is calculated using parameters such as observed and expected heterozygosity ($H_O$ and $H_E$), average minor allele frequency (MAF), inbreeding, and effective population size ($N_e$) (Saravanan *et al.*, 2022). Minor allele frequency refers to the frequency that the rarer allele is observed in the population if working with SNP's (bi-allelic data) and the frequency of the second most common allele when other data is used. Observed heterozygosity is compared to expected heterozygosity, more genetically variable populations will have higher heterozygosity (Saravanan *et al.*, 2022). Population wide observed heterozygosity ($H_O$) is calculated per locus by observing how many individuals in the

population are heterozygous at a specific locus; heterozygosity can also be calculated per individual as a proportion of loci at which an individual is heterozygous using the following formula (Miller *et al.*, 2014; Eusebi *et al.*, 2019):

$$Ho = (N(NM)-O(Hom))/N(NM)$$

Where N(NM) is the number of non-missing genotypes; and O(Hom) is the number of observed homozygotes in the population

Expected heterozygosity ($H_E$) is defined as the probability that two alleles chosen from the population at random will be different and is calculated using the following formula derived from the formula for Hardy-Weinberg equilibrium:

$$H_E=1-\sum p^2$$

Where $p^2$ is the allele frequency of the dominant allele.

Runs of homozygosity is defined as stretches of genome that are homozygous, this parameter can be used as a measure of inbreeding as well as genetic diversity (Purfield *et al.*, 2012; Curik *et al.*, 2014; Peripolli *et al.*, 2016). Runs of homozygosity could indicate identity by descent (IBD) and is caused by as genetic drift, population bottlenecks, inbreeding or intensive natural and artificial selection (Peripolli *et al.*, 2016; Saravanan *et al.*, 2022)

Inbreeding needs to be monitored to prevent or reduce the potential effect of loss of variation, inbreeding depression and an increase in the expression of deleterious alleles (Zhang *et al.*, 2015; Brito *et al.*, 2017). Genomic measures of inbreeding are more accurate than methods using phenotypic data ($F_{PED}$), as accurate pedigree data is needed for $F_{PED}$ and $F_{PED}$ does not reflect data from the proportion of the genome that is not expressed in the phenotype (Barbato *et al.*, 2015; Peripolli *et al.*, 2016). Two genomic measures of inbreeding that is commonly used, include the inbreeding coefficient ($F_{IS}$) and inbreeding calculated from runs of homozygosity ($F_{ROH}$).

$F_{IS}$ is defined as the probability of an individual having two alleles that are identical by descent (IBD) at a specific locus when compared to a base population where all alleles are independent (Marras *et al.*, 2014; Zhang *et al.*, 2015). $F_{IS}$ is calculated using $H_O$ and $H_E$ with the formula:

$$F_{IS} = (H_E - H_O) / H_E.$$

$F_{ROH}$ uses the proportion of the genome covered by runs of homozygosity to calculate inbreeding. Longer sections indicate recent inbreeding while shorter ROHs show more ancient inbreeding or relatedness. The shorter sections in domesticated animals may be caused by

the founder effect that occurs with the formation of a new breed (Purfield *et al.*, 2012; Zhang *et al.*, 2015; Peripolli *et al.*, 2016). The minimum length of a ROH that can be detected depends on the density of loci used. This method has become more popular with the development of medium and high density SNP chips (Marras *et al.*, 2014; Zhang *et al.*, 2015). $F_{ROH}$ is calculated using the formula:

$$F_{ROH} = \sum (L_{ROH} / L_{AUTO})$$

where $L_{ROH}$ is the total length of ROH in the animal and $L_{AUTO}$ is the total length of the autosome covered by SNP's (Purfield *et al.*, 2012; Peripolli *et al.*, 2016).

The effective population size ($N_e$) of a population is defined as the size of a hypothetical population that will have the same amount of genetic diversity as the study population (Barbato *et al.*, 2015; Makina *et al.*, 2015b). $N_e$ estimates are calculated from genomic data, specifically using linkage disequilibrium (LD), and calculates the level of genetic drift in a population. LD is the association of alleles from different loci with each other. This can result from processes such as admixture or genetic drift or from loci that are associated with a trait under selection (Barbato *et al.*, 2015). $N_e$ is estimated using the formula:

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

Where $N_T$ is the past effective population size, (t) is the number of generations ago, c is the recombination rate, $r^2_{adj}$ is the adjusted LD estimation (LD is adjusted for sampling bias), and α is a constant (Barbato *et al.*, 2015).

### *2.7.1 Previous SNP-based studies on the genetic characterisation of goats*

Several studies have been conducted on the genetic characterisation of goat breeds, in various parts of the world and in a great variety of breeds (Table 2.4 and Addendum A). Heterozygosity and inbreeding estimates vary greatly between goat breeds with the average inbreeding ranging from -0.139 to 0.236, while observed heterozygosity ranging from 0.324 to 0.427. Some of these studies included different goat breeds (Nicoloso *et al.*, 2015; Brito *et al.*, 2017; Tarekegn *et al.*, 2019; Chokoe *et al.*, 2020), while others investigated different populations within the same breed (Visser *et al.*, 2016). In most studies the population parameters were defined as part of a larger study on runs of homozygosity, population structure, linkage disequilibrium or inbreeding (Mdladla *et al.*, 2016; Visser *et al.*, 2016; Brito *et al.*, 2017).

Table 2.4 A non-comprehensive list of genetic characterisation studies done in South Africa using the CaprineSNP50 BeadChip and the genetic diversity results obtained

| Breed | Number of individuals | Inbreeding Coefficient - $F_{IS}$ | Observed Heterozygosity - $H_O$ | Expected Heterozygosity - $H_E$ | Authors |
|---|---|---|---|---|---|
| Angora | 48 | 0.009 | 0.324 | 0.333 | Lashmar *et al.*, 2016 |
| Angora | 43 | 0.227 | 0.333 | - | Paim *et al.*, 2019 |
| Angora | 48 | 0.009 | 0.365 | 0.371 | Visser *et al.*, 2016 |
| Boer | 31 | 0.12 | 0.36 | 0.37 | Mdladla *et al.*, 2016 |
| British Alpine | 14 | -0.088 | 0.385 | 0.355 | Lashmar *et al.*, 2016 |
| Kalahari Red | 36 | 0.1 | 0.37 | 0.38 | Mdladla *et al.*, 2016 |
| Nguni | 10 | 0.01 | 0.41 | 0.39 | Mdladla *et al.*, 2016 |
| Saanen | 20 | -0.015 | 0.378 | 0.373 | Lashmar *et al.*, 2016 |
| Savanna | 29 | 0.06 | 0.39 | 0.38 | Mdladla *et al.*, 2016 |
| Toggenburg | 6 | -0.135 | 0.385 | 0.339 | Lashmar *et al.*, 2016 |
| Venda | 25 | 0.04 | 0.41 | 0.4 | Mdladla *et al.*, 2016 |
| Xhosa | 20 | 0.02 | 0.42 | 0.41 | Mdladla *et al.*, 2016 |
| Zulu | 25 | 0.04 | 0.4 | 0.4 | Mdladla *et al.*, 2016 |
| Boer | 24 | 0.014 | 0.384 | 0.39 | Monau *et al.*, 2020b |
| Kalahari Red | 23 | 0.012 | 0.393 | 0.398 | Monau *et al.*, 2020b |
| Tswana | 48 | 0.019 | 0.405 | 0.413 | Monau *et al.*, 2020b |
| Swazi | 48 | 0.011 | 0.383 | 0.387 | Monau *et al.*, 2020b |
| Indigenous | 117 | -- | 0.403 | 0.390 | Chokoe *et al.*, 2020 |

### 2.8 Signatures of selection

Natural selection plays an essential role in the differentiation of breeds that can survive in specific environments. Artificial selection has also been widely applied to livestock to obtain more desirable traits (Brito *et al.*, 2017). Alterations observed across the genome due to selection are more commonly referred to as signatures of selection (Brito *et al.*, 2017; Gouveia *et al.*, 2017; Purfield *et al.*, 2017; Saravanan *et al.*, 2020). This include adaptive traits such as heat tolerance and disease resistance, reproductive traits, production traits and behavioural traits (Moradi *et al.*, 2012; Brito *et al.*, 2017). Signatures of selection are defined as the reduction, loss or change in genetic variation of a genomic region that are next to, or overlaps, the functional gene selected upon (Talenti *et al.*, 2017; Dolebo *et al.*, 2019).

The detection of signatures of selection is important for modern population genetics as it can be used to study evolutionary history and to identify beneficial mutations (Fariello *et al.*, 2013; Saravanan *et al.* 2020). Strong positive selection leads to the rapid fixation of a gene under selection as well as neutral genomic regions around the gene. The rapid fixation of a gene under selection will increases the genetic divergence between populations under selection and populations not under selection (Ajmone-Marsan *et al.*, 2014; Qanbari & Simianer, 2014). A hard selective sweep occurs when selection leads to the fixation of a rare variant (usually rapidly), whereas a soft selective sweep changes the frequency of genetic variants in the population (Andersson, 2013; Purfield *et al.*, 2017; Bertolini *et al.*, 2018). Another form of selection is balancing selection; this actively maintains the diversity in the population for longer than is expected under neutral genetic drift (Horscroft *et al.*, 2019). Regions under selection can be detected by examining markers or haplotypes with allele frequencies that differ from the expected within or between populations (Avila *et al.*, 2018; Gorssen *et al.*, 2021). Identifying regions of the genome under selection is of importance both for conservation and the analyses of complex traits. Loci under selection could reveal functional adaptation, this along with information obtained from neutral regions, are important for conservation decision making (Fariello *et al.*, 2013).

### *2.8.1 Methods for identifying signatures of selection*

Signatures of selection can be identified using two main approaches: intra-population statistics and inter-population statistics. These methods can further be divided into multiple categories and methods (Makina *et al.*, 2015a; Nicolazzi *et al.*, 2015; Bertolini *et al.*, 2018; Saravanan *et al.*, 2020). Intra-population statistics include three main categories namely the site frequency spectrum (SFS), Linkage Disequilibrium (LD) and reduced local variability methodologies. Inter-population statistics include two main methods namely single site differentiation and haplotype based differentiation (Saravanan *et al.*, 2020).

SFS methods are based on the distribution of allele frequencies in a population, this can be used as selective sweeps cause an increase in genetic variants with high frequencies, while reducing genetic variants with low frequencies (Saravanan *et al.*, 2020). LD based methods are based on the principle that selection on a specific gene does not allow for recombination, thus increasing the linkage disequilibrium between the regions. This is useful for detecting regions under partial selection (Saravanan *et al.*, 2020). Reduced local variability methods investigates variation relative to the genome average; this method uses the theory that selective sweeps should have an increase in stretches of homozygous loci (Saravanan *et al.*, 2020).

Single site-differentiation methods include $F_{ST}$ and FLK (an extension of the Lewontin and Krakauer test using a population kinship matrix). $F_{ST}$-based analysis identifies regions of the genome where allele frequencies are significantly different between different populations by calculating the pairwise fixation index between populations (Gouveia *et al.*, 2017; Onzima *et al.*, 2018c; Zheng *et al.*, 2020). Populations with highly different allele frequencies indicate positive selection, FLK compares variances (observed to expected) in the $F_{ST}$ along with a population kinship matrix (F) which accounts for changes in the effective population size (Saravanan *et al.*, 2020). Haplotype based differentiation methods uses the haplotype information in different populations to identify areas of selection (Saravanan *et al.*, 2020).

### *2.8.2 Previous studies on signatures of selection*

Studies on signatures of selection and the identification of genes involved in selection have been conducted on several livestock species including cattle (Gutiérrez-Gil *et al.*, 2015; Zhao *et al.*, 2015; Bhati *et al.*, 2020; Xia *et al.*, 2021; Zhang *et al.*, 2022), sheep (Moradi *et al.*, 2012; McRae *et al.*, 2014; Yurchenko *et al.*, 2019), chickens (Stainton *et al.*, 2014; Nanaei *et al.*, 2022) and pigs (Ai *et al.*, 2013, 2014). Some studies on signatures of selection in goats have been done (Guan *et al.*, 2016; Brito *et al.*, 2017; Talenti *et al.*, 2017; Bertolini *et al.*, 2018). Alberto *et al.* (2018) identified signatures of selection associated with domestication in sheep and goats and found that approximately half of the genomic regions identified in sheep were similarly identified in goats. One study on signatures of selection in South African goats were conducted by Mdladla *et al.* (2018), where the authors used a landscape genomics approach and compared different goat populations to identify regions of the genome associated with adaptation and identified signatures of selection in many different gene regions associated with immunity, metabolism, heat response and water scarcity.

## 2.9 Conservation and management of small populations and at-risk breeds

Sustainable utilization of breeds is essential for efficient conservation and management (Monau *et al.*, 2020a) and supports global food security (Taberlet *et al.*, 2011). Breeds can be considered at risk or endangered due to small population size or low genetic diversity (Lauvie *et al.*, 2011; Harmon and Braude, 2010; Meek *et al.*, 2015). The reduction in the overall genetic diversity of a breed can lead to a loss of rare genetic resources and the extinction of the breed (Taberlet *et al.*, 2011). Other risk factors are if the breed is only found in a single isolated area or has a small number of populations and can cross-breed with other breeds (FAO, 2011). Small populations of any animal species is of conservation concern and needs to be managed to prevent inbreeding and loss of genetic diversity. Small populations are vulnerable to the

random effects of variation in terms of climate, birth rates and sex ratios (Meek *et al.*, 2015). The conservation of such populations requires continuous genetic monitoring over time to ensure the management decisions made will not adversely impact the population by increasing inbreeding or reducing population size below the effective population size (Allendorf, 2017).

Conservation programs therefore aim to increase the population size while maintaining genetic diversity (Lauvie *et al.*, 2011). The FAO (2015) supports four methods for the conservation of breeds: increase population size, manage overall genetic diversity, select for increased productivity and cryopreserve genetic material such as sperm and embryos.

Conservation efforts can be impacted by lack of information about the animal of concern, lack of information about the environment the animals live in, too many stakeholders involved in the decision-making process and lack of a viable conservation plan (Meek *et al.*, 2015). For effective management, conservation and breeding plans are basic information required for recognized (and especially indigenous) breeds (Monau *et al.*, 2020a). This information should include the risk status, effective population size, and genetic diversity and signatures of adaptive selection for each breed (Monau *et al.*, 2020a). The breed of interest should also be compared to related breeds and those in similar environments to establish uniqueness (Monau *et al.*, 2020a). The ideal is that such populations should be enlarged through breeding and spread out to create multiple populations thus reducing the risk of extinction due to diseases or disasters. The reality is that other real-world factors such as land availability, water scarcity and money need to be considered before any management strategies can be implemented (Meek *et al.*, 2015).

# Chapter 3: Materials and Methods

## 3.1 Introduction

As part of the conservation of the Tankwa goat population, a comprehensive genetic characterization study was required. In addition, genomic analyses can identify possible signatures of selection related to the adaptation of the Tankwa goat in the harsh environment. The Northern Cape Department of Agriculture, Environmental Affairs, Rural Development and Land Reform (DAERL) is responsible for the sustainable conservation and management of the Tankwa goats and to this end prioritised and funded the generation of a comprehensive genomic dataset for the population. Official permission for the use of data was obtained from DAERL (Addendum B). As comparison, genomic data of two commercial productions types (dairy and Angora breeds) were provided by the University of Pretoria (UP). Ethical approval for the use of secondary data was obtained from the Ethics Committee of the Faculty of Natural and Agricultural Sciences at the University of Pretoria (NAS446/2019 & NAS350/2020). Research ethical approval for this study was also obtained from the South African National Biodiversity Institute's (SANBI) Research Ethics and Scientific Committee (P2020-33). All genomic data was generated using the Illumina GoatSNP50 BeadChip (Tosser-Klopp *et al.*, 2014).

## 3.2 Materials

### *3.2.1 Description of the Tankwa goat*

Tankwa goats are regarded as a distinct landrace with its origin in the Tankwa Karoo National Park (-32°24'10.21" S, 20°09'55.14" E) in the Northern Cape Province, South Africa. They have survived in a challenging environment with no human intervention for more than 80 years. The environment of the Tankwa Karoo National Park is classified as Succulent Karoo. Average park temperatures over the last 10 years ranged from 4°C in winter to 28°C in summer with an average rainfall of 137 mm per year (World Weather Online).

In 2009, 41 Tankwa goats (24 males and 17 females) were removed from the Park and relocated to the Carnarvon Research Station (-30°95'99.65" S, 22°15'49.44" E), which has a hotter and dryer climate (7°C in winter to 34°Cin summer). The property spans 1000 hectares and do not have running water available throughout the year. The goats have to survive on the available natural resources with no supplementary food or water. The Carnarvon area is classified as Karoo (Acocks, 1975; Figure 3.1).


Figure 3.1 Current location of Tankwa goats in the Tankwa Karoo National Park and Carnarvon, in the Northern Cape Province (source: http://www.southern-africa.arroukatchee.fr/south-africa/map/northern-cape-province.png)

Over the past 10 years, the average yearly rainfall in Carnarvon was 78mm (World Weather Online). Figure 3.2 shows the typical vegetation in Carnarvon during the summer period (December to March).

Figure 3.2 Typical vegetation during the summer in Carnarvon (Anna Kropff, February 2021)

Animals are kept without management intervention and are allowed to breed randomly. The current population size is maintained at close to 300 individuals, with approximately 30 males and 70 females as the core breeding stock. Surplus animals are culled on a yearly basis to maintain the average breeding stock at 100 individuals.

### *3.2.2 Samples selected for study*

Biological samples, that included blood and hair, were collected from the Tankwa goats between 2012 and 2019. In 2012, most animals from both the Tankwa Karoo National Park and the Carnarvon Research Station were sampled. Thereafter all the new offspring born at the Carnarvon Research Station were sampled yearly in either February or August. Blood (10 ml) was collected in Ethylenediaminetetraacetic acid (EDTA) tubes by qualified veterinary staff of the Northern Cape DAERL. All samples were stored at the SANBI Biobank under optimal conditions until analyses.

A subset of 360 Tankwa goat samples representative of all founders for which samples were available (40 individuals) and a total of 40 (20 males and 20 females) offspring per year (2012 to 2019) were selected at random from the total sample database of 1200 collected samples. Samples were collected at random as no information on the relatedness between the individuals were available.

The University of Pretoria provided the genotypes of 48 Angora and 40 dairy goats collected as part of previous studies (Table 3.1). These goats were included for comparison purposes, due to the high level of artificial selection within the respective breeds, focussing on milk and fibre traits instead of adaptive traits.

Table 3.1 Summary of breeds, number of individuals and source of genomic data included in the current study

| Breed / Production Type | Number of Individuals | Provided By |
|---|---|---|
| Tankwa Goats | 360 | DAERL |
| Angora | 48 | UP |
| dairy goats | 40 | UP |

*3.2.3 DNA Extraction and Quantification*

DNA extractions from the Tankwa biological samples were performed at the Centre for Conservation Science (CfCS) Laboratories of the South African National Biodiversity Institute (SANBI) using the Zymo Research Quick-DNA™ Miniprep Kit (Zymo Research - Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa, www.inqababiotec.co.za) following the manufacturer's protocol.

Final elution volumes were 50ul. All samples were quantified using the Qubit® 2.0 fluorometer (Life Technologies [Pty] Ltd, Carlsbad, CA, USA, www.lifetechnologies.com) before SNP genotyping. Samples had DNA concentrations between 40 and 50 ng/ul and 260/280 scores between 1.8 and 2.2.

*3.2.4 SNP Genotyping*

All samples were genotyped using the Illumina GoatSNP50 BeadChip (Tosser-Klopp *et al.*, 2014) at the Biotechnology Platform of the Agricultural Research Council (ARC). The Illumina GoatSNP50 BeadChip was developed using whole-genome sequence data from six different goat breeds with a total of 53 347 SNP's covering 93% of the genome (Tosser-Klopp *et al.*, 2014).

**3.3 Methods**

Separate genomic datasets for the 360 Tankwa, 48 Angora and 40 dairy were prepared and later merged for comparative analyses. Original genomic data was in .MAP and .PED format, and converted into binary files (.BIM, .BED and .FAM) for analysis in PLINK (Purcell *et al.*, 2007) using the *--make-bed* command. Although the focus was on generating Tankwa-specific results, some analyses were also performed for the Angora and dairy datasets for comparative purposes.

### *3.3.1 Quality Control*

Quality control was applied to all three datasets. All non-autosomal SNP's were removed from the datasets. Sample and marker based quality control was performed using PLINK (Purcell *et al.*, 2007). Sample quality control was completed based on a minimum individual call rate of 95%, while marker-based quality control was based on a minimum SNP call rate of 95%, minimum deviation from Hardy-Weinberg equilibrium of p=0.001 and a minimum minor allele frequency (MAF) of 0.02 (Table 3.2).

Table 3.2 Parameters and thresholds used for sample and marker-based quality control as well as the command used in PLINK

| Quality Control parameter | Threshold | PLINK Command |
|---|---|---|
| Removal of non-autosomal SNP's | | --chr |
| Individual call rate | <95% | *--mind* |
| SNP call rate | <95% | *--geno* |
| Hardy-Weinberg equilibrium | >0.001 | *--hwe* |
| Minor allele frequency | <0.02 | *--maf* |

### *3.3.2 Diversity Statistics*

Within population diversity levels were calculated in PLINK (Purcell *et al.*, 2007) for all three datasets. Statistical analyses included average observed ($H_o$) and expected ($H_e$) heterozygosity, minor allele frequency (MAF), runs of homozygosity (ROH) and linkage disequilibrium (LD), using the $r^2$ measurement.

$H_o$ and $H_e$ was calculated using the *--het* function in PLINK (Purcell *et al.*, 2007) which calculates $H_o$ and $H_e$ per individual using the following formula:

$$Ho = (N(NM)-0(Hom))/N(NM)$$

Where N(NM) refers to the number of non-missing genotypes; and O(Hom) refers to the number of observed homozygotes in the population.

LD was calculated using the $r^2$ measurement instead of D' as D' is sensitive to allele frequency and is affected by small population sizes while the $r^2$ measurement is independent of allele frequency (Gurgul *et al.*, 2014). This measurement is a useful tool for determining the degree of genetic diversity within the population (Khanyile *et al.*, 2015; Makina *et al.*, 2015b; Visser *et al.*, 2016; Saravanan *et al.*, 2022). The commands used in PLINK (Purcell *et al.*, 2007) were

listed in Table 3.3. MAF and LD was calculated per chromosome and as an average across the genome, per population.

Table 3.3 PLINK (Purcell *et al.*, 2007) commands used to calculate the diversity statistics

| Parameter | Command |
|---|---|
| Expected ($H_e$) and Observed Heterozygosity ($H_o$) | *--het* |
| Minor Allele Frequency (MAF) | *--freq* |
| Linkage Disequilibrium (LD) | *--r2* |

PLINK (Purcell *et al.*, 2007) uses a sliding window approach to identify stretches of consecutive homozygous SNP's, called runs of homozygosity (ROH). ROH was detected using the parameters listed in Table 3.4. The ROH detected was grouped into different lengths that gives an indication of the time of inbreeding. The classes used were 0.1-2MB; 2.1-4MB; 4.1-8MB; 8.1-16MB and >16MB.

Table 3.4 Parameters and commands using PLINK (Purcell *et al.*, 2007) to detect Runs of Homozygosity (ROH)

| Parameter | Value used | Command |
|---|---|---|
| Minimum number of SNP's | 50 | *--homozyg-snp* |
| Max number of heterozygotes allowed | 1 | *--homozyg-window-het* |
| Max number of missing SNP's allowed | 3 | *--homozyg-window-missing* |
| Max allowed distance between SNP's | 1000 kb | *--homozyg-gap* |
| Sliding window size | 1000 | *--homozyg-window-kb* |
| Window threshold to call a ROH | 0.05 | *--homozyg-window-threshold* |
| Allelic Matching | 0.98 | *--homozyg-match* |

These statistics were calculated for each dataset to be able to compare the diversity levels between the feral and commercial populations.

### *3.3.3 Inbreeding*

Inbreeding levels were calculated using both the individual inbreeding coefficient $F_{IS}$ and $F_{ROH}$ in PLINK (Purcell *et al.*, 2007) for all three datasets.

$F_{IS}$ was calculated using the *--het* function in PLINK (Purcell *et al.*, 2007) which calculates $F_{IS}$ per individual using the following formula:

$$F_{IS} = (H_E - H_O) / H_E$$

where $H_E$ is the expected Heterozygosity and $H_O$ is the observed Heterozygosity

$F_{ROH}$ was calculated for each ROH size class, as well as overall using the formula:

$$F_{ROH} = \sum (L_{ROH} / L_{AUTO})$$

Where $L_{ROH}$ is the total length of ROH in the animal and $L_{AUTO}$ is the total length of the autosome covered by SNP's (Purfield *et al.*, 2012; Peripolli *et al.*, 2016).

### *3.3.4 Effective Population Size – Tankwa Dataset*

Effective population size was estimated using SNeP (Barbato *et al.*, 2015) for the Tankwa population. SNeP estimates the effective population size using genome-wide linkage disequilibrium. The PLINK input file was converted to .MAP and .PED files after quality control (using the *–convert* command) to be used by SNeP. Results of the effective population size over time was visualised and plotted in Microsoft Excel.

### *3.3.5 Principal Component Analysis & Population Structure*

SNP-based relatedness for the Tankwa population was estimated by calculating a genetic relatedness matrix and then estimating eigenvalues and eigenvectors using GCTA version 1.24 (Yang *et al.*, 2011). The PLINK binary files were used as input files to construct the relationship matrix (command *--make-grm*). This relationship matrix was used to calculate the eigenvectors and eigenvalues for the first five principal components (command *--pca 5*). The .eigenvec file was used to plot the PCAs in RStudio (RStudio Team., 2020).

ADMIXTURE version 1.23 (Alexander *et al.*, 2009) was used to determine population structure for the Tankwa population. ADMIXTURE was used to calculate the cross-validation estimates for each K-value from K=1 to K=15 (command *--cv*). The K-value with the smallest CV error and the least variation was chosen as optimal. Structure bar plots were generated using Genesis version 0.2.3 (Buchmann & Hazelhurst, 2014).

### *3.3.6 Population structure and differentiation on the merged dataset*

A sub-sample of 48 Tankwa goats was selected based on year of birth (24 from the founder population and 24 from 2019) of which 50% of each group was male and 50% female, and the results of a pi-Hat relatedness analysis calculated in PLINK (individuals with the lowest pi-Hat

were considered – below 0.5) (Purcell *et al.*, 2007). These parameters were applied to ensure the sub-sample used for further analysis was representative of the population.

The selected sub-sample was then combined with the dairy and Angora populations' data to create a merged dataset for further analysis.

Quality control as explained in 3.3.2.1 was performed on the merged dataset, with the only difference being that no pruning was performed based on MAF values. Principal Component analysis (detail in 3.3.2.5) was performed using the first 5 principal components and population structure was assessed using Admixture (Alexander *et al.*, 2009) with K-values from 3 to 7.

### *3.3.7 Signatures of Selection*

A fixation index ($F_{ST}$) approach was used to identify possible signatures of selection. This approach investigates differences in allele frequencies between populations. $F_{ST}$ values range from 0 to 1, with high $F_{ST}$-values indicating strong selection and low $F_{ST}$-values indicating little or no selection (Purfield *et al.*, 2017). High $F_{ST}$ values indicate selection as the different selection histories between the populations change the allele frequencies in each population (Makina *et al.*, 2015a)

$F_{ST}$ values were calculated per SNP using PLINK (Purcell *et al.*, 2007) by comparing the allele frequencies in the subset of the Tankwa population with the allele frequencies in the combined dataset, as described by Weir & Cockerham (1984), using the *--fst* command. The *--report-variants* command was used to obtain the $F_{ST}$ -values per pair-wise comparison, and the *--base* command was used to specify the Tankwa population as the population of interest (the other populations were compared to the Tankwa population and not to each other).

All negative $F_{ST}$ -values were changed to zero as negative values have no biological meaning (Makina *et al.*, 2015a; Zhao *et al.*, 2015). To reduce the effect of a small sample size a moving average $F_{ST}$ (*ma*$F_{ST}$) approach was used by calculating the average $F_{ST}$ for five adjacent SNP's in a sliding window approach, this ensured only strong signals of selection were identified and reduced background noise (Purfield *et al.*, 2012). The top 0.1% of the $F_{ST}$ values were considered significant, with the two flanking SNP to each side of the significant SNP also considered. The results were visualized on a Manhattan plot using the package ggplot in R-studio (RStudio Team., 2020).

### 3.3.7.1 Gene Annotation

Regions within the genome that were of interest based on the $F_{ST}$ analysis were retrieved from the Ensembl and NCBI genome viewer database (Zerbino *et al.*, 2018). The ARS1 goat reference genome (Bickhart *et al.*, 2017) was used to identify genes with known function which may be associated with signatures of selection; only SNP's associated with genes with known function were further considered. The biological and metabolic functions of the identified genes was investigated to identify pathways that evolved for survival in the harsh climate where the Tankwa goat occurs.

# Chapter 4: Results

## 4.1 Introduction

The genotypic data was generated using the Illumina GoatSNP50 BeadChip (Tosser-Klopp *et al.*, 2014) at the ARC Biotechnology Platform (Veterinary Research Institute, Onderstepoort, South Africa). The main population under investigation consisted of 360 feral Tankwa goats, which were assessed in terms of their population parameters (genetic diversity, effective population size and inbreeding). For comparison purposes, two subsets of highly selected goat populations, namely 48 Angora goats and 40 dairy goats were included. After the calculation of population summary statistics per population, the Tankwa genotypes and additional datasets were merged to investigate population structure and population relatedness. The combined dataset was finally used to study possible signatures of selection.

## 4.2 Diversity parameters per population

### *4.2.1 Quality Control*

The genotypes of all animals were subjected to marker- and sample-based quality control to remove individuals and SNP's with a high level of missing data and to remove SNP's that deviated from Hardy-Weinberg equilibrium (HWE) or had a very low Minor Allele Frequency (MAF).

First, all the non-autosomal SNP's were removed from the datasets. Sample-based quality removed 36 animals (35 from the Tankwa dataset and 1 from the dairy dataset). After this step 325 Tankwa goats, 48 Angora goats and 39 dairy goats remained for further analysis. Marker-based quality control was also completed with most SNP's removed based on MAF for all three datasets. The Tankwa dataset had the most SNP's removed, with 7703 SNP's removed compared to the 3705 SNP's removed from the Angora dataset and the 1720 SNP's removed from the dairy dataset. The detail of the numbers of SNP's removed and remaining per criteria is reported in Table 4.1. The call rates of the various datasets were all above 99.8 after quality control.

Table 4.1 Marker- and sample- based quality control for the three datasets

| Dataset | QC criteria | Original number | Removed | Remaining |
|---|---|---|---|---|
| Tankwa | Individual call rate | 360 Individuals | 35 Individuals | 325 Individuals |
| | SNP call rate | 49 941 SNP's | 1 044 SNP's | 48 897 SNP's |
| | HWE (p-value) | 48 897 SNP's | 978 SNPs' | 47 919 SNP's |
| | MAF | 47 919 SNP's | 5 681 SNP's | 42 238 SNP's |
| Angora | Individual call rate | 48 Individuals | 0 Individuals | 48 Individuals |
| | SNP call rate | 49 941 SNP's | 586 SNP's | 49 355 SNP's |
| | HWE (p-value) | 49 355 SNP's | 323 SNP's | 49 032 SNP's |
| | MAF | 48 897 SNP's | 2 796 SNP's | 46 236 SNP's |
| dairy | Individual call rate | 40 Individuals | 1 Individual | 39 Individuals |
| | SNP call rate | 49 941 SNP's | 679 SNP's | 49 262 SNP's |
| | HWE (p-value) | 49 262 SNP's | 374 SNP's | 48 888 SNP's |
| | MAF | 48 897 SNP's | 667 SNP's | 48 221 SNP's |

The number of SNP's per chromosome that passed quality control ranged from 750 (CHI25 for the Tankwa goats) to 3164 (CHI1 for the dairy goats). On all chromosomes, the dairy goats had the highest number of SNP's remaining, while the Tankwa goats had the least.

### *4.2.2 Population Summary Statistics*

#### 4.2.2.1 Heterozygosity, Minor Allele Frequency and Linkage Disequilibrium

A summary of the average diversity values for the three populations is shown in Table 4.2. The average observed ($H_O$) ranged from 0.349 (Angora) to 0.388 (dairy) and the range for expected ($H_E$) heterozygosity was 0.353 (Angora) to 0.406 (dairy).

Table 4.2 A summary of the number of SNP's that passed quality control (N SNP), average Heterozygosity ($H_E$ and $H_O$), Minor Allele Frequency (MAF) and Linkage Disequilibrium ($r^2$) for the Tankwa, Angora and dairy goat populations

| Breed | N SNP | $H_E$ | $H_O$ | MAF | $r^2$ |
|---|---|---|---|---|---|
| Tankwa | 42 238 | 0.368 | 0.367 | 0.249 | 0.469 |
| Angora | 46 236 | 0.353 | 0.349 | 0.253 | 0.392 |
| dairy | 48 221 | 0.406 | 0.388 | 0.315 | 0.332 |

The average MAF, prior to the removal of all SNP's with a MAF below 0.02, ranged from 0.205 for the Tankwa goat to 0.315 for the dairy goats. After pruning based on MAF, the range was

between 0.249 (Tankwa) and 0.315 (Angora). The dairy population has the highest average MAF across all chromosomes. The average MAF per population is illustrated per chromosome in Figure 4.1.

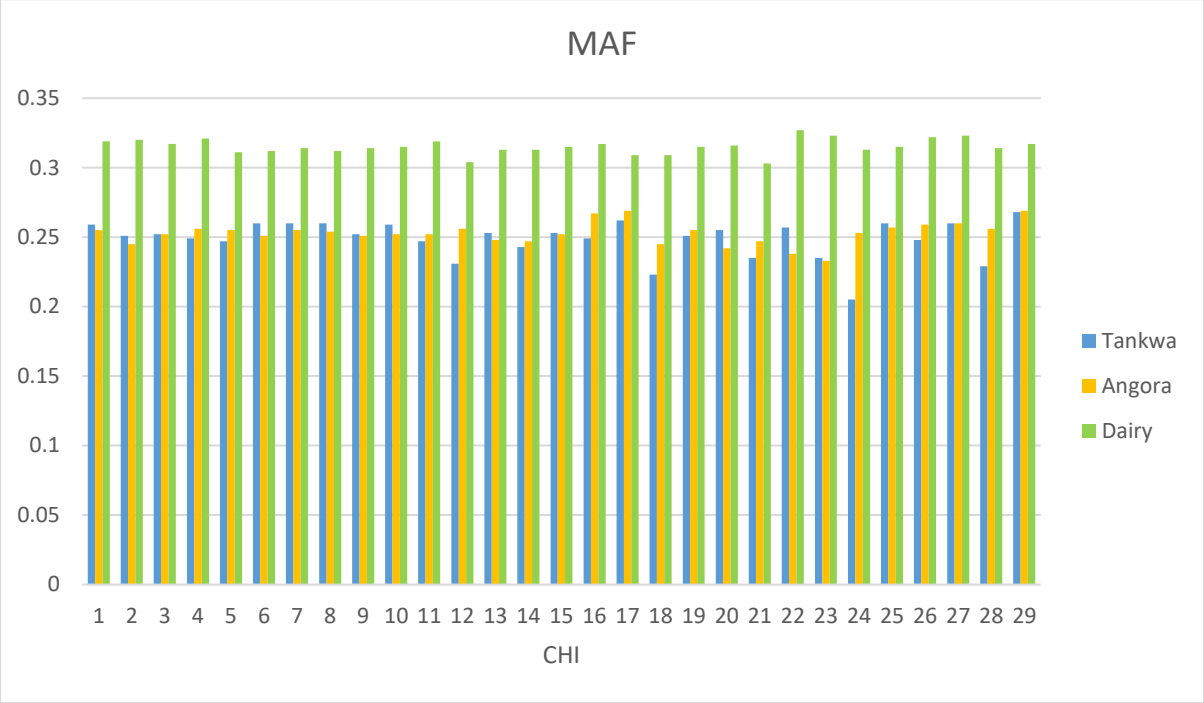


Figure 4.1 MAF per chromosome for the three different populations (Tankwa, Angora and dairy)

Chromosome-specific linkage disequilibrium (LD) estimates using $r^2$ ranged from 0.316 on CHI9 for the dairy goats to 0.525 on CHI 24 for the Tankwa goats. The average LD across the genome ranged from a $r^2$ value of 0.332 for the Angora goats to a $r^2$ value of 0.469 for the Tankwa goats. The number of SNP's amplified, $H_O$, $H_E$, MAF and $r^2$ for all three populations are reported per chromosome and across the genome in Addendum C. The average linkage disequilibrium per chromosome for the three populations is illustrated in Figure 4.2.

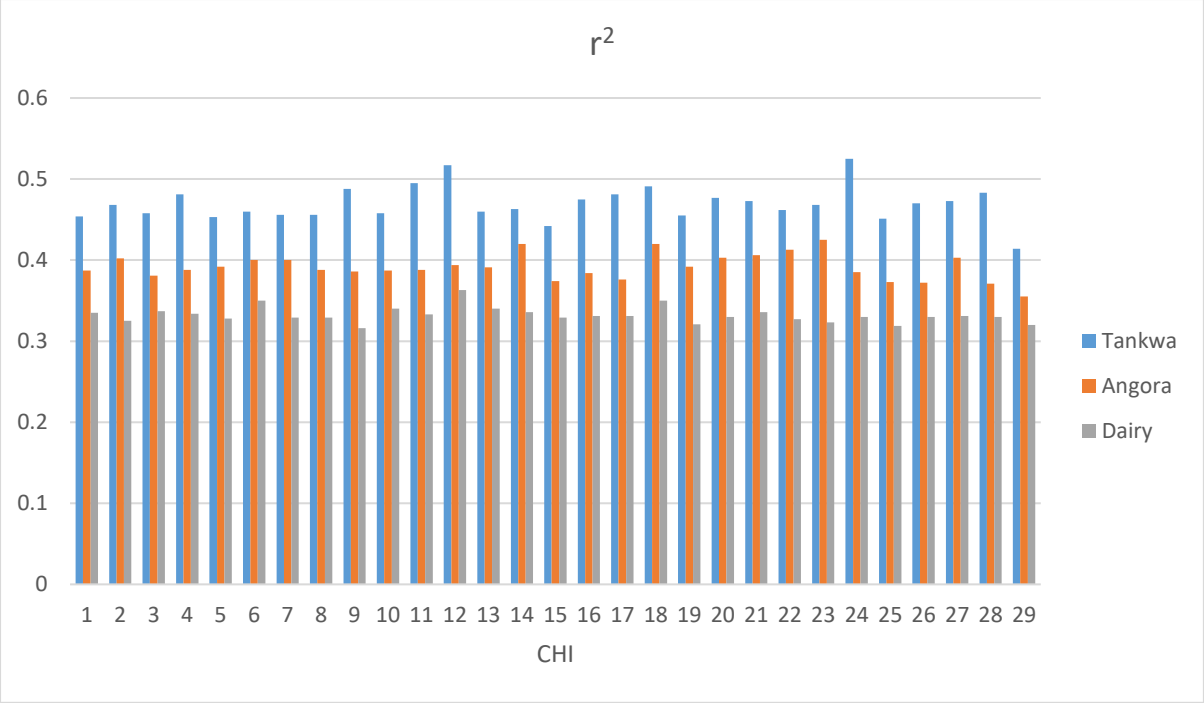


Figure 4.2 Linkage disequilibrium per chromosome for the different populations (Tankwa, Angora and dairy)

### 4.2.2.2 Runs of Homozygosity

Analysis of homozygous fragments identified 1225 runs of homozygosity (ROH) in the Tankwa goats, 1090 ROH in the Angora goats and 784 ROH in the dairy goats. The ROH were classified into groups of 0.1-2 MB, 2.1-4 MB, 4.1-8 MB, 8.1-16MB and >16MB (Figure 4.3). The largest proportion of the ROHs were between 2.1 and 4MB long in all the populations. Fewer ROH in the 8.1-16MB and >16MB categories (40) were observed for the Tankwa goat population, compared to the Angora (173) and dairy goats (198). The shortest ROH of 1.51MB was identified in the Tankwa goat population on CHI6, while the longest ROH of 62.85MB was observed in the dairy goats on CHI20.

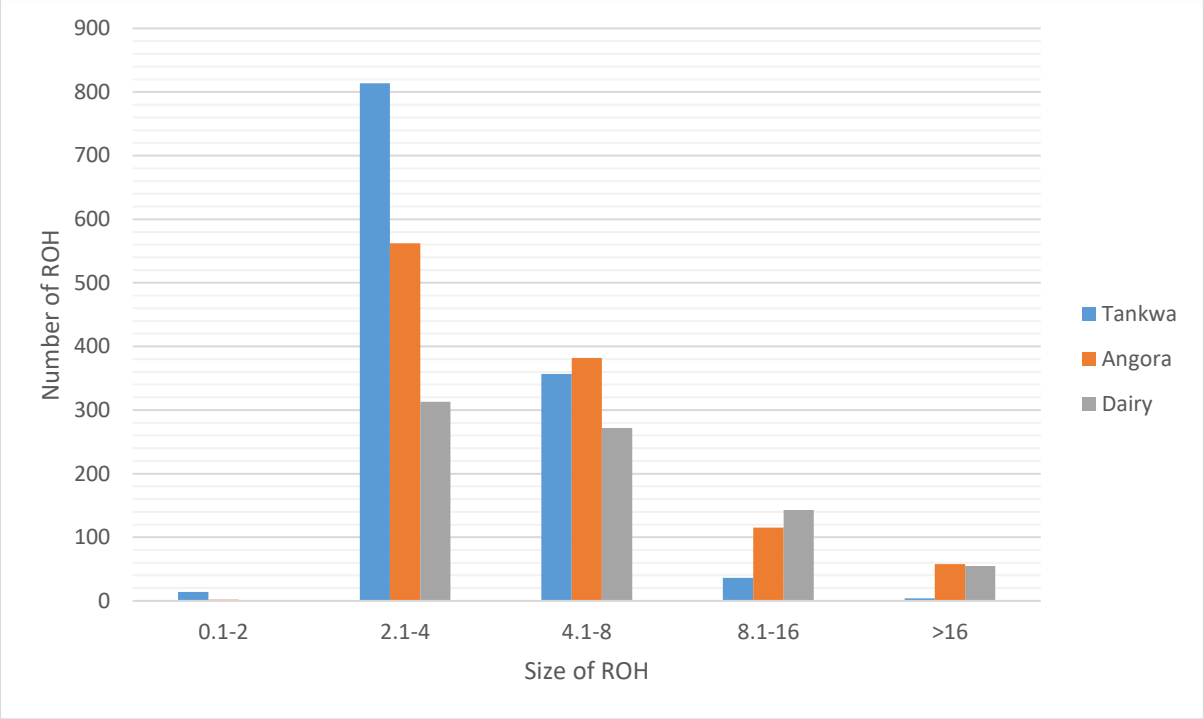


Figure 4.3 The overall length (Mb) and frequency of the identified Runs of Homozygosity measured in three goat populations

The total number of ROHs (irrespective of length) detected on each chromosome per population was calculated and can be seen in Figure 4.4. The Tankwa goat population had significantly more ROH than the other two populations on CHI 6, 7, 8, 12, 13, 16, 25 and 29.

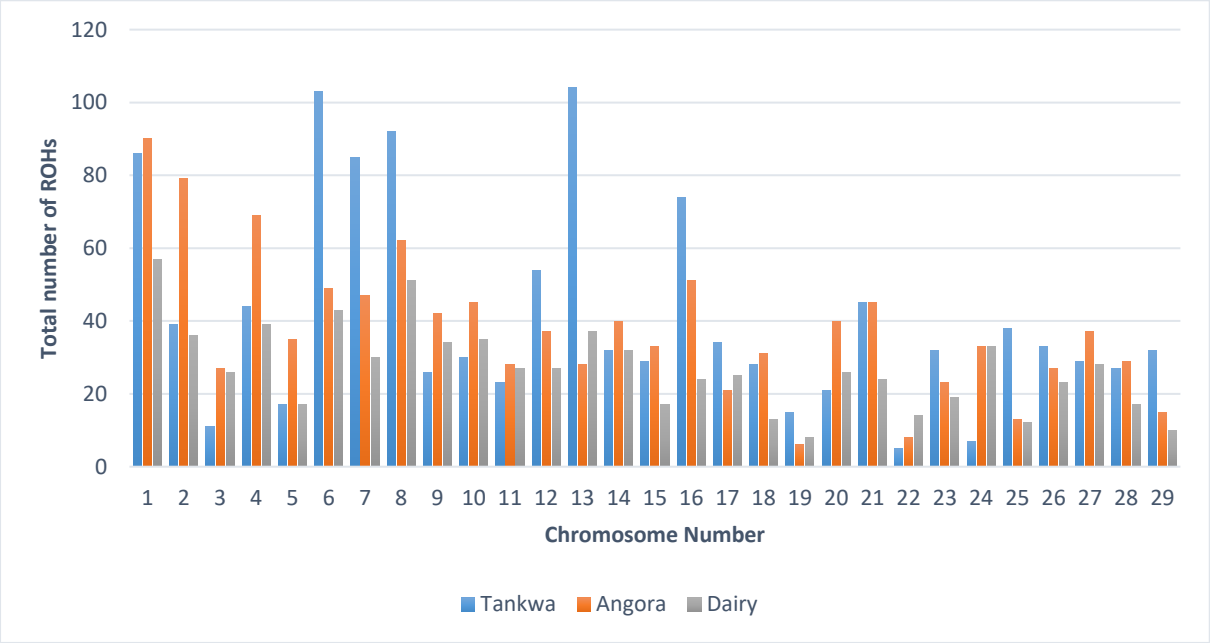


Figure 4.4 The total number of detected ROH per Chromosome

The ROH per size category on each chromosome can be seen in Figure 4.5 for the Tankwa goats and in Addendum D for the other datasets. The highest and lowest number of ROH for the Tankwa were observed on CHI13 and CHI 22 respectively, while the most and least ROH

were identified on CHI1 and CHI19 for both the Angora and dairy goats. For the Tankwa population the shortest ROH (at 1.8Mb) was found on CHI 6 and the longest (at 19.8 Mb) was on CHI 29.

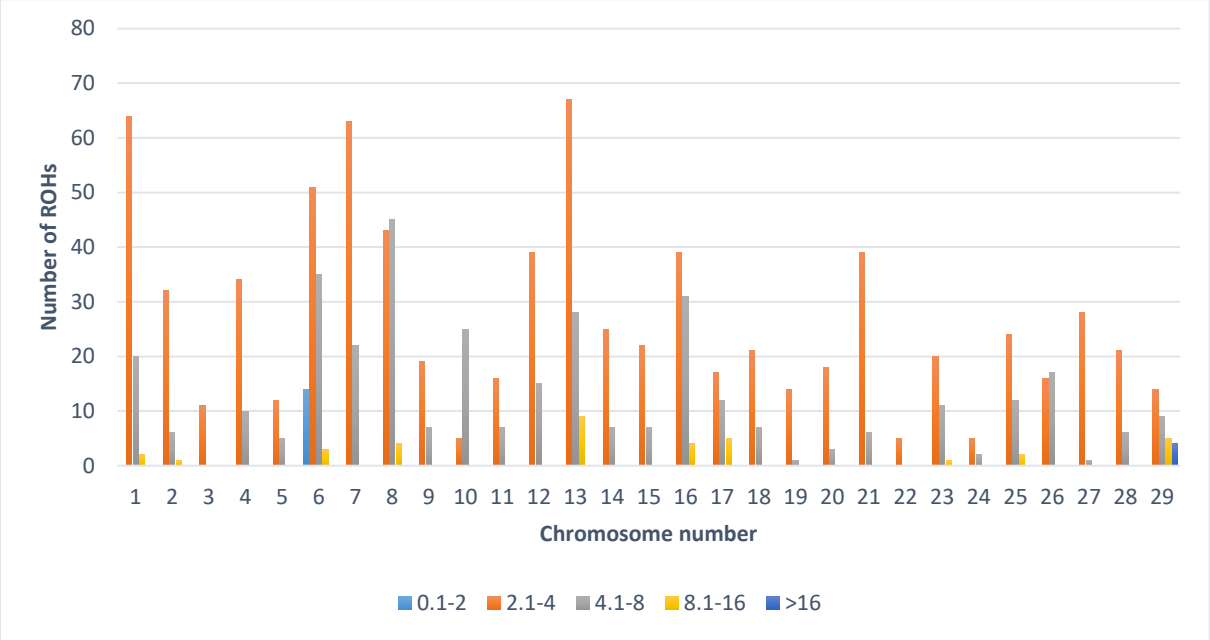


Figure 4.5 The number of detected ROH of each size category per Chromosome for the Tankwa goats

4.2.2.3 Inbreeding parameters

The individual inbreeding coefficients ($F_{IS}$ and $F_{ROH}$) were calculated per individual and as an average across each population (Table 4.2). The average inbreeding coefficients ($F_{IS}$ and $F_{ROH}$) for the population was lowest in the Tankwa goats and highest in the dairy goats. The lowest individual $F_{IS}$ estimate was observed in the Angora population (-0.108) and the highest in the dairy goat population (0.259).

Table 4.3 The average inbreeding ($F_{IS}$ and $F_{ROH}$) per population

| Population | Average $F_{IS}$ | Lowest $F_{IS}$ | Highest $F_{IS}$ | Average $F_{ROH}$ | Lowest $F_{ROH}$ | Highest $F_{ROH}$ |
|---|---|---|---|---|---|---|
| Tankwa | 0.001 | -0.107 | 0.259 | 0.006 | 0.000 | 0.022 |
| Angora | 0.011 | -0.108 | 0.209 | 0.052 | 0.023 | 0.067 |
| dairy | 0.044 | -0.062 | 0.259 | 0.062 | 0.007 | 0.149 |

4.2.2.4 Effective population size

The historic effective population size ($N_e$) of the Tankwa goats was the smallest of the three populations 950 generations ago, followed by the Angora population and then the dairy goats. The dairy goats however had the most significant decrease in $N_e$ over the last 900 years, decreasing by 2000 individuals to a size of 100 individuals, approximately 12 generations ago. The Tankwa goat numbers decreased from 1400 individuals to 60 animals, over the same time period. The historic effective population sizes are shown in Figure 4.6.

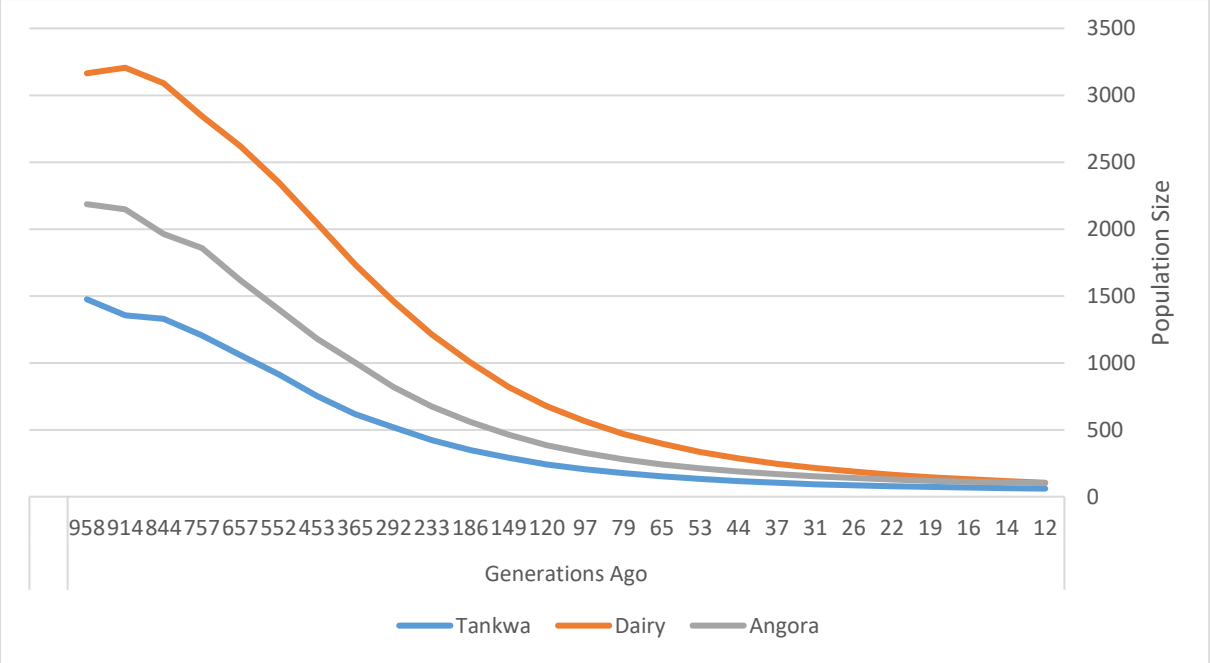


Figure 4.6 Historical effective population size of three goat populations

*4.2.3 Genetic relatedness between individuals and population structure*

To study the genetic relatedness between the individuals in the Tankwa population a principal component analysis (PCA) was performed on the complete sample set of 325 individuals. The first two principal components (PC1 and PC2) were plotted against each other, the resulting plot is shown in Figure 4.7. Principal component 1 accounted for 3.52% of the variation seen in the population, while principal component 2 accounted for 3.05%.

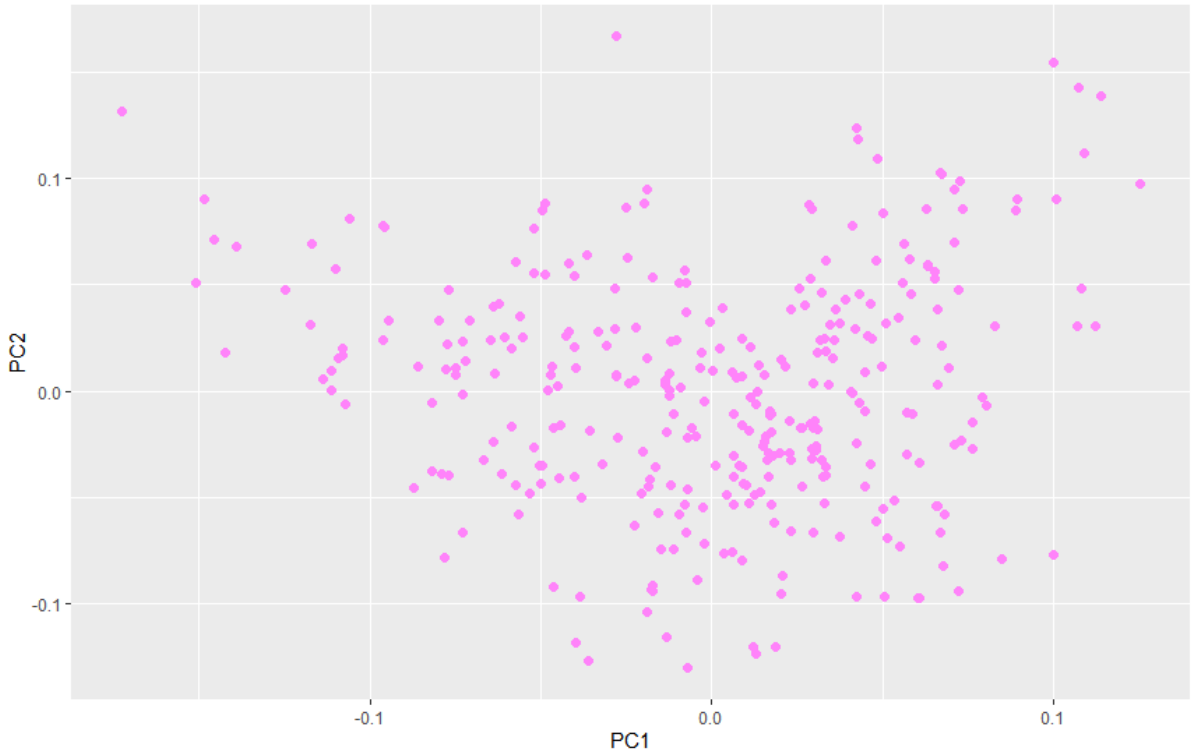


Figure 4.7 Principal component analysis comparing the first two principal components (PC1 and PC2) for the Tankwa population

The results seen in Figure 4.7 show no clearly observable clusters or stratification, but rather a population with high diversity. The admixture analysis supported the PCA with K=3 shown in Figure 4.8.

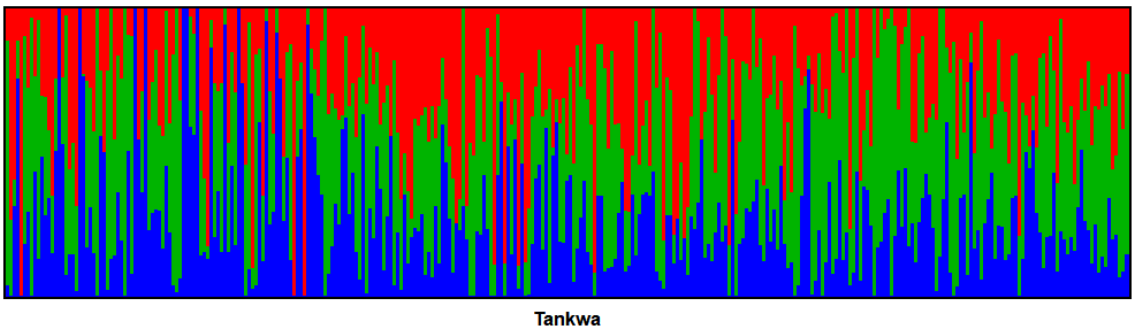


Figure 4.8 Population structure plot showing the ancestral identity for K=3

## 4.3 Merged Dataset

### *4.3.1 Quality Control*

A subset of 48 Tankwa goats was selected (based on year of birth and a pi-Hat value of below 0.5) for comparison with 48 Angora and 40 dairy goats. Marker-based quality control removed 10 922 SNP's, and 39 017 SNP's were used for downstream analysis (Table 4.3). The final dataset had an average call rate of 99.8%

*4.3.2 Genetic Relatedness between individuals*

Principal component analysis (PCA) performed on the merged sample set of 135 individuals. The first two principal components (PC1 and PC2) were plotted against each other, the resulting plot is shown in Figure 4.9. PC1 and PC3 was also plotted with the results shown in Figure 4.10. Principal component 1 accounted for 14.26% of the variation between the populations, principal component 2 accounted for 11.62% and principal component 3 for 3.130%.

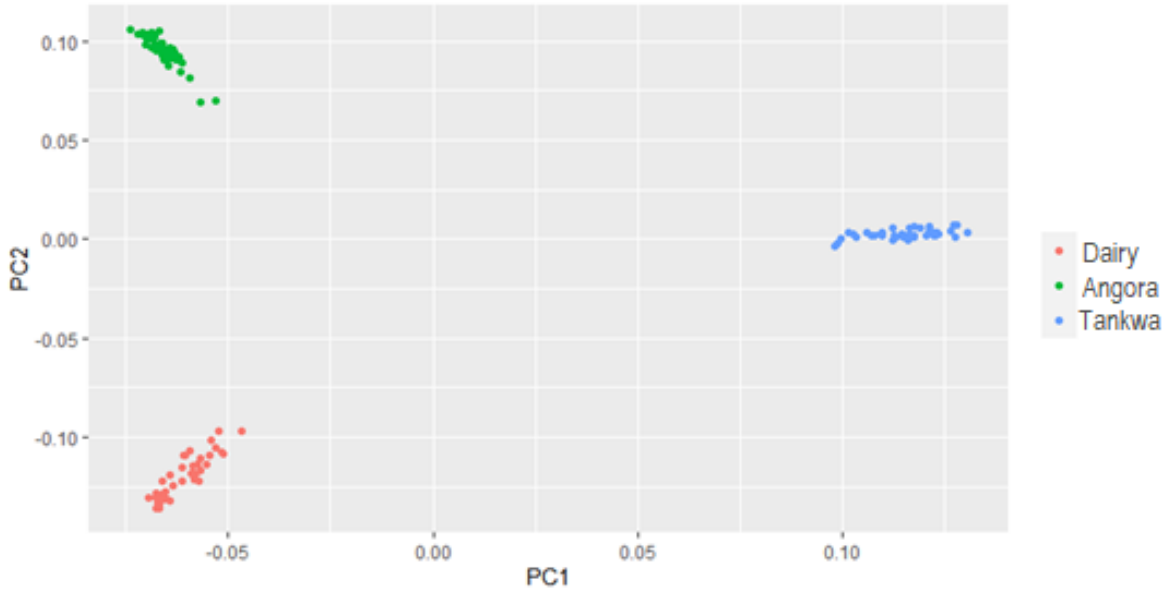


Figure 4.9 Principal component analysis comparing the first two principal components (PC1 and PC2) for the merged dataset of Tankwa, dairy and Angora goat populations

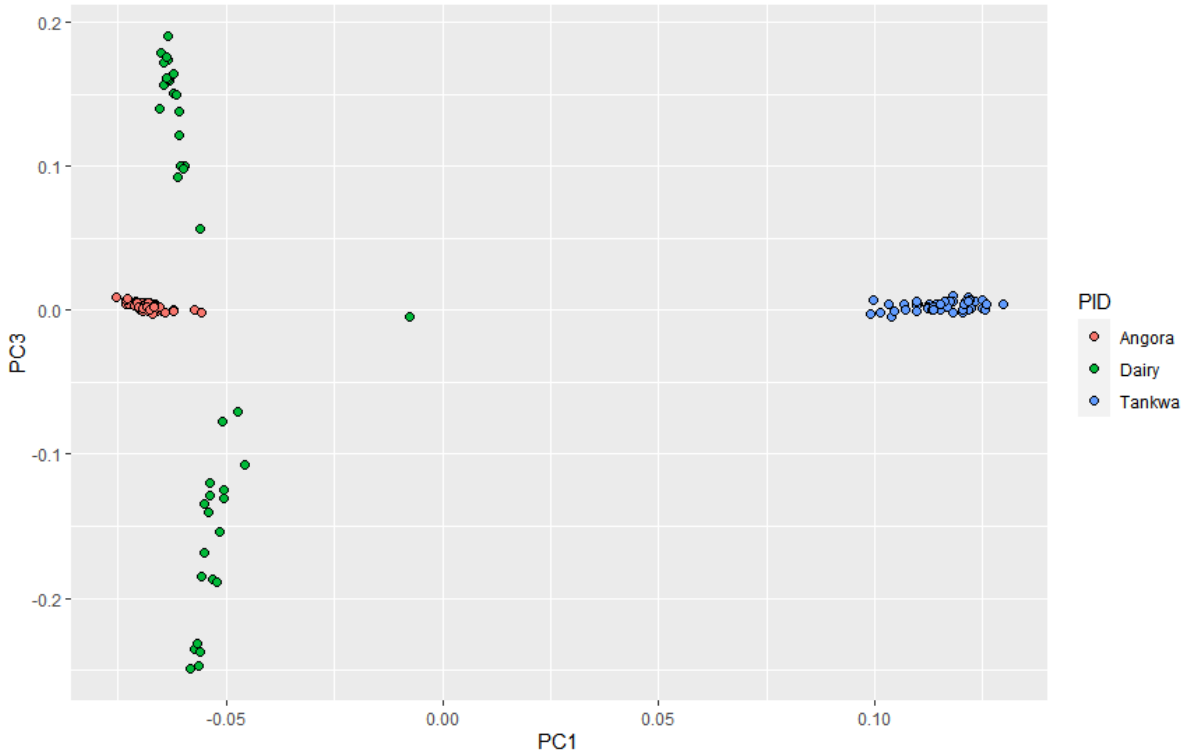


Figure 4.10 Principal component analysis comparing the PC1 and PC3 for the merged dataset of Tankwa, Diary and Angora goat populations

The results indicated in Figure 4.9 show three clearly separated clusters. Each production type formed a distinct cluster when the first two principal components were plotted. However, the dairy goat cluster separated into two loose clusters (one cluster containing Saanen goats and the other containing a mixture of British Alpine and Toggenburg goats) when PC1 and PC3 was plotted. The Tankwa goat cluster remained tightly grouped and separated from the other two clusters.

### *4.3.3 Population Structure*

ADMIXTURE version 1.23 (Alexander *et al.*, 2009) analysis was used to investigate the population structure between the different production types. The cross-validation scores for K=4-7 were very similar, ranging from 0.577 to 0.584. These cross-validation scores with one standard deviation were plotted in Figure 4.11.

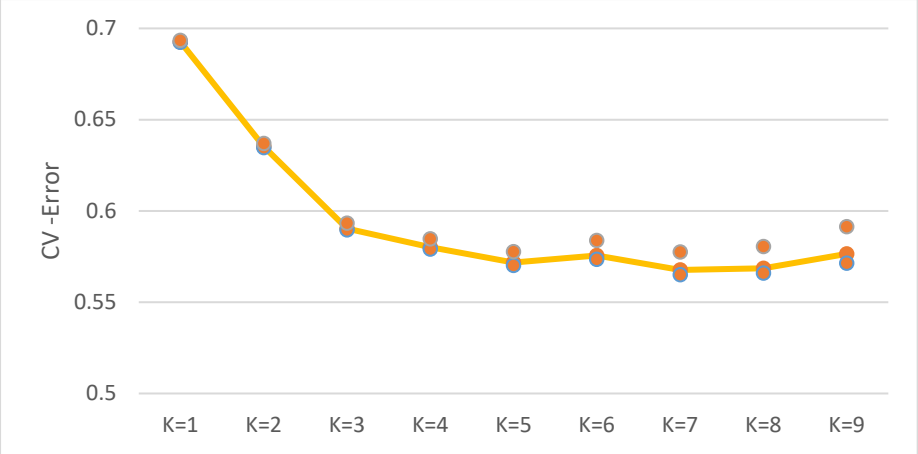


Figure 4.11 Cross-validation errors for the merged dataset of Tankwa, Angora and dairy goat populations

Based on the CV errors, population structure plots were created for K=3 to K=7 (Addendum E) but only the plot for K=7 (the K value with the lowest CV-error) is shown here (Figure 4.12).

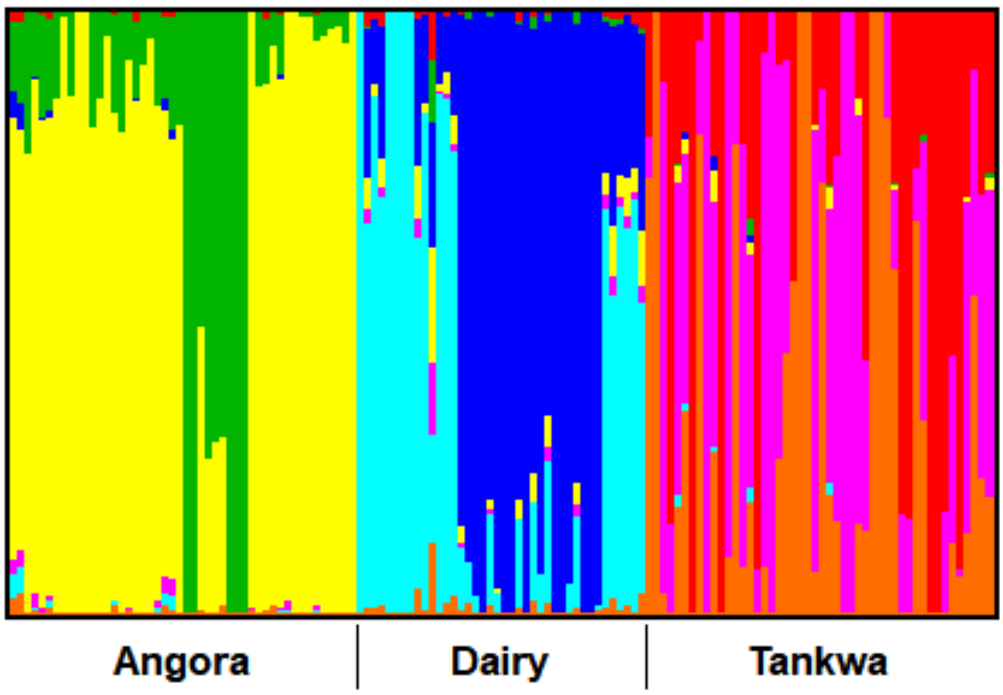


Figure 4.12 Population structure plot for K=7 of three goat populations

The population structure plot shows distinct separation between the three different production types and support the PCA results. The number of ancestral populations in each production type was not clearly resolved using the CV-errors, however the three production types remained separate on a genomic level. This is in agreement with the results obtained during PCA analysis with the samples clusters based on production type. The dairy population was the first to show sub-structure at K=4, followed by the Tankwa and Angora populations at K=5 (Addendum E).

*4.3.4 Signatures of selection*

The fixation index ($F_{ST}$) approach was used to identify specific regions of the genome that contained signatures of selection in this dataset. The pairwise $F_{ST}$ values calculated between the Tankwa population and the other two populations are shown per chromosome in Addendum F.

To reduce the effect of a small sample size a moving average $F_{ST}$ (*ma*$F_{ST}$) approach was used, The number of SNP's with *ma*$F_{ST}$-values in the top 0.1% (above 0.81) were considered possibly significant and are listed in Table 4.4.

Table 4.4 Number of significant SNP's identified using *ma*$F_{ST}$ per chromosome

| CHI | N Significant SNP's |
|---|---|
| 1 | 1 |
| 2 | 6 |
| 3 | 1 |
| 6 | 1 |
| 8 | 1 |
| 10 | 4 |
| 11 | 1 |
| 12 | 4 |
| 14 | 1 |
| 16 | 1 |
| 18 | 9 |
| 20 | 1 |
| 21 | 3 |
| 22 | 3 |
| 23 | 5 |
| 24 | 7 |
| 28 | 1 |

Chromosome 18 contained the most significant SNP's, namely nine. A Manhattan plot was created using the *ma*$F_{ST}$-values to indicate which SNP's had a significant *ma*$F_{ST}$ (the top 0.1% of the *ma*$F_{ST}$ values) (Figure 4.13).

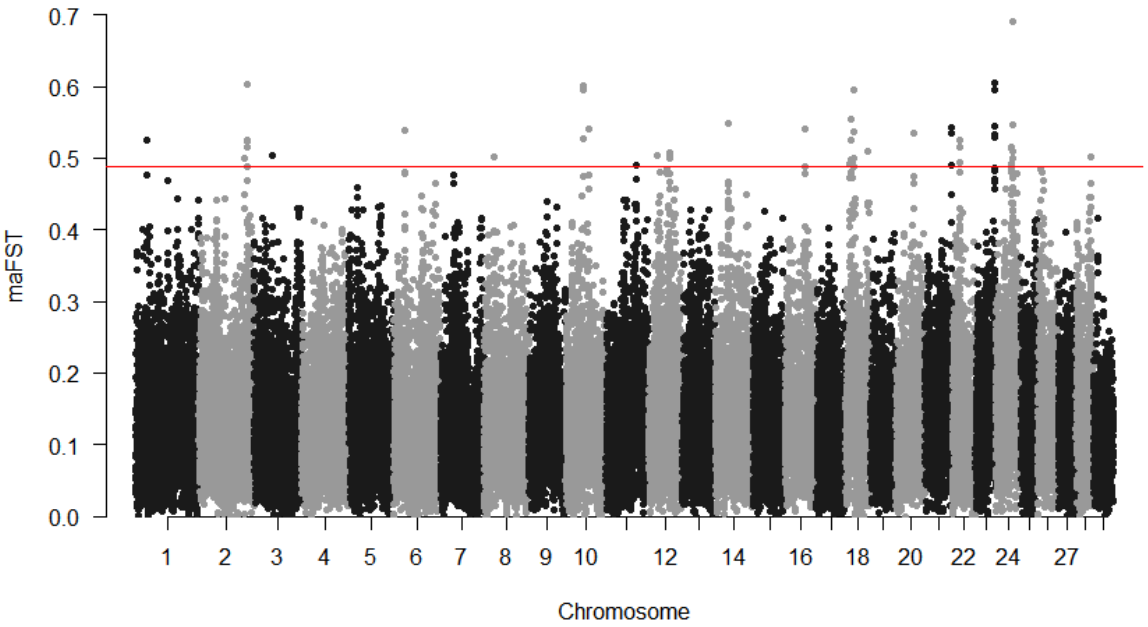


Figure 4.13 Manhattan plot of the *ma*$F_{ST}$-values plotted per chromosome they are associated with

The Ensembl database was used to further investigate the identified significant SNP's as well as the SNP's flanking them. All genes found in the area of the significant SNP were considered possible genes under selection and was then recorded in Addendum G if information was available.

The main functions of these genes were investigated using the Ensembl database and then classified by what part of the body it affects. Table 4.5 list the genes identified that will be discussed further while Addendum H lists all the identified genes and their functions. No genes were found on the database for 18 of the 50 identified SNP's. 49 Possible genes were identified for the remaining SNP's.

Table 4.5 A sub-set of the identified genes and their functions by chromosome number (Ensembl; UniProt; NCBI).

| CHI | Gene | Molecular Function | Biological Process |
|---|---|---|---|
| 6 | MTTP | • Ceramide 1-phosphate transfer activity<br>• Cholesterol transfer activity<br>• Lipid binding<br>• Phosphatidylcholine transfer activity<br>• Phosphatidylethanolamine transfer activity | • Cholesterol homeostasis<br>• Circadian rhythm<br>• Lipoprotein transport<br>• Low-density lipoprotein particle remodelling<br>• Plasma lipoprotein particle assembly<br>• Protein lipidation |

| | | | |
|---|---|---|---|
| | | • Protein heterodimerization activity<br>• Triglyceride transfer activity | • Protein secretion<br>• Triglyceride metabolic process<br>• Triglyceride transport |
| 11 | DRC1 | • None | • Axonemal dynein complex assembly<br>• Cilium-dependent cell motility<br>• Determination of left/right symmetry<br>• Heart development |
| 12 | GJA3 | • Gap junction hemi-channel activity | • Cell communication<br>• Gap junction-mediated intercellular transport<br>• Visual perception |
| 12 | GJB2 | • Calcium ion binding<br>• Gap junction channel activity involved in cell communication by electrical coupling<br>• Identical protein binding | • Cell-cell signalling<br>• Gap junction assembly<br>• Gap junction-mediated intercellular transport<br>• Sensory perception of sound |
| 12 | GJB6 | • Actin filament binding<br>• Beta-tubulin binding<br>• Gap junction channel activity involved in cell communication by electrical coupling<br>• Microtubule binding | • Ear morphogenesis<br>• Gap junction assembly<br>• Gap junction-mediated intercellular transport<br>• Sensory perception of sound |
| 18 | JPH3 | • None | • Exploration behaviour<br>• Learning<br>• Locomotion<br>• Memory<br>• Neuromuscular process controlling balance<br>• Regulation of neuronal synaptic plasticity<br>• Regulation of ryanodine-sensitive calcium-release channel activity |

| | | | |
|---|---|---|---|
| 18 | OPA3 | • None | • Visual perception |
| 22 | GRM7 | • Adenylate cyclase inhibitor activity<br>• Group III metabotropic glutamate receptor activity<br>• Protein dimerization activity<br>• Serine binding | • Axon development<br>• Chemical synaptic transmission<br>• Glycosylation<br>• Sensory perception of sound |
| 24 | ZBTB14 | • DNA-binding transcription repressor activity, RNA polymerase II-specific<br>• RNA polymerase II cis-regulatory region sequence-specific DNA binding | • Cardiac septum development<br>• Coronary vasculature development<br>• Heart valve development<br>• Kidney development |
| 24 | PTPRM | • Cadherin binding<br>• Identical protein binding<br>• Transmembrane receptor protein tyrosine phosphatase activity | • Homophilic cell adhesion via plasma membrane adhesion molecules<br>• Negative regulation of angiogenesis<br>• Negative regulation of endothelial cell migration<br>• Negative regulation of endothelial cell proliferation<br>• Response to xenobiotic stimulus<br>• Retina layer formation<br>• Retinal ganglion cell axon guidance<br>• Signal transduction |

*Ensembl (https://www.ensembl.org/index.html, accessed 11 May 2022; Zerbino *et al.*, 2018); UniProt (https://www.uniprot.org/, accessed 11 May 2022); NCBI Genome data viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/, accessed 11 May 2022)

# Chapter 5: Discussion and Conclusion

## 5.1 Introduction

The Tankwa goat is a feral goat found in the Northern Cape, South Africa. The goats show high adaptability to the harsh climate and lack of water found in the environment where they occur. This goat has been shown to be a unique genetic resource that needs to be conserved. To contribute to the description of these goats the need to genetically characterise the landrace breed was identified. This study therefore aimed to genetically characterise the Tankwa goat, and to identify signatures of selection related to adaptation by using genome wide SNP data generated with the Illumina GoatSNP50 BeadChip.

## 5.2 Genetic characterisation of the Tankwa goat population

The first quality control step was sample based quality control, individuals with low DNA quality and quantity should be removed as samples with low call rates will affect the marker call rates. The average call rates of above 99.8% obtained for the datasets in this study was comparable to other studies using the 50K goat SNP Array. Lashmar *et al.* (2016) obtained average call rates of 99.5% and 99.6% or the dairy and Angora goats respectively, while Monau *et al.* (2018) obtained a call rate of 99.6% for Tswana goats. During the validation phase of the 50K goat SNP chip a call rate of 99.9% was obtained for 10 different breeds (Tosser-Klopp *et al.*, 2014). During quality control the largest number of SNP's were filtered out based on minor allele frequency (MAF) with approximately double the number removed for the Tankwa goats (5681) when compared to the Angora goats (2796); and about eight times the amount compared to the dairy goats (667), due to ascertainment bias. The MAF for the Tankwa goats (0.249) was comparable to other goat breeds. A study on Ugandan goats by Onzima *et al.* (2018) reported MAF ranging from 0.257 to 0.280 for six breeds and in the study by Mdladla *et al.* (2016) the MAF for the Tankwa goats was estimated at 0.24. This fluctuation in number of SNP's with low MAF between breeds can be due to ascertainment bias as the breeds in the dairy dataset were used for the development of the SNP chip. Ascertainment bias is defined as a deviation of the population genetic statistics from the theoretical true value due to non-random selection of individuals (Malomane *et al.*, 2018). In this study ascertainment bias refers to the probability that a SNP included on the commercial chip will also be polymorph in the population for the breeds that were not used in the development of the SNP chip.

The Tankwa goats were genetically characterized by determining the level of linkage disequilibrium (LD), average minor allele frequency (MAF), heterozygosity, effective population size ($N_e$) and inbreeding.

Heterozygosity is an important diversity parameter in population genetics. The Tankwa goats showed an observed heterozygosity (0.367) comparable to other goat breeds (both indigenous and commercial). The dairy and Angora goats in this study showed an observed heterozygosity of 0.388 and 0.349 respectively. Other studies on Angoras in South Africa found heterozygosity values of 0.371, 0.365 and 0.333 (Visser *et al.*, 2011a; Lashmar *et al.*, 2015; Paim *et al.*, 2019), while another study on Saanen from Switzerland found heterozygosity values of 0.386 (Burren *et al.*, 2016).

Only one study determining genetic variation has previously been performed on the Tankwa goats with a small sample size of 20 individuals. This study reported observed and expected heterozygosity levels of 0.35 and 0.33 respectively for the Tankwa population (Mdladla *et al.*, 2016). These values are marginally lower than the results obtained in this study ($H_E$ =0.368 and $H_O$=0.367). This variation is most probably due to the small sample size in the 2016 study. The observed and expected heterozygosity were almost the same in this study and is comparable to other goat breeds. This, as well as the comparable MAF values, shows that sufficient genetic diversity is present in the Tankwa goat population for the continued survival of the breed.

Linkage disequilibrium ($r^2$) is a consequence of various genetic factors, including selection, genetic drift, mutations and non-random mating (Eusebi *et al.*, 2019). The Tankwa goats showed a higher average linkage disequilibrium (0.469) than the other two populations. Mdladla *et al.* (2016) also reported that the Tankwa goats had higher $r^2$ than the veld type goats, this is most probably due to the founder effect and population history of the Tankwa goats. The Tankwa goat is a relatively recently formed breed (about 90 years ago) with a small population that originated from a small number of founders (71).

### 5.3 Estimates of inbreeding

Inbreeding was calculated using both $F_{IS}$ and $F_{ROH}$. For both measures, the Tankwa goats had lower values compared to the other two populations. The average $F_{ROH}$ values were higher than the $F_{IS}$ values, which can be ascribed to the differences between the two measurements. The $F_{IS}$ value can be negative for an individual or population while the $F_{ROH}$ value is always positive leading to a higher average (Onzima *et al.*, 2018b).

Tankwa goats are not intensively managed and are not subjected to any selective breeding, however inbreeding was a concern in the Tanwka goats due to the recent founder effect when the breed was formed, the isolation and the small population size and random mating.

The Tankwa goats had an $F_{IS}$ value of 0.001 and an $F_{ROH}$ value of 0.006, which are lower than the other production types included in this study. Various commercial breeds were studied with $F_{IS}$ values ranging from -0.05 to 0.23 (Nicoloso *et al.*, 2015; Visser *et al.*, 2016; Paim *et al.*, 2019). The study by Paim *et al.*, (2019) included indigenous breeds with $F_{IS}$ values ranging from 0.05 to 0.125. This is consistently lower than the commercial breeds in the same study and higher than the Tankwa goats. Cardoso *et al.*, (2018) studied 25 indigenous breeds and used $F_{ROH}$ as an estimator of inbreeding. For these 25 breeds the $F_{ROH}$ values ranged from 0.02 to 0.66, with most breeds (16) having $F_{ROH}$ values below 0.2. Mdladla *et al.* (2016) also calculated inbreeding for the Tankwa goats and reported a much higher $F_{IS}$ value (0.15). This could probably be attributed to the small samples size (20). Another possible reason could be the unintentional inclusion of closely related animals in the study as no relatedness information was available for selected individuals. In the current study a much larger sample size was used (325).

The Tankwa goat population show low estimates of inbreeding when compared to other indigenous goat populations. This can be attributed to how the population is managed being free roaming with no selective breeding. This can also be due to the history of the formation of the population. It is suspected that the original population was formed by individuals from multiple other breeds that escaped and formed a cohesive herd in the Tankwa National Park about 90 years ago (personal communication with Deon Kriel - DKriel@ncpg.gov.za).

### 5.4 Runs of Homozygosity

Runs of Homozygosity (ROH) can be used as a measure of inbreeding, with the length of the ROH indicating when the inbreeding occurred (Cardoso *et al.*, 2018). The Tankwa goats had more short ROH (<4MB) and very few long ROH (>16MB). This indicates more ancient inbreeding, probably due to the founder effect when the population was formed. This also shows very low recent inbreeding, as was also observed in the inbreeding analysis. This result indicates that the management of the current population is adequate as recent inbreeding is kept low.

The Angora and dairy populations had more longer ROHs (>16) than the Tankwa goats indicating more recent inbreeding in these populations as supported by the inbreeding levels

reported in section 5.1.3. This can probably be explained by the direct artificial selection and emphasis on a few male with high genetic merit in the management of commercial breeds.

### 5.5 Effective Population Size

The historic effective population size ($N_e$) of all three populations decreased over time and 900 generations ago. The dairy population had the highest ancestral population size, due to the different breeds found in this population (Saanen, Toggenburg and British Alpine), while the Tankwa goat had the lowest, due to the more recent formation of the population. The Tankwa goats also have the lowest effective population size at 13 generations ago, most likely due to the small initial population size. This is to be expected as the trend in most domesticated animals is a large historic reduction in effective population size with a more stable smaller size in recent generations (Makina *et al.*, 2015b). This is due to the domestication process and selection for specific traits or production types (Boichard *et al.*, 2015; Brito *et al.*, 2015; Purfield *et al.*, 2017). As long as the management practices used maintain genetic diversity in the population, no further major decline in effective population size is expected (Brito *et al.*, 2015). Effective population size can also affect measures of LD ($r^2$) which could explain the relatively high levels of LD for the Tankwa goats compared to the other goat populations, as a lower $N_e$ leads to high LD levels (Brito *et al.*, 2015).

### 5.6 Genetic relatedness and population structure of the Tankwa goats

The results reported for the Principal Component Analysis (PCA) was supported by the results obtained in the structure analysis. When analysing only the Tankwa goat population, no population substructure was observed as the individuals formed one relatively dispersed cluster. This also indicates that no specific lines with more common genotypes have been formed between generations. The Tankwa goat dataset contained individuals spanning several generations. Each principal component only accounted for a small percentage of the variation (PC1 = 3.517%), which is similar to the study by Monau *et al.* (2018) on various populations of Tswana goats throughout Botswana.

Comparing the three populations to each other, PCA analysis showed clear separation of the three populations when contrasting PC1 and PC2. PC1 accounted for 14.3 % of the variation which is similar to other studies with multiple breeds where PC1 contributed between 5 and 19% of the variation (Brito *et al.*, 2015; Nicoloso *et al.*, 2015; Burren *et al.*, 2016; Mdladla *et al.*, 2016; Onzima *et al.*, 2018c; Paim *et al.*, 2019).The population structure analysis for the three populations supported the PCA results.

The genetic relatedness and population structure results were expected as the three populations analyzed in this study were not closely related and differed in production type. It will be valuable to study the present indigenous populations of goats that occur near the Tanwka Karoo National Park and compare those to the Tanwka goat population using genetic relatedness and structure analysis.

### 5.7 Signatures of Selection and Gene Annotation

Signatures of selection are regions of the genome that have changes in the frequency of alleles due to natural or artificial selection (Bertolini *et al.*, 2018). These changes in the genome can be used to study the genes associated with traits under selection. In this study forty-nine genes under possible selection were identified. Of these, eight genes had no known function with 35 of the genes being part of intracellular transport, signalling pathways, organelle structuring and cellular organization. These genes could affect a variety of processes and would need to be studied further to identify specifically association with environmental adaptation.

The five remaining genes were associated with vision, hearing, development and learning. Of the possible genes identified *GJB2*, *GJB6* and *GJA3* on CHI 12 were the only ones previously associated with adaptation to heat and drought (Kim *et al.*, 2016; Onzima *et al.*, 2018b; Sejian *et al.*, 2019). Genes *GJB2* and *GJB6* are involved in ear morphogenesis and the sensory perception of sound, while *GJA3* is involved in visual perception. These genes were previously identified as genes involved in adaptation to heat stress (Sejian *et al.*, 2019). In addition to GJA3, the gene *OPA3* is also involved in visual perception and was not previously associated with adaptation to heat and drought. The genes associated with vision and hearing could lead to better foraging, however further studies would be needed to support this. The genes associated with vision and hearing could also be under selection for survival against predators such as jackal and caracal as these goats are free roaming and unprotected in the wild. The *DRC1* gene is associated with development functions of the heart and the determination of left/right symmetry. This gene could probably assist in maintaining the relatively constant heart rate and cardiac output which assist goats to survive and perform better than sheep during heat stress (Sejian *et al.*, 2019). The gene *JPN3* is associated with learning, movement, memory and exploratory behaviour. These aspects are all essential to survival and thus of importance for the Tankwa goat.

### 5.8 Population Management Recommendations

This study showed that the Tankwa goat population is currently not at risk of losing genetic diversity with low inbreeding. The genetic characterization of the landrace breed now contributes to the understanding of maintaining key parameters for survival. There are however considerations needed for the optimal genetic management of the population. The FAO (2015) recommends four main methods for the conservation of at-risk breeds: increase population size, manage overall genetic diversity, select for increased productivity and cryopreserve genetic material such as sperm and embryos. These recommendations are already implemented for the conservation of the Tankwa goat population. In addition, the uniqueness of the Tankwa goat was previously established by Kotzé *et al.* (2014).

It is known that small populations are vulnerable to the random effects of variation in terms of climate, birth rates, sex ratios, etc. (Meek *et al.*, 2015). It will be necessary to keep the core herd pure, however cross-breeding these hardy goats with commercial breeds could increase their production capacity and by association their monetary value. This could be important to ensure the continued support for the conservation of this breed. Another risk for small populations or breeds is having all the individuals in one area as disease or disaster could potentially eradicate the whole population (FAO, 2011). The recommendation is to prevent this by placing populations in other areas in the country. This however was not successful with the Tankwa goat as those placed in areas with higher rainfall died due to high parasite loads and disease (Personal communication with Deon Kriel - DKriel@ncpg.gov.za). The drier climate is best suited for this landrace breed. However the original herd is still available in the Tankwa Karoo National Park if genetic rescue is ever needed.

The second recommendation mentioned by the FAO (2015) is the management of genetic diversity. This is already in place and being maintained, every generation that is genotyped on a low-density SNP chip. These genotypes are used to determine heterozygosity and inbreeding in every generation. This is an important management tool as the reduction in overall genetic diversity is a high risk in small populations (Taberlet *et al.*, 2011). Monitoring is needed to ensure that any management decisions do not adversely affect the population (Allendorf, 2017). The final recommendation for the conservation of breeds is the cryopreservation of sperm and embryos. This is in place and ongoing with both sperm analyses done and more than 50 embryos being preserved.

Conservation efforts can also be badly impacted by lack of information about the animals and the environment they live in. This study aims to address some of the information shortages for the Tankwa goats by genetically characterising them, however more research is definitely needed. Studies to answer some of the other question regarding these goats, such as

vegetation present and the browsing habits of the goats, will also assist in the conservation of this breed.

### 5.9 Conclusion

This study was the first comprehensive study on the genetic characterization of the Tankwa goat. In this study, the landrace breed was characterised using the 50K SNP chip with resultant normal levels of heterozygosity and inbreeding determined compared to other breeds. This study showed that the Tankwa goat population is currently not at risk and should be genetically maintained as is.

The Tankwa goats was compared to other production breeds such as the Angora and dairy types and with the PCA and admixture analysis showing clear differentiation, confirming that this is indeed a unique genetic resource.

Possible signatures of selection for environmental adaptation was identified, however further studies would be required to confirm the roles of those genes. Further studies such as GWAS analysis for specific traits (eg. Coat colour) may be useful for determining which genes are truly responsible for the Tankwa goat adaptability to the harsh climate where they easily survive.

# References

Acocks, J.P.H., 1975. Veld types of South Africa, Member of Botswana Survey. South Africa, 40.

Ahmed, S., Kropff, A. S., Jonker, T., & Kotze, A. 2018. Can we source the origin of the feral Tankwa goat of South Africa ? J. Phylogenetics Evoltionary Biol. 6, 2 https://doi.org/10.4172/2329-9002.100020.

Ai, H., Huang, L., & Ren, J. 2013. Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers. PLoS One 8 https://doi.org/10.1371/journal.pone.0056001.

Ai, H., Yang, B., Li, J., Xie, X., Chen, H., & Ren, J. 2014. Population history and genomic signatures for high-altitude adaptation in tibetan pigs. BMC Genomics 15 https://doi.org/10.1186/1471-2164-15-834.

Ajmone-marsan, P., Colli, L., Han, J. L., Achilli, A., Lancioni, H., Joost, S., Lenstra, J. A., Italian Goat Consortium, Econogene Consortium, & Globaldiv Consortium. 2014. The characterization of goat genetic diversity : Towards a genomic approach. Small Rumin. Res. 121, 58–72 https://doi.org/10.1016/j.smallrumres.2014.06.010.

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

Aldezabal, A., & Garin, I. 2000. Browsing preference of feral goats (Capra hircus L.) in a Mediterranean mountain scrubland. J. Arid Environ. 44, 133–142 https://doi.org/10.1006/jare.1999.0573.

Aleena, J., Sejian, V., Bagath, M., Krishnan, G., Beena, V., & Bhatta, R. 2018. Resilience of three indigenous goat breeds to heat stress based on phenotypic traits and PBMC HSP70 expression. Int. J. Biometeorol. 62, 1995–2005 https://doi.org/10.1007/s00484-018-1604-5.

Alexander, D. H., Novembre, J., & Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 https://doi.org/10.1101/gr.094052.109.vidual.

Allendorf, F. W. 2017. Genetics and the conservation of natural populations: allozymes to genomes. Mol. Ecol. 26, 420–430 https://doi.org/10.1111/mec.13948.

Amills, M., Capote, J., & Tosser-Klopp, G. 2017. Goat domestication and breeding : a jigsaw of historical , biological and molecular data with missing pieces. Animla Genet., 631–644 https://doi.org/10.1111/age.12598.

Amills, M., Joost, S., Sonstegard, T. S., Servin, B., Crepaldi, P., Talenti, A., Van Tassell, C. P., Colli, L., Tosser-Klopp, G., Rochat, E., Milanesi, M.,

Guldbrandtsen, B., Stella, A., Palhière, I., Rosen, B. D., Oget, C., Catillo, G., Rothschild, M. F., Bertolini, F., Crisà, A., Marras, G., Kim, E. S., Nicolazzi, E., & Steri, R. 2018. Signatures of selection and environmental adaptation across the goat genome post-domestication. Genet. Sel. Evol. 50, 1–24 https://doi.org/10.1186/s12711-018-0421-y.

Andersson, L. 2013. Molecular consequences of animal breeding. Curr. Opin. Genet. Dev. 23, 295–301 https://doi.org/10.1016/j.gde.2013.02.014.

Avila, F., Mickelson, J. R., Schaefer, R. J., & Mccue, M. E. 2018. Genome-Wide Signatures of Selection Reveal Genes Associated With Performance in American Quarter Horse Subpopulations. Front. Genet. 9 https://doi.org/10.3389/fgene.2018.00249.

Badenhorst, S. 2006. Goats (Capra hircus), the Khoekhoen and Pastoralism: Current evidence from Southern Africa. African Archaeol. Rev. 23, 45–53 https://doi.org/10.1007/s10437-006-9007-0.

Baleseng, L., Kgosikoma, O. E., Makgekgenene, A., Coleman, M., Morley, P., Baker, D., & Bahta, S. 2016. Performance of goats and sheep under communal grazing in Botswana using various growth measures.Pages 320–326 in ICAS VII Seventh International Conference on Agricultural Statistics.

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

Benjelloun, B., Alberto, F. J., Streeter, I., Boyer, F., Coissac, E., Stucki, S., BenBati, M., Ibnelbachyr, M., Chentouf, M., Bechchari, A., Leempoel, K., Alberti, A., Engelen, S., Chikhi, A., Clarke, L., Flicek, P., Joost, S., Taberlet, P., Pompanon, F., & NextGen Consortium. 2015. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats ( Capra hircus ) using WGS data. Front. Genet. 6 https://doi.org/10.3389/fgene.2015.00107.

Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., Palhière, I., Crisà, A., Catillo, G., Steri, R., Amills, M., Colli, L., Marras, G., Milanesi, M., Nicolazzi, E., Rosen, B. D., Tassell, C. P. Van, Joost, S., & Crepaldi, P. 2018. Signatures of selection and environmental adaptation across the goat genome post - domestication. Genet. Sel. Evol. 50, 1–24 https://doi.org/10.1186/s12711-018-0421-y.

Bhati, M., Kadri, N. K., Crysnanto, D., & Pausch, H. 2020. Assessing genomic diversity and signatures of selection in Original Braunvieh cattle using whole-genome sequencing data. BMC Genet. 21 https://doi.org/10.1186/s12864-020-6446-y.

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., Lee, J., Lam, E. T., Liachko, I., Sullivan, S. T., Burton, J. N., Huson, H. J., Nystrom, J. C., Kelley, C. M., Hutchison, J. L., Zhou, Y., Sun, J., Crisà, A., León, F. A. P. De, Schwartz, J. C., Hammond, J. A., Waldbieser, G. C., Schroeder, S. G., Liu, G. E., Dunham, M. J., Shendure, J., Sonstegard, T. S., Phillippy, A. M., Tassell, C. P. Van, & Smith, T. P. L. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat. Genet. 49, 643–650 https://doi.org/10.1038/ng.3802.

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., & Gandini, G. 2015. Challenges and opportunities in genetic improvement of local livestock breeds. Front. Genet. 6, 1–7 https://doi.org/10.3389/fgene.2015.00033.

Boichard, D., Ducrocq, V., & Fritz, S. 2015. Sustainable dairy cattle selection in the genomic era. J. Anim. Breed. Genet. 132, 135–143 https://doi.org/10.1111/jbg.12150.

Bosman, L., van Marle-Köster, E., & Visser, C. 2015. Genetic diversity of South African dairy goats for genetic management and improvement. Small Rumin. Res. 123, 224–231 https://doi.org/10.1016/j.smallrumres.2014.12.003.

Brito, L. F., Jafarikia, M., Grossi, D. A., Kijas, J. W., Porto-Neto, L. R., Ventura, R. V., Salgorzaei, M., & Schenkel, F. S. 2015. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. BMC Genet. 16, 1–15 https://doi.org/10.1186/s12863-015-0220-1.

Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L. R., Cánovas, A., Feng, Z., Jafarikia, M., & Schenkel, F. S. 2017. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC Genomics 18, 1–20 https://doi.org/10.1186/s12864-017-3610-0.

Buchmann, R., & Hazelhurst, S. 2014. The ‘ Genesis ’ Manual. , 0–20.

Burren, A., Neuditschko, M., Signer-Hasler, H., Frischknecht, M., Reber, I., Menzi, F., Drögemüller, C., & Flury, C. 2016. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. Anim. Genet. 47, 727–739 https://doi.org/10.1111/age.12476.

Cardoso, T. F., Amills, M., Bertolini, F., Rothschild, M., Marras, G., Boink, G., Jordana, J., Capote, J., Carolan, S., Hallsson, J. H., Kantanen, J., Pons, A., & Lenstra, J. A. 2018. Patterns of homozygosity in insular and continental goat breeds. Genet. Sel. Evol. 50, 1–11 https://doi.org/10.1186/s12711-018-0425-7.

Chenyambuga, S. W., Hanotte, O., Hirbo, J., Watts, P. C., Kemp, S. J., Kifaro, G. C., Gwakisa, P. S., Petersen, P. H., & Rege, J. E. O. 2004. Genetic Characterization of Indigenous Goats of Sub-saharan Africa Using Microsatellite DNA Markers. Asian-Australasian J. Anim. Sci. 17, 445–452 https://doi.org/10.5713/ajas.2004.445.

Chokoe, T. C., Hadebe, K., Muchadeyi, F. C., Nephawe, K. A., Dzomba, E. F., Mphahlele, T. D., Matelele, T. C., & Mtileni, B. J. 2022. Conservation status and historical relatedness of South African communal indigenous goat populations using a genome-wide single-nucleotide polymorphism marker T. Front. Genet. 13 https://doi.org/10.3389/fgene.2022.909472.

Chokoe, T. C., Mdladla-hadebe, K., Muchadeyi, F., Dzomba, E., Matelele, T., Mpofu, T. J., Nephawe, K., & Mtileni, B. 2020. Genetic Diversity of South African Indigenous Goat Population from Four Provinces Using Genome-Wide SNP Data. Sustainability 12 https://doi.org/10.3390/su122410361.

Chynoweth, M. W., Litton, C. M., Lepczyk, C. A., Hess, S. C., & Cordell, S. 2013. Biology and Impacts of Pacific Island Invasive Species. 9. Capra hircus , the Feral Goat (Mammalia: Bovidae) . Pacific Sci. 67, 141–156

https://doi.org/10.2984/67.2.1.

Colli, L., Milanesi, M., Talenti, A., Bertolini, F., Chen, M., Crisà, A., Daly, K. G., Corvo, M. Del, Guldbrandtsen, B., Lenstra, J. A., Rosen, B. D., Vajana, E., Catillo, G., Joost, S., Nicolazzi, E. L., Rochat, E., Rothschild, M. F., Servin, B., Sonstegard, T. S., Steri, R., Van Tassell, C. P., Ajmone-Marsan, P., Crepaldi, P., Stella, A., & the AdaptMap Consortium. 2018. Genome - wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post - domestication migration routes. Genet. Sel. Evol. 50, 1–20 https://doi.org/10.1186/s12711-018-0422-x.

Curik, I., Ferenčaković, M., & Sölkner, J. 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 166, 26–34 https://doi.org/10.1016/j.livsci.2014.05.034.

Dangi, S. S., Gupta, M., Dangi, S. K., Chouhan, V. S., Maurya, V. P., Kumar, P., Singh, G., & Sarkar, M. 2015. Expression of HSPs: an adaptive mechanism during long-term heat stress in goats (Capra hircus). Int. J. Biometeorol. 59, 1095–1106 https://doi.org/10.1007/s00484-014-0922-5.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J., & Wang, W. 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat. Biotechnol. 31, 135–141 https://doi.org/10.1038/nbt.2478.

Dubeuf, J. P., & Boyazoglu, J. 2009. An international panorama of goat selection and breeds. Livest. Sci. 120, 225–231 https://doi.org/10.1016/j.livsci.2008.07.005.

Dunbar, R. I. M., & Shi, J. 2013. Time as a constraint on the distribution of feral goats at high latitudes. Oikos 122, 403–410 https://doi.org/10.1111/j.1600-0706.2012.20596.x.

El-halawany, N. K., Abd-el-razek, F. M., El-sayed, Y. A., Shawky, A. A., Al-tohamy, A. F., & Abdel-shafy, H. 2017. Genetic Polymorphisms in Exon-3 region of Growth Hormone Gene in the Egyptian Goat Breeds. Egypt. Acad. J. Biol. Sci. 9, 1–8 https://doi.org/10.21608/eajbsc.2017.13662.

Eusebi, P. G., Martinez, A., & Cortes, O. 2019. Genomic Tools for E ff ective Conservation of Livestock Breed Diversity. Diversity 12, 8 https://doi.org/10.3390/d12010008.

FAO. 2011. FAO Animal Production And Health Guidelines.

FAO. 2015. The second report on the state of the world's animal genetic resources for food and agriculture.

FAOStat.

Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. 2013. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. Genetics 193, 929–941

https://doi.org/10.1534/genetics.112.147231.

Garritsen, C., Van Marle-Köster, E., Snyman, M. A., & Visser, C. 2015. The impact of DNA parentage verification on breeding value estimation and sire ranking in South African Angora goats. Small Rumin. Res. 124, 30–37 https://doi.org/10.1016/j.smallrumres.2015.01.008.

Gaughan, J. B., Sejian, V., Mader, T. L., & Dunshea, F. R. 2019. Adaptation strategies : ruminants. Anim. Front. 9 https://doi.org/10.1093/af/vfy029.

Giger-reverdin, S., Domange, C., Broudiscou, L. P., Sauvant, D., & Berthelot, V. 2020. Rumen function in goats , an example of adaptive capacity. J. Dairy Res. 87, 45–51 https://doi.org/10.1017/ S0022029920000060 Received:

Gouveia, J. J. de S., Paiva, S. R., McManus, C. M., Caetano, A. R., Kijas, J. W., Faco, O., Azevedo, H. C., de Araujo, A. M., de Souza, C. J. H., Yamagishi, M. E. B., Carneiro, P. L. S., Lôbo, R. N. B., de Oliveira, S. M. P., & da Silva, M. V. G. B. 2017. Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. Livest. Sci. J. 197, 36–45 https://doi.org/10.1016/j.livsci.2017.01.006.

Guan, D., Luo, N., Tan, X., Zhao, Z., Huang, Y., Na, R., Zhang, J., & Zhao, Y. 2016. Scanning of selection signature provides a glimpse into important economic traits in goats (Capra hircus). Sci. Rep. 6, 36372 https://doi.org/10.1038/srep36372.

Gurgul, A., Semik, E., Pawlina, K., Szmatoła, T., Jasielczuk, I., & Bugno-Poniewierska, M. 2014. The application of genome-wide SNP genotyping methods in studies on livestock genomes. J. Appl. Genet. 55, 197–208 https://doi.org/10.1007/s13353-014-0202-4.

Gutiérrez-Gil, B., Arranz, J. J., & Wiener, P. 2015. An interpretive review of selective sweep studies in Bos taurus cattle populations: Identification of unique and shared selection signals across breeds. Front. Genet. 6 https://doi.org/10.3389/fgene.2015.00167.

Henkel, J., Saif, R., Jagannathan, V., Schmocker, C., Zeindler, F., Bangerter, E., Herren, U., Posantzis, D., Bulut, Z., Ammann, P., Drögemüller, C., Flury, C., & Leeb, T. 2019. Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. PLOS Genet. 15, 1–18 https://doi.org/10.1371/journal.pgen.1008536.

Horscroft, C., Ennis, S., Pengelly, R. J., Sluckin, T. J., & Collins, A. 2019. Sequencing era methods for identifying signatures of selection in the genome. Brief. Bioinform. 20, 1997–2008 https://doi.org/10.1093/bib/bby064.

Khanyile, K. S., Dzomba, E. F., & Muchadeyi, F. C. 2015. Population genetic structure , linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. Front. Genet. 6 https://doi.org/10.3389/fgene.2015.00013.

Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., & Rothschild, M. F. 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity (Edinb). 116, 255–264 https://doi.org/10.1038/hdy.2015.94.

Kotzé, A., Grobler, J. P., Jonker, T., & Dalton, D. L. 2014. The Tankwa Karoo National Park feral goat population : A unique genetic resource. South African J. Anim. Sci. 44, 43–48 https://doi.org/10.4314/sajas.v44i1.6.

Lashmar, S. F., Visser, C., & van Marle-Köster, E. 2015. Validation of the 50k Illumina goat SNP chip in the South African Angora goat. S. Afr. J. Anim. Sci. 45, 2013–2016.

Lashmar, S. F., Visser, C., & van Marle-Köster, E. 2016. SNP-based genetic diversity of South African commercial dairy and fiber goat breeds. Small Rumin. Res. 136, 65–71 https://doi.org/10.1016/j.smallrumres.2016.01.006.

Lauvie, A., Audiot, A., Couix, N., Casabianca, F., Brives, H., & Verrier, E. 2011. Diversity of rare breed management programs : Between conservation and development. Livest. Sci. 140, 161–170 https://doi.org/10.1016/j.livsci.2011.03.025.

Lenstra, J. A., Groeneveld, L. F., Eding, H., Kantanen, J., Williams, J. L., Taberlet, P., Nicolazzi, E. L., Sölkner, J., Simianer, H., Ciani, E., Garcia, J. F., Bruford, M. W., Ajmone-Marsa, P., & Weugend, S. 2012. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. Anim. Genet. 43, 483–502 https://doi.org/10.1111/j.1365-2052.2011.02309.x.

Makina, S. O., Muchadeyi, F. C., Van Marle-Köster, E., Taylor, J. F., Makgahlela, M. L., & Maiwashe, A. 2015a. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genet. Sel. Evol. 47, 92 https://doi.org/10.1186/s12711-015-0173-x.

Makina, S. O., Taylor, J. F., Marle-köster, E. Van, Muchadeyi, F. C., Makgahlela, M. L., Macneil, M. D., & Maiwashe, A. 2015b. Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. Front. Genet. 6 https://doi.org/10.3389/fgene.2015.00337.

Maletsanake, D., Nsoso, S. J., & Kgwatalala, P. M. 2013. Genetic variation from 12 microsatellite makers in an indigenous Tswana goat flock in South-Eastern Botswana. Livest. Res. Rural Dev. 25.

Malomane, D. K., Reimer, C., Weigend, S., Weigend, A., & Sharifi, A. R. 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. BMC Genomics 19 https://doi.org/10.1186/s12864-017-4416-9.

Van Marle-Köster, E., & Visser, C. 2018. Genomics for the advancement of livestock production: A South African perspective. S. Afr. J. Anim. Sci. 48 https://doi.org/10.4314/sajas.v48i5.2.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-marsan, P., Valentini, A., Williams, J. L., & Macciotta, N. P. P. 2014. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 46, 110–121 https://doi.org/10.1111/age.12259.

Marsoner, T., Egarter, L., Manck, F., Jaritz, G., Tappeiner, U., & Tasser, E. 2018. Indigenous livestock breeds as indicators for cultural ecosystem services : A spatial analysis within the Alpine Space. Ecol. Indic. 94, 55–63 https://doi.org/10.1016/j.ecolind.2017.06.046.

Martínez, A., Ferrando, A., Manunza, A., Gómez, M., Landi, V., Jordana, J., Capote, J., Badaoui, B., Vidal, O., Delgado, J. V., & Amills, M. 2012. Inferring the demographic history of a highly endangered goat breed through the analysis of nuclear and mitochondrial genetic signatures. Small Rumin. Res. 104, 78–84 https://doi.org/10.1016/j.smallrumres.2011.11.012.

Masseti, M. 2009. The wild goats Capra aegagrus Erxleben, 1777 of the Mediterranean Sea and the Eastern Atlantic Ocean islands. Mamm. Rev. 39, 141–157 https://doi.org/10.1111/j.1365-2907.2009.00141.x.

Mataveia, G. A., Garrine, C. M. L. P., Pondja, A., Hassen, A., & Visser, C. 2018. Smallholder goat production in the namaacha and moamba districts of southern mozambique. J. Agric. Rural Dev. Trop. Subtrop. 119, 31–41 https://doi.org/10.17170/kobra-2018112825.

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

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

Mdladla, K., Dzomba, E. F., & Muchadeyi, F. C. 2017. Characterization of the village goat production systems in the rural communities of the Eastern Cape, KwaZulu-Natal, Limpopo and North West Provinces of South Africa. Trop. Anim. Health Prod. 49, 515–527 https://doi.org/10.1007/s11250-017-1223-x.

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

Meek, M. H., Wells, C., Tomalty, K. M., Ashander, J., Cole, E. M., Gille, D. A., Putman, B. J., Rose, J. P., Savoca, M. S., Yamane, L., Hull, J. M., Rogers, D. L., Bree, E., Shogren, J. F., Swaisgood, R. R., & May, B. 2015. Fear of failure in conservation : The problem and potential solutions to aid conservation of extremely small populations. Biol. Conserv. 184, 209–217 https://doi.org/10.1016/j.biocon.2015.01.025.

Mhlanga, T. T., Mutibvu, T., & Mbiriri, D. T. 2018. Goat flock productivity under smallholder farmer management in Zimbabwe. Small Rumin. Res. 164, 105–109 https://doi.org/10.1016/j.smallrumres.2018.05.010.

Miller, J. M., Malenfant, R. M., David, P., Davis, C. S., Poissant, J., Hogg, J. T., Festa-Bainchet, M., & Coltman, D. W. 2014. Estimating genome-wide heterozygosity : effects of demographic history and marker type. Heredity (Edinb). 112, 240–247 https://doi.org/10.1038/hdy.2013.99.

Monau, P., Raphaka, K., Zvinorova-Chimboza, P., & Gondwe, T. 2020a. Sustainable Utilization of Indigenous Goats in Southern Africa. Diversity 12, 1–9 https://doi.org/10.3390/d12010020.

Monau, P. I., Visser, C., Muchadeyi, F. C., Okpeku, M., Nsoso, S. J., & van Marle-

Köster, E. 2020b. Population structure of indigenous southern African goats based on the Illumina Goat50K SNP panel. Trop. Anim. Health Prod. 52, 1795–1802 https://doi.org/10.1007/s11250-019-02190-9.

Monau, P. I., Visser, C., Nsoso, S. J., & van Marle-Köster, E. 2018. Phenotypic and genetic characterization of indigenous Tswana goats. S. Afr. J. Anim. Sci. 48 https://doi.org/10.4314/sajas.v48i5.12.

Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., & McEwan, J. C. 2012. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BMC Genet. 13 https://doi.org/10.1186/1471-2156-13-10.

Mwai, O., Hanotte, O., Kwon, Y., & Cho, S. 2015. African Indigenous Cattle : Unique Genetic Resources in a Rapidly Changing World. Asian Australas. J. Anim. Sci. 28, 911–921 https://doi.org/10.5713/ajas.15.0002R.

Naderi, S., Rezaei, H.-R., Pompanon, F., Blum, M. G. B., Negrini, R., Naghash, H.-R., Balkiz, Ö., Mashkour, M., Gaggiotti, O. E., Ajmone-Marsan, P., Kence, A., Vigne, J.-D., & Taberlet, P. 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. Proc. Natl. Acad. Sci. 105, 17659–17664 https://doi.org/10.1073/pnas.0804782105.

Nanaei, H. A., Koopaee, H. K., & Esmailizadeh, A. 2022. Genetic diversity and signatures of selection for heat tolerance and immune response in Iranian native chickens. BMC Genomics 23 https://doi.org/10.1186/s12864-022-08434-7.

Ncube, K. T., Dzomba, E. F., Hadebe, K., Soma, P., Frylinck, L., & Muchadeyi, F. C. 2022. Carcass Quality Profiles and Associated Genomic Regions of South African Goat Populations Investigated Using Goat SNP50K Genotypes. Animals 12, 364 https://doi.org/10.3390/ ani12030364 Academic.

Ncube, K. T., Hadebe, K., Dzomba, E. F., Soma, P., Frylinck, L., & Muchadeyi, F. C. 2020. Relationship between population genomic structure and growth profiles of South African goats under different production systems. Trop, 1277–1286 https://doi.org/10.1007/s11250-019-02128-1.

Nicolazzi, E. L., Biffani, S., Biscarini, F., Orozco ter Wengel, P., Caprera, A., Nazzicari, N., & Stella, A. 2015. Software solutions for the livestock genomics SNP array revolution. Anim. Genet. 46, 343–353 https://doi.org/10.1111/age.12295.

Nicoloso, L., Bomba, L., Colli, L., Negrini, R., Milanesi, M., Mazza, R., Sechi, T., Frattini, S., Talenti, A., Coizet, B., Chessa, S., Marletta, D., D'Andrea, M., Bordonaro, S., Ptak, G., Carta, A., Pagnacco, G., Valentini, A., Pilla, F., Ajmone-Marsan, P., Crepaldi, P., & the Italian Goat Consortium. 2015. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. Genet. Sel. Evol. 47 https://doi.org/10.1186/s12711-015-0140-6.

Onzima, R. B., Gizaw, S., Kugonza, D. R., van Arendonk, J. A. M., & Kanis, E. 2018a. Production system and participatory identification of breeding objective traits for indigenous goat breeds of Uganda. Small Rumin. Res. 163, 51–59 https://doi.org/10.1016/j.smallrumres.2017.07.007.

Onzima, R. B., Upadhyay, M. R., Doekes, H. P., Brito, L. F., Bosse, M., Kanis, E.,

Groenen, M. A. M., & Crooijmans, R. P. M. A. 2018b. Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. Front. Genet. 9, 1–13 https://doi.org/10.3389/fgene.2018.00318.

Onzima, R. B., Upadhyay, M. R., Mukiibi, R., Kanis, E., Groenen, M. A. M., & Crooijmans, R. P. M. A. 2018c. Genome-wide population structure and admixture analysis reveals weak differentiation among Ugandan goat breeds. Anim. Genet. 49, 59–70 https://doi.org/10.1111/age.12631.

Paim, T. do P., Faria, D. A., Hay, E. H., McManus, C., Lanari, M. R., Esquivel, L. C., Cascante, M. I., Alfaro, E. J., Mendez, A., Faco, O., Silva, K. de M., Mezzadra, C. A., Mariante, A., Paiva, S. R., & Blackburn, H. D. 2019. New world goat populations are a genetically diverse reservoir for future use. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-38812-3.

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., & Baldi, F. 2016. Runs of homozygosity : current knowledge and applications in livestock. Anim. Genet. 48, 255–271 https://doi.org/10.1111/age.12526.

Pieters, A., van Marle-Köster, E., Visser, C., & Kotze, A. 2009. South African developed meat type goats: A forgotten animal genetic resource? Anim. Genet. Resour. Inf. 44, 33–43 https://doi.org/10.1017/s1014233900002844.

Purcell, S., Neale, B., Todd-brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. 2007. REPORT PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81, 559–575 https://doi.org/10.1086/519795.

Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13, 70 https://doi.org/10.1186/1471-2156-13-70.

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

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

Rojas-downing, M. M., Nejadhashemi, A. P., Harrigan, T., & Woznicki, S. A. 2017. Climate Risk Management Climate change and livestock : Impacts , adaptation , and mitigation. Clim. Risk Manag. 16, 145–163 https://doi.org/10.1016/j.crm.2017.02.001.

Saravanan, K. A., Panigrahi, M., Kumar, H., & Bhushan, B. 2022. Advanced software programs for the analysis of genetic diversity in livestock genomics : a mini review. Biol. Rhythm Res. 53, 358–368 https://doi.org/10.1080/09291016.2019.1642650.

Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., & Mishra, B. P. 2020. Selection signatures in livestock genome : A review of concepts , approaches and applications. Livest. Sci., 104257 https://doi.org/10.1016/j.livsci.2020.104257.

Sejian, V., Bagath, M., Krishnan, G., Rashamol, V. P., Pragna, P., Devaraj, C., & Bhatta, R. 2019. Genes for resilience to heat stress in small ruminants: A review. Small Rumin. Res. 173, 42–53 https://doi.org/10.1016/j.smallrumres.2019.02.009.

Selolo, T. C., Mashiloane, M. L., Norris, D., Ng'ambi, J. W., & Brown, D. 2015. Morphological differentiation of indigenous goats in different agro-ecological zones of vhembe district, Limpopo province, South Africa. Indian J. Anim. Res. 49, 527–531 https://doi.org/10.5958/0976-0555.2015.00052.7.

Shaji, S., Sejian, V., Bagath, M., Manjunathareddy, G. B., Kurien, E. K., Varma, G., & Bhatta, R. 2016. Summer season related heat and nutritional stresses on the adaptive capability of goats based on blood biochemical response and hepatic HSP70 gene expression. Biol. Rhythm Res. https://doi.org/10.1080/09291016.2016.1232340.

Singh, S. K., Meena, H. R., Kolekar, D. V, & Singh, Y. P. 2012. Climate Change Impacts on Livestock and Adaptation Strategies to Sustain Livestock Production. J. Vet. Adv. 2, 407–412.

Sobrino, B., Brion, M., & Carracedo, A. 2005. SNPs in forensic genetics : a review on SNP typing methodologies. Forensic Sci. Int. 154, 181–194 https://doi.org/10.1016/j.forsciint.2004.10.020.

Stainton, J. J., Haley, C. S., Charlesworth, B., Kranis, A., Watson, K., & Wiener, P. 2014. Detecting signatures of selection in nine distinct lines of broiler chickens. Anim. Genet. 46, 37–49 https://doi.org/10.1111/age.12252.

Stella, A., Nicolazzi, E. L., Tassell, C. P. Van, Rothschild, M. F., Colli, L., Rosen, B. D., Sonstegard, T. S., Crepaldi, P., Tosser-Klopp, G., Joost, S., & AdaptMap Consortium. 2018. AdaptMap : exploring goat diversity and adaptation. Genet. Sel. Evol. 50, 1–7 https://doi.org/10.1186/s12711-018-0427-5.

Taberlet, P., Coissac, E., Pansu, J., & Pompanon, F. 2011. Conservation genetics of cattle, sheep, and goats. C. R. Biol. 334, 247–254 https://doi.org/10.1016/j.crvi.2010.12.007.

Talenti, A., Bertolini, F., Pagnacco, G., Pilla, F., Ajmone-Marsan, P., Rothschild, M. F., Crepaldi, P., & The Italian Goat Consortium. 2017. The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. Mamm. Genome 28, 114–128 https://doi.org/10.1007/s00335-017-9678-7.

Tarekegn, M. G., Wouobeng, P., Jaures, K. S., Mrode, R., Edea, Z., Liu, B., Zhang, W., Mwai, O. A., Dessie, T., Tesfaye, K., Strandberg, E., Berglund, B., Mutai, C., Osama, S., Wolde, A. T., Birungi, J., Djikeng, A., & Meutchieye, F. 2019. Genome-wide diversity and demographic dynamics of Cameroon goats and their divergence from east African, north African, and Asian conspecifics. PLoS One 14 https://doi.org/10.1371/journal.pone.0214843.

Thornton, P. K., Boone, R. B., & Ramirez-villegas, J. 2015. Climate Change Impacts on Livestock.

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., Donnadieu-Tonon, C., Eggen, A., Heuven, H. C. M., Jamli, S., Jiken, A. J., Klopp, C., Lawley, C. T., McEwan, J., Martin, P., Moreno, C. R., Mulsant, P.,

Nabihoudine, I., Pailhoux, E., Palhière, I., Rupp, R., Sarry, J., Sayre, B. L., Tircazes, A., Wang, J., Wang, W., Zhang, W., & the International Goat Genome Consortium. 2014. Design and characterization of a 52K SNP chip for goats. PLoS One 9, 1–8 https://doi.org/10.1371/journal.pone.0086227.

Visser, C. 2019. A review on goats in southern Africa: An untapped genetic resource. Small Rumin. Res. 176, 11–16 https://doi.org/10.1016/j.smallrumres.2019.05.009.

Visser, C., Lashmar, S. F., Van Marle-Köster, E., Poli, M. A., & Allain, D. 2016. Genetic diversity and population structure in South African, French and Argentinian Angora goats from genome-wide SNP data. PLoS One 11 https://doi.org/10.1371/journal.pone.0154353.

Visser, C., Van Marle-Köster, E., Bovenhuis, H., & Crooijmans, R. P. M. A. 2011a. QTL for mohair traits in South African Angora goats. Small Rumin. Res. 100, 8–14 https://doi.org/10.1016/j.smallrumres.2011.05.007.

Visser, C., van Marle-Köster, E., & Friedrich, H. 2011b. Parentage verification of South African Angora goats, using microsatellite markers. South African J. Anim. Sci. 41, 250–255 https://doi.org/10.4314/sajas.v41i3.7.

Visser, C., Van Marle-Köster, E., Snyman, M. A., Bovenhuis, H., & Crooijmans, R. P. M. A. 2013. Quantitative trait loci associated with pre-weaning growth in South African Angora goats. Small Rumin. Res. 112, 15–20 https://doi.org/10.1016/j.smallrumres.2012.11.035.

Webb, E. C., & Mamabolo, M. J. 2004. Production and reproduction characteristics of South African indigenous goats in communal farming systems. S. Afr. J. Anim. Sci. 34, 236–239.

Weir, B. S., & Cockerham, C. C. 1984. Atistics for the analysis of population structure. Evolution (N. Y)., 1358–1370 https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.

Willing, E. M., Dreyer, C., & van Oosterhout, C. 2012. Estimates of genetic differentiation measured by fst do not necessarily require large sample sizes when using many snp markers. PLoS One 7, 1–7 https://doi.org/10.1371/journal.pone.0042649.

World Weather Online.

Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., Sun, H., Liu, X., Lyu, S., Wang, X., Li, Z., Yang, P., Xu, J., Ding, X., Shi, Q., Wang, E., Ru, B., Xu, Z., Lei, C., Chen, H., & Huang, Y. 2021. Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. BCM Genomics 22 https://doi.org/10.1186/s12864-020-07340-0.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. 2011. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 https://doi.org/10.1016/j.ajhg.2010.11.011.

Yurchenko, A. A., Deniskova, T. E., Yudin, N. S., Dotsev, A. V, Khamiruev, T. N., Selionova, M. I., Egorov, S. V, Reyer, H., Wimmers, K., Brem, G., Zinovieva, N. A., & Larkin, D. M. 2019. High-density genotyping reveals signatures of

selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. BMC Genomics 20, 294 https://doi.org/10.1186/s12864-019-5537-0 RESEARCH.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., Laird, M. R., Lavidas, I., Liu, Z., Loveland, J. E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D. N., Newman, V., Nuhn, M., Ogeh, D., Ong, C. K., Parker, A., Patricio, M., Riat, H. S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S. E., Kostadima, M., Langridge, N., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Aken, B. L., Cunningham, F., Yates, A., & Flicek, P. 2018. Ensembl 2018. Nucleic Acids Res. 46, D754–D761 https://doi.org/10.1093/nar/gkx1098.

Zhang, Q., Calus, M. P. L., Guldbrandtsen, B., Lund, M. S., & Sahana, G. 2015. Estimation of inbreeding using pedigree , 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet. 16 https://doi.org/10.1186/s12863-015-0227-7.

Zhang, S., Yao, Z., Li, X., Zhang, Z., Liu, X., Yang, P., Chen, N., Xia, X., Lyu, S., Shi, Q., Wang, E., Ru, B., Jiang, Y., Lei, C., Chen, H., & Huang, Y. 2022. Assessing genomic diversity and signatures of selection in Pinan cattle using whole - genome sequencing data. BMC Genomics 23 https://doi.org/10.1186/s12864-022-08645-y.

Zhao, F., McParland, S., Kearney, F., Du, L., & Berry, D. P. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet. Sel. Evol. 47 https://doi.org/10.1186/s12711-015-0127-3.

Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., Pan, X., Gong, M., Zhang, Y., Guo, Y., Wang, Y., Liu, J., Cai, Y., Chen, Q., Okpeku, M., Colli, L., Cai, D., Wang, K., Huang, S., Sonstegard, T. S., Esmailizadeh, A., Zhang, W., Zhang, T., Xu, Y., Xu, N., Yang, Y., Han, J., Chen, L., Lesur, J., Daly, K. G., Bradley, D. G., Heller, R., Zhang, G., Wang, W., Chen, Y., & Jiang, Y. 2020. The origin of domestication genes in goats. Sci. Adv. 6 https://doi.org/10.1126/sciadv.aaz5216.

**Addendum A** A non-comprehensive list of genetic characterisation studies done using the CaprineSNP50 BeadChip and the genetic diversity results obtained

| Breed | Country | Number of individuals | Inbreeding Coefficient - $F_{IS}$ | Observed Heterozygosity - $H_O$ | Expected Heterozygosity - $H_E$ | Authors |
|---|---|---|---|---|---|---|
| Afar | Ethiopia | 49 | 0.02 | 0.383 | 0.391 | Tarekegn *et al.*, 2019 |
| Alpine | Canada | 403 | 0.031 | 0.385 | 0.388 | Brito *et al.*, 2017 |
| Ambo | Ethiopia | 119 | 0.02 | 0.371 | 0.379 | Tarekegn *et al.*, 2019 |
| Angora | Argentina | 23 | 0.073 | 0.4 | - | Paim *et al.*, 2019 |
| Angora | Argentina | 30 | -0.047 | 0.414 | 0.397 | Visser *et al.*, 2016 |
| Angora | France | 26 | -0.003 | 0.378 | 0.38 | Visser *et al.*, 2016 |
| Angora | South Africa | 48 | 0.009 | 0.324 | 0.333 | Lashmar *et al.*, 2016 |
| Angora | South Africa | 43 | 0.227 | 0.333 | - | Paim *et al.*, 2019 |
| Angora | South Africa | 48 | 0.009 | 0.365 | 0.371 | Visser *et al.*, 2016 |
| Angora | United States | 29 | 0.143 | 0.37 | - | Paim *et al.*, 2019 |
| Argentata dell'Etna | Italy | 25 | 0.02 | 0.41 | 0.41 | Nicoloso *et al.*, 2015 |
| Barki | Egypt | 52 | 0.02 | 0.401 | 0.41 | Tarekegn *et al.*, 2019 |
| Bionsa dell'Adamello | Italy | 24 | 0.02 | 0.4 | 0.4 | Nicoloso *et al.*, 2015 |
| Boer | Australia | 61 | 0.047 | 0.365 | 0.356 | Brito *et al.*, 2017 |
| Boer | Canada | 67 | 0.057 | 0.363 | 0.357 | Brito *et al.*, 2017 |
| Boer | South Africa | 31 | 0.12 | 0.36 | 0.37 | Mdladla *et al.*, 2016 |
| Boer | United States | 17 | 0.165 | 0.36 | - | Paim *et al.*, 2019 |
| British Alpine | South Africa | 14 | -0.088 | 0.385 | 0.355 | Lashmar *et al.*, 2016 |
| Camosciata delle Alpi | Italy | 31 | 0.02 | 0.4 | 0.4 | Nicoloso *et al.*, 2015 |
| Caninde | Brazil | 19 | 0.236 | 0.329 | - | Paim *et al.*, 2019 |
| Cashmere | Australia | 48 | 0.021 | 0.384 | 0.372 | Brito *et al.*, 2017 |
| Cashmere | China | 108 | -0.02 | 0.369 | 0.363 | Tarekegn *et al.*, 2019 |
| Central highland | Camaroon | 94 | 0.03 | 0.341 | 0.352 | Tarekegn *et al.*, 2019 |
| Ciociara Grigia | Italy | 19 | 0.06 | 0.39 | 0.4 | Nicoloso *et al.*, 2015 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Colorada Pampeana | Argentina | 11 | 0.072 | 0.4 | - | Paim *et al.*, 2019 |
| Criolla Formosena | Argentina | 13 | 0.124 | 0.378 | - | Paim *et al.*, 2019 |
| Criollo de los Llanos | Argentina | 13 | 0.093 | 0.391 | - | Paim *et al.*, 2019 |
| Criollo Neuquino | Argentina | 17 | 0.05 | 0.41 | - | Paim *et al.*, 2019 |
| Criollo Riojano | Argentina | 6 | 0.099 | 0.389 | - | Paim *et al.*, 2019 |
| Dell'Aspromonte | Italy | 24 | 0.06 | 0.38 | 0.4 | Nicoloso *et al.*, 2015 |
| Di Teramo | Italy | 23 | -0.06 | 0.38 | 0.35 | Nicoloso *et al.*, 2015 |
| Djallonke | Camaroon | 33 | 0.05 | 0.348 | 0.366 | Tarekegn *et al.*, 2019 |
| Girgentana | Italy | 24 | 0.004 | 0.36 | 0.36 | Nicoloso *et al.*, 2015 |
| Gumez | Ethiopia | 42 | 0.01 | 0.376 | 0.38 | Tarekegn *et al.*, 2019 |
| Iranian goat | Iran | 9 | 0.08 | 0.392 | 0.422 | Tarekegn *et al.*, 2019 |
| Kalahari Red | South Africa | 36 | 0.1 | 0.37 | 0.38 | Mdladla *et al.*, 2016 |
| Keffa | Ethiopia | 51 | 0.06 | 0.353 | 0.374 | Tarekegn *et al.*, 2019 |
| LaMancha | Canada | 81 | 0.039 | 0.384 | 0.382 | Brito *et al.*, 2017 |
| LaMancha | United States | 11 | 0.114 | 0.382 | - | Paim *et al.*, 2019 |
| Long-eared Somali | Ethiopia | 48 | 0.01 | 0.378 | 0.382 | Tarekegn *et al.*, 2019 |
| Maltese | Italy | 16 | 0.06 | 0.36 | 0.37 | Nicoloso *et al.*, 2015 |
| Moroccan Goat | Morocco | 30 | 0.06 | 0.388 | 0.411 | Tarekegn *et al.*, 2019 |
| Moxoto | Brazil | 18 | 0.218 | 0.337 | - | Paim *et al.*, 2019 |
| Nguni | South Africa | 10 | 0.01 | 0.41 | 0.39 | Mdladla *et al.*, 2016 |
| Nicastrese | Italy | 25 | 0.07 | 0.38 | 0.4 | Nicoloso *et al.*, 2015 |
| North-west Highland | Camaroon | 166 | 0.08 | 0.335 | 0.363 | Tarekegn *et al.*, 2019 |
| Nubian | Canada | 54 | 0.057 | 0.338 | 0.335 | Brito *et al.*, 2017 |
| Nubian | Ethiopia | 47 | 0.07 | 0.366 | 0.395 | Tarekegn *et al.*, 2019 |
| Orobica | Italy | 24 | 0.01 | 0.35 | 0.35 | Nicoloso *et al.*, 2015 |
| Rangeland | Australia | 66 | 0.009 | 0.413 | 0.411 | Brito *et al.*, 2017 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Saanen | Canada | 318 | 0.033 | 0.379 | 0.382 | Brito *et al.*, 2017 |
| Saanen | Costa Rica | 28 | 0.044 | 0.413 | - | Paim *et al.*, 2019 |
| Saanen | Italy | 24 | -0.001 | 0.41 | 0.41 | Nicoloso *et al.*, 2015 |
| Saanen | South Africa | 20 | -0.015 | 0.378 | 0.373 | Lashmar *et al.*, 2016 |
| Sarda | Italy | 32 | 0.06 | 0.39 | 0.41 | Nicoloso *et al.*, 2015 |
| Savanna | South Africa | 29 | 0.06 | 0.39 | 0.38 | Mdladla *et al.*, 2016 |
| Spanish | United States | 19 | 0.011 | 0.427 | - | Paim *et al.*, 2019 |
| Toggenburg | Canada | 53 | 0.046 | 0.353 | 0.336 | Brito *et al.*, 2017 |
| Toggenburg | South Africa | 6 | -0.135 | 0.385 | 0.339 | Lashmar *et al.*, 2016 |
| Tswana | Botswana | 48 | 0.009 | 0.419 | 0.423 | Monau *et al.*, 2018 |
| Tswana | South Africa | 20 | 0.03 | 0.4 | 0.41 | Mdladla *et al.*, 2016 |
| Valostana | Italy | 24 | 0.05 | 0.36 | 0.37 | Nicoloso *et al.*, 2015 |
| Valpassiria | Italy | 24 | 0.02 | 0.4 | 0.4 | Nicoloso *et al.*, 2015 |
| Venda | South Africa | 25 | 0.04 | 0.41 | 0.4 | Mdladla *et al.*, 2016 |
| Xhosa | South Africa | 20 | 0.02 | 0.42 | 0.41 | Mdladla *et al.*, 2016 |
| Zulu | South Africa | 25 | 0.04 | 0.4 | 0.4 | Mdladla *et al.*, 2016 |

**Addendum B** Official permission letter for the use of data was obtained from Northern Cape DAERL

agriculture, land reform & rural development
Department:
agriculture, land reform & rural development
NORTHERN CAPE PROVINCE
REPUBLIC OF SOUTH AFRICA

**Office of the Head of Department**
162 George Street, Kimberley, Northern Cape, South Africa, 8300, Tel: (053) 838 9118
Fax: (086) 771 8745, E-mail: ZMogorosi@ncpg.gov.za, web: www.agrinc.gov.za

Our Reference

Enquiries: Ms R. Burgess

Date: 16 September 2020

Prof A Kotze
Manager: Research and Scientific Services
South African National Biodiversity Institute (SANBI)
NATIONAL ZOOLOGICAL GARDEN
P. O. Box 754
**PRETORIA**
0001

TO WHOM IT MAY CONCERN:

The Northern Cape Department of Agriculture, Land Reform and Rural Development hereby grants permission for Ms Anna Sophia Kropff to use the 50K SNP data generated for 350 Tankwa goats in her MSc Study titled Genetic characterisation of the Tankwa Goat (*Capra hircus*) using genome-wide SNP data and the identification of genes associated with climate adaptation. Thank you.

Sincerely

Ms R Burgess: Manager: Research

Supported / ~~Not supported~~

Dr Kegakilwe: Chief Director Technical Services

**Addendum C:** Number of SNP's, Observed and Expected Heterozygosity, Minor Allele Frequency (MAF) and Linkage Disequilibrium (LD) calculated as the average per chromosome for the Tankwa Goats (Table1) Angora Goats (Table 2) and dairy Goats (Table 3).

Table 1: Population-specific diversity parameters per chromosome for the Tankwa goat population

| CHI | N SNP | CHI (Mb) | $H_E$ | $H_O$ | MAF | $r^2$ |
|---|---|---|---|---|---|---|
| 1 | 2764 | 157 | 0.378 | 0.378 | 0.259 | 0.454 |
| 2 | 2403 | 136 | 0.370 | 0.371 | 0.251 | 0.468 |
| 3 | 1993 | 120 | 0.372 | 0.377 | 0.252 | 0.458 |
| 4 | 2017 | 120 | 0.371 | 0.368 | 0.249 | 0.481 |
| 5 | 1900 | 119 | 0.367 | 0.370 | 0.247 | 0.453 |
| 6 | 2037 | 117 | 0.381 | 0.383 | 0.260 | 0.460 |
| 7 | 1888 | 108 | 0.379 | 0.379 | 0.260 | 0.456 |
| 8 | 2013 | 112 | 0.376 | 0.373 | 0.260 | 0.456 |
| 9 | 1602 | 91 | 0.371 | 0.375 | 0.252 | 0.488 |
| 10 | 1803 | 101 | 0.374 | 0.372 | 0.259 | 0.458 |
| 11 | 1784 | 106 | 0.370 | 0.370 | 0.247 | 0.495 |
| 12 | 1401 | 87 | 0.356 | 0.353 | 0.231 | 0.517 |
| 13 | 1421 | 83 | 0.370 | 0.367 | 0.253 | 0.460 |
| 14 | 1627 | 94 | 0.362 | 0.356 | 0.243 | 0.463 |
| 15 | 1394 | 81 | 0.371 | 0.371 | 0.253 | 0.442 |
| 16 | 1352 | 79 | 0.367 | 0.369 | 0.249 | 0.475 |
| 17 | 1266 | 71 | 0.377 | 0.377 | 0.262 | 0.481 |
| 18 | 1037 | 67 | 0.345 | 0.331 | 0.223 | 0.491 |
| 19 | 1023 | 62 | 0.373 | 0.379 | 0.251 | 0.455 |
| 20 | 1265 | 71 | 0.375 | 0.368 | 0.255 | 0.477 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 21 | 1219 | 69 | 0.354 | 0.357 | 0.235 | 0.473 |
| 22 | 1022 | 60 | 0.373 | 0.370 | 0.257 | 0.462 |
| 23 | 894 | 48 | 0.353 | 0.348 | 0.235 | 0.468 |
| 24 | 1048 | 62 | 0.333 | 0.332 | 0.205 | 0.525 |
| 25 | 750 | 42 | 0.372 | 0.372 | 0.260 | 0.451 |
| 26 | 896 | 26 | 0.364 | 0.365 | 0.248 | 0.470 |
| 27 | 791 | 27 | 0.379 | 0.379 | 0.260 | 0.473 |
| 28 | 761 | 28 | 0.356 | 0.354 | 0.229 | 0.483 |
| 29 | 867 | 29 | 0.379 | 0.377 | 0.268 | 0.414 |
| Average | | | 0.368 | 0.367 | 0.249 | 0.469 |

*The column headings are: CHI – Chromosome number, N SNP – number of SNP's that passed quality control, CHI (Mb) – Chromosome length, $H_E$ – Expected Heterozygosity, $H_O$ – Observed Heterozygosity, MAF – Minor allele frequency, $r^2$ – Linkage disequilibrium

Table 2: Population-specific diversity parameters per chromosome for the Angora goat population

| CHI | N SNP | CHI (Mb) | $H_E$ | $H_O$ | MAF | $r^2$ |
|---|---|---|---|---|---|---|
| 1 | 3044 | 157 | 0.353 | 0.348 | 0.255 | 0.387 |
| 2 | 2625 | 136 | 0.347 | 0.350 | 0.245 | 0.402 |
| 3 | 2201 | 120 | 0.351 | 0.354 | 0.252 | 0.381 |
| 4 | 2264 | 120 | 0.355 | 0.351 | 0.256 | 0.388 |
| 5 | 2091 | 119 | 0.355 | 0.347 | 0.255 | 0.392 |
| 6 | 2201 | 117 | 0.357 | 0.346 | 0.251 | 0.400 |
| 7 | 2020 | 108 | 0.357 | 0.358 | 0.255 | 0.400 |
| 8 | 2186 | 112 | 0.355 | 0.351 | 0.254 | 0.388 |
| 9 | 1760 | 91 | 0.352 | 0.356 | 0.251 | 0.386 |
| 10 | 1957 | 101 | 0.353 | 0.347 | 0.252 | 0.387 |
| 11 | 1990 | 106 | 0.355 | 0.351 | 0.252 | 0.388 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 12 | 1629 | 87 | 0.355 | 0.355 | 0.256 | 0.394 |
| 13 | 1508 | 83 | 0.354 | 0.356 | 0.248 | 0.391 |
| 14 | 1768 | 94 | 0.346 | 0.337 | 0.247 | 0.420 |
| 15 | 1529 | 81 | 0.352 | 0.346 | 0.252 | 0.374 |
| 16 | 1506 | 79 | 0.360 | 0.347 | 0.267 | 0.384 |
| 17 | 1373 | 71 | 0.369 | 0.360 | 0.269 | 0.376 |
| 18 | 1157 | 67 | 0.349 | 0.349 | 0.245 | 0.420 |
| 19 | 1110 | 62 | 0.361 | 0.354 | 0.255 | 0.392 |
| 20 | 1382 | 71 | 0.344 | 0.331 | 0.242 | 0.403 |
| 21 | 1314 | 69 | 0.350 | 0.346 | 0.247 | 0.406 |
| 22 | 1065 | 60 | 0.342 | 0.341 | 0.238 | 0.413 |
| 23 | 939 | 48 | 0.334 | 0.341 | 0.233 | 0.425 |
| 24 | 1230 | 62 | 0.353 | 0.354 | 0.253 | 0.385 |
| 25 | 782 | 42 | 0.361 | 0.351 | 0.257 | 0.373 |
| 26 | 980 | 26 | 0.357 | 0.347 | 0.259 | 0.372 |
| 27 | 869 | 27 | 0.355 | 0.339 | 0.260 | 0.403 |
| 28 | 846 | 28 | 0.357 | 0.347 | 0.256 | 0.371 |
| 29 | 910 | 29 | 0.367 | 0.369 | 0.269 | 0.355 |
| Average across Genome | | | 0.353 | 0.349 | 0.253 | 0.392 |

*The column headings are: CHI – Chromosome number, N SNP – number of SNP's that passed quality control, CHI (Mb) – Chromosome length, $H_E$ – Expected Heterozygosity, $H_O$ – Observed Heterozygosity, MAF – Minor allele frequency, $r^2$ – Linkage disequilibrium

Table 3: Population-specific diversity parameters per chromosome for the dairy goat population

| CHI | N SNP | CHI (Mb) | $H_E$ | $H_O$ | MAF | $r^2$ |
|---|---|---|---|---|---|---|
| 1 | 3164 | 157 | 0.408 | 0.389 | 0.319 | 0.335 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | 2746 | 136 | 0.410 | 0.405 | 0.320 | 0.325 |
| 3 | 2263 | 120 | 0.407 | 0.373 | 0.317 | 0.337 |
| 4 | 2341 | 120 | 0.410 | 0.400 | 0.321 | 0.334 |
| 5 | 2162 | 119 | 0.402 | 0.377 | 0.311 | 0.328 |
| 6 | 2334 | 117 | 0.404 | 0.388 | 0.312 | 0.35 |
| 7 | 2118 | 108 | 0.404 | 0.388 | 0.314 | 0.329 |
| 8 | 2270 | 112 | 0.403 | 0.377 | 0.312 | 0.329 |
| 9 | 1853 | 91 | 0.404 | 0.390 | 0.314 | 0.316 |
| 10 | 2031 | 101 | 0.405 | 0.396 | 0.315 | 0.34 |
| 11 | 2076 | 106 | 0.409 | 0.399 | 0.319 | 0.333 |
| 12 | 1684 | 87 | 0.396 | 0.372 | 0.304 | 0.363 |
| 13 | 1580 | 83 | 0.405 | 0.382 | 0.313 | 0.34 |
| 14 | 1846 | 94 | 0.404 | 0.379 | 0.313 | 0.336 |
| 15 | 1578 | 81 | 0.406 | 0.388 | 0.315 | 0.329 |
| 16 | 1537 | 79 | 0.407 | 0.402 | 0.317 | 0.331 |
| 17 | 1425 | 71 | 0.399 | 0.379 | 0.309 | 0.331 |
| 18 | 1216 | 67 | 0.402 | 0.392 | 0.309 | 0.35 |
| 19 | 1183 | 62 | 0.408 | 0.382 | 0.315 | 0.321 |
| 20 | 1447 | 71 | 0.407 | 0.389 | 0.316 | 0.33 |
| 21 | 1380 | 69 | 0.395 | 0.384 | 0.303 | 0.336 |
| 22 | 1136 | 60 | 0.414 | 0.394 | 0.327 | 0.327 |
| 23 | 1016 | 48 | 0.412 | 0.397 | 0.323 | 0.323 |
| 24 | 1275 | 62 | 0.405 | 0.388 | 0.313 | 0.33 |
| 25 | 830 | 42 | 0.404 | 0.374 | 0.315 | 0.319 |
| 26 | 1019 | 26 | 0.409 | 0.392 | 0.322 | 0.33 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 27 | 888 | 27 | 0.411 | 0.390 | 0.323 | 0.331 |
| 28 | 891 | 28 | 0.405 | 0.371 | 0.314 | 0.33 |
| 29 | 932 | 29 | 0.409 | 0.395 | 0.317 | 0.32 |
| Average across Genome | | | 0.406 | 0.388 | 0.315 | 0.332 |

*The column headings are: CHI – Chromosome number, N SNP – number of SNP's that passed quality control, CHI (Mb) – Chromosome length, $H_E$ – Expected Heterozygosity, $H_O$ – Observed Heterozygosity, MAF – Minor allele frequency, $r^2$ – Linkage disequilibrium

**Addendum D**: Analysis of runs of homozygosity per chromosome

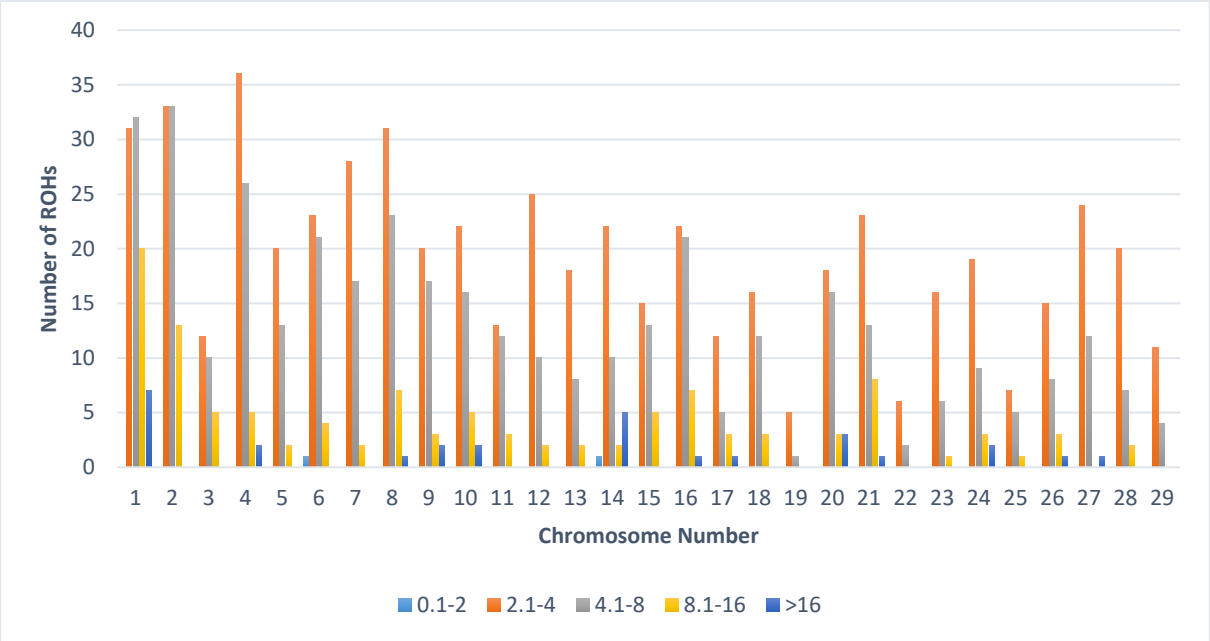


Figure 1 The number of detected ROH of each size category per Chromosome for the Angora goats

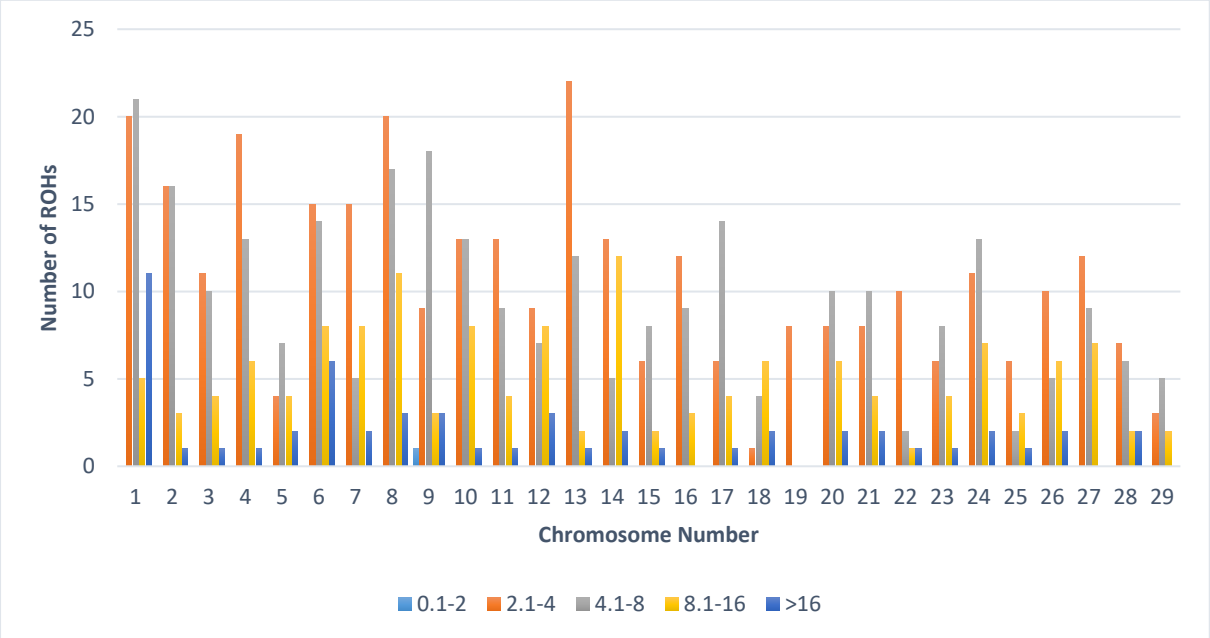


Figure 2 The number of detected ROH of each size category per Chromosome for the dairy goats

**Addendum E**: Structure Plots for the merged dataset

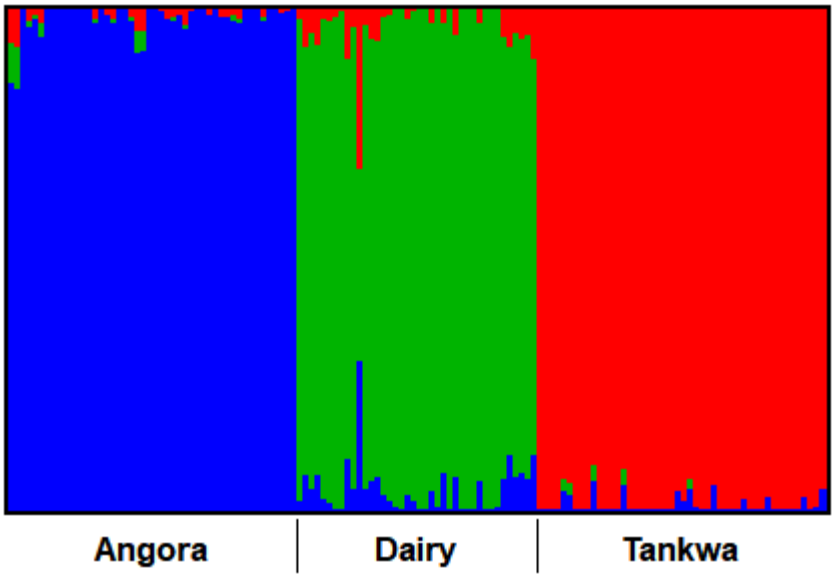


Figure 1 Population structure plot for K=3 of three goat populations

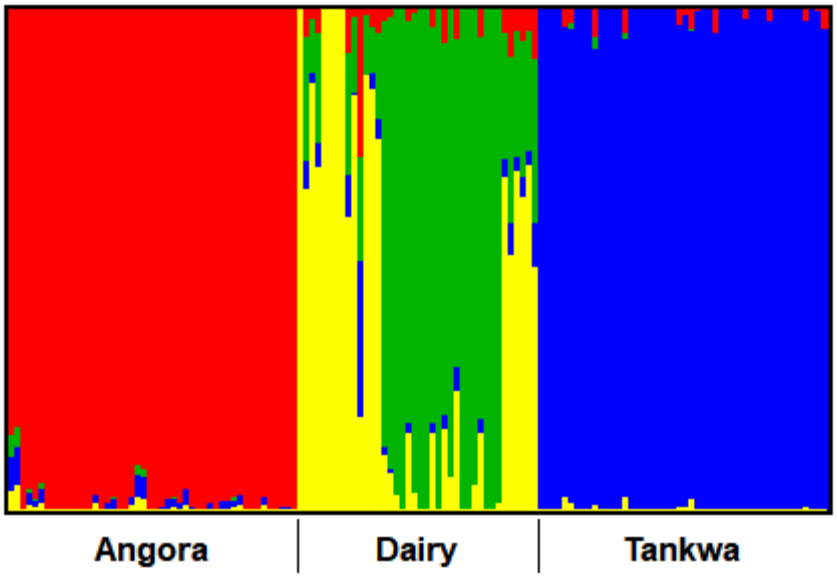


Figure 2 Population structure plot for K=4 of three goat populations

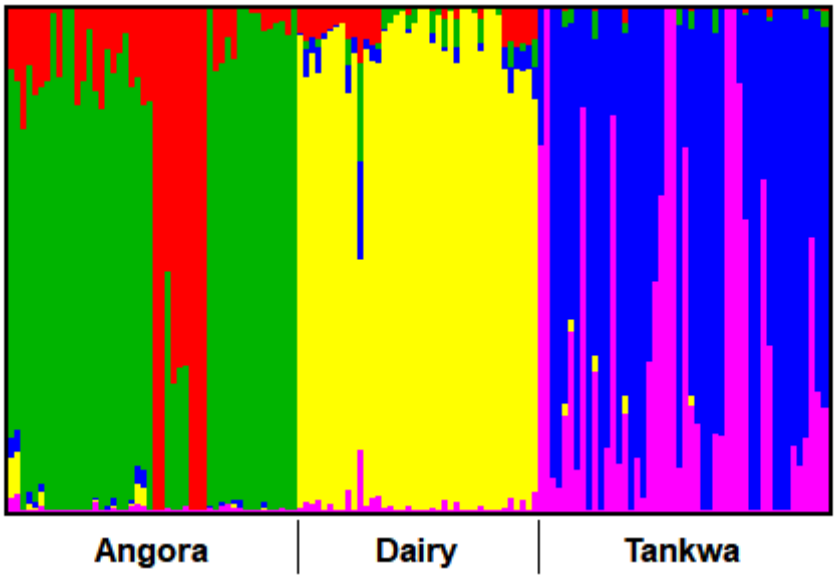


Figure 3 Population structure plot for K=5 of three goat populations

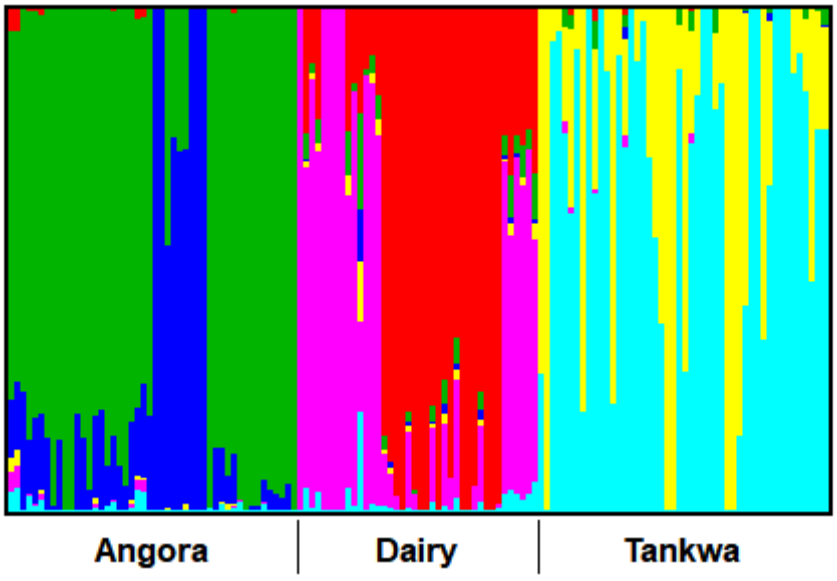


Figure 4 Population structure plot for K=6 of three goat populations

**Addendum F**: The average pairwise $F_{ST}$ per chromosome and across the genome comparing the Tankwa goats to the dairy and Angora goats

| CHI | # of SNP's | $F_{ST}$ |
|---|---|---|
| 1 | 3256 | 0.137 |
| 2 | 2829 | 0.154 |
| 3 | 2380 | 0.138 |
| 4 | 2415 | 0.140 |
| 5 | 2243 | 0.142 |
| 6 | 2435 | 0.134 |
| 7 | 2191 | 0.133 |
| 8 | 2351 | 0.136 |
| 9 | 1894 | 0.137 |
| 10 | 2098 | 0.137 |
| 11 | 2138 | 0.149 |
| 12 | 1749 | 0.163 |
| 13 | 1649 | 0.147 |
| 14 | 1911 | 0.144 |
| 15 | 1639 | 0.134 |
| 16 | 1592 | 0.150 |
| 17 | 1469 | 0.127 |
| 18 | 1291 | 0.163 |
| 19 | 1227 | 0.132 |
| 20 | 1495 | 0.146 |
| 21 | 1430 | 0.160 |
| 22 | 1169 | 0.140 |

| | | |
|---|---|---|
| 23 | 1047 | 0.158 |
| 24 | 1323 | 0.175 |
| 25 | 855 | 0.138 |
| 26 | 1044 | 0.149 |
| 27 | 928 | 0.135 |
| 28 | 914 | 0.159 |
| 29 | 977 | 0.111 |
| Average across the genome | | 0.143 |

**Addendum G** A list of each significant SNP, the chromosome where the SNP is located and all possible genes associated with it

| SNP | CHI | Gene |
|---|---|---|
| snp7474-scaffold127-4095845 | 2 | TLK1, GORASP2 |
| snp3648-scaffold1113-1209905<br>snp3647-scaffold1113-1168279 | 2 | CCDC141 |
| snp34851-scaffold415-1591800 | 6 | MTTP, C4orf54, TRMT10A, C4orf17 |
| snp57278-scaffold912-887689<br>snp57279-scaffold912-928024<br>snp57280-scaffold912-957614 | 10 | DMXL2, GLDN |
| snp25149-scaffold259-3189246 | 10 | RAB8B, APH1B, CA12 |
| snp17810-scaffold185-123982 | 11 | OTOF, DRC1, SELENOI, HADHB, HADHA, GAREM2, RAB10 |
| snp36024-scaffold431-5222024 | 12 | ZMYM5, ZMYM2, GJA3, GJB2, GJB6 |
| snp14829-scaffold1599-1445421 | 16 | MMEL1, PRXL2B |
| snp41898-scaffold546-1841171 | 18 | GCSH, C16orf46, ATMIN, CENPN, CMC2, PKD1L2 |
| snp59705-scaffold99-452145 | 18 | JPH3, KLHDC4, SLC7A5, CA5A, BANP |
| snp35685-scaffold43-3345768<br>snp35684-scaffold43-3315236<br>snp35682-scaffold43-3228757 | 18 | ABCC12 |
| snp18301-scaffold1857-80950 | 18 | FOSB, RTN2, PPM1N, VASP, OPA3, GPR4, EML2 |
| snp7648-scaffold1276-837684<br>snp7647-scaffold1276-801231<br>snp7646-scaffold1276-763056 | 21 | DLK1 |
| snp50705-scaffold730-1805048 | 22 | GRM7 |

| | | |
|---|---|---|
| snp50704-scaffold730-1762368 | | |
| snp54538-scaffold833-507630<br>snp54541-scaffold833-631584<br>snp54542-scaffold833-694185 | 24 | ZBTB14, EPB41L3 |
| snp54574-scaffold833-1955034<br>snp54575-scaffold833-1991748<br>snp54576-scaffold833-2042027<br>snp54577-scaffold833-2084137 | 24 | PTPRM |

**Addendum H** A list of the identified genes and their functions by chromosome number (Ensembl; UniProt; NCBI).

| CHI | Gene | Molecular Function | Biological Process |
|---|---|---|---|
| 2 | TLK1 | • ATP-binding<br>• Protein serine/threonine kinase activity | • Intracellular protein transport<br>• Intracellular signal transduction<br>• Regulation of chromatin assembly or disassembly |
| 2 | GORASP2 | • None | • Organelle organization |
| 2 | CCDC141 | • Unknown | • Unknown |
| 6 | MTTP | • Ceramide 1-phosphate transfer activity<br>• Cholesterol transfer activity<br>• Lipid binding<br>• Phosphatidylcholine transfer activity<br>• Phosphatidylethanolamine transfer activity<br>• Protein heterodimerization activity<br>• Triglyceride transfer activity | • Cholesterol homeostasis<br>• Circadian rhythm<br>• Lipoprotein transport<br>• Low-density lipoprotein particle remodelling<br>• Plasma lipoprotein particle assembly<br>• Protein lipidation<br>• Protein secretion<br>• Triglyceride metabolic process<br>• Triglyceride transport |
| 6 | C4orf54 | • Unknown | • Unknown |
| 6 | TRMT10A | • tRNA (guanine(9)-N(1))-methyltransferase activity<br>• tRNA binding | • tRNA N1-guanine methylation |
| 6 | C4orf17 | • Unknown | • Unknown |
| 10 | DMXL2 | • Unknown | • Unknown |
| 10 | GLDN | • Protein binding involved in heterotypic cell-cell adhesion | • Clustering of voltage-gated sodium channels<br>• Microvillus organization |
| 10 | RAB8B | • GTP binding<br>• GTPase activity<br>• GDP binding<br>• Signalling receptor binding | • Antigen processing and presentation<br>• Protein import into peroxisome membrane |

| | | | |
|---|---|---|---|
| 10 | APH1B | • Endopeptidase activator activity<br>• Protein-macromolecule adaptor activity | • Amyloid-beta formation<br>• Notch receptor processing<br>• Notch signalling pathway<br>• Protein processing |
| 10 | CA12 | • Carbonate dehydratase activity<br>• Zinc ion binding | • Chloride ion homeostasis |
| 11 | OTOF | • Calcium ion binding | • Synaptic vesicle exocytosis |
| 11 | DRC1 | • None | • Axonemal dynein complex assembly<br>• Cilium-dependent cell motility<br>• Determination of left/right symmetry<br>• Heart development |
| 11 | SELENOI | • Phosphotransferase activity, for other substituted phosphate groups | • Glycerophospholipid biosynthetic process |
| 11 | HADHB | • Acetyl-CoA C-acetyltransferase activity<br>• lncRNA binding | • Cellular response to lipopolysaccharide<br>• Fatty acid beta-oxidation<br>• Gene expression |
| 11 | HADHA | • Enoyl-CoA hydratase activity<br>• Long-chain-3-hydroxyacyl-CoA dehydrogenase activity<br>• NAD+ binding | • Cardiolipin acyl-chain remodelling<br>• Fatty acid beta-oxidation<br>• Response to insulin |
| 11 | GAREM2 | • Unknown | • Unknown |
| 11 | RAB10 | • GDP binding<br>• GTP binding<br>• GTPase activity<br>• Myosin V binding | • Antigen processing and presentation<br>• Endoplasmic reticulum tubular network organization<br>• Endosomal transport<br>• Establishment of protein localization to endoplasmic reticulum membrane<br>• Regulated exocytosis |

| | | | |
|---|---|---|---|
| 12 | ZMYM5 | • Zinc ion binding | • Negative regulation of transcription by RNA polymerase II |
| 12 | ZMYM2 | • Ubiquitin conjugating enzyme binding<br>• Zinc ion binding | • None |
| 12 | GJA3 | • Gap junction hemi-channel activity | • Cell communication<br>• Gap junction-mediated intercellular transport<br>• Visual perception |
| 12 | GJB2 | • Calcium ion binding<br>• Gap junction channel activity involved in cell communication by electrical coupling<br>• Identical protein binding | • Cell-cell signalling<br>• Gap junction assembly<br>• Gap junction-mediated intercellular transport<br>• Sensory perception of sound |
| 12 | GJB6 | • Actin filament binding<br>• Beta-tubulin binding<br>• Gap junction channel activity involved in cell communication by electrical coupling<br>• Microtubule binding | • Ear morphogenesis<br>• Gap junction assembly<br>• Gap junction-mediated intercellular transport<br>• Sensory perception of sound |
| 16 | MMEL1 | • Metal ion binding<br>• Metalloendopeptidase activity | • None |
| 16 | PRXL2B | • Prostaglandin-F synthase activity | • Prostaglandin biosynthetic process |
| 18 | GCSH | • None | • Glycine decarboxylation via glycine cleavage system |
| 18 | C16orf46 | • Unknown | • Unknown |
| 18 | ATMIN | • DNA-binding transcription activator activity, RNA polymerase II-specific<br>• Dynein complex binding<br>• Transcription cis-regulatory region binding | • Motile cilium assembly<br>• Positive regulation of gene expression<br>• Positive regulation of non-motile cilium assembly |
| 18 | CENPN | • None | • Chromosome segregation |

| | | | |
|---|---|---|---|
| | | | • Kinetochore assembly |
| 18 | CMC2 | • Unknown | • Unknown |
| 18 | PKD1L2 | • Calcium ion binding<br>• Carbohydrate binding | • None |
| 18 | JPH3 | • None | • Exploration behaviour<br>• Learning<br>• Locomotion<br>• Memory<br>• Neuromuscular process controlling balance<br>• Regulation of neuronal synaptic plasticity<br>• Regulation of ryanodine-sensitive calcium-release channel activity |
| 18 | KLHDC4 | • Unknown | • Unknown |
| 18 | SLC7A5 | • L-leucine transmembrane transporter activity<br>• L-tryptophan transmembrane transporter activity<br>• Thyroid hormone transmembrane transporter activity | • L-leucine import across plasma membrane<br>• L-tryptophan transmembrane transport |
| 18 | CA5A | • Carbonate dehydratase activity<br>• Zinc ion binding | • None |
| 18 | BANP | • DNA binding | • Cell cycle<br>• Positive regulation of transcription, DNA-template |
| 18 | ABCC12 | • ABC-type transporter activity<br>• ATP binding | • None |
| 18 | FOSB | • DNA binding<br>• DNA-binding transcription factor activity | • Regulation of transcription by RNA polymerase II<br>• Cellular response to calcium ion |

| | | | |
|---|---|---|---|
| | | • RNA polymerase II cis-regulatory region sequence-specific DNA binding | |
| 18 | RTN2 | • None | • Gene expression<br>• Intracellular protein transmembrane transport<br>• Negative regulation of amyloid-beta formation<br>• Regulation of glucose import |
| 18 | PPM1N | • Magnesium ion binding<br>• Manganese ion binding<br>• Protein serine/threonine phosphatase activity | • None |
| 18 | VASP | • Actin binding<br>• Profilin binding<br>• SH3 domain binding | • Actin polymerization or depolymerisation<br>• Axon guidance<br>• Neural tube closure<br>• Positive regulation of actin filament polymerization<br>• Protein homotetramerization |
| 18 | OPA3 | • None | • Visual perception |
| 18 | GPR4 | • G protein-coupled receptor activity | • Adenylate cyclase-activating G protein-coupled receptor signalling pathway<br>• Angiogenesis involved in wound healing<br>• Glomerular mesangial cell development<br>• Negative regulation of angiogenesis<br>• Phospholipase C-activating G protein-coupled receptor signalling pathway<br>• Positive regulation of inflammatory response |

| | | | |
|---|---|---|---|
| | | | • Positive regulation of Rho protein signal transduction<br>• Regulation of cell adhesion<br>• Regulation of vascular permeability<br>• Response to acidic pH |
| 18 | EML2 | • Microtubule binding | • Negative regulation of microtubule polymerization<br>• Regulation of microtubule nucleation |
| 21 | DLK1 | • Calcium ion binding | • Negative regulation of Notch signalling pathway |
| 22 | GRM7 | • Adenylate cyclase inhibitor activity<br>• Group III metabotropic glutamate receptor activity<br>• Protein dimerization activity<br>• Serine binding | • Axon development<br>• Chemical synaptic transmission<br>• Glycosylation<br>• Sensory perception of sound |
| 24 | ZBTB14 | • DNA-binding transcription repressor activity, RNA polymerase II-specific<br>• RNA polymerase II cis-regulatory region sequence-specific DNA binding | • Cardiac septum development<br>• Coronary vasculature development<br>• Heart valve development<br>• Kidney development |
| 24 | EPB41L3 | • Actin binding<br>• Structural molecule activity | • Cortical actin cytoskeleton organization<br>• Regulation of cell growth |
| 24 | PTPRM | • Cadherin binding<br>• Identical protein binding<br>• Transmembrane receptor protein tyrosine phosphatase activity | • Homophilic cell adhesion via plasma membrane adhesion molecules<br>• Negative regulation of angiogenesis<br>• Negative regulation of endothelial cell migration |

| | |
|---|---|
| | • Negative regulation of endothelial cell proliferation<br>• Response to xenobiotic stimulus<br>• Retina layer formation<br>• Retinal ganglion cell axon guidance<br>• Signal transduction |

*Ensembl (https://www.ensembl.org/index.html, accessed 11 May 2022; Zerbino *et al.*, 2018); UniProt (https://www.uniprot.org/, accessed 11 May 2022); NCBI Genome data viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/, accessed 11 May 2022)