POPULATION GENETIC STRUCTURE OF SMALL HOLDER DAIRY CATTLE HERDS IN SOUTH AFRICA USING SNP MARKERS

BY

MAAKE MPHAPANTSI ELDRED

MINI-DISSERTATION

Submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE

In

AGRICULTURE (ANIMAL PRODUCTION)

In the

FACULTY OF SCIENCE AND AGRICULTURE

(School of Agriculture and Environmental Science)

at the

UNIVERSITY OF LIMPOPO

SUPERVISOR: Prof. C.B. Banga

CO-SUPERVISOR: Dr O Tada

2020

Declaration

I declared that the POPULATION GENETIC STRUCTURE OF SMALL HOLDER DAIRY CATTLE HERDS IN SOUTH AFRICA USING GENOME-WIDE SNP MARKERS (mini-dissertation) hereby submitted to the University of Limpopo, for the degree of Master of Science in Agriculture has not previously been submitted by me for a degree at this or any other university, that it is my work in design and in execution, and that all material contained herein has been duly acknowledged.

…… ……..

Signature Date

Acknowledgments

It’s by God’s grace that all was possible until this day. I would like to thank the Mighty Lord of Mount Zion for giving me strength and wisdom throughout this study.

As always, support comes from different spheres of life. I would like to express my deepest gratitude to Prof. Banga from the Agricultural Research Council – Animal Production (ARC - AP) for giving me an opportunity to work with him, I was indeed the chosen one among many. Special thanks to Ms. Sanarana for helping from the study design to data analysis, her generous nature made it easy and possible for me to learn more during the course of the study. Much appreciation to Dr. Tada, for accepting me and mentoring me as the student at the University of Limpopo.

To the researchers and my fellow colleguesat at ARC – AP, Animal Breeding and Genetics unit, thank you for your comments and inputs that have helped me to complete the study. I cannot forget Mr. Tshilate's for his assistance in laboratory work. Special thanks to ARC dairy research technicians your efforts contributed lots. Countless thanks to the smallholder dairy farmers of the Eastern Cape, Free State, Gauteng, KwaZulu-Natal, and North West for availing their animals for this study, “tsa lona di le lokele”. I would like to also express my gratitude to the University of Limpopo, Faculty of Science and Agriculture Executive Deans's office for the support. To Mr. Tlabela, Banga-drive memories will keep me sane forever, true brotherhood was forged there.

This study wouldn’t be a success without the financial support from the National Research Foundation (NRF), Agricultural Research Council (ARC) and the University of Limpopo (UL).

Dedication

This thesis is dedicated to my grandparents William & Mmamaropeng Ramatsoma, Mapula Chaisa Ramoshaba and my late grandfather Mapheya Maake. My parents Mohale Maake and Nyedisane Ramatsoma. My siblings Mokgadi, Masilo, Mapula and my nephew and niece Molate, and Amogelang Maake. To my brother Kalauba Aubrey Maake, I will forever be indebted to you. Lastly, I would like to dedicate this thesis to my unborn child (Kopano Mpho) and his mother.

Abstract

The smallholder dairy sector in South Africa is characterized by a low input production system and poor animal productivity. Research has been carried out to benchmark cow productivity on smallholder dairy herds; however, there is a paucity of information on the current status of breeding practices and the genetic consititution of cattle used in this production system. This information is vital for the development of sound and sustainable breeding programs for SHD production, which can have an enormous positive impact on food security and rural livelihoods. Thus, the aim of this study was to evaluate the levels of genetic diversity and population structure in South African smallholder dairy (SHD) herds using single nucleotide polymorphism (SNP) markers. A total of 192 animals from SHD dairy herds were genotyped using the GeneSeek® Genomic Profiler (GGP) 150K-BeadChip. Four specialized dairy breeds included the Ayrshire(n = 200), Holstein(n = 231), Jersey (n = 224) and Nguni (n = 209) were used as the reference populations. The mean MAF values ranged from 0.30 Ayshire (AYR), Jersey (JER), and Nguni (NGI) to 0.31 Holstein (HOL) and SHD between the populations. There were slight differences in the levels of genetic diversity ranged between 0.39 (JER and NGI) to 0.40 (AYR, HOL, and SHD). A moderate level of inbreeding (0.02) was observed in the SHD population, which results in high genetic diversity among this herds. Principal Component Analysis (PCA) revealed four homogeneous clusters comprising of AYR, HOL, JER, NGI, and a heterogeneous cluster of the SHD. The heterogeneity observed in the SHD population indicates widespread crossbreeding. The model-based cluster analysis corresponded with the PCA and pointed out the predominance of HOL, JER, with marginal gene flow from the AYR and NGI. These results have provided a useful insight into the genetic structure and prevailing breeding practices on South African SHD herds.

Keywords: Genetic diversity, PCA, Smallholder, South Africa

Table of Contents

CHAPTERS

LIST OF FIGURES

Figure……...Page

LIST OF TABLES

Table……Page

LIST OF ABBREVIATIONS

ARC	Agricultural Research Council
AYR	Ayrshire population
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FAO	Food and Agriculture Organisation
FAO-UN	Food and Agriculture Organisation of the United Nations
FIS	Inbreeding coefficient of individuals within a population
He	Expected heterozygosity
Ho	Observed heterozygosity
HOL	Holstein population
HWE	Hardy-Weinberg equilibrium
JER	Jersey population
MAF	Minor allele frequency
NGI	Nguni population
PCA	Principal Components Analysis
SA	South African
SADC	Southern African Development Community
SDS	Sodium Dodecyl Sulfate
SE	Standard Error
SHD	Smallholder Dairy herds
SNP	Single Nucleotide Polymorphism

STE	Sodium-Tris-EDT

CHAPTER 1: GENERAL INTRODUCTION

There is growing importance in livestock farming, in the smallholder sector of developing countries, driven by an unprecedented increase in the demand for livestock products (Delgado *et al.,* 2001). Global demand for dairy products is projected to increase by 22% by the year 2027 (FAO, 2018). Most of the escalation in milk production (80%) to meet this rise in demand is anticipated to emanate from developing countries (FAO, 2018). Smallholder dairy production has the potential to contribute significantly towards meeting this demand, while also promoting food security, financial security and creating employment in the entire dairy chain (Bennett *et al.,* 2006; Baiphethi & Jacobs, 2009). However, inadequate infrastructure, limited technical capacity and harsh environmental conditions result in poor animal productivity, this limits the exploitation of this production potential in this sector (Bennett *et al.,* 2006; Getachew, 2015).

A general lack of structured breeding programs is a major factor contributing to such impaired livestock productivity. This is exacerbated by poor access to well-adapted high-quality germplasm and systems for supporting sound breeding decisions or appropriate genetic improvement programs (Gorbach *et al*., 2010). Furthermore, most smallholder dairy farmers practice indiscriminate natural mating, using animals of unknown genetic value and generally do not have systems to inform their breeding decisions.

Systematic crossbreeding or knowledge of optimal admixture levels is essential for the improvement of livestock productivity in the smallholder sector of developing countries (Gibson, 2008). Knowledge of the performance of different genotypes in their specific environment is a major prerequisite to sound and systematic crossbreeding, as it will assist in the selection of the most suitable purebreds/crossbreds for improved performance. Insight on the existing genetic admixture levels and their performance, as well as information on the population structure, will form the basis for evaluation and selection of animal genotypes that will perform best in the smallholder environment. Breeding programs for the smallholder sector should focus on developing already existing ecotypes that are more productive and resilient to harsh

environments. For this reason, admixture analysis studies are increasingly being conducted in developed and developing countries (Gorbach *et al.,* 2010; Bray *et al.,* 2014; Strucken *et al.*, 2017; Mujubi *et al*., 2019). Information generated by admixture analysis studies of cattle breeds is useful when deciding the most optimal, for example, crossbreeding strategies to improve phenotypic performance by exploiting heterosis (Kelleher *et al*., 2017).

Recently, the use of genetic markers, particularly single necleotide polymorphisms (SNPs) in determining breed composition of cattle has attracted great interest (Mujubi et al., 2019). Previous studies have demonstrated the utility of SNP markers in providing highly reliable estimates of inbreeding, gene diversity, and levels of admixture in developing countries (Makina *et al.,* 2014; Strucken *et al.*, 2017; Mujubi *et al.*, 2019). This has provided guidelines for breed improvement, through appropriate utilization and conservation of livestock genetic resources.

The current study was carried out to investigate the population genetic structure and levels of admixture in the smallholder dairy cattle population of South Africa using the SNPs markers. This was an important step in generating information on the association between the different genetic groups of smallholder cattle, which will aid in the development of sound and sustainable breeding programs.

1.1 Problem statement

The South African smallholder dairy production system can easily be define by poor animal productivity and low input production (Mapekula *et al.,* 2010). The non-existence of genetic improvement programs and a lack of systems for supporting sound breeding decisions are some of the major factors contributing to impaired animal productivity in this production system (Abin *et al*., 2018). Due to paucity of knowledge of the performance of different breeds in this environment, and lack of genetic improvement programmes, smallholder farmers resort to poorly adapted exotic breeds or indiscriminate crossbreeding (Muntswu *et al.*, 2016).

This has led to the prevalence of genetically admixed animals on smallholder herds. There is, however, a general lack of knowledge on the performance of the different genotypes (i.e. admixture levels) in this environment. An evaluation of the different genotypes in the smallholder production system is a prerequisite to any efforts to develop appropriate breeding programs for this environment (Marshall *et al*., 2011).

As a result of sparse performance data and pedigree structure, the establishment of reference populations in smallholder dairy cattle populations remains the primary challenge to the determination of breed composition, diversity and estimation of genomic breeding values (Gorbach *et al.*, 2010; van Marle-Kőster *et al.*, 2015). This study will provide information that will help to guide breed improvement programs to meet current production needs in the smallholder dairy sector of South Africa.

1.2. The rationale of the study

Smallholder dairying has the potential to alleviate poverty, provide sustainable livelihoods and enhance household food and nutritional security (FAO, 2011). Currently, in South Africa, 27.6% of households with more than eight household members or with three or more children reported having inadequate access to food (StatsSA, 2019). Stats SA recommended subsistence (smallholder) farming as an important player in reducing the vulnerability to hunger in rural and urban food insecure households. Included in subsistence farming is smallholder dairy farming, which has huge room for job creation starting from milking cows to milk processing.

A major problem in the smallholder dairy production system is a mismatch between the available genotypes and the environment in which the animals perform (Chagunda *et al.,* 2018). This is exacerbated by poor performance and pedigree recording in smallholder dairy herds, which leads to a lack of knowledge on the degree of genetic variability and breed (Gorbach *et al.,* 2010). Currently there is lack of knowledge on population structure and admixture levels of smallholder dairy herds in South Africa. The current study was motivated by the need to determine the predominant genotypes available in the smallholder dairy cattle population, which will form the basis for their evaluation in this production system. Information from the study is expected to assist in decision making when designing and implementing breeding programs for the smallholder dairy sector.

1.2.1 Aim of the study

The primary aim of the study was to evaluate the population genetic structure, admixture levels and the prevalent genotypes in the smallholder dairy cattle population of South Africa using single nucleotide polymorphism (SNP) markers.

1.2.2 Study objectives

The objectives of the study were to:

I. Determine genetic diversity and level of inbreeding in smallholder dairy herds of South Africa
II. Determine population structure among smallholder dairy herds of south Africa
III. Determine the levels of admixture in smallholder dairy cattle herds of South Africa with reference to the established commercial dairy and the indigenous Nguni dual-purpose populations

1.2.3 Hypotheses

It was hypothesized that;

I. There is less of genetic diversity in smallholder dairy herds of South Africa
II. Smallholder dairy cattle in South Africa are clustered under one sub-population.
III. There is no admixture among smallholder dairy cattle of South Africa.

CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

Livestock breeding is an important agricultural sector worldwide, and it is linked to the historical, social, cultural, and climatic features per country or region (Sermyagin *et al.,* 2018). South Africa has its own diverse livestock breeds that have adapted to the prevailing environment and management conditions. The country has unique rich diversity when it comes to available livestock resources, vegetation, climatic regions and cultures (Van Marle-Koster & Visser, 2018). The South African livestock sector is composed of the highly developed commercial sector that reaps the benefits of modern technologies and a developing sector that includes emerging and smallholder farmers (van Marle-Koster and Visser, 2018).

The South African dairy industry is the fourth largest agricultural industry in the country, in terms of the value of agricultural production (MPO, 2016). Commercial farming and non-commercial (smallholder) farming constitute this industry. The number of non-commercial farms has declined consistently since 1997. The decrease in smallholder farms might be due to a lack of support services, such as extension and inadequate infrastructure. The smallholder dairy sector in South Africa has the potential to contribute to food security and household income. Poor cow productivity on smallholder dairy herds is, however, a major concern (Abin *et al*., 2018). Smallholder dairy farmers in South Africa have generally not adopted technologies to enhance cow productivity, such as artificial insemination and performance recording (Muntswu *et al*., 2016). This might be due to the fact that most of the farmers rear animals for savings and insurance, which means that maintaining large livestock numbers is more important than increasing animal productivity (Marshall *et al*., 2019). A major constraint to poor cow productivity in the smallholder sector of South Africa is a lack of breeding programmes. Farmers lack knowledge of the appropriate genotypes to increase profitability under their conditions, ofcoarse majority of them cannot afford appropriate genotypes.

The development of breeding programs in the smallholder dairy sector of the developing world has been difficult to implement because of poor pedigree and performance recording (Gorbach *et al*., 2010; Ojango *et al*., 2014; Changuda *et al*., 2018). Pedigree data has been the main source of information for determining breed

composition. The development of molecular technologies such as dense SNP markers and reduced costs of genotyping/DNA sequencing offers an alternative path to developing breeding programmes in the smallholder farming sector (Meuwissen *et al*., 2016; Mujibi *et al*., 2019). Thus, the use of genetic markers in determining breed composition of livestock has gained much interest in recent years, especially in developing countries where there is a general lack or incomplete pedigree records (Gorbach *et al*., 2010; Changuda *et al*., 2018; Ahmed *et al*., 2019). These markers provide knowledge on breed composition, which is primarily essential for systematic crossbreeding. This review discusses the smallholder dairy production system and the opportunities for developing breeding programs for South African smallholder dairy cattle using genomic technology.

2.2. The South African (SA) smallholder dairy sector

Almost 15.6% of South African households are involved in smallholder farming to supplement food for their households (StatsSA, 2019). South African smallholder agriculture has been identified as a vehicle through which the goals of poverty reduction and rural development can be achieved (Pienaar & Traup, 2015). Currently, there are about 1,3 million smallholder livestock farmers and 67% of these are stagnant in terms of progressing to emerging commercial operations (DAFF, 2017b; van Marle-Koster & Visser, 2018). The stagnant is cause by poor government policies and poor support from government (Chikazunga and Paradza, 2012). Smallholder farmers generally lag behind in the adoption of modern technologies, the major reason being lack of skills and infrastructure, as well as the cost of adopting such technology.

In the South African context, a smallholder dairy herd (SDH) may be defined as a farm that produces less than 500 litres of milk a day, irrespective of the number of cows or size of the farm (Manzana *et al*., 2014). Smallholder farmers make a small contribution to the mainstream dairy industry in South Africa. This contrasts sharply with other African countries such as Ethiopia, Kenya and Tanzania where 98%, 80% and 99% respectively, of the milk sold, is produced on smallholder herds (Swai & Karimuribo, 2011; Bereda *et al.,* 2013; Odero-Waitituh, 2017). Almost all the milk produced by SDH does not enter the South African commercial market. It is mainly consumed at household level or sold in the immediate vicinity (MilkSA, 2017). The majority of cows produce less than 10 litres a day and the average herd size is less than 15 cows. Poor

cow productivity on SHD herds may be attributable, to a large extent, to the non-existence of genetic improvement programs and lack of systems to support sound breeding decisions (Muntswu et al., 2017).

2.3. Breeds and breeding management on smallholder dairy herds

The majority of smallholder dairy farmers in South Africa (nearly 100%) use natural service (Muntswu *et al.,* 2017), contrary to the commercial dairy sector which predominantly uses artificial insemination. The general use of natural service might be due to lack of skills, poor infrastructure and facilities required to apply artificial insemination. The major dairy breeds used in the commercial sector are Holstein, Jersey, Ayrshire and Guernsey (Banga, 2009). SDH, on the other hand, comprise mainly of crossbreds (75%), followed by Holstein (21%) and Jersey (4%) (Muntswu *et al.,* 2017). These mostly cross between indigenous and exotic breeds, and in many cases, they are not specifically bred for milk production (Mapekula *et al.,* 2011; Tanyanyiwa *et al.,* 2017). They are not bred for milk production, because some SHD farmers rear animals for savings and insurance, not for milk production *per se*. Some farmers, however, do keep improved breeds with an idea that a successful dairy enterprise should use improved breed types. Thus, leading SHD farmers to buy exotic breeds and cross them with indigenous breeds to increase survivability, milk production, and adaptation of their cattle. Their crossbreeding is, however, indiscriminate, due to poor or lack of performance and pedigree recording (Gorbach *et al*., 2010). Reliable pedigree and phenotypic data are essential for genetic improvement. Such information is either poor or not available on SHD herds, meaning that there are no breeding programs. Commercial farmers, on the other hand, use estimated breeding values (EBV) to select superior cows and bulls for breeding (van Marle-Koster and Visser, 2018).

2.4. Characteristics of the major breeds used in South African smallholder dairy herds

2.4.1. Crossbreds

South African indigenous cattle like the Nguni and Afrikaner breeds have poor milk productivity, hence the need for crossbreeding with specialized dairy breeds. In South Africa, smallholder dairy cattle have only been characterized phenotypically (Grobler

et al., 2008; Tanyanyiwa *et al*., 2017; Muntswu *et al*., 2017). The levels of admixture and prevailing genotypes in the South African SHD system is unreported. There is also limited literature on this production system.

Figure 2.1 Crossbred bull from smallholder dairy farm in the Free State province, capture during data collection

2.4.2 Holstein

The Holstein is perhaps the most recognized breed of dairy cattle and the most common dairy breed in South Africa. Holstein cows have distinctive black and white or red and white markings (Prendiville *et al*., 2011). The breed is known for high milk production but has less butterfat and protein-based on percentage in the milk, compared to other breeds (Horan *et al*., 2004). Holstein cows originated in the Netherlands approximately 2,000 years ago (Lopreiato *et al*., 2019). Two breeds of cattle, black animals from the Batavians (present-day Germany) and white animals from Friesland (present-day Holland), were crossed to create a new breed of cattle. Originally, the breed was known as Holstein-Friesian but is now known more simply as Holstein (Prendiville *et al*., 2011). Friesian cattle still exist today but are separate

from the Holstein breed. There are Friesian breeds from the United Kingdom, New Zealand, and Holland and these animals tend to be smaller bodied than Holstein cattle.

Figure 2.2 Holstein cow from smallholder dairy herd in the Eastern Cape, captured during data collection

2.4.3 Jersey

The Jersey breed was developed on Jersey Island, one of a series of the small Channel Islands in the channel between England and France, just off the coast of Normandy, France (www.jerseycanada.com). It is rumoured that some of the foundation genetics for the Jersey breed came from Africa. This theory can best be explained by Jersey's strong tolerance to heat and high humidity conditions. It is a small breed that is fawn brown in colour. Because of their colour and the shape of their eyes, Jersey cows are often described as "deer-like".

Figure 2.3 Jersey cow from smallholder dairy herd in Free State, captured during data collection

On Jersey Island the dairy rations were primarily forage-based, thus requiring a cow that could efficiently convert grasses and legumes into milk and milk solids (Prendiville *et al*., 2011). Jersey owners placed emphasis on developing a breed of cows with very high solid levels in milk (www.jerseycanada.com). This selection over generations has created a cow with extraordinary levels of butterfat relative to the other common breeds of dairy cattle today (Capper and Cady, 2012). For much of the first six decades of the 20th century, Jersey Island was the source of breeding stock to start Jersey populations all over the globe. The breed has been particularly noteworthy in New Zealand, Australia, Denmark, the United States, South Africa, Great Britain, and Canada (Prendiville *et al*., 2011).. There are no available records indicating when the Jersey breed was brought to South Africa. However, the first Jerseys were imported by Mr. Adrian van der Byl of Roodebloem Estate, Woodstock, Cape, from Jersey Island, in the early 1880s, with 1881 as the most probable date (www.jerseysa.co.za).

2.4.4 Nguni Breed

Figure 2.4 Nguni bull from Eastern Cape Smallholder dairy herd, captured during data collection

The Nguni breed is a well-known transboundary, indigenous Southern African cattle breed with a small to medium frame size, which is highly dependent on the nutritional condition (Scholtz, 2005). The breed can better be identified by its unicoloured or multi-coloured (black, brown, white, red, grey and black and tanor brindle) coat. The breed's ability to adapt to harsh environmental conditions makes it the number one choice for smallholder dairy herds in poor grazing farmlands and parasite infested areas (Mapholi et al., 2014). Many cattle farmers grew interested in the Nguni breed due to its ability to produce and reproduce under harsh environmental conditions, their natural immunity against endemic diseases (Mapholi *et al*., 2014) and its ability as a dam line in a terminal crossbreeding (Scholtz, 1988).

2.5. Productivity of smallholder dairy herds in South Africa

The productivity of smallholder dairy cows is low (4 093 kg of milk per 305days) compared to their commercial sector counterparts which produce 6 921 kg of milk per

305 days of lactation (Abin *et al*., 2018). Cows in the commercial sector produce more milk than those on smallholder herds because they optimise production performance of their cows, which is a pre-requisite for profitable and sustainable farming. Abin *et al*. 2018 found that commercial production system cows produced 40.8, 41.7, and 42.5%, more milk, fat, and protein (kg), respectively, than those in the smallholder system. Poor pedigree and performance records and the absence of genetic evaluation and improvement programs contribute to impaired cow productivity in the smallholder production system.

2.6. Genomic tools

Research has been conducted on the production performance and milk quality of smallholder dairy cattle of South Africa, in comparison to their commercial counterparts (Abin *et al*., 2018). There is, however, limited comprehensive research on genetic aspects of these cattle. A better understanding of genetic characteristics of cattle on smallholder dairy herds can be achieved through the use of genomic tools. Recent developments in molecular genetics and bioinformatics, such as whole genome sequencing technology, have enabled the development of genome wide single nucleotide polymorphism (SNP) arrays for many livestock including cattle (Bovine Consortium *et al.,* 2009). This has identified more than ten million SNPs which could explain a high percentage of phenotypic variation in cattle (Makina *et al.,* 2016). The availability of these massive millions of SNP markers has resulted in the development of the two initial genome assemblies. Today various commercial SNP bead chips are available for cattle through three leading companies (AffymetrixTM, Illumina®, Neogen's GeneSeek®) (Nicolazzi *et al*., 2015). The new assemblies will aid in improving genome continuity, remapping reads, and improving marker order, which might influence SNP selection for the development of SNP genotyping platforms in the future (Lashmar *et al*., 2019). The commercial bead chips that are currently available for cattle are summarized in Table 1, adapted from Nicolazzi *et al*., (2015).

Table 2.1 List of available SNP BeadChip panels for cattle.

Company	Beadchip	Number of SNPs
Affymatrix®	Axiom® Genome Bos1	648 875
Geneseek®	Geneseek Dairy Ultra LD V2 GGP-LD	7 049
	➢ Version 1 (GGP9K)	8 610
	➢ Version 2 (GGP20K)	19 721
	➢ Version 3	26 151
	GGP-indicus	35 090
	GGP-HD	76 879
	GGP-150K	139 480
Illumina®	Golden Gate Bovine 3K	2 900
	Bovine LD	
	➢ Version 1	6 909
	➢ Version 1.1	6 912
	➢ Version 2	7 931
	Bovine SNP50	
	➢ Version 1	54 001
	➢ Version 2	54 609
	Bovine HD	777 962

These chip panels allow simultaneous high throughput interrogation of large numbers of loci with high measurement precision (Matukumalli *et al.,* 2009). This presents an opportunity to study South African smallholder dairy cattle in order to establish their population structure, as well as determine their genetic make-up.

2.7. Utilization of SNP markers

Genetic improvement of cattle was, in the past, previously based on quantitative analysis of performance and pedigree data, and microsatellite markers were applied mainly for population genetics (Sanarana *et al.*, 2016; Madilindi *et al.,* 2018). Microsatellite markers have also been useful to identify quantitative trait loci (QTL) with effects on several economically important traits in cattle (Boichard *et al.,* 2003; Casas *et al.,* 2003; Hu *et al.,* 2006). Genotyping for microsatellite markers is labour-intensive and allele calls are laboratory-specific (Williams *et al.,* 2009) and these anonymous markers provide no information on the genes underlying QTL. In recent years, single nucleotide polymorphisms (SNPs) which are more dense and abundant than microsatellites, occurring at a frequency of about one SNP per kb in humans and about one SNP per 500 base pairs (bp) in mice (Lindblad-Toh *et al.,* 2000) and cattle (Heaton *et al.,* 2001) have became popular. A single-nucleotide polymorphism (SNP) is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is commonly present within a population (e.g. > 1 %) (Kumar *et al*., 2019). Despite being bi-allelic and so having a lower information content than microsatellite markers, the availability of high throughput SNP genotyping platforms makes it feasible to undertake high-density scans using large numbers of SNP markers (Wiggans., *et al* 2009).

In South Africa, the utility of the Bovine SNP sets was examined by Qwabe *et al.* (2013). With the primary findings that 56% of the 54 609 called SNPs from the bovine SNP50 beadchip were polymorphic among the 91 cattle belonging to four cattle breeds, with an average minor allele frequency of 0,23 across the entire set. It was then concluded that the Bovine SNP50K set array is applicable in South African cattle populations, provided that the DNA quality meets the required quality of infinium assay. It was further concluded that Bovine SNP array will be useful for genomic studies across Angus, Holstein, Nguni, Bonsmara, Drakenberger and Afrikaner cattle that are widely used in South Africa for dairy and beef production.

2.7.1. Application of genomic tools in genetic diversity and population structure studies

Knowledge about genetic diversity and population structure is useful for designing effective strategies to improve the production, management, and conservation of farm animal genetic resources (Edea *et al*., 2014). This is particularly useful in a production system where breeding management and strategies do not exist. Genetic diversity can be referred to as the variation in the amount of genetic information within and among individuals of a population or species (www.biodiversity.org). There are several statistical approaches to studying genetic diversity and population structure. It is normally measured by the frequency of genotypes and alleles, the proportion of polymorphic loci, and the observed and expected heterozygosity (Nei, 1973). To measure diversity within populations, the expected diversity or gene diversity is the most widely used parameter. Other measures of genetic diversity include allelic diversity (number of alleles segregating in the population) (Toro *et al.,* 2009). A high number of alleles implies more genetic variation (Nei, 1973). When using allelic diversity, which depends largely on the sample size of the population, it is important to sample population that is equal, because the detected alleles may increase with increased population size (Toro *et al.,* 2009).

Previous studies on genetic diversity, inbreeding and population structure of South African cattle from 16 breeds have been carried out using SNPs, since the inception of the Beef Genomics Project (BGP) and Dairy Genomics Project (DGP) in 2015, 2016, respectively (van Marle-koster and Visser, 2018). There is no previous study on genetic diversity and population structure, to determine the prevailing or genetic make-up of the animals in the smallholder dairy population of South Africa.

Population structure can be determined using the ADMIXTURE software (Alexander *et al.,* 2009) which implements a model-based clustering method for inferring population structure from genotypic data. The software has the ability to assign individuals to populations and assumes a model in which there are K populations, where each population is characterized by a set of allele frequencies at each locus. The ADMIXTURE software is usually used to assign individuals correctly to a population, especially when the phenotypic differentiation between breeds/populations is difficult to detect or when genealogical data is absent (Alexander *et al.,* 2009).

Makina *et al*. (2014) observed some level of admixture among the indigenous and locally-developed South African breeds and supported the clustering of the breeds according to their history of origin. It was found that 5% of SA Nguni cattle were admixed with the Afrikaner breed, while 5% of Drakensberger cattle showed signs of admixture with Nguni, Bonsmara, and Angus. Information of this nature can assist in preserving genetic diversity, improving and developing breeding programs (Alexander *et al.,* 2009; Gorbach *et al.,* 2010; Makina *et al.,* 2014).

2.7.2. Measures of population genetic structure

The parameters normally used to define population genetic structure are observed and expected heterozygosity (Ho and He, respectively), genetic distance (D), amount of structuring between subpopulations (FST), and gene flow (Nem, where m is the migration rate).

2.7.2.1 Heterozygosity

Mean heterozygosity, calculated across a number of loci, is a valuable parameter used to estimate the degree of genetic variation within a population. Population structuring occurs when genotype frequencies deviate from Hardy–Weinberg expected proportions (Groeneveld *et al*., 2010). If either inbreeding or selection occurs, then populations can be considered “structured” in some way.

2.7.2.2 Genetic Distance

When two populations are genetically isolated, both mutation and genetic drift lead to differentiation in the allele frequencies at selectively neutral loci (Dash *et al*., 2019). As the amount of time that two populations are separated increases, the difference in allele frequencies between them should also increase, until each population is completely fixed for separate alleles (Kelleher *et al*., 2017). Therefore, calculation of genetic distance (D) between two populations provides a relative estimate of the time elapsed since these populations have existed as a single random mating unit (Scutari *et al*., 2016). Small estimations of distance among completely isolated populations indicate that they have only been separated for a short period of time (Dash *et al*., 2019). Alternatively, in the absence of isolation, small values of genetic distance may indicate population structure (i.e., subpopulations in which there is random mating, but between which there is a reduced amount of gene flow).

2.7.2.3 F-Statistics

F-statistics, developed by Wright (1965), represent the basic method to measure the amount of subdivision in a population. They can be viewed as a measure of the correlation of alleles within individuals, and they are related to inbreeding coefficients (Kelleher *et al*., 2017). An inbreeding coefficient is a measure of the non-random association of alleles within an individual. As such, F-statistics describe the amount of inbreeding-like effects within subpopulations, among subpopulations, and within the entire population. In particular, the FST index (or RST, as estimated for microsatellite data) is an estimator of the amount of structuring of a population into subpopulations.

Fst is one of the most commonly used metrics for detecting signatures of selection in animals (Maiorano et al., 2018). Researchers use Fst as a tool for identifying patterns of genetic variation at a locus among populations relative to that within populations (Pintus *et al*., 2013; Maiorano *et al*., 2018). The fixation index (Fst) is an estimate of population differentiation, based on genetic polymorphism data, and it is calculated using the relationship between inbreeding and heterozygosity (Pintus *et al*., 2013)

2.7.2.4 Migration

If there is no migration (gene flow) occurring between two populations or demes, eventually alternate alleles will become fixed and will reach 1 (Dash *et al*., 2019). Alternatively, it has long been known that if migration, measured in terms of Nem, is >1 (where Ne is the effective population size and m is the proportion of migrants per generation or migration rate), the allele frequencies in the subpopulations remain homogenized (Wright, 1931). If, however, migration is present but Nem < 1, an equilibrium based on the rate of mutation, migration, and genetic drift will be established.

2.7.2.5 Phylogeography

Recently, a relatively new discipline named phylogeography has been applied to investigate the principles and processes governing the geographic distributions of genealogical lineages within and among closely related extant species (Kelleher *et al*., 2017). Phylogeographic studies focus on understanding the contribution of historical versus contemporary ecological processes in shaping present-day species distributions. Phylogeographic inferences are based on DNA sequences sampled from

the same locus in many individuals collected throughout the geographic range of a species. Statistical analyses are based on coalescence theory that employs a sample of individuals from a population to trace all alleles of a gene shared by all members of the population to a single-ancestral copy (Pintus *et al*., 2013). This uses sophisticated model-driven approaches that answer specific questions for inferring population history. Such studies can provide substantially new insights into the processes responsible for shaping the spatial patterns of genetic variation within and among populations as well as their distributions.

2.8. Importance of population structure and genetic diversity

Population genetic structure refers to any pattern in the genetic makeup of individuals within a population (Ojango *et al*., 2014) . It allows for information about an individual to be inferred from other members of the same population. Sbordoni *et al. (*2010) described population structure as a fundamental guideline to understanding the evolution of animals, simply because it represents the outcome of history and adaptation to their environment. Therefore, genetic diversity and population structure studies can be used to identify genomic regions that have adaptive and productive significance in admixed populations.

Genetic diversity and population structure are thus two important aspects of defining any livestock population (Kumar *et al*., 2019). They help in genetic improvement through the manipulation of breeding plans based on existing diversity, aimed at improving the adaptation of these populations to local environmental conditions (Groeneveld *et al*., 2010). Therefore, population structure and characterisation can assist in identifying the prevailing genotypes in the smallholder dairy sector (Edea *et al*., 2014). This can further assist in conservation of unique characteristics within this system.

It is important to investigate the levels of genetic diversity of a population, as genetic diversity represents the raw material essential for breeding and has practical implications for implementation of genomic selection A number of studies have been carried out in African countries to determine genetic diversity, structure and level of admixture (Gorbarch *et al*., 2010; Ojango *et al*., 2014). Mujibi *et al*. (2019), were able, with the aid of genomic data, to estimate the breeding value of smallholder dairy herds and implemented genetic improvement program without pedigree information. They

found that smallholder dairy breed types with exotic blood between 75 and 85% are the most appropriate genotypes in Tanzanian environment. Ojango *et al*. (2019) further predicted Genomic Estimated breeding value of smallholder system using the G matrix in the absence of pedigree information. This was examined in Tanzania and Kenya, where Kim & Rothschild, (2014) reported that smaller farms in that region use admixed populations of Holstein-Friesian, Norwegian Red (or Ayrshire), and Guernsey cattle. Therefore, Genomic selection can be useful in the improvement of milk production and milk components in smallholder dairy sector of South Africa.

2.9. Importance of population admixture analysis

Admixture remains the only form of gene flow between populations of different ancestry. It can be defined as the process whereby two or more genetically and phenotypically distant populations with diverse allele frequencies copulate and breed new offspring, called a mixed or hybrid population (Shriver *et al.,* 2003; Ding *et al.,* 2011). A classic example of a commercial admixed population is the famous Brahman beef breed population, which was produced by cross-breeding the Kankrej cattle population and Guzerat, Ongole, Gir, and Krishna Valley cattle populations (Bonsma, 1980).

In developing countries, smallholder dairy farming is mainly based on the use of crossbred cows that combine local adaptation traits of indigenous breeds with the high milk yield potential of exotic dairy breeds (Strucken *et al.,* 2017). In small farms of Africa, cattle have been maintained by crossbreeding to increase survivability under severe environmental conditions (Kim & Rothschild, 2014). Thus, genetic variation between breeds for most quantitative traits manifest opportunities to combine breeds in order to improve productivity (Van Vleck *et al*., 1986). This also offers the opportunity to increase the adaptability of cattle from other geographic regions (Kim & Rothschild, 2014). However, indiscriminate crossbreeding in these regions produces highly admixed animals with large variability in production (Ojango *et al*., 2014).

Population admixture analysis studies are gaining popularity in smallholder dairy systems of developing regions, and have been conducted in countries such as Brazil, Ethiopia, India, Kenya and Tanzania (Ojango *et al*., 2014; Panetto *et al*., 2017; Strucken 2017; Mujibi *et al*, 2019). Knowledge of admixture levels in crossbred populations, along with the information on population structure, is immensely important

in the genetic improvement of livestock populations. In Tanzania, Mujibi *et al*. (2019) were able to identify the breed composition that is most appropriate for the majority of smallholder farms. This demostrated that farmers who rely on the intensive feeding system are best suited to a breed comprising 75% of the Holstein breed. Hence, a baseline information that will allow farmers to plan their crossbreeding on the level of known exotic genetics existing in their farming systems.

Admixture analysis, therefore, allows the identification of exact breed composition in animals, and this can be associated with the productivity of an individual animal. Thus, appropriate recommendations can be made to farmers and other stakeholders interested in maximizing animal productivity through the matching of environmental conditions with appropriate genotypes. This highlights the importance of carrying out such studies for the South African smallholder dairy system.

Crossbreeding of indigenous cattle with exotic breeds has created a new unaccounted population that makes-up most of the dairy cattle in the smallholder dairy sector of developing countries today (Blench & MacDonald, 2000; Strucken *et al.*, 2017). This has been necessitated by increased demand for milk, fostering a new wave of crossbreeding in Africa (Strucken *et al*., 2017). The exotic breeds from North America and Europe, are known for their high milk production capability. Under appropriate conditions, crossbreeding and the use of crossbred cattle can yield significant increases in smallholder income. There are, however, no genetic improvement programs to facilitate crossbreeding in this production system. Identification of breed composition and association with individual productivity is a pre-requisite to the establishment of genetic evaluation programs. Recently, Ojango *et al.* (2014) evaluated crosses between indigenous cattle and exotic dairy breeds such as Holstein, Friesian, Ayrshire, and Jersey in East Africa and recommended the best levels of admixture for the smallholder dairy production system

2.10. Conclusion

There is room for improving the productivity of the smallholder dairy production system in South Africa (Abin *et al.,* 2018). However, in-depth genetic information on the population structure and prevailing genotypes in the production system is not available. Such information is a prerequisite to designing programmes to optimise the utilisation of genetic resources, such as selection, crossbreeding, breed improvement

and conservation. Molecular genetics techniques, in conjunction with conventional animal breeding methods, could be used to design such programmes, which can result in genetic gains.The availability of genomic tools presents an opportunity to study smallholder dairy cattle breeds, at the genomic level, in order to determine the population structure and prevailing genotypes. Population genetic structure and admixture analysis of smallholder dairy cattle herds in South Africa is thus the first step towards developing sound genetic improvement programs.

CHAPTER 3: MATERIALS AND METHODS

3.1. Study cohorts

Figure 3.1 *Map of South Africa showing the location of smallholder herds comprising the study population*

All farms that were targeted were those that participate in the National Dairy Animal Recording and Improvement Scheme of the Agricultural Research Council (ARC). The study population comprised of smallholder dairy herds from five South African provinces including the Eastern Cape, Gauteng, Free State, KwaZulu-Natal, and North West.

3.2. Sample collection

A total of 192 unrelated animals (males = 19, females = 173) were selected from the smallholder herds. Farmers were interviewed about the nature of breed types that they used for breeding and their responses are summarized in Table 3.1.

Table 3.1 Animal breeds sampled from SA smallholder farms as defined by the farmers

Breed	Females	Males	Total
Holstein	92	22	114
Jersey	33	17	50
Nguni	7	0	7
Unknown	17	4	21

3.3. Sample collection and DNA extraction

Following the breed types identified by farmers in their herds, animals were restrained in a crush pen or milking parlour for hair sampling. About 30 to 40 hair samples were plucked from the tail and placed into an envelope. The samples were sent to the ARC Biotechnology Platform for genotype processing.

DNA was extracted from the hair samples at the ARC's Biotechnology Platform, using the Chemegen DNA extraction kit, according to the manufacturer's purification protocol (Chemegen, 2016). The protocol was adapted for hair samples and sodium chloride-tris EDTA (STE) was added together with sodium dodecyl sulfate (SDS) and Proteinase K to digest the hair follicles. The samples were further incubated at 56°C for 4 hours 30 minutes until lysis was complete (Qwabe *et al.* 2013).. The integrity of the genomic DNA was quantified using both the Quibit® 2.0 Fluorometer and the Nanodrop spectrophotometer. The concentration of the DNA was diluted whenever it was highly concentrated (>150 ng).

3.4. SNP Genotyping

The extracted DNA of individuals was genotyped using the GeneSeek® Genomic Profiler (GGP) 150K- BeadChip at the ARC's Biotechnology Platform. Genotyping was performed using the standard Infinium array protocol, which features 141 722 SNP probes distributed across the whole bovine genome (Illumina, Inc. San Diego, CA, USA). Approximately 10 µl of DNA was loaded into each well of the Beadchip for genotyping. Each sample was whole-genome amplified for 20 hours at 37°C. The samples were then fragmented, precipitated and re-suspended in an appropriate hybridization buffer. The samples were hybridized on the prepared GGP 150K Beadchip for 20 hours at 48°C. Following the hybridization, non-specifically hybridized samples were removed by washing, while the remaining specifically hybridized loci were processed for the single base extension reaction, stained and imaged on an Illumina iScan Reader (Qwabe *et al*. 2013).. Genotypic data generated from the iScan system were loaded into the Illumina Genome studio version 1.9.0 software, which uses algorithms to perform primary data analysis including raw data normalization, clustering, and genotype calling. A final custom report with genotype information of all the 192 animals was created from the genome studio using PLINK input report 2.1.1, which created a Ped (Pedigree file) and Map (SNP panel file) file for downstream analyses.

3.5. Genotypic data and quality control (QC)

A number of 655 animals representing the major commercial dairy breeds, included Holstein (n = 231), Jersey (n = 224), Ayrshire (n = 200) and Nguni (n = 203) were used as reference populations for the determination of genetic diversity, population structure and admixture analysis with the smallholder dairy herd dataset. These were genotyped with the 50K bovine SNP BeadChip consisting of 54 609 SNPs. The quality control and data editing was performed across all populations and summarized in Table 3.2 using PLINK software (Purcell *et al*., 2007). The datasets of the 50K and 150K SNP panels were merged. The Nguni dataset was then used as the common denominator and only autosomal SNPs 38 446 SNPs were in common for further analysis. Basic genotype statistics were performed to eliminate animals with low call rate (<0.90) and SNPs with <0.1 minor allele frequency (MAF) and those that deviated from the Hardy-Weinberg equilibrium (HWE).

Table 3.2 Quality control and filtration summary of five datasets.

Population	No. samples	Samples < 0.90 callrate	No.SNPs < 0.95 callrate	No.SNPs <0.1 MAF	HWE <0.001	SNPs remained	Genotyping rate (%)
Ayrshire (AYR)	200	0	68	9254	104	29 020	100
Holstein (HST)	222	9	306	6 364	177	31 599	95.9
Jersey (JER)	222	2	221	11 932	175	26 118	99.1
Nguni (NGU)	203	7	217	12 746	779	24 704	96.6
Smallholder (SHD)	189	3	200	4565	245	33 436	98.4

Furthermore, a total of 1 035 SNPs with high linkage disequilibrium (LD) were pruned following the Plink command ----indep-pairwise 50 5 0.5. This was done to eliminate any effects that might be caused by the ascertainment bias between the populations.

3.6. Genetic diversity

The allele frequencies were used to estimate the level of genetic diversity parameters per population. The expected heterozygosity (He), observed (Ho) heterozygosity, and the inbreeding coefficient (*Fis*). The mean MAF was estimated across all the breeds from the allele frequencies.

3.7. Population structure and Admixture analysis

The Principal Component Analysis (PCA) was used to investigate the level of relatedness of the populations using the eigenvalues and eigenvectors constructed from the Genome-wide Complex Trait Analysis (GCTA) software (Yang *et al*., 2011). The PCA plots were visualized as PC1 vs 2 and 1 vs 3 on Microsoft excel 2016. Furthermore, Admixture 1.3.0 software (Alexander *et al*., 2009) was used to investigate the population structure of the smallholder dairy sector in South Africa. The population structure was evaluated using model-based clustering, ADMIXTURE software. This uses the cross-validation (CV) error to guide the selection of distinct ancestries best supported by the data based on the probable *K* value which is the number of clusters. The preferable *K* is the one that exhibits a low cross-validation error in comparison to the other *K* values. The low cross-validation error (0.62) was detected at *K* = 5 as shown in Figure 4.3. This was used to determine the number of clusters as five populations.

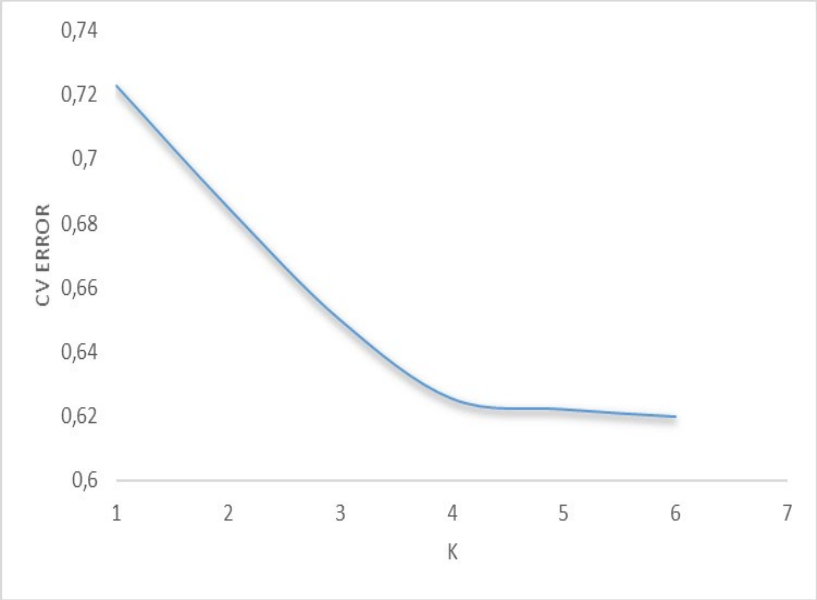

Figure 3.2 Cross-validation plot for five populations presented in this study (*K* = 5)

CHAPTER 4: RESULTS

4.1. Genetic diversity

The average genotype call rate across the breeds was 98%, this ranged from 95.9 (HOL) to 100% (AYR). After the SNP quality control, the SHD population retained more SNPs followed by the HOL than AYR, JER, and NGI. Hence, SHD and HOL had slightly high MAF (0.31) compared to AYR, JER, and NGI (0.30). When all the quality-controlled datasets of SHD with the reference populations were merged and pruned, only 13 891 SNPs remained with no general pattern of where the missing genotypes occurred along the genome.

The genetic diversity was estimated separately for each dataset as shown in Table 4.1. The results revealed slight differences in genetic diversity between the populations. The genetic diversity, as measured by H_O ranged from 0.39 for NGI and SHD to 0.40 for AYR, HOL, and JER. Thus, slightly high gene diversity was observed in the SHD, AYR, and HOL. Also, H_O was slightly high than the H_E in the SHD. The average inbreeding coefficient (F_{IS}) did not show any substantial average inbreeding in any of the breeds, it ranged from -0.004 (NGI) to 0.02 (SHD)

Table 4.1 Genetic diversity and the inbreeding of the smallholder and other dairy populations in South Africa

Population	N	MAF	H_O	H_E	F_{is}
SHD	189	0.31(0.11)	0.39(0.10)	0.40(0.09)	0.02(0.07)
AYR	200	0.30(0.12)	040(0.10)	0.40(0.10)	-0.02(0.03)
HOL	222	0.31(0.09)	0.40(010)	0.40(0.09)	-0.01(0.04)
JER	222	0.30(0.12)	0.40(0.10)	0.39(0.10)	-0.02(0.05)
NGI	202	0.30(0.12)	0.39(0.10)	0.39(0.10)	-0.004(0.04)

N:number of animals, MAF:minor allele frequency, H_O: observed heterozygosity, H_E: expected heterozygosity, F_{is}: inbreeding coefficient, SHD: smallholder dairy herds, AYR: Ayshire, HOL: Holstein, JER: Jersey, NGI: Nguni

4.2. Principal component analysis (PCA)

The PCA based on combined datasets is shown in Figure 4.1. The plot separated the NGI cattle breed from the dairy breeds. The NGI population formed a distinct cluster, away from the rest of other populations, with slight dispersion of the SHD population. The proportion of the SHD population mainly dispersed between the HOL and JER populations. There was a tight cluster between the HOL and the AYR populations with large significant number of SHD population individuals.

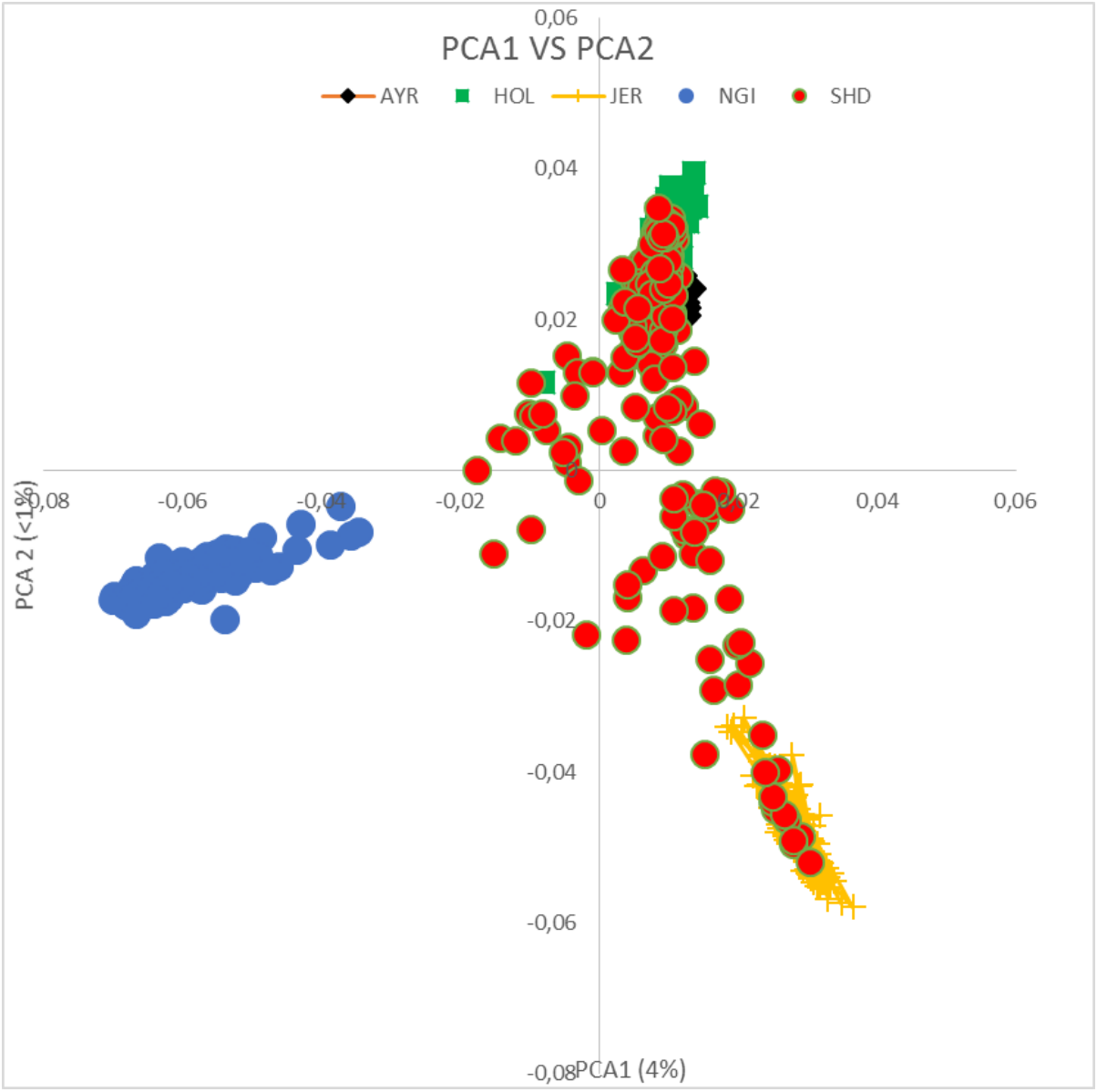

Figure 4.1 Principal component analysis plot constructed for PC1 and PC2

The populations were further visualized in PC1 and PC3 and represented in Figure 4.2. This accounted for a 4% variation, separating the HOL and AYR populations.

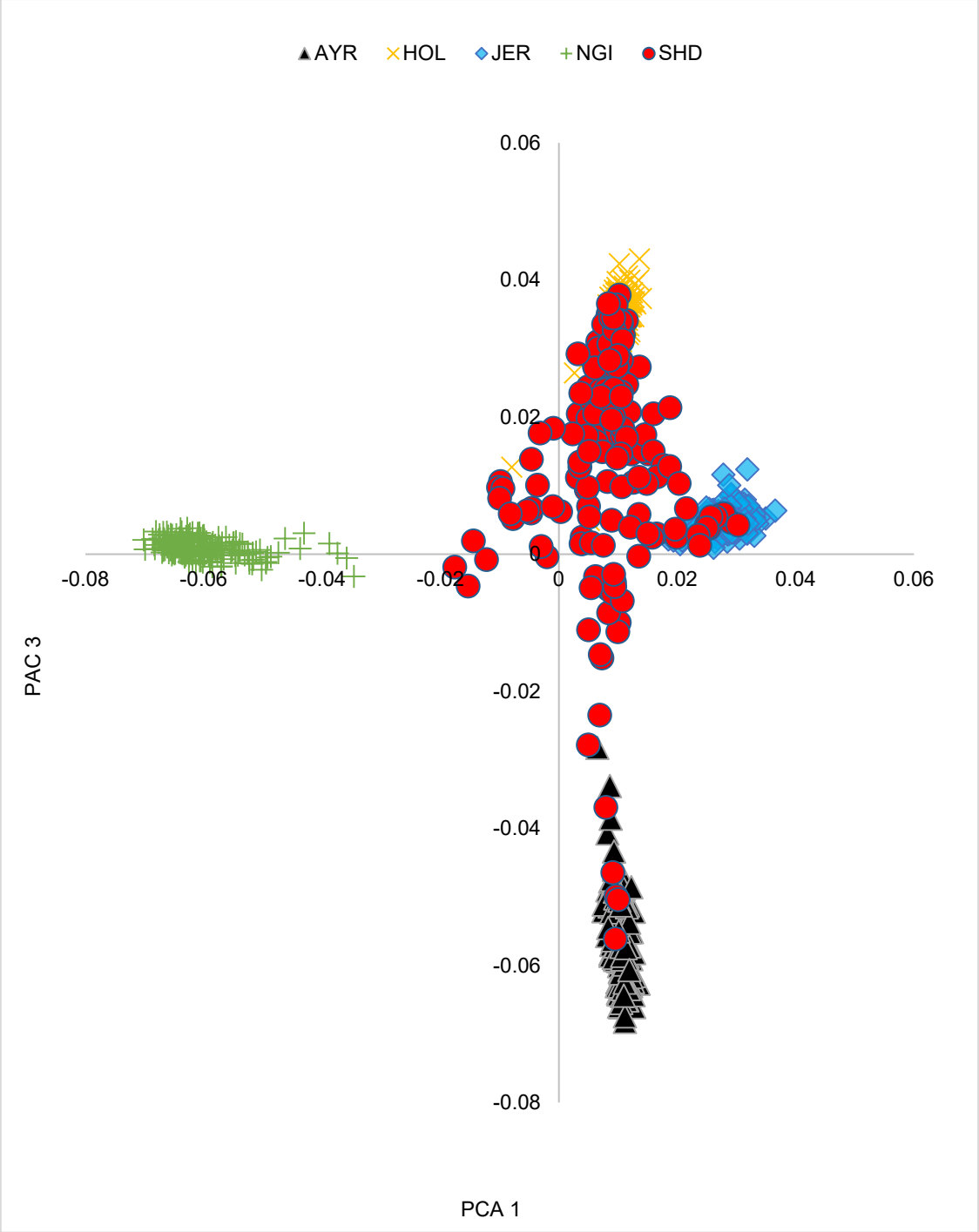

Figure 4.2 Principal component analysis plot constructed for PC1 and PC3

The PC 1 and 3 indicated five distinct populations, with the SDH population distributed among the Holstein, Jersey, and AYR populations. Most individuals from the SHD population did not form a distinct cluster, in contrast to the other four populations. A large majority of these individuals were closely related to the Holstein and Jersey, with a few clustering closely with the AYR. This indicates that most of the SHD populations

are predominantly crossbred, with HOL and JER being the major breeds used in crossbreeding. Farmers were asked about the nature of breeds they have in their farms, non of them mention AYR bloodline or crossbreds. But PCA results obtained in this study shows AYR infusion in some of the animals.

4.3. Admixture analysis

The admixture was performed to determine the existing genetic make-up of the smallholder population from *K* = 2 to *K* = 6 in Figure 4.3.

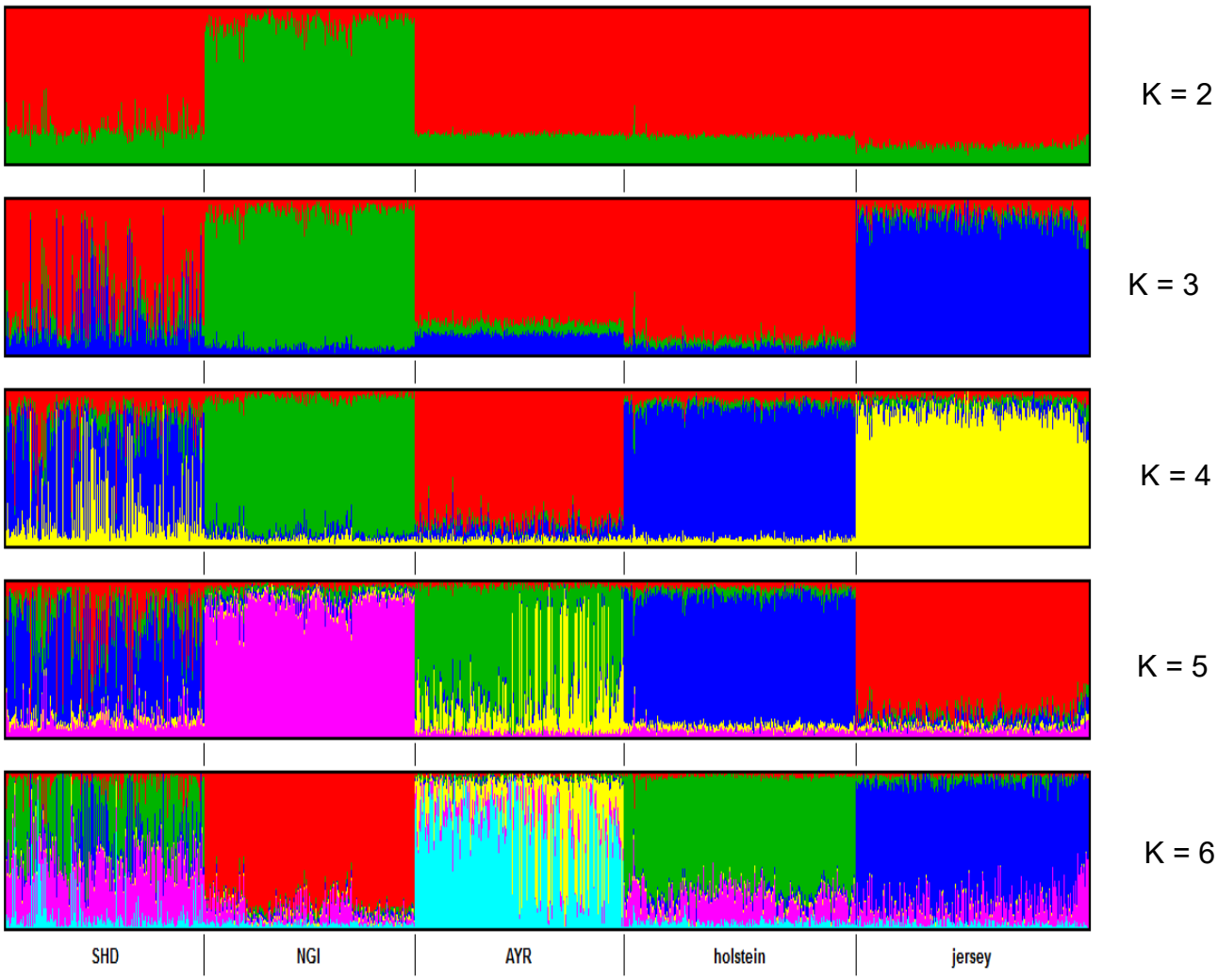

Figure 4.3 Admixture bar plots of breed compositions (K = 2 to K = 6), with K representing the optimal number of discrete breeds

The NGI population formed a distinct cluster in *K* = 2, with one population representing the other four populations. The gene-flow from the Jersey population was observed in *K* = 3. This corresponded with the PCA results, where NGI clustered separately from other populations. At this stage, the SHD population showed signs of admixture

between HOL and JER, with little introgression from the Nguni population. HOL and AYR clustered together which was consistent with PC1 vs PC2. At K = 4, HOL and AYR separated into distinct populations, and the SHD population showed admixture of HOL, JER, and AYR, with little gene-flow from NGI population.

CHAPTER 5: DISCUSSION

5.1. Introduction

Smallholder dairy farming in South Africa is based on the use of crossbreds with a high milk yield potential of exotic dairy breeds (Muntswu *et al*., 2016). The use of the pedigree recording is rare in such systems, which makes it difficult to make informed breeding decisions (Gorbach *et al*., 2010). Genomic data has been used to capture genetic diversity and population structure, in order to develop appropriate recommendations to the farmers and others intending to maximize productivity of these systems (Ojango *et al*., 2014). This has also opened up opportunities for developing genetic improvement programmes in the smallholder sector of developing countries. This study has provides new knowledge on genetic structure of smallholder dairy cattle populations in South Africa using the genomic technology approach.

5.2. Genetic diversity

A higher level of polymorphism was observed in the SHD followed by HOL compared to AYR, JER and NGU populations. The high level of heterogeneity in the smallholder dairy cattle population may be an indication of widespread crossbreeding. This was further verified by a relatively higher MAF value (0.31±0.10) in the HOL and SHD populations, comparison to the others (AYR, JER and NGI) (0,30±0,12). MAF is the frequency estimate of the least common allele per breed. The mean MAF estimate from the present study was higher than that reported previously in South Africa by Qwabe *et al*. (2013) for HOL (0.22) and NGI (0.21) breeds. Also, higher than those observed in the Rwandan cattle population (0,29) (Chagunda *et al*., 2018) and Indian cattle population (0.24) (Ahmed *et al*., 2019).

The genetic diversity within the populations was estimated based on the observed (H_O), expected (H_E) heterozygosity and the inbreeding (F_{IS}). The observed heterozygosity described as the percentage of loci heterozygous per individual, was marginally lower in the NGI and SHD populations (0.39±0.10) than in the specialized commercial dairy populations (0.40±0.10). This could be due to forces such as inbreeding resulting in deficit of heterozygotes (Ojango *et al*., 2011). However, the H_O detected in this study is higher than those obtained in the Rwandan (Chagunda *et al*., 2018) and the Indian (Ahmed *et al*., 2019) dairy cattle populations that had the values

of 0.35 and 0.38, respectively. In JER, Ho was lower than the HE while in AYR, HOL and NGI the H_E was equal to the H_O which is likely associated with random mating (Mburu & Hanotte, 2005). The SHD population had slightly higher inbreeding than other populations supporting the lower values of Ho than H_E obtained in this study whereas on other populations the level of inbreeding observed was lower predicting the possibilities of random matings. The higher values of MAF, Ho and H_E for HOL and NGI breeds obtained in this study compared to the values reported by Makina *et al*. (2014) for South African HOL (0.31) and NGU (0,24) can be associated with the increase in sample population size and thus an increase in allele frequencies over the years. The higher the number of animals genotyped, the higher the MAF values (McClure *et al.*, 2018).

5.3. Principal Component Analysis

Results from PCA analysis showed that the SHD population comprises predominantly of crossbred individuals, derived mainly from the Holstein and Jersey. The smallholder population displayed a heterogeneous cluster, an indication of a sub-population rather than a distinct population. The NGU population separated distinctly from the specialised dairy breeds (AYR, HOL, and JER), and formed a homogeneous cluster that was closely related to only a few animals from the SHD population. This points out the limited use of indigenous breed (NGU) in the crossbreeding practiced on smallholder dairy herds which is similar to the reports by Mujibi *et al*. (2019), further indicated that populations with 75 and 85% crossbreds as best performing breeds in the majority of smallholder dairy herds in Tanzania. This is in contrast to the widespread use of indigenous breeds in crossbreeding that has been observed in smallholder dairy production systems of other African countries (Ojango *et al*., 2016; Chagunda *et al*. 2018). Thus, the limited use of indigenous breeds in crossbreeding on smallholder dairy herds in South Africa might be compromising cow performance. Some of the smallholder farmers in South Africa market their milk to processing companies that pay on the basis of milk volume and solids content. Such farmers might be crossbreeding HOL and JER in order to complement the high milk production of the former and high solids production of the latter breed, which is an increasingly common practice on commercial herds. This might explain the predominance of crosses involving these two breeds in the SHD population.

5.4. Admixture Analysis

Admixture analysis was conducted to further elucidate the genetic make-up of cattle from the SHD population. A high level of admixture was observed in the SHD dairy population, with HOL and JER being the predominant contributory breeds, and with a little infusion from the AYR and NGU. In contrast to these findings, infusion of indigenous *Bos indicus* breeds has been observed in crossbred cattle on smallholder herds in Rwanda (Chagunda *et al*., 2018), Tanzania (Mujibi *et al*., 2019) and India (Ahmed *et al*., 2019). Although at $K = 6$ the SHD population showed some infusion of unaccounted genotypes, it is not clear whether these bloodlines represent exotic or indigenous breeds. This can be investigated further, using other dairy and other beef breeds available in South Africa as reference populations.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

The current study has generated new knowledge on the genetic diversity, population structure and prevailing breeding management practices on smallholder dairy herds in South Africa. The results of this study will be related to performance data to identify appropriate levels of gene admixture that would support maximum productivity and adaptability of cattle under prevailing production systems. This will form the basis for sound and systematic crossbreeding, which will lead to suitable selection of purebreds/crossbreds for improved cattle performance.

Results of the study show that smallholder dairy farmers in South Africa are mostly farming with crosses of Holstein and Jersey breeds, and there is a fairly high level of genetic diversity and low level of inbreeding in this production system. There is, however, limited use of the indigenous Nguni breed in the crossbreeding. This may mean that cattle on smallholder dairy herds are generally compromised on traits related to adaptability to the harsh environmental conditions in this production system. The poor cow productivity on smallholder dairy herds, reported in previous studies, may be partly attributable to a possible mismatch between the genotypes and production environment. It is recommended that further research be conducted to evaluate the performance of the various admixture levels in the SHD production system. Smallholder dairy farmers also need to be made aware of the importance of utilizing indigenous breeds in their crossbreeding programmes.

CHAPTER 7: REFERENCE AND APPENDIX

REFERENCES

Abin, S., Visser, C. and Banga, C.B., 2018. Comparative performance of dairy cows in low-input smallholder and high-input production systems in South Africa. Tropical animal health and production, pp.1-6.

Alexander, D.H., Novembre, J. and Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19(9), pp.1655-1664.

Ahmad, S.F., Panigrahi, M., Chhotaray, S., Pal, D., Parida, S., Bhushan, B., Gaur, G.K., Mishra, B.P. and Singh, R.K., 2019. Revelation of genomic breed composition in a crossbreed cattle of India with the help of Bovine50K BeadChip. Genomics.

Baiphethi, M.N. & Jacobs, P.T., 2009. The contribution of subsistence farming to food security in South Africa. Agrekon, 48(4), pp.459-482.

Banga, C.B., 2009. The development of breeding objectives for Holstein and Jersey cattle in South Africa (Doctoral dissertation, University of the Free State).

Bennett, A., Lhoste, F., Crook, J. & Phelan, J., 2006. The future of small scale dairying. Livestock report.

Bereda, A., Yilma, Z & Nurfeta, A., 2013. Handling, processing and utilization of milk and milk products in Ezha district of the Gurage zone, Southern Ethiopia. Journal of Agricultural Biotechnology and Sustainable Development, 5(6): 91-98.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M.Y. and Levéziel, H., 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genetics Selection Evolution, 35(1), pp.77.

Bonsma, J.C., 1980. Cross-breeding, breed creation and the genesis of the Bonsmara. Livestock Production - a Global Approach. pp. 90–110 Tafelberg, Cape Town

Blench, R. and MacDonald, K., 2006. The origins and development of African livestock: archaeology, genetics, linguistics and ethnography. Routledge.

Bray, T.C., Hall, S.J.G. and Bruford, M.W., 2014. Admixture analysis in relation to pedigree studies of introgression in a minority B ritish cattle breed: the L incoln R ed. *Journal of Animal Breeding and Genetics*, *131*(1), pp.19-26.

Capper, J.L. and Cady, R.A., 2012. A comparison of the environmental impact of Jersey compared with Holstein milk for cheese production. Journal of dairy science, 95(1), pp.165-176.

Casas, E., Shackelford, S.D., Keele, J.W., Koohmaraie, M., Smith, T.P.L. and Stone, R.T., 2003. Detection of quantitative trait loci for growth and carcass composition in cattle. Journal of animal science, 81(12), pp.2976-2983.

Chagunda, M., Dusingizimana, T., Kamana, O., Cheruiyot, E., Mujibi, F.D. & Mwai, O., 2018. Use of high density single nucleotide polymorphism (SNP) arrays to assess genetic diversity and population structure of dairy cattle in smallholder dairy systems: the case of Girinka Programme in Rwanda. Frontiers in genetics, 9, pp.438.

DAFF Department of Agriculture, Forestry and Fisheries, 2017b. Economic Review of the South African Agriculture 2016/17 Directorate Statistics and Economic Analysis. Pretoria: DAFF.1

Dash, S., Singh, A., Bhatia, A.K., Jayakumar, S., Sharma, A., Singh, S., Ganguly, I. and Dixit, S.P., 2018. Evaluation of bovine high-density SNP genotyping Array in indigenous dairy cattle breeds. Animal biotechnology, 29(2), pp.129-135.

Delgado, C.L., Rosegrant, M.W. and Meijer, S., 2001, January. Livestock to 2020: The revolution continues. In annual meetings of the International Agricultural Trade Research Consortium (IATRC), Auckland, New Zealand (pp. 18-19).

Ding L., Wiener H., Abebe T., Altaye M., Go R. C., Kercsmar C., *et al*.,2011. Comparison of measures of marker informativeness for ancestry and admixture mapping. BMC Genomics 12:622. 10.1186/1471-2164-12-622

Edea, Z., Dadi, H., Kim, S.W., Park, J.H., Shin, G.H., Dessie, T., Kim, K.S., 2014. Linkage disequilibrium and genomic scantodetect selectiveloci incattle populations adapted to different ecological conditions in Ethiopia. J. Anim. Breed. Genet. 131, 358–366.

FAO, 2011. Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. No.9. In: Commission on Genetic Resources for Food and Agriculture Assessments. FAO, Rome.

FAO, 2018. http://dx.doi.org/10.1787/agr_outlook-2018-en.

Getachew, T., 2015. Genetic diversity and admixture analysis of Ethiopian Fat-tailed and Awassi sheep using SNP markers for designing crossbreeding schemes (Doctoral dissertation, University of Natural Resources and Life Sciences, Vienna).

Gorbach, D.M., Makgahlela, M.L., Reecy, J.M., Kemp, S.J., Baltenweck, I., Ouma, R., Mwai, O., Marshall, K., Murdoch, B., Moore, S. & Rothschild, M.F., 2010. Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. Journal of animal breeding and genetics, 127(5), pp.348-351.

Grobler, S.M., Scholtz, M.M., Bester, J., Mamabolo, J.M & Ramsay, K.A., 2008. Dairy production systems in the emerging and communal sectors of South Africa: results from a structured survey. Applied Animal Husbandry and. Rural Development, (1): 25-30.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E. & Weigend, S., 2010. Genetic diversity in farm animals – a review. Animal Genetics, 41, 6-31.

Heaton, M.P., Grosse, W.M., Kappes, S.M., Keele, J.W., Chitko-McKown, C.G., Cundiff, L.V., Braun, A., Little, D.P. & Laegreid, W.W., 2001. Estimation of DNA sequence diversity in bovine cytokine genes. Mammalian Genome, 12(1), pp.32-37.

Horan, B., Mee, J.F., Rath, M., O'connor, P. and Dillon, P., 2004. The effect of strain of Holstein-Friesian cow and feeding system on reproductive performance in seasonal-calving milk production systems. Animal Science, 79(3), pp.453-467.

Hu, Z.L., Fritz, E.R. & Reecy, J.M., 2006. Animal QTL db: a livestock QTL database tool set for positional QTL information mining and beyond. Nucleic Acids Research, 35(suppl_1), pp.D604-D609.

Kelleher, M.M., Berry, D.P., Kearney, J.F., McParland, S., Buckley, F. and Purfield, D.C., 2017. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. Animal, 11(1), pp.15-23.

Kim, E.S. and Rothschild, M.F., 2014. Genomic adaptation of admixed dairy cattle in East Africa. Frontiers in genetics, 5, p.443.

Kosgey, I., Baker, R., Udo, H. & Van Arendonk, J.A.M. 2006. Successes and failures of small ruminant breeding programmes in the tropics: a review. Small Ruminant. Research, 61:13–28.

Kumar, H., Panigrahi, M., Chhotaray, S., Pal, D., Bhanuprakash, V., Saravanan, K.A., Shandilya, R., Parida, S. & Bhushan, B., 2019. Identification of breed-specific SNP panel in nine different cattle genomes. Biomedical Research, 30(1), pp.78-81.

Lashmar, S.F., Muchadeyi, F.C. & Visser, C., 2019. Genotype imputation as a cost-saving genomic strategy for South African Sanga cattle: A review. South African Journal of Animal Science, 49(2), pp.262-280.

Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Laviolette, J.P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P. and Shah, N., 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nature genetics, 24(4), p.381.

Lopreiato, V., Minuti, A., Trimboli, F., Britti, D., Morittu, V.M., Cappelli, F.P., Loor, J.J. and Trevisi, E., 2019. Immunometabolic status and productive performance differences between periparturient Simmental and Holstein dairy cows in response to pegbovigrastim. Journal of dairy science, 102(10), pp.9312-9327.

Madilindi, M.A., Banga, C.B., Bhebhe, E., Sanarana, Y.P., Nxumalo, K.S., Taela, M.G., Magagula, B.S. and Mapholi, N.O., 2019. Genetic diversity and relationships among three Southern African Nguni cattle populations. Tropical animal health and production, pp.1-10.Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E., MacNeil, M.D. & Maiwashe, A., 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel.

Makina, S.O., Whitacre, L.K., Decker, J.E., Taylor, J.F., MacNeil, M.D., Scholtz, M.M., van Marle-Köster, E., Muchadeyi, F.C., Makgahlela, M.L., Maiwashe, A., 2016. Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide. Genet. Select. Evol. 48, 88.

Manzana, N.P., McCrindle, C.M., Sebei, P.J. and Prozesky, L., 2014. Optimal feeding systems for small-scale dairy herds in the North West Province, South Africa. Journal of the South African Veterinary Association, 85(1), pp.01-08.

Mapekula, M., Chimonyo, M., Mapiye, C. & Dzama, K., 2010. Milk utilisation patterns in the low-input production systems in South Africa. Tropical animal health and production, 42(7), pp.1413-1419.

Mapekula, M., Chimonyo, M., Mapiye, C. & Dzama, K., 2011. Fatty acid, amino acid and mineral composition of milk from Nguni and local crossbred cows in South Africa. Journal of food composition and analysis, 24(4-5), pp.529-536.

Mapholi, N.O., Marufu, M.C., Maiwashe, A., Banga, C.B., Muchenje, V., MacNeil, M.D., Chimonyo, M. and Dzama, K., 2014. Towards a genomics approach to tick (Acari: Ixodidae) control in cattle: a review. Ticks and tick-borne diseases, 5(5), pp.475-483.

Marshall, K., Quiros-Campos, C., Van der Werf, J.H.J. & Kinghorn, B., 2011. Marker-based selection within smallholder production systems in developing countries. Livestock Science, 136(1), pp.45-54.

Marshall, K., Gibson, J.P., Mwai, O., Mwacharo, J.M., Haile, A., Getachew, T., Mrode, R. and Kemp, S.J., 2019. Livestock genomics for developing countries–african examples in practice. Frontiers in genetics, 10, p.297.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S. and Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS one, 4(4), p.e5350.

Mburu, D. and Hanotte, O., 2005. A practical approach to microsatellite genotyping with special reference to livestock population genetics. A manual prepared for the IAEA/ILRI training course on molecular characterisation of small ruminant genetic resources of Asia.

McClure, M.C., McCarthy, J., Flynn, P., McClure, J. C., Dair, E., O'Connell, D. K., & Kearney, J. F., 2018. SNP Data Quality Control in a National Beef and Dairy Cattle System and Highly Accurate SNP Based Parentage Verification and Identification. *Frontiers in Genetics*, *9*, pp 84

Meuwissen, T., Hayes. B & Goddard, M., 2016. Genomic selection: A paradigm shift in animal breeding, Animal Frontiers, Volume 6, Issue 1.

Mujibi, F.D., Rao, J., Agaba, M., Nyambo, D., Cheruiyot, E.K., Kihara, A., Zhang, Y. and Mrode, R., 2019. Performance evaluation of highly admixed Tanzanian Smallholder Dairy Cattle using SNP derived kinship matrix. Frontiers in genetics, 10.

Muntswu, A.E., Banga, C.B. & Norris, D., 2016. Breeding and management practices of smallholder and emerging dairy herds in South Africa. Milk SA bulletin, pp 36.

Muntswu, E.A., Chitura, T., Abin, S.A. and Banga, C.B., 2017. Characterization of emerging and smallholder dairy production systems in South Africa. In: Proc of the 50th annual congress of the SA society for animal science.

MPO. https://www.mpo.co.za/wp-content/uploads/2017/05/LACTO_DATA_MEI_16.pdf

Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences, 70(12), pp.3321-3323.

Nicolazzi, E.L., Biffani, S., Biscarini, F., Orozco ter Wengel, P., Caprera, A., Nazzicari, N. and Stella, A., 2015. Software solutions for the livestock genomics SNP array revolution. Animal genetics, 46(4), pp.343-353.

Odero-Waitituh, J. A., 2017. Smallholder dairy production in Kenya; a review. Livestock Research for Rural Development, Volume 29, 139.

Ojango, J.M., Mpofu, N., Marshall, K. & Andersson-Eklund, L., 2011. Quantitative methods to improve the understanding and utilization of animal genetic resources. In: Animal Genetics Training Resources, version 3. Ojango, J.M., Malmfors, B & Okeyo, A.M (Eds). International Livestock Research Institute, Nairobi, Kenya and Swedish University of Agricultural Science, Uppsala, Sweden.

Ojango, J.M., Marete, A., Mujibi, F.D.N., Rao, E.J.O., Poole, E.J., Rege, J.E.O., Gondro, C., Weerasinghe, W.M.S.P., Gibson, J.P. & Okeyo Mwai, A., 2014. A novel use of high density SNP assays to optimize choice of different crossbred dairy cattle genotypes in smallholder systems in East Africa. American Society of Animal Science.

Ojango, J.M., Mrode, R., Rege, J.E.O., Mujibi, D., Strucken, E.M., Gibson, J. & Mwai, O., 2019. Genetic evaluation of test-day milk yields from smallholder dairy production

systems in Kenya using genomic relationships. Journal of dairy science, 102(6), pp.5266-5278.

Panetto, J.D.C., Silva, M.V.G.B., Leite, R.M.H., Machado, M.A., Bruneli, F.A.T., REIS, D.D.L., Peixoto, M.G.C.D. and VERNEQUE, R.D.S., 2017. Red Sindhi cattle in Brazil: population structure and distribution. Embrapa Gado de Leite-Artigo em periódico indexado (ALICE).

Pienaar, L.& Traub, L., 2015. Understanding the smallholder farmer in South Africa: Towards a sustainable livelihoods classification (No. 1008-2016-79955).

Prendiville, R., Pierce, K.M., Delaby, L. and Buckley, F., 2011. Animal performance and production efficiencies of Holstein-Friesian, Jersey and Jersey× Holstein-Friesian cows throughout lactation. *Livestock Science*, *138*(1-3), pp.25-33.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, 81(3), pp.559-575.

Qwabe, S.O., van Marle-Köster, E. and Visser, C., 2013. Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. Tropical animal health and production, 45(2), pp.511-516.

Sanarana, Y., Visser, C., Bosman, L., Nephawe, K., Maiwashe, A. and van Marle-Köster, E., 2016. Genetic diversity in South African Nguni cattle ecotypes based on microsatellites markers. Tropical Animal Health and Production, 48 (2), pp379–385

Scholtz, M.M., 1988. Selection possibilities of hardy beef breeds in Africa: The Nguni example. In 3. Congres Mondial de Reproduction et Selection des Ovins et Bovins a Viande, Paris (France), 19-23 Jun 1988. INRA.

Scholtz, M.M., 2005. The role of research and the seed stock industry in the in situ conservation of livestock genetic resources. In 4th All Africa Conference on Animal Agriculture, Arusha, Tanzania pp. 311-316.

Scutari, M., Mackay, I. and Balding, D., 2016. Using genetic distance to infer the accuracy of genomic prediction. PLoS genetics, 12(9), p.e1006288.

Sermyagin, A.A., Gladyr, E.A., Plemyashov, K.V., Kudinov, A.A., Dotsev, A.V., Deniskova, T.E. & Zinovieva, N.A., 2018. Genome-wide association studies for milk production traits in Russian population of Holstein and black-and-white cattle. In Proceedings of the Scientific-Practical Conference" Research and Development-2016" (pp. 591-599). Springer, Cham.

Shriver, M.D., Parra, E.J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A., Cameron, N. and Baron, A., 2003. Skin pigmentation, biogeographical ancestry and admixture mapping. *Human genetics*, *112*(4), pp.387-399.

StatsSA, 2019. Towards measuring the extent of food security in South Africa: An examination of hunger and food inadequacy.

Strucken, E.M., Al-Mamun, H.A., Esquivelzeta-Rabell, C., Gondro, C., Mwai, O.A. & Gibson, J.P., 2017. Genetic tests for estimating dairy breed proportion and parentage assignment in East African crossbred cattle. Genetics Selection Evolution, 49(1), p.67.

Swai, E. S. & Karimuribo, E. D., 2011. 'Smallholder Dairy Farming in Tanzania: Current Profiles and Prospects for Development', Outlook on Agriculture, 40(1), pp. 21–27. doi: 10.5367/oa.2011.0034.

Tanyanyiwa, F.K., Kolanisi, U., Chimonyo, M. & Banga, C., 2017. Identification of Entrepreneurial Characteristics of Emerging Smallholder Dairy Farmers: A Case Study of Groblersdal and Matatiele Local Municipalities, South Africa. J Hum Ecol, 57(3), pp.108-117.

Toro, M.A., Fernández, J. & Caballero, A., 2009. Molecular characterization of breeds and its use in conservation. Livestock Science, 120(3), pp.174-195.

van Marle-Kőster, E., Visser, C., Makgahlela, M. and Cloete, S.W., 2015. Genomic technologies for food security: A review of challenges and opportunities in Southern Africa. Food Research International, 76, pp.971-979.

van Marle-Köster, E. and Visser, C., 2018. Genetic improvement in South African livestock: can genomics bridge the gap between the developed and developing sectors?. Frontiers in genetics, 9, p.331.

Van Vleck, L. D., Pollak, E. J., & Oltenacu, E. A. B., 1986. Genetics for the Animal Sciences. New York, NY: W.H. Freeman and Co.

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Schenkel, F.S. and Van Tassell, C.P., 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of dairy science, 92(7), pp.3431-3436.

Williams, J.L., Dunner, S., Valentini, A., Mazza, R., Amarger, V., Checa, M.L., Crisa, A., Razzaq, N., Delourme, D., Grandjean, F. & Marchitelli, C., 2009. Discovery, characterization and validation of single nucleotide polymorphisms within 206 bovine genes that may be considered as candidate genes for beef production and quality. Animal genetics, 40(4), pp.486-491.

Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics, 88(1), pp.76-82.

Zwane, A.A., Maiwashe, A., Makgahlela, M.L., Choudhury, A., Taylor, J.F. and van Marle-Köster, E., 2016. Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds. South African Journal of Animal Science, 46(3), pp.302-312.

https://www.canr.msu.edu history_of_dairy_cow_breeds_holstein

https://milksa.co.za/research/dairy-rd-in-sa/variation-herd-size-and-milk-production-south-african-farms-relation

http://www.elsenburg.com. 2017. The Dairy Farming Handbook Section

APPENDIX

Smallholder herds inbreeding frequency report

FID	IID	O(HOM)	E(HOM)	N(NM)	F
1	1	19750	2.005e+04	33357	-0.0225
1	H11	19690	2.006e+04	33370	-0.02765
2	2	20176	2.007e+04	33399	0.007678
2	H12	20839	2.008e+04	33405	0.0571

3	3	20853	2.008e+04	33409	0.05796
3	H13	20204	2.009e+04	33421	0.008722
4	4	19951	2.008e+04	33406	-0.009496
4	H14	20794	2.009e+04	33417	0.05315
5	5	21533	2.007e+04	33395	0.1097
5	H15	19837	2.008e+04	33414	-0.01845
6	7	20067	1.986e+04	33028	0.01605
6	H21	19884	2.009e+04	33422	-0.01531
7	12	20382	2.008e+04	33400	0.02303
7	H22	20331	2.008e+04	33411	0.01866
8	13	20245	2.007e+04	33389	0.01329
8	H23	19999	2.008e+04	33416	-0.006429
9	14	19668	2.004e+04	33334	-0.0276
9	H24	20743	2.009e+04	33430	0.04874
10	15	20107	2.008e+04	33406	0.00218
10	H31	20487	2.009e+04	33421	0.02999
11	16	20923	2.007e+04	33394	0.06394
11	H32	20183	2.009e+04	33422	0.007057
12	17	19695	1.997e+04	33229	-0.02089
12	H33	22816	2.008e+04	33409	0.2053
13	38	21504	2.006e+04	33375	0.1086
13	H34	20677	2.007e+04	33394	0.04549
14	19	20038	2.007e+04	33397	-0.002581
14	H35	20493	2.008e+04	33414	0.03069
15	20	19999	2.007e+04	33399	-0.005568
15	H41	19359	2.009e+04	33426	-0.05485
16	28	21885	2.007e+04	33388	0.1365
16	H42	19479	2.009e+04	33426	-0.04587
17	29	19614	2.008e+04	33406	-0.03479
17	H43	19395	2.009e+04	33422	-0.05199

18	30	19790	2.005e+04	33366	-0.01986
18	H44	20323	2.009e+04	33423	0.01755
19	31	20362	2.006e+04	33369	0.02302
19	H45	19429	2.008e+04	33409	-0.04887
20	32	19985	2.006e+04	33379	-0.005774
20	H51	20196	2.009e+04	33426	0.007915
21	33	19560	2.007e+04	33398	-0.03849
22	36	20342	2.008e+04	33402	0.01996
22	49	19352	1.984e+04	33011	-0.03712
23	37	19720	2.008e+04	33407	-0.02694
23	H54	20189	2.009e+04	33418	0.007799
24	39	20030	1.989e+04	33079	0.01088
24	H55	20228	2.004e+04	33342	0.01408
25	40	19554	2.007e+04	33389	-0.03856
25	H61	23709	2.008e+04	33411	0.2721
26	41	20482	2.007e+04	33400	0.0306
26	H62	20129	2.009e+04	33426	0.00286
27	42	21056	2.008e+04	33413	0.07302
27	H63	21583	2.009e+04	33427	0.1118
28	43	20703	2.008e+04	33408	0.0468
28	H64	22657	2.009e+04	33422	0.1926
29	44	20513	2.008e+04	33409	0.03244
29	H71	20066	2.009e+04	33424	-0.001726
30	45	20834	1.999e+04	33244	0.06399
30	H72	20572	2.009e+04	33428	0.03603
31	46	20732	2.008e+04	33401	0.04926
31	H73	21286	2.009e+04	33428	0.08953
32	47	21436	2.008e+04	33403	0.102
32	H74	21257	2.009e+04	33420	0.08774
33	50	20111	2.008e+04	33407	0.002385

33	H75	19800	2.009e+04	33418	-0.02139
34	51	19368	2.008e+04	33416	-0.05372
34	H81	19377	2.009e+04	33419	-0.0532
35	52	19888	2.007e+04	33399	-0.01398
35	H82	20159	2.009e+04	33421	0.005383
36	53	19682	1.994e+04	33172	-0.01974
36	H83	19539	1.997e+04	33221	-0.03243
37	61	19781	2.006e+04	33379	-0.02115
37	H84	22469	2.008e+04	33413	0.179
38	62	19869	2.007e+04	33393	-0.01512
38	H85	20124	2.009e+04	33425	0.002573
39	63	22360	2.007e+04	33395	0.1718
39	H89	21068	2.009e+04	33427	0.07328
40	64	22235	2.007e+04	33395	0.1624
40	H91	21171	2.009e+04	33428	0.0809
41	65	19985	2.007e+04	33398	-0.00667
41	H92	20693	2.009e+04	33424	0.04529
42	66	20709	1.98e+04	32944	0.06906
42	H93	22506	2.009e+04	33417	0.1816
43	67	21513	2.006e+04	33382	0.1089
43	H94	19397	2.009e+04	33422	-0.05181
44	68	21396	2.007e+04	33398	0.09925
44	H101	20416	2.009e+04	33421	0.02459
45	69	19677	2.008e+04	33402	-0.02995
45	H102	23549	2.009e+04	33416	0.2598
46	70	21029	2.007e+04	33397	0.07182
46	H103	20136	2.009e+04	33423	0.003546
47	71	20494	2.007e+04	33392	0.03188
47	H104	20153	2.009e+04	33430	0.004495
48	72	20354	1.986e+04	33035	0.0377

48	H105	19985	2.007e+04	33398	-0.00669
49	73	21863	2.001e+04	33297	0.1392
49	H111	20068	2.008e+04	33413	-0.001097
50	74	20409	2.008e+04	33407	0.02473
50	H112	23479	2.009e+04	33422	0.2543
51	75	20596	2.008e+04	33409	0.03869
51	H113	20434	2.008e+04	33416	0.0262
52	76	19422	2.007e+04	33397	-0.04882
52	H114	20009	2.009e+04	33426	-0.0061
53	77	19196	2.008e+04	33407	-0.06633
53	H115	20463	2.009e+04	33424	0.02805
54	78	21293	2.006e+04	33383	0.09224
54	H121	20201	1.996e+04	33208	0.01786
55	79	19545	2.007e+04	33389	-0.03923
55	H122	20641	2.009e+04	33428	0.04121
56	80	19917	2.008e+04	33407	-0.01212
56	H123	19782	2.006e+04	33372	-0.02073
57	81	20977	2.007e+04	33397	0.0679
57	H124	19398	2.009e+04	33418	-0.05162
58	82	19956	2.008e+04	33403	-0.009067
58	H125	19608	2.009e+04	33428	-0.03627
59	83	19299	2.007e+04	33395	-0.05798
59	H131	22411	2.008e+04	33405	0.1751
60	84	20507	1.99e+04	33109	0.04581
60	H133	22154	1.997e+04	33222	0.1647
61	85	20193	2.007e+04	33393	0.009255
61	H134	21728	2.009e+04	33420	0.1231
62	86	19264	2.008e+04	33409	-0.0612
62	H141	20384	2.009e+04	33424	0.02211
63	87	19405	2.005e+04	33349	-0.04811

63	H142	20530	2.009e+04	33421	0.03317
64	88	20384	2.008e+04	33409	0.02281
64	H143	19997	2.009e+04	33424	-0.006919
65	89	19372	2.005e+04	33353	-0.0506
65	H144	19831	2.009e+04	33424	-0.01934
66	90	19646	2.008e+04	33403	-0.03233
66	H151	19743	1.991e+04	33117	-0.01239
67	91	19692	2.007e+04	33398	-0.02862
67	H152	19868	2.007e+04	33397	-0.0154
68	92	19816	2.006e+04	33382	-0.01857
68	H153	20401	2.009e+04	33420	0.02357
69	93	19651	2.008e+04	33416	-0.03253
69	H154	20394	2.009e+04	33430	0.02258
70	94	21553	2.007e+04	33389	0.1115
70	H155	20467	2.009e+04	33426	0.02822
71	95	19992	2.007e+04	33400	-0.006222
71	H161	20139	2.008e+04	33413	0.00416
72	96	20415	1.993e+04	33156	0.03658
72	H162	19142	1.984e+04	32999	-0.05287
73	97	20704	2.007e+04	33391	0.04763
73	H163	19562	2.007e+04	33390	-0.03812
74	98	19435	2.007e+04	33396	-0.04785
74	H164	21250	2.008e+04	33403	0.08803
75	H165	19441	2.009e+04	33421	-0.04844
75	KD_1405	20530	2.007e+04	33396	0.03442
76	H171	20215	2.009e+04	33420	0.009655
76	KD_1402	23101	2.007e+04	33396	0.2274
77	H172	20082	2.009e+04	33418	-0.0002955
77	KD_1430	21516	2.007e+04	33386	0.1089
78	H173	20687	2.009e+04	33433	0.04442

78	KD_calf_1	19788	1.999e+04	33256	-0.01492
79	H174	19880	2.009e+04	33418	-0.01544
79	KD_calf_2	19242	2.006e+04	33378	-0.06155
80	H181	20379	2.009e+04	33423	0.02176
80	JC_Bmiller	20127	2.007e+04	33396	0.004112
81	H182	20206	2.009e+04	33424	0.008731
81	JC_Dkwoo	19625	2.007e+04	33400	-0.03375
82	H183	20697	2.009e+04	33426	0.04545
82	JC_C107	20135	2.007e+04	33401	0.004508
83	H184	20525	2.009e+04	33429	0.03245
83	JC_C120	19884	2.006e+04	33381	-0.01342
84	H193	20095	2.007e+04	33390	0.00188
84	JC_Dsantj	19664	1.994e+04	33166	-0.02053
85	DB_N1	20749	2.006e+04	33373	0.05192
85	H194	20034	2.008e+04	33413	-0.003753
86	DB_Calf_2	19960	2.006e+04	33367	-0.007229
86	H195	19948	2.008e+04	33406	-0.009856
87	FSD_2007	20149	2.008e+04	33402	0.005512
87	H196	20537	2.009e+04	33417	0.0339
88	FSC_3	19777	2.003e+04	33321	-0.01881
88	H197	22207	2.008e+04	33414	0.1592
89	21	21521	2e+04	33284	0.1142
89	H198	20017	2.009e+04	33419	-0.005209
90	22	20126	2.001e+04	33284	0.008989
90	H199	20312	2.009e+04	33419	0.01692
91	H1211	19610	2.009e+04	33422	-0.03588
92	H25	20079	2.009e+04	33417	-0.0004864
93	27	19556	2.006e+04	33379	-0.038
93	H147	19363	2.009e+04	33427	-0.0546
94	48	18139	1.865e+04	31051	-0.04157

94	H1411	20382	2.009e+04	33419	0.02219
95	1401	20064	2.007e+04	33391	-0.0003999
95	H148	22327	2.009e+04	33417	0.1681
96	99	19605	1.993e+04	33150	-0.02438
96	H129	20021	1.999e+04	33261	0.002143