Genetic Diversity and Population Structure of the Indigenous Sheep in Kenya Based on Microsatellite Analysis

Denis Misiko Mukhongo

A thesis submitted in partial fulfillment for the degree of Master of Science in Biotechnology in the Jomo Kenyatta University of Agriculture and Technology

2010

DECLARATION

This is my original work and has never been presented for a degree or for any other award in any other University whatsoever.

Signature ………………………… Date …………………………….

Denis Misiko Mukhongo

This thesis has been submitted for examination with our approval as university supervisors.

Signature ………………………….. Date ……………………………..

Prof. Anne W. Muigai (PhD)
JKUAT, Kenya.

Signature ………………………….. Date ………………………………

Dr. Miika Tapio (PhD)
MTT Agrifood Research Finland Institute, Finland.

Signature ………………………….... Date ………………………………..

Dr. Okeyo Mwai (PhD)
ILRI, Kenya.

DEDICATION

To my Mother **Mrs. Gladys N. L.W. Mukhongo**, who stood by me in all situations and felt my pain as hers, who inspired and propelled me to greater heights of achievements, who would never allow me to quit without trying, who dared me to hope for the best and whose unconditional love continuously lifts me…

Thank you 'Mama'

'Asandi Maayi'

ACKNOWLEDGEMENTS

I thank you my God for seeing me through tumultuous times, through out this course and all my life. I sincerely thank my parents, Mr and Mrs Allan Joseph Mukhongo, for inspiration, encouragement, and selflessness that made all this work worthwhile. To my siblings, my dear brothers, my sisters you are such a blessing. I am honoured to be part of our family.

My greatest and sincere thanks to my supervisors for providing me with the excellent mentoring that I received. Prof. Anne Muigai, I will remember you more like my mother, for your availability, patience and expediency in handling everything, you are just priceless. May the Lord reward you Professor, you gave me a one in a lifetime opportunity. Dr. Miika Tapio, you continue to amaze me, 'Asante' for all the time you spent explaining the genetic concepts encountered while analysing my data. You are my role model. Dr. Okeyo Mwai thanks so much, for showing me the ropes, and taking me through those difficult beginner steps especially from a breeder's perspective.

To my colleagues and all my Laboratory mates (God bless you), thank you for the company you provided and the companionship through out my work. Moses Ogugo and Isaiah Obara, you are just the best, thanks for the financial boosts when I was desperately in need. James Audho, thanks for the excellent mentoring on animal handling skills you volunteered while we moved round sampling, you are such an expert!

My appreciations too to the International Federation of Agricultural Research for their funding that made this very vital research work possible.

TABLE OF CONTENTS

DECLARATION II

DEDICATION II

ACKNOWLEDGEMENTS IV

TABLE OF CONTENTS V

LIST OF TABLES X

LIST OF FIGURES XII

LIST OF APPENDICES XIII

LIST OF ABBREVIATIONS AND ACRONYMS XIV

ABSTRACT XVI

CHAPTER ONE: INTRODUCTION 1

1.1.0 BACKGROUND 1

1.2.0 ASSUMPTIONS OF THE STUDY 8

1.3.0 HYPOTHESES OF THE STUDY 8

1.4.0 OBJECTIVES OF THE STUDY 9

1.4.1 General Objective 9

1.4.2 Specific Objectives 9

1.5.0 JUSTIFICATION AND SIGNIFICANCE OF THE STUDY 10

CHAPTER TWO: LITERATURE REVIEW 11

2.1 ORIGIN AND DOMESTICATION OF SHEEP 11

2.2 SHEEP GENETIC RESOURCES IN KENYA 14

2.3.0 IMPORTANCE OF SHEEP GENETIC RESOURCES 15

2.4.0 MANAGEMENT OF ANIMAL GENETIC RESOURCES 17

2.5.0 CONSERVATION OF GENETIC DIVERSITY 19

2.5.1 Farm animal genetic diversity 20

2.5.2 Loss of genetic diversity 21

2.5.3 Methods of conservation of genetic resources 24

2.5.3.1 *In-situ* conservation 24

2.5.3.2 *Ex-situ* conservation 24

2.6.0 MOLECULAR MARKERS USED IN DIVERSITY STUDIES 25

2.6.1 Random Amplified Polymorphic DNA (RAPD) 28

2.6.2 Restriction Fragment Length Polymorphism (RFLP) 28

2.6.3 Amplified Fragment Length Polymorphism (AFLP) 29

2.6.4 Single Nucleotide Polymorphisms (SNPs) 31

2.6.5 Diversity arrays technology (DArT) 34

2.6.6 Mitochondrial DNA (mtDNA) 36

2.6.7 Y-chromosome 37

2.6.8 Copy Number Variations (CNVs) 38

2.6.9 Microsatellites 40

2.6.9.1 Classification of Microsatellites 41

2.6.9.2 Mutation rates of Microsatellites 42

2.6.9.3 Mutation patterns of Microsatellites 43

2.6.9.4 Theoretical models of Microsatellite Mutation 44

CHAPTER THREE: MATERIALS AND METHODS 47

3.1 SAMPLING 47

3.1.1 The sampling sites 47

3.1.2 The samples 47

3.1.3 Sheep sampled in this project 48

3.1.3.1 Blackhead Somali 48

3.1.3.2 Dorper 49

3.1.3.3 Red Maasai 50

3.2 DNA EXTRACTION 54

3.2.1 DNA Extraction from Blood 54

3.2.1.1 Cleaning of FTA Filter Papers for Polymerase Chain reaction 55

3.2.1.2 DNA Extraction from Peripheral Blood Lymphocytes in urea 55

3.3 DETERMINATION OF THE DNA CONCENTRATION 56

3.4 PCR AMPLIFICATION AND MICROSATELLITE GENOTYPING 56

3.5 AGAROSE GEL ELECTROPHORESIS 58

3.6 MICROSATELLITE GENOTYPING 59

3.7 STATISTICAL ANALYSES USED IN THE STUDY 60

3.7.1 Genetic Diversity 60

3.7.2 Population Structure and Admixture 61

3.7.3 Hardy-Weinberg Equilibrium 63

3.7.4 Population Genetic Differentiation 64

3.7.5 Genetic Distances and Phylogenetic relationships 65

3.7.6 Analysis of molecular variance (AMOVA) 67

3.7.7 Genetic Bottleneck 68

3.7.8 Genetic Linkage Disequilibrium 70

CHAPTER FOUR: RESULTS 71

4.1 GENETIC VARIABILITY 71

4.2 POPULATION DIFFERENTIATION 77

4.3 HARDY WEINBERG EQUILLIBRIUM 87

4.4 NEI'S GENETIC DISTANCES 90

4.5 PHYLOGENETIC RELATIONSHIPS 90

4.6 LINKAGE DISEQUILIBRIUM 94

4.7 ANALYSIS OF MOLECULAR VARIANCE (AMOVA) 94

4.8 GENETIC BOTTLENECK 95

CHAPTER FIVE 97

5.1 DISCUSSION 97

5.1.1 Genetic Diversity 97

5.1.2 Genetic Bottleneck 100

5.1.3 Genetic Distances and Phylogenetic Relationship 101

5.1.4 Population Genetic Differentiation 104

5.1.5 Hardy-Weinberg Equillibrium 108

5.1.6 Linkage Disequilibrium 110

5.2 IMPLICATIONS FOR THE SAKE OF CONSERVING THE SHEEP STUDIED 112

5.3 CONCLUSIONS... 117

5.4 RECOMMENDATIONS... 118

CHAPTER SIX: REFERENCES 119

APPENDICES 164

LIST OF TABLES

Table 1: Sub-Saharan African sheep genetic resources………….……………………..3

Table 2: Sampling sites and their GPS positions……………..………………………..48

Table 3: The 15 autosomal microsatellite loci used in the study………………..……..57

Table 4: Number of Alleles sampled from Kenyan Sheep genotyped ……………….73

Table 5: Number of Animals Sampled, Mean heterozygosity, Mean number of Alleles and standard deviations for each of the fifteen populations studied…………74

Table 6: Genetic diversity per locus and population…………………………..…….75

Table 7: Allelic richness per locus and population ……………………………….……76

Table 8: Weir and Cockerham 1984 Multilocus estimates for diploid data after Jackknifing over all loci………………………………………………………78

Table 9: Pairwise tests of Population differentiation between fifteen Kenyan sheep populations……………………………………………………………………..79

Table 10: Within population inbreeding estimates (f_{IS}) in fifteen Kenyan sheep populations across 15 markers………...……………………………………..80

Table 11: Proportion of membership of each of the fifteen sheep populations in each of the four inferred clusters using the program **STRUCTURE** (at ***K*****=4**)………………...……………………………………………………83

Table 12: Proportion of membership of the re-analysed sheep population clusters…………………………………………………………………………86

Table 13: Results of the Fishers exact test on Hardy-Weinberg Equillibrium for all breeds and loci………………………………………………………...…..89

Table 14: Pairwise population matrix of Nei's genetic distances (***D_A***) for the fifteen Kenyan sheep populations……………………………………………………92

Table 15: Analysis of Molecular Variance……………………………………………94

Table 16: The Wilcoxon test for genetic bottlenecks…………………………………95

Table 17: Locus by population table of the estimated null allele frequencies………..96

LIST OF FIGURES

Figure 1: A Kenyan map showing sampling sites……………………………………48

Figure 2: Phenotypic features of sheep genetic resources sampled…………………..51

Figure 3: Graphs showing how the probable number of clusters, ***K*** was determined...81

Figure 4: Population partitioning of the sheep populations as suggested by **STRUCTURE**……………………………………………………........85

Figure 5: An Independent Component Analysis plot for the sheep populations……....87

Figure 6: An unrooted neighbour joining phylogenetic tree showing the relationship among the sheep populations studied……………………..…………………93

LIST OF APPENDICES

APPENDIX I: Protocol for Nucleic acid measurement……………….....……….162

APPENDIX II: Genotypic linkage disequilibrium *p* – values.………………... 164

LIST OF ABBREVIATIONS AND ACRONYMS

AMOVA	Analysis of Molecular Variance
AnGR	Animal Genetic Resources
BC	Before Christ
CNV	Copy Number Variation
D_A	Modified Cavalli-Sforza genetic distance
DAGRIS	Domestic Animals Genetic Resources Information System
DArT	Diversity Array Technology
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetra-acetic acid
FAnGR	Farm Animal Genetic Resources
FAO	Food and Agricultural Organization of the United Nations
G_{ST}	Coefficient of Genetic Differentiation
GOK	Government of Kenya
GPS	Global Positioning System
H_E	Expected Heterozygosity
H_O	Observed Heterozygosity
HWE	Hardy-Weinberg equilibrium
ICA	Independent Component Analysis
ILRI	International Livestock Research Institute
LD	Linkage disequilibrium
MNA	Mean Number of Alleles

mtDNA	Mitochondrial Deoxyribonucleic Acid
PCR	Polymerase Chain Reaction
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
S.D.	Standard Deviation
SNP	Single Nucleotide Polymorphisms
θ_{ST}	Standard variance in allele frequencies among populations

ABSTRACT

The knowledge of the genetic relationship and admixture among neighbouring populations is crucial for conservation efforts. The aim of this study was to analyze the genetic diversity and population structure of the indigenous sheep in Kenya using 15 microsatellite markers. Blood samples from 582 individuals were obtained and genotyped across 15 microsatellite markers. Values of expected heterozygosity and Mean number of alleles ranged from 0.596 to 0.807 and 6.67 to 9.33 respectively depending on the population. Most of the investigated populations showed a significant heterozygote deficiency caused by a moderately high level of inbreeding indicated by f_{IS} (0.109). The observed genetic diversity was found to be high in the nucleus flocks as opposed to those kept by the famers. Genetic differentiation between breeds was moderate (θ_{ST} = 0.101) but significant. The genetic distances obtained reflect the historical knowledge of these breeds and some patterns of ancestral and recent gene flow between neighbour populations arise. According to genetic relationships, multivariate and structure analyses four population clusters were detected which were majorly based on geographical proximity and interbreeding among the populations. These results indicate that the levels of admixture observed warrant the institution of conservation measures if the genetic purity and integrity of the indigenous sheep in Kenya is to be maintained. However, a more encompassing study that includes all the regions in the country known to harbor the indigenous sheep, as well as a larger sample size with more microsatellite markers genotyped is necessary to enable a comprehensive understanding of the dynamics of genetic introgression in the country.

CHAPTER ONE: INTRODUCTION

1.1.0 Background

Sheep are ruminant members of the *Bovidae* family, Subfamily *Caprinae* and they belong to the genus *Ovis*, which contains several species i.e. *Ovis musimon*, *Ovis ammon, Ovis canadensis, Ovis dalli, Ovis nivicolla, Ovis orientalis*, and *Ovis aries* (Nadler *et al*., 1973). Domestic sheep (*Ovis aries*) are quadrupedal, cloven-hoofed mammals kept as livestock, they are the most numerous of their genus, and are most likely to have descended from the wild mouflon (Ryder, 1983). Bökönyi (1976), reports that sheep were among the first animals to be domesticated at about 9 000 BC because of their relatively inexacting food requirements, ease of handling, versatility of their products such as meat, blood, milk and wool and also their important role as sacrificial animals in many indigenous religions. Goats and sheep, often referred to as small ruminants or ovicaprids, are important livestock in all ecological zones that is arid, semi-arid, sub-humid, humid, highland and under various agricultural systems in tropical Africa, they are complementary to cattle and camels in their production cycles and generally do not compete directly with them for feed (FAO, 1991).

Sheep are widespread across the world, as they are adapted to different climatic conditions and econiches (Ryder, 1983). African sheep can be classified as thin-tailed, fat-tailed or fat-rumped (Mason and Maule, 1960) and thin-tailed are sometimes further

segregated into hairy or woolled types (Epstein, 1971). The Eastern African Sheep are classified as either fat-tailed or fat-rumped (Rege *et al*., 1996).

Type	Sub-type	Region	Breeds
Thin-Tailed Hair Sheep	African long-legged Tropical Dwarf	West African/Sahelian Zaire Angola Sudan Sudan Desert West African Zaire Angola Sudan	Fulani, Maure, Touareg, Zaire Long-legged, Baluba Angola long-legged, Zunu Fellata;Zaghawa Sudan Desert Djallonk'e Southern Bahu Angola Maned South Sudanese
Fat-Tailed Hair Sheep	East African Fat-Tailed African Long-Fat-Tailed	Ethiopia East African East African Southern African	Ethiopian Red Maasai, East African Blackheaded Ruanda-Burundi, Tanzania Long-Tailed Hottentot, Africander, Madagascar, Malawi, Sabi, Mondombes, Nguni, Tswana
Fat-Rump Hair Sheep Fat-Tailed Fur Sheep Thin-Tailed Coarse Wool Sheep		East Africa Southern Africa West Africa East Africa	Somali, Southern Africa Karakul Macina, Barbarin, Toubou, Dan'e Zaila Arrit, Dongala
Fat-Tailed Coarse Wool Sheep	East African Fat-Tailed	East African Southern Africa	Ethiopian, Somali Arab Woolled Persian
Fine Wool Sheep		Southern Africa	South African Merino, South African Mutton Merino; Ronderib Merino
Developed Breeds	Long-Legged Hair Sheep Fat-Rump Hair Sheep Coarse Wool Sheep Fine Wool	West Africa Sudan Desert West Africa Southern Africa West and Central Africa Southern Africa Southern Africa	Toronk'e, Vogan, Waral'e Ingessana, Meidob Nungua Blackhead Wiltiper, Van Rooy Okouma, Permer Bezuidenhout Africander, Dormer, Dorper, Vandor, White Woolled Mountain Afrino, Dohne Merino, Letelle Merino, Walrich Mutton Merino

Table 1: Sub-Saharan African sheep genetic resources (Rege *et al* ., 1996)

Food and Agricultural Organization (FAO) (2004a), documents that of the world's sheep population about one third is found in developed and two thirds in developing countries. The indigenous animal genetic resources are believed to preserve much of the current global genetic diversity with millions of people directly depending on them for their livelihood (Rege & Gibson 2003). Devendra (2002), reports that more production units own goats and sheep in African farming systems than any other species of domestic livestock except poultry. Small ruminants are particularly important for resource-poor households and are often the property of underprivileged groups, such as women and children, within those households (Kosgey *et al.,* 2008).

Kosgey *et al.,* (2006a, b) indicates that in Kenya, small ruminants are kept for both tangible benefits like cash income from animals, milk and meat sales and for home consumption and intangible benefits like savings, an insurance against emergencies, cultural and ceremonial purposes. A study by Kosgey *et al*., (2006b) ranked regular cash income as the most important purpose of small ruminants for both smallholders and pastoral extensive farmers. As reported by Ministry of Livestock and Fisheries Development report (2003) in Kenya, sheep supply an estimated 15–20 % of the red meat consumed in the country. The Sheep and Goat Annual Report in Kenya (2003), records that the Indigenous fat-tailed sheep breeds found in Kenya include the Red Maasai and East African Somali Blackhead Persian, which are found in virtually any part of the country. Arid and semi arid land forms 85 % of Kenya's diverse ecological zones and it is also the home to most of the indigenous sheep genotypes (Gathuka 1986).

FAO (2007), reports that 20 % of all breeds with reported population data are at risk of extinction, however, the population status of many breeds is still unknown, and the problem may thus be underestimated. The main threat to continuing utilization of indigenous animal genetic resources is the transformation of traditional systems into external input-oriented systems, which typically use standard common production breeds that displace local breeds. The indiscriminate crossbreeding with exotic breeds is also rapidly compromising the genetic integrity of local populations (FAO, 2000). The introduction of exotic breeds is being practiced on the premise that the exotic breeds are more superior to indigenous breeds because they produce more and better quality meat and milk. These exotic breeds however, require specialized inputs, including housing and feeds that are not widely accessible to majority of the resource-poor farmers resulting in reduced profitability (Devendra, 2002). The erosion of animal genetic resources is a long-term threat to ensuring food security and rural development. Without strategically planned interventions, using both *in situ* and *ex situ* conservation, erosion will continue and may in fact accelerate (FAO, 2000). Developing countries have notably higher rates than others and socio-political, economic, and agro-ecological determinants are leading to the most dramatic genetic erosion of AnGR important to the livelihoods of many of the world's poor (Anderson, 2003).

Most indigenous breeds are managed through traditional husbandry only (Bouzat *et al.*, 1998a); therefore, they are to a great degree subjected to the process of natural selection. As a consequence, these breeds have become locally adapted to a wide range of environments, showing high levels of phenotypic variability and good fitness under

natural conditions (DAGRIS, 2006). As reported by Baker *et al*., (1999;2003) the Red Maasai sheep of East Africa are both more resistant and tolerant (resilient) to naturally acquired and artificial infections with gastro-Intestinal nematode parasites, particularly the highly pathogenic nematode parasite *Haemonchus contortus* than other breeds with which they have been compared, notably the Dorper. The potential for the sustainable utilization of these resources, believed to have adaptive advantages in their respective production systems, largely depends on our ability to quantify current and future diversity levels, estimate risk of extinction of breeds, prioritize breeds and allocate conservation funds according to their relative contribution to future and current diversity based on suitable diversity measures (Rege and Gibson 2003; Simianer *et al*., 2003; Bennewitz and Meuwissen 2005). Characterization of genetic structure and variation of local populations is an important step towards identifying unique and valuable genetic resources. The use of Microsatellites has become a standard method to estimate neutral genetic diversity in livestock (Baumung *et al*., 2004).

Microsatellites are DNA regions with variable numbers of short tandem repeats flanked by unique sequences. They are polymorphic loci present in nuclear and organellar DNA that consist of tandemly repeating units that range in size from 1-6bp. Some advantages of microsatellites relative to other traditional molecular markers include: high level of polymorphism and genomic abundance, random distribution throughout the genome, ease of reproducibility, highly codominant and reasonably amenable to automation. This type of marker has disadvantages such as occurring at subsets of loci as non-detectable alleles (so called “null alleles”), difficulty to interprete

the PCR products as alleles, and size homoplasy (Schlötterer, 2004). Microsatellites have been successfully applied to parentage and relatedness testing in horses (Bowling *et al.*, 1997; Marklund *et al.*, 1994), and their usefulness for estimating genetic distances among closely related populations has been documented. With the advent of molecular techniques an increasing number of studies have focused on the genetic characterization of domestic breeds through molecular genetic markers. These studies included comparisons of genetic variability among breeds (Hanslik *et al.*, 2000), phylogenetic studies on the historical origin of domestic species (Edwards *et al.*, 2003; MacHugh *et al.*, 1997) among others.

Muigai (2003) characterized eighteen indigenous sheep breeds of Eastern, Western, Northern and Southern Africa, amongst which was one Red Maasai. This was part of a larger study aimed at characterizing the genetic diversity of the sheep of Africa, Europe, Asia and the Near East. This study was followed by the characterization of indigenous sheep from nine Kenyan populations sampled from west of the Rift Valley (Migori, Homa Bay/Kendu Bay, Transmara and Kakamega) and east of the Rift Valley (Kajiado, Naivasha/Olmagogo, Mutara, Laikipia East and Laikipia West) (Muigai *et al.,* (2009).

This current study set out principally to study the genetic diversity and population structure of the uncharacterized indigenous and exotic sheep in Kenya.

1.2.0 Assumptions of the Study

i. The gene-pool of indigenous sheep in Kenya is inhomogeneous on national level, but there are sufficient pedigree links between neighbouring flocks so that the whole gene pool can be characterized using moderately dense sampling.

ii. Some indigenous sheep populations in Kenya (particularly those kept by farmers) harbour a significant proportion of foreign ancestry while the others (particularly those in the nucleus herds) show a low-level of introgression with the exotic breeds.

iii. The indigenous sheep in Kenya are not significantly inbred in comparison to those containing foreign ancestry and differences in the genetic diversities are insignificant.

1.3.0 Hypotheses of the Study

1. H_A The indigenous sheep in Kenya have a high level of genetic diversity.

   H_O The indigenous sheep in Kenya have a low level of genetic diversity.

2. H_A The indigenous sheep in Kenya show a high level of genetic admixture with exotic genotypes.

   H_O The indigenous sheep in Kenya show a low level of genetic admixture with exotic genotypes.

3. H_A The indigenous sheep in Kenya show a high level of population structure.

   H_O The indigenous sheep in Kenya show a low level of population structure.

1.4.0 Objectives of the Study

1.4.1 General Objective

To characterize the indigenous sheep in Kenya for their rational management and conservation.

1.4.2 Specific Objectives

i. To determine the population structure and genetic relationship of the indigenous sheep in Kenya.
ii. To determine the genetic diversity within and between the indigenous sheep in Kenya.
iii. To assess the level of genetic admixture of the indigenous sheep breeds/ populations in Kenya with exotic breeds.

1.5.0 Justification and significance of the study

Indigenous animal genetic resources have developed some unique features that are key to sustainable agricultural practice i.e. disease resistance, the ability to walk for long distances in search for pasture and water among others. These unique indigenous genetic resources are however being lost in most parts of the country due to indiscriminate crossbreeding and introduction of exotic or other indigenous sheep genotypes. In addition to the exotic breeds, successive droughts, civil strife, pressures for economic development are threatening these unique genetic resources before they are characterized to quantify their diversity and population structure so as to be documented before they become extinct. Therefore the information generated from this study could be found vital in the rational management and conservation of these indigenous sheep genetic resources.

CHAPTER TWO: LITERATURE REVIEW

2.1 Origin and Domestication of Sheep

The Mesolithic man began the process of sheep domestication approximately 8 000 to 9 000 years ago (Ryder, 1984) and this ongoing process has resulted in the establishment of more than 14 000 breeds (Scherf, 2000). These breeds are defined by phenotypic variation in coat colour, environmental tolerance, wool characteristics and food production traits.

Domestic sheep have played an important role in human history, but the origins of the modern domestic sheep *(Ovis aries)* are not well understood (Ryder, 1984). Based on archaeological evidence, sheep were probably first domesticated in the Fertile Crescent region of Southwest Asia approximately c 9 000 years ago (Ryder, 1983 and 1984). The earliest zooarchaeological evidence for domestic sheep comes from sites in present-day Iran, Turkey, and Cyprus (Hahn, 1896; Uerpmann 1979; Ryder, 1984; Clutton-Brock, 1999; Vigne *et al.,* 2003).

Sheep are thought to have evolved from the goat-antelope, *Rupicaprini* represented by the *Capriconis* of Southeast Asia, this has been supported by paleontology and behavioral evidence (Geist, 1971). During the late Pleistocene period, goats and sheep formed an interbreeding population (Payne, 1968) however, blood antigens, blood proteins and chromosome structure showed large differences between sheep and goats (Lay *et al.*, 1971). The fact that present-day interbreeding between sheep and goats is not very successful makes such a link unlikely (Dain, 1970; Curtain, 1971; Lindley *et al.*, 1971). Wild sheep have survived in large numbers despite the

presence of man (Ryder, 1984). The home of the wild sheep is the mountain ranges of Central Asia, from where sheep spread westward into Europe and eastwards into North America during the Pleistocene period (Ryder, 1983).

There are six species of wild sheep in existence, which could have given origin to our domestic breeds (Ryder, 1983, 1984). The most important of these are the Argali (*Ovis ammon*), the Urial (*Ovis orientalis*), the Mouflon (*Ovis musimon*) and the Bighorn (*Ovis canadensis*). All domestic breeds of sheep are thought to have descended from the Mouflon (*O. musimon*), although the Urial (*O. orientalis*) may have contributed to European breeds. Domesticated sheep have 2n = 54 chromosomes, the same chromosome number as the European mouflon, the Asiatic mouflon, the Bighorn and the Dall sheep. The snow sheep of eastern Siberia has 52 pairs of chromosomes, the Argali, 56 and the Urial, 58 (Ryder, 1984; FAO, 2000). It is believed that the Urial was domesticated first, since Urial remains have been found around the area where domestication appeared to have begun (Ryder, 1984). The Urial is thought to have arrived in Europe first and later the Mouflon. The two species are thought to have mixed (Ryder, 1983). It is also believed that the Urial gave rise to “wool” sheep and that “hair” sheep originated from the Mouflon (Zeuner, 1963). Others believe that only a single wild species contributed to the gene pool of present day domestic sheep. This theory has been supported by chromosome counts and blood protein analysis (Schmidtt and Ulbright, 1968; Ryder, 1984).

In recent studies by Hiendleder *et al.,* (1998; 2002), the origin of domesticated sheep was investigated using mitochondrial DNA (mtDNA) where sixteen mtDNA

haplotypes among 243 domesticated sheep of European, Asian and Central Asian origin were identified. None of these haplotypes were present in Urial or Argali, thus excluding these two species as ancestors of present-day domesticated sheep. However, some of these haplotypes were found in wild populations of Mouflon, strongly indicating that this species has contributed to the genetic pool of domesticated sheep. There are a number of different theories regarding the origins of domestic sheep. Sheep (*Ovis aries*) were among the earliest livestock species to be domesticated. As ruminants, they provided humankind with a means of digesting via fermentation, a substantial proportion of the fibrous material produced by grasslands, which single stomach or monogastric species are less able to digest. Sheep (*O. aries*) evolved in Eurasia in the early Pleistocene period about 2.5 million years ago (Ryder, 1983). The first sheep that appeared in the Villafranchian period were as large as oxen. By the end of this period, approximately three million years ago, the first true sheep replaced these 'oxen.'

African sheep are thought to be of Near-Eastern origin (Epstein, 1954, 1971; Ryder, 1984; Marshal, 2000). The earliest sheep in Africa were thin-tailed and hairy and introduced to East Africa through North Africa (Marshal, 2000). The second wave of sheep introduction to Africa included fat-tailed sheep entering North Africa *via* the Isthmus of Suez straits and East Africa *via* straits of Bab-el-Mandeb (Ryder, 1984). Fat-rumped sheep entered East Africa much later (Epstein, 1954, 1971; Ryder, 1984).

Accordingly, African sheep have been traditionally described and classified based on their tail type (Epstein, 1971; Ryder, 1984). Sheep are therefore not native to Africa. To date, studies into the process of domestication and the structure and

relationship between modern *Ovis* populations have relied on autosomal microsatellites and mitochondrial (mt) DNA. Allelic polymorphism at microsatellite loci has proven particularly successful for distinguishing closely related breeds (Arranz *et al.,* 2001; Diez-Tascon *et al.,* 2000) and for precisely assessing population structure even where breed classification is a poor predictor of genetic similarity (Tapio *et al.,* 2005).

2.2 Sheep Genetic Resources in Kenya

FAO (2004) documents that the population of small ruminants in Africa and Asia are increasing, this is due to their better adaptation to the prevailing conditions and suitability to small-scale farms. In sub-Saharan Africa, small ruminants play a crucial role in sustaining agricultural production. According to Lebbie and Ramsay (1999), small ruminants account for 62 % of the total number of domesticated ruminant livestock in this region, with 34 % and 28 % for goats and sheep, respectively. Of the sheep population, 57 % are found in the drier and fragile arid and semi-arid zones. Arid and semi arid lands (ASALs) cover 80 % of the total land surface in Kenya and provide subsistence economy to 25 % of the population who are mainly pastoralists and agropastoralists (GOK, 2002a).

According to FAO (2000), in Kenya the small ruminant population is estimated at approximately 15 million and they are kept by over 75 % of farmers (Peters, 1988). A report by the MLFD (2003) shows that sheep make up 47 % of this population and their number is growing at a rate of 3.6 % annually. In Kenya, sheep are predominantly kept under pastoral production systems in arid and semi-arid lands (Kinyamario and Ekeya,

2001) where according to Kosgey (2004), livestock forms a pivotal part of the socio-cultural life of the rural people. The two types of indigenous sheep found in Kenya are fat-tailed and fat-rumped hair sheep (Mason, 1996). The significance of these breeds lies in their adaptation to the stressful environmental conditions and the traditional husbandry systems (Rege, 1994). There is evidence that they are resistant to diseases endemic to Africa such as haemonchosis, trypanosomiasis and blue-tongue (Baker *et al*., 1999; Wanyangu *et al*., 1993).

2.3.0 Importance of Sheep Genetic Resources

Globally, domestic Animal Genetic Resources (AnGR) supply some 30 % of total human requirements for food and agricultural production (FAO, 1999). Delgado *et al.,* (1999) shows that the poor and landless derive a higher proportion of household income from livestock sources than do those with greater wealth living in the same communities. As pointed out by Ledin (1997), "*Small ruminants are generally kept by poor people and are often tended by women, who seldom have any influence on the situation. Small ruminants and the people who keep them are held in low esteem and given few priorities in development*". As such, sheep and goats are often the forgotten livestock in many pastoral systems. For example, in his description of main livestock raised by pastoral groups, Blench (2001) mentioned only cattle for Maasai and only dromedary camels for Raika, with no mention of sheep and goats. Tibetan and Mongolian pastoralists are associated with yaks, but sheep and goats are much more numerous and, in many regions, are more important than yaks (Naess *et al.,* 2004).

Kosgey (2004), records that small ruminants (i.e. sheep and goats), especially the indigenous breeds, are widespread and important to the subsistence, economic and social livelihoods of a large human population in developing countries. As reported by Meadows *et al.,* (2005) sheep are a highly adaptable and versatile domestic species, this has made them a critically important resource in human societies around the world.
Gicheha *et al.,* (2006) reports that sheep production in Kenya has a significant role both at household and national levels. Sheep contribute 15-20 % of the red meat consumed in the country (MLFD 2003). As indicated by Baker and Rege (1994), sheep play a vital role in the utilization of vast natural pastureland where crop production is not feasible.

Devendra (2002) points that these animals have lower feed and capital requirements than larger species, making them suited to smallholder producers. They also have lower generation intervals, higher prolificacy, small size, and are better able to utilize a wide range of feedstuffs, including crop residues that are of little value otherwise (Holst, 1999; Pelant *et al.,* 1999). Sheep are an important protein source in the diets of the poor and help to provide extra income and support survival for many farmers in the tropics and subtropics. For example, sheep contribute 15-20 % of the red meat consumed in the Kenya (MLFD 2003). Being multipurpose animals producing meat, milk, skins, wool/hair, financing, insurance and risk aversion (Upton, 1985; Jaitner *et al.,* 2001), payment of bride price and use as gifts (Grandin *et al.,* 1991), as a status symbol or sign of wealth and as a form of ''currency'' in which social obligations are expressed (Rege, 1994) sheep form (next to cattle) the most important group of ruminants in both temperate and tropical agriculture.

The implications of these additional roles of livestock on biological productivity are often disregarded in favour of technical facets such as nutrition and reproduction (Gatenby, 1986; Bosman *et al.,* 1997), probably due to the difficulty of measuring and valuing them (Roeleveld, 1996). According to Devendra (2005), the importance and extent of the contributions of small ruminants, especially to the poor in the rural areas, are inadequately understood, this means that these valuable genetic resources continue to be generally neglected. The lack of data on the values of indigenous AnGR contributes to the undervaluation of the values and as a result, the erosion of biodiversity (Rege and Gibson 2003; Wollny 2003) which would not otherwise occur.

2.4.0 Management of Animal Genetic Resources

The term Animal Genetic Resources (AnGR) includes all animal species, breeds and strains (with their wild relatives) that are of economic, scientific and cultural interest to humankind in terms of food and agricultural production for the present or in the future.

Management of these AnGR involves the sum total of technical, policy, and logistical operations involved in understanding (characterization), using and developing (utilization), maintaining (conservation), accessing, and sharing the benefits of animal genetic resources (FAO, 1998).

Scarpa *et al.,* (2003a) points out that management of animal genetic resources (AnGR) requires many decisions that would be facilitated by the availability of information regarding the economic value of breeds, traits and alternative breeding

and/or conservation strategies. As concluded by FAO (1997), the conservation and sustainable development of animal genetic resources (AnGR) requires a shift towards a broad focus on the many 'adaptive' breeds that survive well in the low external input agriculture typical of developing countries'.

Four main livestock keeping systems can be identified (Waters-Bayer and Bayer, 1992):

i. Full-time livestock keepers who depend primarily on livestock for their livelihoods (they may be nomadic, sedentary or transhumant).
ii. Livestock-keepers who do some cropping but livestock remain their main means of living (may be transhumant or settled).
iii. Crop farmers who also keep animals and usually stay in one place all year round.
iv. The landless that keep some livestock often as a subsidiary activity and live on the edge of villages, towns or cities.

Women livestock keepers often fall into the small stock keepers or the landless livestock keeper categories depending upon their land endowment and right of use within the household.

Approximately 64 % of the world's small ruminants are kept in smallholdings (de Haan *et al.,* 1996) and the number of animals under this system is increasing (FAO, 2004). In smallholder production systems, small ruminants are kept under a wide range of minimal husbandry conditions (Holst, 1999). Ownership of and management decisions on small ruminants differ between communities (Kosgey, 2004), but well-defined breeding strategies are rare.

As reported by Olivier *et al.,* (2002) smallholder farmers are resource-limited and hence are risk averse. The poor tend to be both producers and consumers of livestock products and services and different livestock may have multiple livelihood functions (Dorward *et al.,* 2001). Unlike commercial farmers, they tend to keep animals for multiple needs, and not only as an economic enterprise, i.e., tangible benefits and intangible benefits are important (Kosgey, 2004). As indicated by Amer *et al.,* (1998) increased productivity is not usually their goal, especially when production risks are high. Consequently, production levels are often lower than the economic optimum and technical maximum. And usually *ad hoc* sales of animals are practised to meet emergency needs for cash (Kosgey, 2004).

Kosgey and Okeyo (2007), points out that the potential for small ruminants to contribute more to the livelihoods of people in low-input, smallholder production systems is great, but for that to be achieved, efficient and sustainable genetic improvement programs are required to boost output and profitability, and meet the peoples' needs for the animal products. On the contrary Kosgey *et al .,* (2006) indicates that despite the large numbers and the importance of small ruminants in developing countries, information on sustainable genetic improvement programs under small-holder production circumstances is scarce especially for the adapted breeds.

2.5.0 Conservation of genetic diversity

Conservation of farm AnGR incorporates preservation, maintenance, improvement and sustainable utilization (FAO, 1986). The primary focus of farm AnGR

conservation is on the conservation of breeds, including management for better utilization (breeding programs) and conserving those at risk, with the aim of minimizing the loss of diversity among breeds (Barker, 2001).

2.5.1 Farm animal genetic diversity

Frankham *et al.,* (2002) defines genetic diversity as the variety of alleles and genotypes present in a population and this is reflected in morphological, physiological and behavioural differences between individuals and populations. From a functional point of view, genetic diversity can be classified as neutral, deleterious or adaptive (Hedrick, 2001). Generally, neutral variants are used for conservation applications, but deleterious and adaptive variations are also important in the contexts of population survival and economically important traits in domestic plants and animals (Caballero and Toro, 2005). One of the most important considerations in the conservation of Farm Animal Genetic Resources (FAnGR) is the criterion to use in choosing appropriate breeds or populations for conservation, among the many available, so that maximum genetic diversity is maintained.

Many species of animals have been domesticated since the initial domestication of the first ruminants. Much of the existing diversity today has emerged as a consequence of settlement in different regions of the world leading to physical isolation, and as a result of human development of these species causing reproductive isolation. The resulting breeds and strains are locally adapted to suit different ecological

environments and human needs. These breeds and strains constitute the global animal genetics resource.

Genetic variation at the population level consists of sets of alleles whose frequencies may vary between populations. Variation between populations is important, being derived from local adaptation processes thus implying increased fitness for that particular environment. Random processes such as founder effects, demographic bottlenecks, mutations, genetic drift and selection cause this variability. It is instructive therefore to maintain both the within and between population components of variation through conservation efforts. This ensures the better utilization of present resources while preserving the potential to face future challenges in terms of unexpected changes and needs (FAO, 1995).

2.5.2 Loss of genetic diversity

Genetic diversity can be lost, with a reduction occurring both within and between populations. However, this reduction in diversity is generally expressed in terms of loss of breeds and strains. It is estimated that at least 30-40 % of all domestic animal genetic resources are at risk of extinction (FAO, 1994). This may be an underestimate since most domestic animals have not been characterized and their status is unknown. Factors that diminish genetic diversity include:

i. Genetic bottlenecks, which occurs when a population encounters a severe temporary reduction in size due to environmental or demographic events such as natural disasters, drought, war or disease outbreaks. The few

individuals that survive contain the overall diversity of the resultant population, thus some genetic diversity may be lost in the process.

ii. Random genetic drift leads to fixation of some alleles at certain loci, thus reducing variability after many generations of drift.

iii. Inbreeding allows the expression of harmful rare recessive alleles in homozygous state with resultant harmful effect on the offspring especially a reduction in reproductive fitness, thus increasing the risk of extinction (Soulé, 1980).

iv. Anthropological activities such as habitat destruction, pollution, overexploitation and introduction of exotic species represent the greatest threat to animal diversity. The number of individuals in the breed or the populations will reduce, with consequent loss of genetic variation.

v. Genetic uniformity through directional selection for a few economical traits and artificial selection of some reproductive individuals (through artificial insemination and embryo transfer) will contribute to a reduction of genetic diversity (Blench, 2003).

The danger is that the number of individuals in a breed or population may become so small that the population or breed is not viable. Conservation, on the other hand, refers to all human activities including strategies, plans, policies and actions undertaken to ensure that the diversity of farm animal genetic resources is being maintained to contribute to food and agricultural production and productivity, now and in the future. The underlying justification for conservation is to protect the Farm Animal

Genetic Resources (FAnGR) against risks coupled with imperfect knowledge of what attributes FAnGR currently possess and what economic, social and cultural needs local, regional and global society will have in the short-, medium- and long-term future.

Conservation entails the following:

a) Defining utilization strategies for existing breeds to maximize production and improve efficiency of sustainable use.

b) Placing of optimum strategies for assessment and preservation of the genetic diversity in breeds not currently favored for commercial use. This also requires the heightening of the awareness of the diverse contributions made by the indigenous adapted breeds to sustainable development of agrosystems.

c) Identifying and listing all breeds and the remaining animal genetic resources for objective priority setting. Description and characterization of breeds will help identify unique qualities and potential contributions. This aids in postulating which breeds have a potential for a great variety of use in future.

d) Monitoring the population statistics for each breed and regularly reporting those at risk of extinction.

e) Storage of samples of many breeds generally in the form of frozen semen, ova and embryo to enable future regeneration of lost populations of animals. This is normally the last resort for conservation after extinction is imminent.

2.5.3 Methods of conservation of genetic resources

2.5.3.1 *In-situ* conservation

This involves maintenance of breeds in their production systems and agro-ecological zones of origin. The *In situ* conservation approaches should be preferred for conservation where maintenance and management of the FAnGR is the best available livelihood option for the farmers involved .These methods are advantageous in that they allow natural evolution of the breeds to continue. However where populations are small, selection and drift may lead to unfavourable changes and subsequent loss of fitness.

2.5.3.2 *Ex-situ* conservation

This type of conservation involves the preservation of live animals and or samples in places other than their natural habitats. Two main approaches are used; *Ex-situ in vivo Ex-situ in vitro.*

Ex-situ in vivo conservation where live animals, as a sample of a breed are kept away from their natural habitat in farm parks and collections where there is more control in the management of the population. *Ex situ in vivo* conservation in institutional or communally owned herds or flocks can successfully be used to support conservation of FAnGR that have current value. Virtually all examples of *ex situ in vivo* conservation of FAnGR (in the developing world) are designed to support current use by farmers or expected use in the near future or are populations being maintained for research purposes. Creating conservation herds involves crossing rare breeds and then breeding

them to maintain variability. However, although useful genes may be preserved, breed identity is usually sacrificed.

Ex-situ in vitro conservation where cryopreservation makes use of freezing to store useful breeds in form of semen, ova, embryo or tissue, which can be used to regenerate organisms. *In vitro* conservation is usually recommended to provide a secure back-up to protect against a variety of threats that can drive FAnGR to extinction faster than monitoring can identify the threat and also faster than alternative conservation approaches can respond.

2.6.0 Molecular markers used in diversity studies

Molecular genetics is highly informative for investigating the relationship between animal populations as well as documenting the levels of genetic variation resident within breeds (Meadows *et al.,* 2005). A molecular marker is a fragment of DNA sequence that is associated to a part of the genome. They are used in molecular biology and biotechnology experiments to identify a particular sequence of DNA.

However, due to the existence of various molecular biology techniques to produce them, a large variety exists, from which choices will have to be made according to the purpose. Two main points have to be considered when using molecular markers for genetic studies. From the molecular technique point of view, the genotyping procedure should be simple, reliable, reproducible and cost efficient in order to generate the required amount and quality of genotyping data. From the population genetic analysis point of view, characteristics such as the dominance relationships, information content,

neutrality, map positions or genetic independence of markers are important (Vignal *et al.,* 2002).

In recent years, molecular markers and especially DNA-based markers have been extensively used in many areas such as gene mapping and tagging (Kliebenstein *et al*., 2001; Karp and Edwards, 1997), analysis of genetic diversity (Erschadi *et al*., 2000; Palacios *et al*., 1999) or genetic relatedness. In population genetics, protein-based markers (allozymes) were the first markers developed and widely used (Hamrick and Godt, 1990). For example, Mwacharo *et al*., (2002) used blood protein polymorphisms to characterise Kenyan sheep populations. DNA-based methodologies are now the method of choice to differentiate closely related organisms. Moreover, the use of DNA-based markers allows efficient comparisons because genetic differences are detectable at all stages of development of the organism unlike allozymes which may show age-dependent changes.

A good molecular marker for a population geneticist should have the following features:

i. Show Mendelian inheritance i.e. should transmit from one generation to another.

ii. Be polymorphic i.e. present several alleles at the locus being investigated (multiallelic).

iii. Be codominant i.e. allow the discrimination between homozygotes and heterozygotes.

iv. Be neutral i.e. all alleles have the same fitness.

v. Not epistatic i.e. one can determine the genotype of a phenotype irrespective of the genotype of the other loci.
vi. Be independent of the environment i.e. no phenotypic plasticity.
vii. Have sufficient occurrence in the genome.
viii. Be evenly distributed throughout the genome.
ix. Be highly reproducible.

The most frequently used markers in population genetics are allozymes (biochemical), Randomly Amplified polymorphic DNA (RAPD; Williams *et al*., 1990), Restriction Fragment Length Polymorphism (RFLP; Botstein *et al*., 1980), Amplified Fragment Length Polymorphism (AFLP; Zabeau and Vos 1993) minisatellite finger prints, Single Sequence Repeats (SSR; Tautz and Renz, 1989) and Single Nucleotide Polymorphisms (SNP).The choice of a specific molecular marker depends on its suitability to answer a particular question. For this purpose the main difference among molecular markers is their degree of dominance and amount of obtainable data. Codominant markers enable an easy estimation of allele frequencies hence are suitable to study and or estimate gene flows between populations. Dominant markers can estimate genotypes but not the allele frequencies. Dominant markers are preferably used as fingerprints (Mueller and Wolfenbarger, 1999; Hongtrakul *et al*., 1997; Weising *et al*., 1995) and can be helpful in the identification of clones.

2.6.1 Random Amplified Polymorphic DNA (RAPD)

This method is based on the Polymerase Chain Reaction (PCR) using short (usually 10 nucleotide) primers of arbitrary sequences. The amplification protocol differs from the standard PCR conditions (Erlich, 1989) in that only a single random oligonucleotide primer is employed and no prior knowledge of the genome subjected to analysis is required. Each product is derived from a region of the genome that contains two short segments in inverted orientation, on opposite strands, which are complementary to the primer and sufficiently close together for the amplification to work. The amplification products are resolved on agarose gels and polymorphisms serve as dominant genetic markers, which are inherited in a Mendelian fashion (Williams *et al.,* 1990; Carlson *et al.,* 1991).

Polymorphism of amplified fragments are caused by: (i) base substitutions or deletions in the priming sites, (ii) Insertions that render priming sites too distant to support amplification, or (iii) Insertions or deletions that change the size of the amplified fragment.

The disadvantage for using RAPD is the dominant nature of the marker which makes it impossible for the identification of heterozygotes.

2.6.2 Restriction Fragment Length Polymorphism (RFLP)

A restriction fragment length polymorphism is defined by the existence of alternative alleles associated with restriction fragments that differ in size from each other. Restriction fragment length analyses uses restriction enzymes (RE) to cut DNA at

specific 4-6 bp recognition sites (Dowling *et al.,* 1990). Sample DNA is cut (digested) with one or more RE's and resulting fragments are separated according to molecular size using gel electrophoresis (Avise, 1994). Molecular size standards are used to estimate fragment size. Differences result from base substitutions, additions, deletions or sequence rearrangements within RE recognition sequences (Avise, 1994). Restriction fragment length polymorphism (RFLP) is most suited to studies at the intraspecific level or among closely related taxa. Presence and absence of fragments resulting from changes in RE's recognition sites are used identifying species or populations. RFLPs are codominant, typically neutral, genetic markers.

The main advantage of RFLP analysis over PCR-based protocols is that no prior sequence information, nor oligonucleotide synthesis, is required. Furthermore, in some cases, it may not be feasible to develop a PCR protocol to detect a particular form of allelic variation. The detection of a RFLP, in and of itself, does not provide information as to the mechanism by which it was created. RFLPs can be generated by all of the mechanisms through which DNA variation can occur. The simplest RFLPs are those caused by single base-pair substitutions. However, RFLPs can also be generated by the insertion of genetic material, such as transposable elements, or by tandem duplications, deletions, translocations, or other rearrangements.

2.6.3 Amplified Fragment Length Polymorphism (AFLP)

Amplified-fragment Length polymorphism (Zabeau and Vos, 1993) is a whole-genome fingerprinting method based on selective amplification of restriction fragments

(Vos *et al.,* 1995). The AFLP reaction is a multistep procedure which in an elegant manner combines the power of Polymerase Chain Reaction (PCR) with the informativeness of restriction enzyme analysis. The procedure includes the preparation of an AFLP template where genomic DNA is digested with two restriction endonucleases which produce cohesive fragment ends and cut DNA with different frequencies (rare cutter [RC] and frequent cutter [FC]). Following digestion, genomic restriction fragments are modified by ligation of synthetic, double-stranded oligonucleotide adapters (RC adapter and FC adapter) with ends complementary to those of the restriction fragments. Thus, after the ligation step, genomic restriction fragments have termini of known sequences. Such an AFLP template is submitted to a highly stringent PCR amplification with primers fully complementary to their targets (RC primer and FC primer). PCR amplification of restriction fragments is achieved by using the adapter and restriction site sequence as target sites for primer annealing. The selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotides flanking the restriction sites. The amplified products are separated on a sequencing gel and can be visualized using radioactive or fluorescent labelling. AFLP usually yields more complex banding patterns than most of the available DNA fingerprinting methods, which may likely increase discrimination between populations under study. The strategy of using two restriction enzymes and selective amplification provides an extraordinary flexibility in designing the typing protocols optimal for the given species and the chosen detection system as well. These features, combined with

the possibility for automation and a high-throughput analysis, make AFLP an interesting alternative to the mostly used whole-genome fingerprinting techniques. The disadvantage for using AFLP is the dominant nature of the marker which makes it impossible for the identification of heterozygotes.

2.6.4 Single Nucleotide Polymorphisms (SNPs)

A Single Nucleotide Polymorphism is a change in the nucleotide composition of a DNA sequence at a single site, SNPs occur at a frequency of approximately 0.3–1 SNP/kb throughout the human genome (Marth *et al.,* 2001) and apparently at equal frequencies in other mammalian species. For a base position with sequence alternatives in genomic DNA to be considered as an SNP, the least frequent allele should have a frequency of 1 % or greater. Although in principle, at each position of a sequence stretch, any of the four possible nucleotide bases can be present, SNPs are usually biallelic in practice. One of the reasons for this is the low frequency of single nucleotide substitutions at the origin of SNPs, estimated to being between 1×10^{-9} and $5 \text{ x } 10^{-9}$ per nucleotide and per year at neutral positions in mammals (Martinez-Arias *et al.,* 2001). SNPs can serve as genetic markers for identifying disease genes by linkage studies in families, linkage disequilibrium in isolated populations, association analysis of patients and controls, and loss of- heterozygosity studies in tumors (Risch *et al.,* 1996).

Clear identification and characterization of large numbers of SNPs is necessary before we can begin to use them extensively as genetic tools. The methods commonly used for SNPs (mutation) detection are:

i. Identification of Single strand conformation polymorphism: The fragment spanning the putative SNP is PCR amplified, denatured and run on a non-denaturing polyacrylamide gel. During the gel run, the single stranded fragments adopt a secondary structure according to sequence. Fragments bearing SNPs are identified as a result of their aberrant migration pattern and confirmed by sequencing. This is a widely used and relatively simple technique with a success rate ranging from 70-95 %, it is labour intensive and has relatively low throughput.

ii. Heteroduplex analysis: This relies on the formation of a Heteroduplex during re-annealing of denatured strands of a PCR product derived from an individual heterozygous for the SNP. The Heteroduplex can be detected as a band shift on a gel or by differential retention on a high performance liquid chromatography column (HPLC). This method is credited for simplicity, low cost, and high rate of detection.

iii. Direct DNA sequencing: This is the most favoured high-throughput method for SNP detection. Once the sequencing reactions have been completed, a single Applied Biosystems 3700 capillary system can generate sequence from 1500 DNA fragments of 500 bp in 48 hours with minimal human intervention. Dye-terminator sequencing chemistry will detect 95 % of heterozygotes; the more expensive, labour intensive dye-primer chemistry will detect 100 %.

iv. Variant detector arrays: This is a relatively recent SNP detection method. This technique allows identification of SNPs by hybridization of a PCR product to

oligonucleotides arrayed on a glass chip and measuring the difference in the hybridization strength between the matched and mismatched oligonucleotides.

Although less informative than multiallelic microsatellites, the biallelic SNPs possess considerable advantages:

i. They have lower mutation rates hence they are extremely stable (Sachidanandam *et al.,* 2001).
ii. More robust in laboratory handling and data interpretation (Krawczak, 1999).
iii. Suitability for standardized representation of genotyping results as a digital DNA signature (Fries and Durstewitz, 2001).
iv. Suitability for various genotyping techniques and high potential for automation (Kruglyak, 1997).
v. Mutations observed as SNPs are abundant and widespread in many species' genomes (coding and non-coding regions), and they evolve in a manner well described by simple mutation models, such as the infinite sites model.
vi. SNP alleles are almost exclusively identical-by-decent (IBD), preventing scoring errors associated with homoplasy.

One disadvantage is that any SNP has lower information content, compared with a highly polymorphic microsatellite. But this disadvantage can be compensated for by use of a higher number of markers. Since genotyping error rates tend to be lower for SNPs (Kennedy *et al.,* 2003; Bonin *et al.,* 2004); larger numbers of markers can be run jointly

eliminating what is generally the most important cost element of genotyping (Kennedy *et al.,* 2003; Anderson and Garza, 2006).

2.6.5 Diversity arrays technology (DArT)

A DArT marker is a segment of genomic DNA, the presence of which is polymorphic in a defined genomic representation. DArT is a hybridization-based genotyping technology which is currently implemented on the microarray platform to rapidly and simultaneously identify and type DNA polymorphism (Kilian *et al.,* 2003). It is one of the recently developed molecular techniques and has been used in rice (Jaccoud *et al.,* 2001), barley (Wenzl *et al.,* 2004), eucalyptus (Lezar *et al.,* 2004), Arabidopsis (Wittenberg *et al.,* 2005), cassava (Xia *et al.,* 2005), wheat (Akbari *et al.,* 2006; Semagn *et al.,* 2006), and pigeon-pea (Yang *et al.,* 2006). DArT is a microarray hybridization-based technique that enables the simultaneous typing of several hundred polymorphic loci spread over the genome (Jaccoud *et al.,* 2001; Wenzl *et al.,* 2004).

For each individual DNA sample being typed, genomic representations are prepared by restriction enzyme (e.g., *PstI* and *TaqI*) digestion of genomic DNA followed by ligation of restriction fragments to adapters. The genome complexity is then reduced by PCR using primers with complementary sequences to the adapter and selective overhangs. The fragments from representations are cloned, and cloned inserts are amplified using vector-specific primers, purified and arrayed onto a solid support (microarray) resulting in a “discovery array.” Labeled genomic representations prepared from the individual genomes included in the pool are hybridized to the discovery array

(Jaccoud *et al.,* 2001). Polymorphic clones (DArT markers) show variable hybridization signal intensities for different individuals. Such clones can subsequently be assembled into a “genotyping array” for routine genotyping.

DArT technique has a number of advantages:

i. It does not need prior sequence information for the species to be studied; this makes the method applicable to all species regardless of how much DNA sequence information is available for that species.
ii. It is a high throughput, quick, and highly reproducible method.
iii. It is cost effective, with an estimated cost per data point tenfold lower than SSR markers (Xia *et al.,* 2005).
iv. The genetic scope of analysis is defined by the user and easily expandable.
v. It is not covered by exclusive patent rights, but on the contrary open-source (i.e., it is designed for open use and shared improvement).

This technique however, has limitations:

i. DArT is a microarray-based technique that involves several steps, including preparation of genomic representation for the target species, cloning, and data management and analysis. The latter requires dedicated software’s such as DArTsoft and DArTdb. The establishment of DArT system, therefore, is highly likely to demand an extensive investment both in laboratory facility and skilled manpower.

ii. DArT assays for the presence (or amount) of a specific DNA fragment in a representation. Hence, DArT markers are primarily dominant (present or absent) or differences in intensity, which limits its value in some applications.

2.6.6 Mitochondrial DNA (mtDNA)

Although mtDNA only represents a very small fraction of coding potential of the genome, analysis of its variation in populations has provided a disproportionate amount of information about evolution, origins, and patterns of migration (Torroni *et al.,* 1993a, 1993b; Shields *et al.,* 1993). Two aspects of mtDNA make it particularly useful in evolutionary studies; first, it is maternally transmitted and thus does not undergo recombination (Giles *et al.,* 1980). Therefore, mtDNA mutations have accumulated sequentially along radiating maternal lineages, which have diverged as females migrated into different geographic regions. Second, the mtDNA sequence evolution rate is much higher than that of nuclear genes (Miyata *et al.,* 1982; Wallace *et al.,* 1987), resulting in the ability to discriminate between even relatively recently separated populations (Torroni *et al.,* 1994a). This genetic marker represents a rapidly evolving DNA sequence that is informative for answering population-level questions. However, the high information content, a result of high mutation rates, comes at a price and inferences drawn from mtDNA sequences are further limited by the fact that the mtDNA genome comprises a single maternally inherited locus.

2.6.7 Y-chromosome

The Y chromosome determines maleness by causing the development of the testis. It is an unusual segment of the genome since, apart from two small regions in which pairing and exchange take place with the X chromosome, it is haploid, and escapes from recombination. The Y chromosome as put by Jobling and Tyler (2003) is not essential for the life of an individual (males have it, but females do well without it), one-half consists of tandemly repeated satellite DNA and the rest carries few genes. These unique properties of the Y chromosome have important consequences for its mutation processes, its genes, and its population genetics. Y chromosomes pass down from 'father' to 'son' largely unchanged, (this leads to less effective population size), except by the gradual accumulation of mutations.

By examining the differences between modern Y chromosomes (as DNA polymorphisms) we can attempt to reconstruct a history of paternal lineages (Jobling *et al.,* 1995). Y-chromosome DNA, has proven to be a useful tool in phylogeography (Jobling *et al.,* 2003), once a mutation occurs on the Y chromosome, it is a slow process for it to spread in the population. The markers on the non-recombinant part of the Y chromosome allow the reconstruction of intact haplotypes, which are not likely to be eroded by recombination and recurrent mutation and which are, therefore, highly informative for tracing ancient population migrations. Given the fact that Y chromosomes have a smaller effective population size than do autosomes, Y-chromosome–specific polymorphic markers are probably the best genetic tool to study

early migrations as bottleneck events that are often associated with such migrations become more pronounced.

2.6.8 Copy Number Variations (CNVs)

Deletions, insertions, duplications and complex multi-site variants (Freeman *et al.,* 2004), collectively termed copy number variations (CNVs) or copy number polymorphisms (CNPs), are found in all humans (Feuk *et al.,* 2006) and other mammals examined (Freeman *et al.,* 2006). Mileyko *et al.,* 2008 points out that Copy number variations (CNVs) are an important and widespread component of within and between population genetic variations.

CNVs refer to DNA segments that are 1 kb or larger and present at variable copy numbers in comparison with a reference genome (Feuk *et al.,* 2006). A CNV can be simple in structure, such as tandem duplication, or may involve complex gains or losses of homologous sequences at multiple sites in the genome. CNVs influence gene expression, phenotypic variation and adaptation by disrupting genes and altering gene dosage (Mc Carroll *et al.,* 2006, Buckland *et al.,* 2003), and can cause disease, as in microdeletion or microduplication disorders (Inoue *et al.,* 2002; Shaw-Smith *et al.,* 2004), or confer risk to complex disease traits such as HIV-1 infection and glomerulonephritis (Gonzalez *et al.,* 2005, Aitman *et al.,* 2006). CNVs often represent an appreciable minority of causative alleles at genes at which other types of mutation are strongly associated with specific diseases: CHARGE syndrome (Jongmans *et al.,* 2006)

and Parkinson's and Alzheimer's disease (Singleton *et al.,* 2003; Rovelet-Lecrux *et al.,* 2006).

Furthermore, CNVs can influence gene expression indirectly through position effects, predispose to deleterious genetic changes, or provide substrates for chromosomal change in evolution (Lupski *et al.,* 2005; Feuk *et al.,* 2006). Recent data suggests that copy number variation could account for up to 4 megabases (Mb) of normal genetic differences, compared to roughly 2.5 Mb for SNP variation (Tuzun *et al.,* 2005).

CNVs can be identified by examination of either probe intensity differences (Redon *et al.,* 2006;Huang *et al.,* 2004) or based on deviations from Hardy– Weinberg equilibrium (Mc Carroll *et al .,* 2007) or expected Mendelian transmission (Mc Carroll *et al .,* 2007; Conrad *et al.,* 2006). The probe intensity approach allows detection of CNV gains and losses. The Hardy–Weinberg or Mendelian transmission approaches only identify CNV losses, not gains. A limiting factor in all of these types of analyses is that previous generations of SNP arrays were optimized for allelic discrimination rather than copy-number measurement.

Therefore, CNV discovery has been mainly restricted to that of large and simple biallelic variants with high confidence. Commonly used CNV calling algorithms differ both in the number of samples used as reference (either one or a pool of references) and their calling criteria. As a result, they vary in the number and size of CNVs called. In fact, there are relatively few CNVs shared among studies carried out to date. The proportion of overlapping CNVs varies from 25 % (Eichler *et al.,* 2006) (based on the

size of overlap of deletions, when deletions were detected in the same samples using different methodologies), to 45 % (Zogopoulos *et al.,* 2007) (when assessing the occurrence of any CNV overlap between two surveys). A CNV region (CNVR) is an artificial grouping of CNVs overlapping or in close proximity to each other (Redon *et al.,* 2006; Scherer *et al.,* 2007).

Without standard samples (Scherer *et al.,* 2007) and independent validation of identified CNVRs (e.g. by quantitative PCR), it is difficult to assess whether this variation is due to differences in selectivity and sensitivity of these analyses, or whether it reflects different abilities to recognize specific classes of CNVs.

2.6.9 Microsatellites

They are moderately repetitive and composed of arrays of short (2 – 6 bp) repeats found in vertebrates, insect and plant genomes. Copy numbers are characteristically variable within a population, typically with mean array sizes on the order of 10 – 100 kb.

Microsatellites have been the most widely applied class of molecular markers used in genetic studies, with applications in many fields of genetics including genetic conservation, population genetics, molecular breeding and paternity testing. This range of application is due to the fact that microsatellite unlike other markers, are co-dominant, multi-allelic, highly reproducible, have high resolution and are based on the polymerase chain reaction (PCR).These markers (microsatellites) are widely distributed in the

eukaryotic and prokaryotic genome and they have high rates of mutation than the rest of the genome (Field and Wills, 1998; Jarne and Lagoda, 1996; Toth *et al*., 2000).

The Food and Agriculture Organization of the United Nations recommends the use of microsatellites as genetic markers in domestic animal diversity studies (FAO, 1998).

Progress in the use of microsatellites has encountered set backs due to the high costs of developing specific primers. However, many studies have shown that primer pairs designed for one species can be used for other species of the same genus (Isagi and Suhandano, 1997; Cipriani *et al*., 1999) or even for different genera of the same family (White and Powell, 1997; Roa *et al*., 2000; Zucchi *et al*., 2002). Transferability reduces costs when working on taxa with low microsatellite frequencies or from which microsatellites are difficult to isolate. Variation within autosomal microsatellites has been used successfully to make inferences about population history (Forbes *et al.,* 1995; Walling *et al.,* 2004) and to examine the relationship between sheep breeds from Europe (Arranz *et al.,* 1998, 2001; Diez-Tascon *et al.,* 2000; Tapio *et al.,* 2003) and Asia (Chu *et al.,* 2003).

2.6.9.1 Classification of Microsatellites

This is done as per the type of repeat sequence as perfect, imperfect, interrupted or composite. In perfect microsatellite, the repeat sequence is not interrupted by any base not belonging to the motif (e.g. TATATATATATATATA) while in an imperfect microsatellite there is a pair of bases between the repeated motifs that does not match the

motif sequence (e.g. TATATACTATATA). In the case of an interrupted microsatellite there is a small sequence within the repeated sequence that does not match the motif sequence (e.g. TATATACGTGTATATATATA) while in a composite microsatellite the sequence contains two adjacent distinctive sequence repeats (e.g. TATATATATAGTGTGTGT).

2.6.9.2 Mutation rates of Microsatellites

It is known that the mutation rate of microsatellites is much higher than that of other parts of the genome, ranging from 10^{-2} to 10^{-6} nucleotides per locus per generation, but the mutational dynamics of the genomic regions is still not well understood (Schlötterer, 2000 and Sia *et al*., 2000). Several mechanisms have been suggested to explain the high mutation rate of microsatellites, including errors during recombination, unequal crossing-over and polymerase slippage during DNA replication or repair (Strand *et al*., 1993).

In regard to the inclusion of errors during recombination, Levinson and Gutman (1987) found that strains of *Escherichia coli* with or without a functional recombination system had a similar mutation rate, suggesting that recombination is not the predominant mechanism in the generation of microsatellite variability. When unequal crossing-over occurs, there can be drastic changes such as the loss or gain of a large number of repeats. This is because when microsatellite repetitive regions are present, a hairpin (the dark region) can be formed during synapsis, which means that only parts, usually unequal in length, of each chromosome will be exchanged and one chromosome will receive a

larger fragment because of the larger number of microsatellite repeats exchanged, the homologues chromosome receiving a smaller number of repeats.

During DNA replication or repair, DNA polymerase slippage can occur in which one DNA strand temporarily dissociates from the other and rapidly rebinds in a different position, leading to base-pairing errors and continued lengthening of the new strand and an increase in the number of repeats (i.e. additions) in the allele if the error occurs on the complementary strand or a decreased number of repeats (i.e. deletions) if the error occurs on the parent strand. High rates of slippage have been demonstrated but these appear to lead to only small changes in the number of repeats (Hentschel, 1982; Streisinger and Owen, 1985; Schlötterer and Tautz, 1992). Slippage can destabilize microsatellites either because there is no effective repair system for DNA loops or because of alterations in DNA polymerase or its cofactors that result in increased slippage rates. Mutations in the genes of the DNA repair system substantially increase (up to 700 times) microsatellite instability in *E. coli* (Bichara *et al*., 2000), yeast (Strand *et al*., 1993; Sia *et al*., 1997) and mammallian cells (Kolodner and Marsischky, 1999) while mutations affecting the DNA polymerase correction domain produce less drastic effects (Sia *et al*., 2000).

2.6.9.3 Mutation patterns of Microsatellites

Since mutations by definition are rare events, even for microsatellites, there are few empirical data on the type of mutations. According to Primmer *et al*., (1998), it appears that most mutations involve the addition or deletion of a single repeat, with

fewer mutations involving two to several repeats. A study by Xu *et al*., (2000) identified 236 mutations at tetranucleotide microsatellites, where the ancestral and derived states were known. A total of 85 % mutations involved a single repeat and 95 %, less than three repeats. The largest mutation was a five repeat expansion. It is however unclear whether this result would hold for other microsatellite repeat motifs (di- and trinucleotides; Ellegren, 2000a). As reviewed by Ellegren (2000b), the frequency of nonstepwise mutations seems to vary considerably between taxonomic groups, with estimates ranging from 4 % to 74 %. There is a strong body of evidence that the maximal possible size of microsatellite alleles is constrained. For instance, perfect dinucleotide alleles rarely exceed 30 repeats. A restricted number of possible allelic states will definitely lead to additional size homoplasy. Neither of the two extreme mutation models proposed by population geneticists (Infinite alleles model and Stepwise mutation model) appears to perfectly account for the observed patterns of microsatellite mutations. Their mutation pattern probably lies somewhere in between these two extreme models. Furthermore, neither of these extreme models nor their offshoots (*K*-allele model, two-phase model) can account for asymmetries in the mutation patterns or constraint on allele size

2.6.9.4 Theoretical models of Microsatellite Mutation

Mutational models are used to derive the expected number of alleles in a population from the observed heterozygosity and also in the statistical analyses of

genetic variation, but all models have some disadvantages when applied to microsatellite data.

In general, four models can be used;

i. Infinite alleles model (IAM)

In this model, each mutation randomly creates a new allele. Applying this model to microsatellite loci, mutations alter the number of repeats. For example, an allele with 10 repeats is considered to be as closely related genetically to an allele with 15 repeats as to one with 16 repeats, i.e. proximity in terms of the number of repeats does not indicate a greater phylogenetic relationship. This is Wright's (1931) classical model in which he uses *F*-statistics.

ii. K-alleles model (KAM)

Crow and Kimura (1970) proposed the KA model in 1970, which assumes that if there are exactly *k* possible alleles in a given locus then the probability of a given allele mutating into any other is ц/k - 1, where 'ц' is the mutation rate.

iii. Stepwise mutation model (SMM)

When a microsatellite locus mutates, it gains or loses a repeat. This implies that two alleles differing by only one motif are more related (i.e. share a more recent common

ancestor) than alleles differing by several repeats. Slatkin (1995) proposed a genetic differentiation measure $\boldsymbol{R_{ST}}$ similar to Wright's (1951) $\boldsymbol{F_{ST}}$ *and* Nei's (1973) $\boldsymbol{G_{ST}}$ but based on the SM model. The SM model is usually preferred when estimating relations between individuals and population structure, except in the presence of homoplasy (i.e. when two alleles are identical by state but not by descent). Homoplasy may seriously influence population studies involving high mutation rates and large population sizes together with strong allele size constraints (Estoup *et al*., 2002). The model described by Slatkin (1995) is based on traits with continuous distribution, number of base pairs or number of repeats, and it groups individuals according to the number of repeats.

iv. Two phase model (TPM)

Di Rienzo *et al*., (1994) introduced this model as an extension of the SM model for studies on microsatellites. It states that most mutational events result in an increase or decrease of one repeat unit, though infrequent alterations of a large number of repeats also occur.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Sampling

3.1.1 The sampling sites

Twelve Kenyan sheep populations were sampled from Bungoma, Lokichoggio, Vipingo, West Pokot, Kapiti (Dorper and Red Maasai), Moyale, Garissa, Lamu and Kajiado (Olkiramatian (Blackhead Somali and Red Maasai) and Loitoktok). Three more populations were obtained from the laboratory's samples bio-bank at the International Livestock Research Institute. These samples had been originally collected from Homabay, Mombasa (both in Kenya) and Somali (see figure 1).

In order to ensure that the sampled individuals were not closely related, only one to three animals from each flock were sampled and the genealogy of these animals obtained from the flock owner or herdsman. Blood samples were collected from at least 30 individuals per sheep population. Each representative member of the population was photographed and the Global Positioning System (GPS) coordinates of the sampling sites was recorded using a GPS receiver.

3.1.2 The samples

Approximately 5 ml of blood was collected from each animal by venipuncture using EDTA Vacutainer® (Becton Dickinson) and 5 ml of Magic Buffer® was added. Magic Buffer® preserves the cell in a way that it does not break and that way the DNA is preserved in nucleus where it would stay stable for years. Each tube was gently mixed by inversion, marked with sex of the animal, the date and a sampling code. The tubes

containing the blood were placed in cool boxes containing ice packs and transported to the International Livestock Research Institute (ILRI) laboratory in Nairobi where they were stored at 4 ^{O}C.

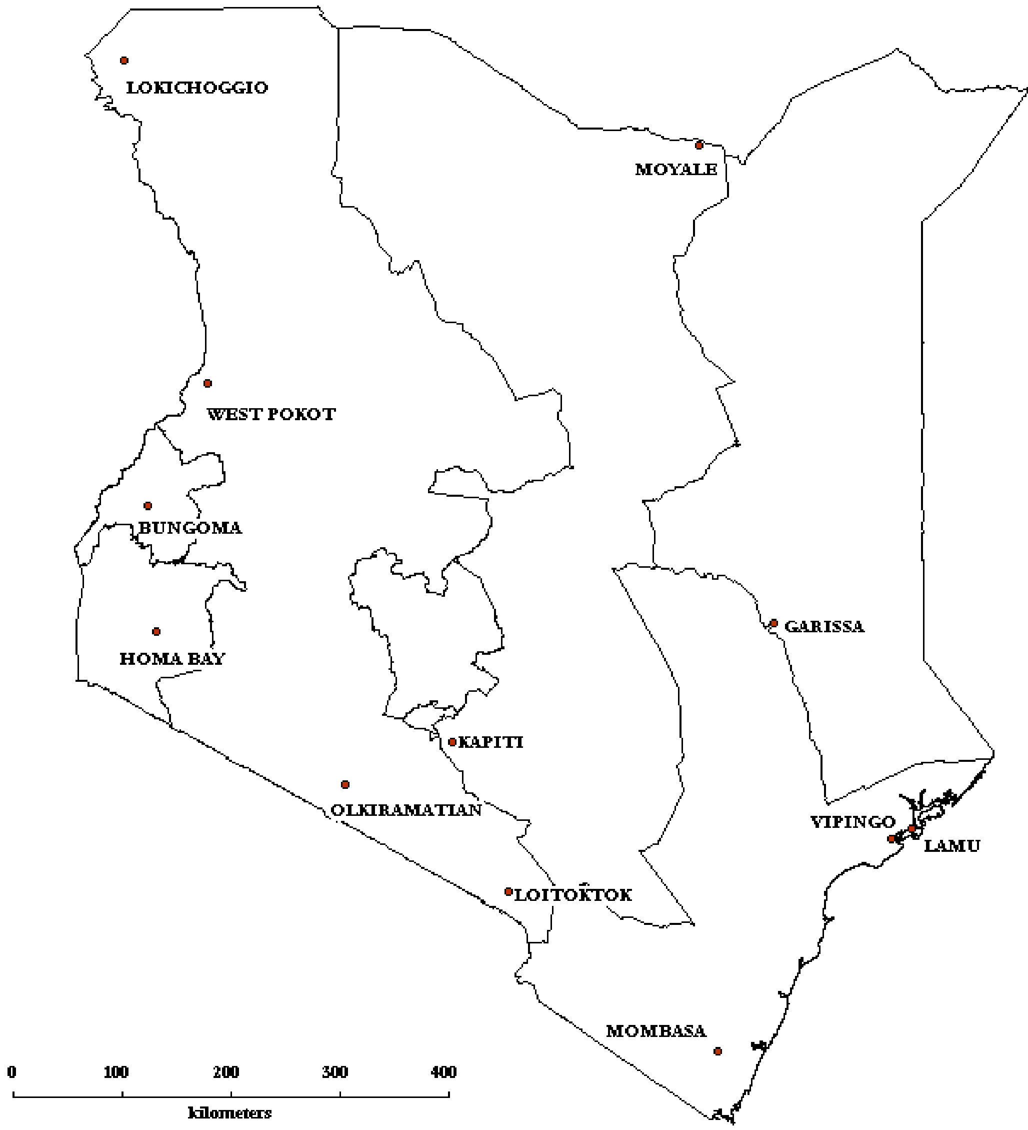


Figure 1: A Kenyan map showing the sampling sites

Table 2: Sampling sites, Population acronyms (in brackets), GPS positions, Breed and number of sheep sampled

Site Name	Latitude	Longitude	Breed	# Ewes sampled	# Rams sampled
BUNGOMA(BGM)	+0.4592	+34.5163	East African fat-tailed	27	13
GARISSA(GAR)	-0.5458	+39.6831	Blackhead Somali	29	11
WEST POKOT(WP)	+1.4900	+35.0188	East African fat-tailed	31	10
VIPINGO(VIP)	-2.3239	+40.7267	Red Maasai	30	12
LAMU(LAM0	-2.2337	+40.9071	Blackhead Somali	30	10
LOKICHOGGIO(LOK)	+4.1829	+34.3231	Blackhead Somali	24	8
LOITOKTOK(LTK)	-2.5784	+36.9474	Red Maasai	31	12
MOYALE(MOY)	+3.5129	+39.0631	Blackhead Somali	32	11
KAPITI(KAP)	-1.5678	+36.9322	Red Maasai and Dorper	52	20
OKIRAMATIAN(OKM)	-1.8652	+36.1656	Red Maasai and Blackhead Somali	53	23
HOMABAY(HOM)	-0.5873	+34.5941	East African fat-tailed	29	11
SOMALI(SBH)	+7.5800	+47.4400	Blackhead Somali	30	10
MOMBASA(REMA)	- 4.1028	+39.2737	Red Maasai	24	9
Total				**422**	**160**

Average number of sheep sampled per population = 39 Total number of sheep sampled = 582

3.1.3 Sheep sampled in this project

A brief description of the sheep genetic resources used in this project is given below:

A total of fifteen sheep populations were targeted with the total number of sheep sampled being five hundred and eighty two with an average of thirty-nine animals per population.

3.1.3.1 Blackhead Somali

These originated in the arid regions of East Africa in what is now Somalia (Mason 1988). They are a type of hair sheep with a white body and black head and neck with the two colors sharply distinguished. They have a fat rump, short legs, and a compact conformation and both sexes are polled (Thomas 1991).

The short, smooth coat of the Blackhead Somali sheep indicates that the breed characteristics have evolved from the crossbred progeny of fat-tailed and thin-tailed sheep. The fat rump classifies them with other fat tailed sheep, while the short coat of the Somali may therefore be attributed to their thin-tailed African ancestors (Wilson, 1991). The range of distribution of fat-rumped sheep of Africa includes the whole of Somalia, the lowlands of eastern Ethiopia including Afar, Dire Dawa and Ogaden; the north of Kenya and along the Ethiopian border as far as the Toposa area of the Sudan (Mason 1988). In East Africa the fat-rumped sheep is essentially an inhabitant of dry areas (FAO 1991). The pigmentation of the head is probably due to both selection practiced by the breeders and natural selection (against intense solar radiation). The

breed name "Blackhead Persian" of South Africa is a misnomer, because this breed did not originate from Persia but from Somaliland and developed through crossing with local domesticated stock and selective breeding (Epstein, 1971).

The characteristic features of the Blackhead Somali are the fat rump and the black head; horns are absent, although scurs also occur; ears can be very short to moderately long; the improved breed in South Africa has thick neck of good proportion to body; body is broad, deep and reasonably long with broad withers and back; straight back; prominent chest standing out vertically; well developed and freely hanging dewlap; tail comprises three parts: the first broad and firm close to the rump, not hanging down and not tapering; the second is curved upwards and rests against the centre of the first, tapers towards apex, which should be level with the back, and shows a clean black skin area; the third hangs from the apex of the second, is 5-8 cm long and curved with short smooth hair (Wilson, 1991). These can be seen from figures 2c, 2g, 2i and 2l.

3.1.3.2 Dorper

Dorper is a composite breed widely occurring in Kenya. Composites or synthetics are breeds that are developed by intentional crossbreeding aiming to combine desirable genetic material from several sources (Rege *et al.,* 1996).

Dorper was developed in the Grootfontein area of South Africa from 1942 onwards, by interbreeding F1 crosses between the Black Head Persian females (originally from Somalia) and Dorset Horn males breeds (Milne, 2000). A breed society

was established in 1950 in the Republic of South Africa. The Dorsian (White variety) was affiliated to Dorper in 1964 (Wilson, 1991). Dorper is found over most of the southern African countries including South Africa, Namibia, Zimbabwe, Angola, Botswana and also in Kenya and Tanzania with smaller numbers in West African countries. The habitat is mainly semi-arid and arid, and the breed is mainly maintained in commercial ranches (Wilson, 1991). The Dorper has a reputation for adaptability to harsh, arid conditions (Cloete *et al.,* 2000) and was first imported into Kenya from South Africa in the early 1960s.

Dorper is characterised by strong and long head; broad and strong shoulder; wide and deep chest; straight legs; straight and long back-line; well developed udder; even distribution of fat in carcass; coat colour is white with black confined to head and neck; short and loose coat of hair or wool (Wilson, 1991). These can be seen from figure 2a.

3.1.3.3 Red Maasai

Red Maasai sheep are found in Northern Tanzania and South-central Kenya (Rege *et al*., 1996). They are principally owned by the Maasai pastoralists, but similar sheep are owned by many other communities in Kenya that include the Nandi, Samia and Bukusu. In Northern and Central Tanzania by the Gogo and in the drier parts of Karamoja (DAGRIS, 2006). They perform well in altitude that range from 500 to 1500 m above sea level, with climatic conditions of semi-arid with bimodal rainfall. However, higher altitude areas e.g. Nandi district and its environs harbour highly

performing strains of East African fat-tailed populations and this is because the production systems are either pastoral or agro-pastoral (Wilson, 1991).

Red Maasai sheep are characterised by variable coat colour but are mostly red-brown in colour even though black ones are also common. These can be seen from figures 2b, 2e, 2h, 2f and 2k. The hair coat is relatively long and it has a relatively large body size of approximately 70 cm height. The forehead is short and broad. The head profile is convex in males and straight in females. Pads of fat occur on the front of face and behind the poll in males, but are less common in females than in other fat-tailed sheep. The horns are often present in both sexes (Wilson, 1991).

Figure 2(a-l) Phenotypic features of the sheep genetic resources sampled

a) Dorper ewes at Kapiti - ILRI Nucleus herd

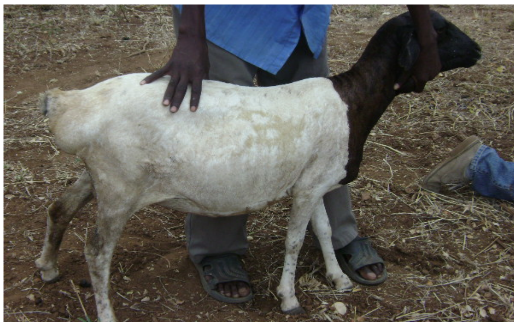
b) A Red Maasai ram at Kapiti - ILRI Nucleus herd

c) Sheep in Lokichoggio

d) Sheep in Lamu

e) Red Maasai sheep in Vipingo-a nucleus herd

f) East African Fat-Tailed sheep in West Pokot


g) A Blackhead Somali ewe in Garissa

h) East African Fat-Tailed sheep in Bungoma

i) Blackhead Somali sheep in Moyale township

j) Sheep in Olkiramatian-Kajiado
(Note the colour of ewe and lamb)

k) A Ram in Loitoktok-Kajiado

l) A ewe in Olkiramatian-Kajiado

3.2 DNA Extraction

3.2.1 DNA Extraction from Blood

A day before DNA extraction, the tube was stored at room temperature. The tube was mixed gently by inverting to homogenize the contents, before a 700 µl aliquot of the sample was transferred into an eppendorf® tube. An equal volume of distilled water (700 µl) was then added, contents vortexed and centrifuged at 10 000 rpm for ten minutes, the supernatant discarded and a similar amount of distilled water added. The pellet was vortexed until it disaggregated, then centrifuged at 10 000 rpm for ten minutes and the supernatant discarded.

A 200 µl aliquot of lyses buffer (160 mM sucrose, 80 mM EDTA (pH 8.0), 100 mM Tris (pH 8.0), 0.5 % SDS) and 2 µl proteinase K solution (20 mg/ml) was added to the white cells pellet and homogenized by inverting the tubes several times. To allow digestion, the tube was incubated at 58°C for four hours and 100 µl of 4.5M NaCl added. Thereafter, 225 µl of chloroform was added and the contents mixed to uniformity. The tube was then centrifuged at 14 000 rpm for ten minutes. The upper (aqueous) phase was transferred to a new microfuge tube and an equal amount of isopropanol added in order to precipitate the DNA. The tube was then centrifuged at 14 000 rpm for ten minutes, after which the supernatant was discarded and the DNA pellet collected. The pellet was washed with 500 µl of 70 % ethanol, incubated for fifteen

minutes at room temperature, centrifuged at 14 000 rpm for ten minutes and the alcoholic solution discarded. The pellet was then air dried and a 100 µl of Tris-EDTA added on the dried pellet and incubated at 37°C overnight. The solution was mixed by pipetting several times and the DNA quantified using Nanodrop© ND-1000 spectrophotometer (Inqaba Biotec South Africa). The stock DNA was stored at ⁻20°C.

3.2.1.1 Cleaning of FTA Filter Papers for Polymerase Chain reaction

Ten discs were punched from the filter paper containing dried blood using Harris Micro-Punch™ 1.20 mm (Whatman International Ltd UK) and placed into a 1.5 ml eppendorf tube. 300 µl of FTA® Purification Reagent, Whatman® (Whatman International Ltd UK) was then added into the eppendorf tube and then incubated for five minutes at room temperature without shaking. The purification reagent was then aspirated, this step was repeated three times. 300 µl of Tris-EDTA buffer were added into the 1.5 ml eppendorf and incubated at room temperature for five minutes without shaking. The Tris-EDTA buffer was then aspirated and the discs were kept at 4°C. For PCR Amplification, a disc was then put into each of the individual wells of the microtitre plate.

3.2.1.2 DNA Extraction from Peripheral Blood Lymphocytes preserved in urea

Samples of the Red Maasai population from Mombasa had been stored in urea hence they were extracted using the illustra™ blood genomic Prep Mini Spin Kit (GE Healthcare UK Limited) according to the Manufacturer's instructions. Briefly this

protocol entailed the following: Blood cells were lysed by use of a chaotropic salt in the lysis solution in the presence of proteinase K. Then the genomic DNA was bound on the columns' silica membrane as denatured proteins were collected in the flow through. The membrane bound genomic DNA was cleaned by washing off proteins and other contaminants using wash buffers. Finally, with the aid of a low ionic strength buffer, the genomic DNA was eluted.

3.3 Determination of the DNA concentration

DNA concentration was determined by the use of Nanodrop© ND-1000 spectrophotometer (Inqaba Biotec South Africa) as detailed in Appendix I.

3.4 PCR Amplification and Microsatellite Genotyping

A total of fifteen microsatellite markers, (*BM8125, DYMS1, HSC, HUJ616, ILSTS005, MAF209, MCM42, OARFCB11, OARFCB20, OARFCB226, OARHH47, OARJMP29, OARVH72, SRCRSP9* and *TGLA53*) were used for genotyping.

Information on these markers is shown in Table 3.

Table 3: The 15 autosomal Microsatellite loci used in the study

Locus	Ch.	Allele Size (base pairs)	Annealing Temp ^{0}C	*Dye	Forward 5`-3` primer	Reverse 3`-5` primer	Reference
DYMS1	23	140-230	58	Ned	AACAACATCAAACAGTAAGAG	CATAGTAACAGATCTTCCTACA	Weimann *et al.,* (2002)
MCM42	9	86-109	55	Pet	CATCTTTCAAAAGAACTCCGAAAGTG	CTTGGAATCCTTCCTAACTTTCGG	Hulme *et al.,* (1994)
OARFCB20	2	92-112	55	6Fam	AAATGTGTTTAAGATTCCATACAGTG	GGAAAACCCCCATATATACCTATAC	Buchanan *et al* ., (1994
ILSTS005	7	181-216	55	6Fam	GGAAGCAATGAAATCTATAGCC	TGTTCTGTGAGTTTGTAAGC	Brezinsky *et al..,* (1993)
SRCRSP9	12	80-150	55	6Fam	AGAGGATCTGGAAATGGAATC	GCACTCTTTTCAGCCCTAATG	Bhebhe *et al.* , (1994)
HSC	Unassigned	267-301	60	6Fam	CTGCCAATGCAGAGACACAAGA	GTCTGTCTCCTGTCTTGTCATC	Achmann *et al..,* (1998)
OARFCB11	2	121-143	58	Vic	GCAAGCAGGTTCTTTACCACTAGCACC	GGCCTGAACTCACAAGTTGATATATCTATCAC	Buchanan and Crawford (19
OARJMP29	24	96-150	63	Pet	GTATACACGTGGACACCGCTTTGTAC	GAAGTGGCAAGATTCAGAGGGGAAG	Crawford *et al.,* (1995)
OARHH47	18	139-153	63	Vic	TTTATTGACAAACTCTCTTCCTAACT CCACC	GTAGTTATTTAAAAAAATATCATACCTCTTAAG G	Henry *et al*., (1993)
MAF209	17	109-135	58	Vic	TCATGCACTTAAGTATGTAGGATGCTG	GATCACAAAAAGTTGGATACAACCGTGG	Buchanan and Crawford (19
TGLA53	12	121-147	55	Vic	GCTTTCAGAAATAGTTTGCATTCA	ATCTTCACATGATATTACAGCAGA	Crawford *et al.,* (1995)
OARVH72	25	121-135	55	Ned	CTCTAGAGGATCTGGAATGCAAAGCTC	GGCCTCTCAAGGGGCAAGAGCAGG	Crawford *et al.,* (1993)
HUJ616	13	268-282	TD	6Fam	TTCAAACTACACATTGACAGGG	GGACCTTTGGCAATGGAAGG	Barendse *et al*., (1994)
OARFCB226	2	119-153	58	Vic	CTATATGTTGCCTTTCCCTTCCTGC	GTGAGTCCCATAGAGCATAAGCTC	Buchanan and Crawford (19
BM8125	17	116-122	55	Ned	CTCTATCTGTGGAAAAGGTGGG	GGGGGTTAGACTTCAACATACG	Crawford *et al*., (1995)

***Dye: Ned - Yellow, 6Fam - Blue, Pet - Red and Vic -Green**

The PCR amplification was performed in a total volume of 10 µl containing 20 ng DNA template, 40 ng each of labeled forward (PET-Red, VIC-Green, 6FAM-Blue and NED-Yellow) and unlabelled reverse primers, 5 µl of ReddyMix™ PCR Master Mix (ABgene, UK) and 3.6 µl of distilled water. The amplifications were performed using a GeneAMP®PCR System 9800. The PCR reactions included a five minute denaturation step at 95°C, followed by 35 cycles of thirty seconds at 94°C, one minute at 54-64°C annealing temperature, one minute at 72°C and a final extension step at 72°C for ten minutes.

3.5 Agarose Gel electrophoresis

The positive amplification of PCR products was confirmed using agarose gel electrophoresis based on negatively charged DNA migrating in the agarose matrix towards anode when placed in an electrolyte under the influence of an electric current.

One hundred ml agarose matrix was made using 1.0 g of agarose powder (Invitrogen, USA) was mixed with 100 ml of 1X Tris-acetate-EDTA (TAE) (0.09 M Tris, 0.09 M Acetic acid and 0.002 EDTA, pH 8.3) (Sambrook *et al*., 1989) electrophoresis buffer, and then heated in a microwave oven until completely melted. 5 µl of GelRed™ Nucleic Acid Gel Stain (Biotium-Hayward, CA) was added to the gel to facilitate visualization of DNA after electrophoresis. After cooling the solution to about 60ºC, it was poured into a casting tray containing a sample comb and allowed to polymerize at room temperature. After the gel had polymerized and the comb removed, the gel (still in its casting tray) was inserted horizontally into the electrophoresis chamber and covered

with buffer. 2.0 µl of DNA samples were then pipetted into the sample wells, in addition, 2.0 µl aliquot of PCR marker (Promega, Madison, USA) was pipetted into the first well of each lane as a ladder. The lid and power leads were placed on the apparatus, and a current of between 80-100 volts applied for forty-five minutes.

The gel was then visualised and the image captured using UVIpro® Version 12.4 gel image software (UVitec, Cambridge, UK).

3.6 Microsatellite genotyping

The internal standards were prepared by adding 12 µl of LIZ (Orange) standard to 1ml of HiDi® formamide and mixed by pipetting. 1 µl of the diluted PCR products was pipetted into individual wells of the microtitre plate to which 9 µl of the standard/formamide mix was added and mixed. The plate was then placed on GeneAMP® PCR System 9700 and the contents denatured. After which the plate was immediately placed on ice to avoid renaturation of the DNA. The denatured PCR fragments were separated using an ABI 3730 (Applied Biosystems, Warrington, UK) automated capillary DNA sequencer. GeneMapper® software (version 3.7, Applied Biosystems), was used to perform allele calling using the third order least squares method for fragment sizing.

3.7 STATISTICAL ANALYSES USED IN THE STUDY

3.7.1 Genetic Diversity

Genetic diversity has been defined as the variety of alleles and genotypes present in a population and this is reflected in morphological, physiological and behavioural differences between individuals and populations (Frankham *et al*. 2002). From a functional point of view, genetic diversity can be classified as neutral, deleterious or adaptive (Hedrick, 2001). Generally, neutral variants are used for conservation applications, but deleterious and adaptive variation are also important in the contexts of population survival and economically important traits in domestic plants and animals.

The indicators of the genetic polymorphism within the populations under study are the sample size corrected mean number of alleles (MNA) or mean allelic richness, $\mathbf{r}_{[g]}$ (El Mousadik and Petit, 1996) and the expected heterozygosity, H_E (Nei, 1978). The $\mathbf{r}_{[g]}$ is the average expected number of alleles in a specified ***(g)*** sized sample (restricted to be equal or smaller than the smallest actual sample size), while the unbiased expected heterozygosity is the probability that two random alleles are different in a population.

MICROSATELLITE TOOLKIT (http://animalgenomics.ucd.ie/sdepark/ms-toolkit/) was used to calculate ***r***, allele frequencies and heterozygosity.

H_E was calculated as shown below

(Hart and Clark, 1997)

$$H_E = 1 - \sum_{i=1}^{m} (f_i)^2$$

where m is the number of alleles at the target locus and fi is the frequency of the ith allele at the target locus.

3.7.2 Population Structure and Admixture

The program **STRUCTURE** (Pritchard *et al.,* 2000) was used to analyze the genetic structure of the populations. This program infers the number of populations into which the analyzed genotypes can be divided. **STRUCTURE** estimates the natural logarithm of the probability that a given genotype $\boldsymbol{X}$ is part of a given population $\boldsymbol{K}$: $\boldsymbol{ln\ Pr(X\ /\ K)}$. This ensures that the groups are, as representatively as possible, samples from a single population. As put by Alvarez *et al.,* (2004), the program assumes that there are $\boldsymbol{K}$ populations with unknown gene frequency distribution at each locus $\boldsymbol{p_{kl}}$ for the $\boldsymbol{k = 1 \ldots K}$ populations and $\boldsymbol{l = 1 \ldots L}$ loci contributing to the gene pool of the target population. Alleles at each locus are sampled independently for each individual, conditional on the proportion $\boldsymbol{q_i}$ of its genotype in a given population. **STRUCTURE** program uses the Markov Chain Monte Carlo method to separately estimate the posterior probability distribution of each parameter (particularly $\boldsymbol{q_i}$ and $\boldsymbol{q_{kl}}$) in an integrated way over all the other parameters. Pritchard *et al.,* (2000) demonstrated that the **STRUCTURE** genetic cluster analysis method can accurately infer individual ancestries with extensive simulations. **STRUCTURE** assumes that the marker loci are unlinked and at linkage equilibrium with one another within populations. It also assumes Hardy-Weinberg Equilibrium within populations.

The most probable number of genetic partitions in the populations, $\boldsymbol{K}$, was determined by evaluating the significance of the posterior probabilities (Pritchard *et al.* 2007), examining the measures of consistency across multiple runs as well as the distribution of their standard deviations across these runs at each $\boldsymbol{K}$ value and by using

the method described by Evanno *et al*. (2005) that examines **Δ*K***, an *ad hoc* quantity related to the change in posterior probability between runs of different ***K***. After partitioning the populations, the analysis was repeated for each partition and when a population was split between partitions for subsequent **STRUCTURE** runs, we assigned it to the one with the highest ancestral contribution. The program was ran using the admixture model and correlated allele frequencies option, which are considered most appropriate for detecting structure among populations that are likely to be similar due to migration or shared ancestry (Falush *et al.,* 2003; Pritchard *et al*. 2007). Five replicates were performed at each ***K*** (***K*** = 1–17) using a burn in time of 100 000 followed by 400 000 iterations. The output from **STRUCTURE** was then sent to **STRUCTURE HARVESTER** (Earl, D.A. 2009) which helped in plotting the graph according to Evanno *et al.,* (2005) and Pritchard (2000) for ***K*** estimation. **STRUCTURE HARVESTER** also assisted in the preparation of the input files for **CLUMPP** (Jakobsson & Rosenberg 2007). Since consistency across runs seems to be an informative method for assessing species structure across breeds (Parker *et al.,* 2007), **CLUMPP** was used to compute similarities over runs for the different values of ***K***, using LARGEK GREEDY algorithm to compute the similarity function ***G'***. For each ***K***, the mean Q-matrix over all runs was used. The output from **CLUMPP** was then used as input by the cluster visualization program **DISTRUCT** (Rosenberg, 2004).

3.7.3 Hardy-Weinberg Equilibrium

In the absence of migration, mutation, natural selection, and assortative mating, genotype frequencies at any locus are a simple function of allele frequencies when assuming infinite population. Expected genotype frequencies under "Hardy-Weinberg equilibrium" (HWE) (according Hardy, 1908 and Weinberg, 1908) can be summarized for biallelic locus by the equation below

$$p^2 + 2pq + q^2 = 1$$

Where p and q are the frequencies of alleles for a given locus in the population, p^2 and q^2 represent the proportions of each homozygote genotype, and $2pq$ represents the proportion of heterozygotes. Thus, a population should be in equilibrium where the expected frequencies of all genotypes (homozygotes and heterozygotes) can be predicted by probability from the individual allele frequencies. If the observed genotype frequencies are different to those expected from the Hardy–Weinberg equilibrium (HWE) model, the locus is said not to conform to Hardy–Weinberg expectations.

Testing for Hardy–Weinberg expectations is a fundamental genetic approach for determining whether allele frequencies for a given gene locus in a given population are maintained in a constant manner from one generation to the next. Deviations from Hardy-Weinberg expectations in this study were estimated using the Fisher's exact test as implemented by **GENEPOP** software (Rice, 1989). This was done assuming 1000 batches with 5000 iterations per batch.

3.7.4 Population Genetic Differentiation

When a population comprises of isolated subpopulations, there is less heterozygosity than there would be if it was undivided. The decline in heterozygosity due to subdivision within a population has usually been quantified using an index known as Wright's *F* statistic, also known as the fixation index. The fixation index ranges from 0 (indicating no differentiation between the overall population and its subpopulations) to a theoretical maximum of 1. θ_{ST} is the correlation between two alleles chosen at random within a subpopulation relative to the alleles sampled at random from the total population. Thus, θ_{ST} measures the heterozygote deficit relative to its expectation under Hardy-Weinberg equilibrium (Hartl and Clark, 1997).

Nei (1977) defined the fixation indices for multiple alleles as: $\theta_{ST} = (H_T - H_S) / H_T$, where H_S is the average subpopulation Hardy–Weinberg heterozygosity and the total population heterozygosity $H_T = 1 - \sum \bar{p}^2_i$ for any number of alleles, where $\bar{p}_i$ is the average frequency of allele i over subpopulations.

Cockerham and Weir (1987) defined an θ_{ST} related to probabilities of identities:

$\theta_{ST} = (f_0 - f_1) / (1 - f_1)$,

where f_0 is the probability of identity in state for pairs of genes between individuals within subpopulations and f_1, between subpopulations.

Nei (1973) defined a multiallelic analogue of F_{ST} among a finite number of subpopulations, called the coefficient of gene differentiation, G_{ST} as being the ratio:

$G_{ST} = D_{ST} / H_T = (H_T - H_S) / H_T$, where D_{ST} is the average gene diversity between populations.

The main difference between F_{ST} and G_{ST} is that the estimation of heterozygosities in G_{ST} rely on allele frequencies only (Nei, 1987), whereas to estimate the H_S the individual genotypes have to be known. F_{ST} estimates were obtained using the program **FSTAT 2.9.3**

3.7.5 Genetic Distances and Phylogenetic relationships

The calculation of genetic distances between two populations gives a relative estimate of the time that has passed since the populations have existed as a single cohesive unit (Avise, 1994).

The standard genetic distance of Nei (1972, 1978) remains one of the most commonly used genetic distances. For populations X and Y with r loci and m alleles per locus, the standard distance D_S is defined as

$D_S = -\ln (J_{xy}/\sqrt{(J_{xx}J_{yy})})$ where

$$J_{xy}=\sum_{i=1}^{m}\sum_{j=1}^{r} x_{ij}\, y_{ij}/r,$$

$$J_{xx}=\sum_{i=1}^{m}\sum_{j=1}^{r} x^2_{ij}/r,$$

$$J_{yy}=\sum_{i=1}^{m}\sum_{j=1}^{r} y^2_{ij}/r,$$

where x_{ij} is the frequency of the ith allele at the jth locus in population X, and y_{ij} is the frequency of the ith allele at the jth locus in population Y. The parametric value of D_S between two populations that became separated t generations in the past is approximated as shown below

D_S=2μt, where μ is the infinite alleles' mutation rate at the loci examined. This expression assumes that fragmentation of the ancestral population was instantaneous and complete, and that each population has had a constant effective size equal to the effective size of the original ancestral population. Note that D_S increases linearly with time from zero to infinity and will have a value proportional to the mutation rate (Nei, 1978).

The D_A distance of Nei (Nei *et al.,* 1983) is a modification of the original Cavalli-Sforza chord distance (1967) and can be estimated as

$$D_A = 1 - \sum_{i=1}^{m}\sum_{j=1}^{r}[(\sqrt{x_{ij}\,y_{ij}})/r]$$

where x_{ij} is the frequency of the *i*th allele at the jth locus in population *X,* and y_{ij} is the frequency of the ith allele at the jth locus in population *Y*.

Its maximum value of 1.0 is achieved when two populations share no alleles at any loci. The D_A distance has proven to be useful for reconstructing phylogenies (Takezaki and Nei 1996).

However, available evidence suggests that D_S has the highest coefficient of variation, and that D_A has the lowest (Nei, 1987; Kalinowski, 2002). This appears to be because D_S has a high interlocus variance. Apparently, this is due to it not having a maximum value. This may be why D_A is more successful at estimating the topology of phylogenetic trees than D_S (Takezaki and Nei, 1996).

Neighbor Joining method (Saitou and Nei, 1987) which produces a unique final tree under the principle of minimum evolution was used for drawing clustering

dendograms. This method does not necessarily produce the minimum-evolution tree, but computer simulations have shown that it is quite efficient in obtaining the correct tree topology and is applicable to any type of evolutionary distance data.

The **PHYLIP** package (Felsenstein, 2004) helped in the construction of the phylogenetic relationships of the sheep using Nei's D_A genetic distances (Nei *et al.,* 1983) calculated by Microsatellite Analyzer (Dieringer and Schlötterer, 2002) and the consensus tree drawn by the **SplitsTree** program (Huson and Bryant, 2006).

3.7.6 Analysis of molecular variance (AMOVA)

Analysis of Molecular Variance (AMOVA) is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation. A variety of molecular marker data (for example, RFLP or AFLP), direct sequence data, or phylogenetic trees based on such molecular data may be analyzed using this method (Excoffier, *et al.,* 1992).

AMOVA treats any kind of data as a Boolean vector ***pi***, that is a 1×n matrix of 1s and 0s, 1 indicating the presence of a marker and 0 its absence. Euclidean distances between pairs of vectors are then calculated by subtracting the Boolean vector of one haplotype from another (Excoffier *et al.,* 1992). For all pairwise arrangements of Boolean vectors, squared Euclidean vectors are calculated, arranged into a matrix and partitioned into submatrices corresponding to subdivisions within the population. The **AMOVA** analysis was done using **GeneAlex 6.1** (Peakall and Smouse, 2007).

3.7.7 Genetic Bottleneck

A genetic bottleneck is an evolutionary event in which a significant percentage of a population or species is killed or otherwise prevented from reproducing. It is possible that when populations become small, they lose genetic variability, which may make them susceptible to extinction e.g. due to lack of adaptive flexibility (Beardmore, 1983).

Secondly, populations that have reduced recently may be more susceptible to demographic stochasticity causing emergence of deleterious phenotypes caused by recessive alleles now occurring as homozygous genotypes (Lande, 1988; Mills and Smouse, 1994). Unfortunately, it is often difficult to identify losses of variability because levels of genetic variability prior to a population decline are generally unknown (Spencer *et al.,* 2000). In such cases, indirect assessments of the magnitude and severity of genetic bottlenecks are frequently made by contrasting levels of genetic variability between related taxa. Although such approaches can provide important insights into the role of historic population declines on the levels of genetic variability in small populations, prebottleneck estimates based on variability in other species may not represent historic levels accurately (Amos and Harwood, 1998).

A more direct method for identifying populations that have lost genetic variability during a recent bottleneck event is to examine the contemporary population for evidence of an excess of loci with an abundance of heterozygosity relative to the observed number of alleles. This transient excess of heterozygosity occurs following the loss of rare alleles from a population, and persists until a new mutation-drift equilibrium is reached. This condition occurs because the rare alleles that were lost contributed little to the

overall heterozygosity. By testing for a heterozygote excess, populations that have recently lost genetic variability may be identified independent of any assumptions about the severity of a bottleneck or prebottleneck levels of genetic variability (Cornuet and Luikart, 1996).

In this study allele frequency data was tested for heterozygosity excess or deficiency using the program **BOTTLENECK** (Cornuet and Luikart, 1997).

The **BOTTLENECK** tests for the departure from mutation drift equilibrium based on heterozygosity excess or deficiency. It compares heterozygosity expected (H_E) at Hardy-Weinberg equilibrium to the heterozygosity expected (H_E) at mutation drift equilibrium in the same sample, that has same size and same number of alleles. Calculations were performed assuming two phase mutations (TPM) and using Wilcoxon Signed rank test. TPM was favoured since Infinite Allele Model - *F* statistics underestimates diversity causing high bottlenecks while Stepwise Mutation Model overestimates mutation.

3.7.8 Genetic Linkage Disequilibrium

Linkage disequilibrium (LD) refers to the non-random association of marker alleles within a genomic region in which the presence of an allele of one marker predicts the presence of an allele of a nearby marker. If two alleles at two different loci are in LD, combinations of alleles within haplotypes occur at frequencies that differ from that expected under the hypothesis of independence. An association between the genetic variation at a locus and a phenotype indicates that either the genetic variation at that locus directly affects the phenotype of interest or the locus is in LD with the causal mutation (Mueller, 2004). As indicated by Heutink and Oostra (2002), evolutionary forces (mutation, genetic drift, migration, and selection), recombination and gene conversion can shape linkage disequilibrium. Testing for the presence of LD and measuring its value are two important instruments of statistical genetics that have recently received a great deal of attention (Sabatti and Risch, 2002). In this study the Fisher's exact test was used to determine linkage disequilibrium using **GENEPOP** (Raymond and Rousset, 1997).

CHAPTER FOUR: RESULTS

4.1 Genetic variability

From the fifteen sheep populations and 582 samples studied, a total of 173 alleles were found for all the 15 loci genotyped (Table 4). The number of alleles per locus ranged from 7 (*BM8125*) to 17 (*HUJ616* and *OARHH47*) (Table 4). The population with the highest number of alleles was the Kapiti Red Maasai with 140 while the Olkiramatian Red Maasai population had the least number of alleles (100) (Table 4). The majority of the markers were highly polymorphic, in only three markers (*BM8125, MCM42*, and *SRCRSP9*) the detected numbers of alleles were less than ten (Table 4). The mean $\boldsymbol{H_E}$ across all loci was 0.724±0.03, while the mean $\boldsymbol{H_O}$ was 0.648±0.02 (Table 5). The mean number of alleles (MNA) per population ranged from 9.33±2.55 in the Kapiti Red Maasai population to 6.67±2.94 in the Olkiramatian Red Maasai population (Table 5).

Table 6 shows the Gene diversity in each population over the microsatellite typed and this ranged as follows: Bungoma (0.546-0.852), Somali (0.525-0.804), Garissa (0.261-0.912), Kapiti Dorper (0.679-0.898), Kapiti Red Maasai (0.74-0.887), Kijipwa (0.407-0.856), Lamu (0.511-0.853), Lokichoggio (0.638-0.873), Loitoktok (0.404-0.848), Moyale (0.568-0.878), Olkiramatian Blackhead Somali (0.53-0.85), Olkiramatian Red Maasai (0.317-0.841), Mombasa Red Maasai (0.531-0.849), West Pokot (0.555-0.803) and Homa bay (0.464-0.876). Surprisingly, both the lowest and the highest value of gene diversity over all populations and loci was observed in the Garissa

population, with observed values of 0.261 (*BM8125*) and while the highest 0.912 (*OARFCB226*), respectively.

Based on the 27 diploid individuals, the allelic richness per locus and per population were as shown in Table 7. The Kapiti Red Maasai population recorded the highest with 8.563 followed by the Lokichoggio population with 8.354, the Red Maasai population from Olkiramatian was the least rich with 6.182 (Table 7). The loci allelic richness ranged from 12.58 (*HUJ616*) to 6.332 (*SRCRSP9*) (Table 7). In the sample, loci *OARFCB226* in the Kapiti Dorper recorded the highest allelic richness (13.674) whereas the *OARFCB11* showed the least allelic richness (1.994) in the Olkiramatian Red Maasai (Table 7).

Table 4: Number of Alleles sampled from fifteen Kenyan Sheep populations genotyped using 15 microsatellite markers

Locus/Popln	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	REMA	WP	HOM	Total
BM8125	6	6	3	6	6	5	5	6	6	7	6	5	6	7	6	7
DYMS1	7	9	5	10	10	9	8	9	7	5	12	7	10	8	9	13
HSC	10	12	9	7	10	9	11	10	11	10	9	11	9	8	7	12
HUJ616	9	9	7	9	15	8	11	11	11	9	7	9	11	7	10	17
ILSTS005	9	8	10	5	8	5	9	8	8	9	8	6	6	8	8	10
MAF209	9	9	6	7	9	7	7	9	6	8	3	6	8	9	10	10
MCM42	5	4	2	7	7	4	5	8	3	4	7	5	3	5	6	8
OARFCB11	4	7	11	8	11	5	9	12	10	8	4	2	7	7	9	12
OARFCB20	7	7	11	10	8	7	8	11	10	7	10	3	10	10	9	13
OARFCB226	8	7	14	14	10	7	15	11	9	13	5	12	6	7	6	16
OARHH47	10	8	11	5	14	11	11	8	8	12	8	10	11	11	9	17
OARJMP29	8	5	5	7	10	5	4	3	8	7	8	4	5	6	5	10
OARVH72	8	4	9	7	7	7	4	7	6	8	8	6	8	7	4	10
SRCRSP9	4	3	6	7	7	6	6	8	4	6	4	5	6	4	5	8
TGLA53	9	9	8	8	8	9	7	8	10	10	9	9	9	8	8	10
Total	**113**	**107**	**117**	**117**	**140**	**104**	**120**	**129**	**117**	**123**	**108**	**100**	**115**	**112**	**111**	**173**

NOTE: BGM- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**-Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

Table 5: Number of Animals Sampled, Mean heterozygosity, Mean number of Alleles and standard deviations for each of the fifteen populations studied

Population	Sample size	Loci typed	H_E	H_E SD	H_O	H_O SD	MNA	MNA SD
BGM	40	15	0.6861	0.0288	0.6186	0.0199	7.53	2.00
SBH	40	15	0.7021	0.0248	0.6373	0.0200	7.13	2.42
GAR	40	15	0.7089	0.0491	0.5641	0.0203	7.80	3.34
KAPD	30	15	0.8068	0.0159	0.8519	0.0168	7.80	2.27
KAPRM	42	15	0.8045	0.0131	0.7713	0.0167	9.33	2.55
VIP	42	15	0.7100	0.0293	0.6286	0.0193	6.93	1.98
LAM	40	15	0.6728	0.0335	0.6150	0.0199	8.00	3.09
LOK	32	15	0.7941	0.0204	0.7051	0.0208	8.60	2.29
LTK	43	15	0.7020	0.0380	0.6217	0.0191	7.80	2.46
MOY	43	15	0.7555	0.0240	0.6717	0.0185	8.20	2.43
OKMBHS	38	15	0.7230	0.0236	0.6415	0.0201	7.20	2.46
OKMRM	38	15	0.5956	0.0645	0.5120	0.0210	6.67	2.94
REMA	33	15	0.7580	0.0240	0.6626	0.0213	7.67	2.35
WP	41	15	0.7058	0.0227	0.5911	0.0198	7.47	1.77
HOM	40	15	0.7254	0.0307	0.6206	0.0198	7.40	1.96
Mean	**38.8**	**15**	**0.7230**	**0.0300**	**0.6480**	**0.0200**	**7.70**	**2.42**

NOTE: BGM- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**- Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

Table 6: Genetic diversity per locus and population

Loci/Popln	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	REMA	WP	HOM
BM8125	0.567	0.713	0.261	0.772	0.740	0.407	0.511	0.642	0.404	0.837	0.802	0.317	0.655	0.677	0.464
DYMS1	0.640	0.616	0.441	0.834	0.788	0.760	0.535	0.838	0.554	0.606	0.776	0.640	0.757	0.592	0.755
HSC	0.846	0.860	0.722	0.849	0.871	0.821	0.853	0.803	0.870	0.886	0.790	0.832	0.848	0.797	0.832
HUJ616	0.764	0.789	0.744	0.864	0.846	0.735	0.565	0.864	0.814	0.749	0.530	0.708	0.826	0.726	0.824
ILSTS005	0.759	0.804	0.826	0.679	0.823	0.756	0.829	0.725	0.797	0.803	0.834	0.740	0.639	0.785	0.836
MAF209	0.828	0.719	0.782	0.840	0.780	0.711	0.755	0.835	0.692	0.786	0.601	0.549	0.787	0.778	0.841
MCM42	0.548	0.567	0.463	0.805	0.746	0.530	0.517	0.831	0.483	0.568	0.672	0.721	0.531	0.555	0.649
OARFCB11	0.546	0.626	0.831	0.821	0.851	0.753	0.554	0.880	0.825	0.764	0.736	0.102	0.805	0.767	0.684
OARFCB20	0.781	0.798	0.865	0.883	0.852	0.745	0.820	0.873	0.848	0.800	0.850	0.053	0.841	0.803	0.854
OARFCB226	0.583	0.792	0.912	0.898	0.839	0.717	0.832	0.857	0.588	0.809	0.736	0.780	0.720	0.658	0.658
OARHH47	0.671	0.668	0.872	0.758	0.887	0.822	0.581	0.735	0.819	0.878	0.757	0.827	0.839	0.788	0.876
OARJMP29	0.694	0.660	0.550	0.810	0.757	0.652	0.667	0.638	0.765	0.670	0.734	0.659	0.725	0.664	0.681
OARVH72	0.661	0.525	0.829	0.704	0.749	0.672	0.597	0.840	0.584	0.673	0.599	0.581	0.818	0.609	0.578
SRCRSP9	0.565	0.665	0.779	0.778	0.740	0.730	0.686	0.818	0.682	0.712	0.681	0.601	0.751	0.609	0.636
TGLA53	0.852	0.745	0.785	0.795	0.805	0.856	0.800	0.754	0.818	0.806	0.764	0.841	0.849	0.801	0.734
Mean	**0.687**	**0.703**	**0.710**	**0.806**	**0.805**	**0.711**	**0.673**	**0.796**	**0.703**	**0.756**	**0.724**	**0.597**	**0.7594**	**0.707**	**0.727**

NOTE: **BGM**- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

Table 7: Allelic richness per locus and population based on 27 diploid individuals

Loci/Popln	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	REMA	WP	HOM	Mean
BM8125	5.886	5.813	2.675	5.999	5.913	4.272	4.347	5.974	5.715	6.983	5.998	4.132	5.995	6.618	5.875	6.835
DYMS1	6.130	8.191	4.385	10.000	8.889	8.308	7.719	8.799	6.200	4.861	10.704	6.719	9.923	7.174	7.668	10.764
HSC	9.772	10.827	7.596	7.000	9.589	8.147	10.59	9.486	10.813	9.722	8.617	10.518	8.903	7.544	6.990	10.695
HUJ616	8.206	8.306	6.540	8.800	12.788	7.254	9.257	10.64	9.887	8.471	6.393	8.231	10.508	6.203	8.912	12.58
ILSTS005	8.263	7.589	9.356	4.800	7.475	5.000	8.538	7.684	7.825	8.097	7.992	5.838	5.606	7.507	7.683	9.585
MAF209	8.632	8.213	5.643	6.900	8.226	6.200	6.865	8.687	5.475	7.444	3.000	5.399	7.908	8.230	9.143	9.400
MCM42	4.665	3.866	2.000	6.900	6.776	3.286	4.247	7.967	2.864	3.847	6.676	4.978	2.969	4.533	5.025	6.826
OARFCB11	3.643	6.906	10.844	7.799	9.917	4.957	7.492	11.642	9.332	7.856	4.000	1.994	6.938	6.533	7.665	10.666
OARFCB20	6.674	6.725	9.983	9.891	7.957	6.707	7.469	10.509	9.023	6.991	9.526	2.421	9.424	9.256	8.673	10.175
OARFCB226	6.637	6.896	13.450	13.674	9.607	6.379	12.922	10.509	8.048	10.949	4.711	10.496	5.938	6.528	5.674	12.381
OARHH47	8.560	7.503	10.455	5.000	12.265	10.631	9.479	7.643	7.813	11.032	7.130	9.332	10.392	9.825	8.966	12.353
OARJMP29	7.184	4.866	4.643	6.991	8.500	4.641	3.675	3.000	7.464	6.308	7.584	4.000	4.990	5.280	4.572	7.734
OARVH72	7.041	3.572	8.530	6.883	6.858	6.147	3.995	7.000	5.443	7.361	7.023	5.131	7.606	6.129	3.959	7.827
SRCRSP9	3.968	3.000	5.967	6.892	6.161	5.285	5.025	7.974	3.864	5.593	3.711	4.683	5.986	3.659	4.572	6.332
TGLA53	8.247	8.593	7.753	7.799	7.517	8.842	6.665	7.799	9.460	9.602	8.888	8.854	8.786	7.848	7.859	9.248
Mean	**6.900**	**6.724**	**7.321**	**7.689**	**8.563**	**6.404**	**7.219**	**8.354**	**7.282**	**7.674**	**6.797**	**6.182**	**7.458**	**6.858**	**6.882**	

NOTE: **BGM**- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

4.2 Population differentiation

As shown in Table 8, the Mean estimates of *F*-statistics were: $\boldsymbol{F_{IT}}$ = 0.199±0.02, $\boldsymbol{f_{IS}}$ = 0.109±0.0109, $\boldsymbol{\theta_{ST}}$ = 0.101±0.019, $\boldsymbol{G_{ST}}$ = 0.096 and $\boldsymbol{G_{ST}}$'= 0.102. The $\boldsymbol{G_{ST}}$ and $\boldsymbol{\theta_{ST}}$ values were similar in all the fifteen loci but the $\boldsymbol{\theta_{ST}}$ showed slightly higher values. The $\boldsymbol{G_{ST}}$**'** a sample size independent estimator of $\boldsymbol{G_{ST}}$ gave similar values to $\boldsymbol{\theta_{ST}}$. They indicate that 9.6 % - 10.2 % of the total genetic variation was due to population differences. All loci had values that were significant to the $\boldsymbol{\theta_{ST}}$. Fourteen markers (*BM8125*, *DYMS1*, *HSC*, *HUJ616*, *ILSTS005*, *MAF209*, *MCM42*, *OARFCB11*, *OARFCB20*, *OARFCB226*, *OARHH47*, *OARJMP29*, *OARVH72* and *TGLA53*) had significant values to the heterozygote deficit within populations ($\boldsymbol{f_{IS}}$). The global heterozygote deficit $\boldsymbol{F_{IT}}$ was significantly contributed to by the same fourteen markers as in $\boldsymbol{f_{IS}}$ above.

The $\boldsymbol{\theta_{ST}}$ values and their significance levels for all population pairs are presented in Table 9. The highest $\boldsymbol{\theta_{ST}}$ value (0.127) was between the Olkiramatian Red Maasai and Vipingo populations while the least value (0.026) was between the Mombasa Red Maasai and Loitoktok populations. The Lokichoggio and Olkiramatian Red Maasai populations appeared to be the most differentiated of all the populations studied.

As shown in Table 10, the $\boldsymbol{f_{IS}}$ values per locus in each population ranged as: the highest values were found in Olkiramatian Red Maasai (0.629 in *OARFCB226*) and Homa bay (0.601 in *ILSTS005*). The highest population inbreeding coefficient was found in the Garissa population (0.206) and the Kapiti Dorper population had the least inbreeding coefficient (-0.057). All populations except the Kapiti Dorper and Red Maasai populations showed an overall significant deficit of heterozygotes.

Table 8: Weir and Cockerham 1984 multilocus estimates for diploid data based on Jackknife resampling over all loci (the number in the parenthesis indicates the standard error)

Locus	f_{IS}	θ_{ST}	F_{IT}	G_{ST}	G_{ST}'	H_T	H_S
BM8125	0.0744*	0.1559**	0.2187**	0.149	0.158	0.687	0.585
DYMS1	0.1207**	0.1275**	0.2328**	0.114	0.121	0.762	0.675
HSC	0.0352*	0.0363**	0.0703**	0.034	0.037	0.862	0.832
HUJ616	0.1163**	0.1065**	0.2104**	0.10	0.106	0.84	0.756
ILSTS005	0.3123**	0.0730**	0.3625**	0.073	0.078	0.837	0.776
MAF209	0.1042**	0.1354**	0.2255**	0.125	0.133	0.86	0.752
MCM42	0.1212**	0.1067**	0.2150**	0.10	0.107	0.68	0.612
OARFCB11	0.0952**	0.1950**	0.2716**	0.184	0.194	0.861	0.703
OARFCB20	0.0683**	0.1088**	0.1696**	0.102	0.109	0.866	0.778
OARFCB226	0.1004**	0.0844**	0.1763**	0.08	0.085	0.824	0.758
OARHH47	0.1201**	0.1057**	0.2131**	0.10	0.107	0.873	0.785
OARJMP29	0.0719*	0.0668**	0.1339**	0.072	0.077	0.742	0.688
OARVH72	0.1116**	0.1460**	0.2413**	0.136	0.144	0.773	0.668
SRCRSP9	0.0098^{ns}	0.0473**	0.0566^{ns}	0.047	0.051	0.73	0.696
TGLA53	0.1516**	0.0256**	0.1733**	0.025	0.027	0.821	0.8
Overall:	**0.109 (0.019)****	**0.101(0.012)****	**0.199 (0.02)****	**0.096**	**0.102**	**0.801**	**0.724**

f_{IS}, Within-population inbreeding estimate; F_{IT}, total inbreeding estimate; θ_{ST}, measure of population differentiation

Statistical significance: *- p<0.05, ** - p<0.01, ***- p<0.001 ^{ns}- non-significant based on 10 000 randomisations (after Bonferroni corrections)

Table 9: Pairwise tests of Population differentiation between fifteen Kenyan sheep populations

POPLN	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	REMA	WP	HOM
BGM	-	***	***	***	***	***	***	***	***	***	***	***	***	***	***
SBH	0.066	-	***	***	***	***	***	***	***	***	***	***	***	***	***
GAR	0.057	0.042	-	***	***	***	***	***	***	***	***	***	***	***	***
KAPD	0.073	0.058	0.059	-	***	***	***	***	***	***	***	***	***	***	***
KAPRM	0.058	0.045	0.058	0.031	-	***	***	***	***	***	***	***	***	***	***
VIP	0.058	0.054	0.056	0.064	0.050	-	***	***	***	***	***	***	***	***	***
LAM	0.054	0.041	0.041	0.074	0.060	0.055	-	***	***	***	***	***	***	***	***
LOK	0.077	0.063	0.083	0.059	0.050	0.083	0.087	-	***	***	***	***	***	***	***
LTK	0.037	0.033	0.039	0.062	0.050	0.047	0.037	0.072	-	***	***	***	***	***	***
MOY	0.041	0.033	0.038	0.047	0.041	0.054	0.034	0.058	0.027	-	***	***	***	***	***
OKMBHS	0.056	0.057	0.051	0.070	0.062	0.085	0.059	0.066	0.055	0.043	-	***	***	***	***
OKMRM	0.090	0.110	0.084	0.122	0.115	0.127	0.108	0.122	0.090	0.098	0.074	-	***	***	***
MOMBRM	0.043	0.031	0.038	0.053	0.036	0.031	0.045	0.059	0.026	0.032	0.058	0.098	-	***	***
WP	0.048	0.045	0.056	0.052	0.036	0.051	0.064	0.069	0.038	0.042	0.071	0.115	0.033	-	
HOM	0.066	0.071	0.076	0.069	0.054	0.046	0.082	0.073	0.059	0.062	0.085	0.115	0.051	0.048	-

F_{ST} Values (below diagonal) and their tests of significance (above the diagonal). ***= p<0.001

NOTE: **BGM**- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

Table 10: Within population inbreeding estimates (f_{IS}) in fifteen Kenyan sheep populations across 15 markers

Loci/Popln	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	REMA	WP	HOM
BM8125	-0.058	0.022	-0.148	-0.252	-0.158	-0.17	0.071	0.465**	0.136	0.25**	0.442***	0.004	0.168	-0.081	0.192
DYMS1	0.095	0.126	0.011	0.067	-0.146	0.342**	0.252	0.067	0.12	-0.1	0.129	0.198	0.239	0.155	0.239
HSC	-0.094	0.025	0.066	-0.099	-0.066	0.188	0.003	0.144	0.038	0.003	0.034	-0.012	0.214	0.112	0.008
HUJ616	0.117	0.049	0.26	-0.081	0.212*	-0.134	0.027	0.169	0	0.162	0.006	0.121	0.193	0.529***	0.059
ILSTS005	0.506***	0.294**	0.334***	0.018	0.566***	0.37**	0.066	0.095	0.241*	0.334***	0.054	0.254	0.288	0.472***	0.601***
MAF209	-0.027	0.047	0.585***	-0.112	-0.037	-0.004	0.106	0.064	0.093	0.142	0.08	0.137	0.23	0.217	0.019
MCM42	0.088	0.25	0.352	-0.118	-0.213	0.056	0.081	0.457***	0.182	0.099	0.178	0.015	0.201	0.296	0.076
OARFCB11	0.45*	0.117	0.067	-0.136	0.021	0.146	-0.037	-0.03	0.098	0.056	0.142	1.0**	0.134	-0.017	0.269
OARFCB20	0.2	0.086	0.306***	0.018	0.133	0.009	-0.037	0.034	-0.014	0.07	-0.021	0.5	-0.08	-0.032	0.21*
OARFCB226	-0.029	-0.105	0.177*	-0.114	-0.135	0.004	-0.022	0.088	0.09	0.022	0.357**	0.629***	-0.094	0.333**	0.354**
OARHH47	-0.23	0.326**	0.427***	-0.143	0.114	0.363***	0.054	0.278	-0.022	0.258**	-0.008	0.013	0.097	0.041	0.087
OARJMP29	0.243	0.015	0.227	0.013	0.025	0.196	0.1	-0.274	0.058	-0.146	0.247	0.121	0.289	0.009	-0.065
OARVH72	0.205	-0.048	0.186	0.195	0.333**	-0.241	0.079	0.182	0.443*	0.033	0.077	0.004	0.037	0.118	0.004
SRCRSP9	-0.15	0.098	-0.026	-0.243	-0.062	-0.012	0.198	-0.07	0.08	0.086	0.111	-0.05	0.072	-0.041	0.096
TGLA53	0.149	0.127	-0.019	0.162	-0.006	0.36***	0.375***	0.047	0.317***	0.291**	-0.137	0.133	0.001	0.33***	0.013
All	**0.099*** **	**0.094** **	**0.206*** **	**-0.057**	**0.042**	**0.116*** **	**0.087** **	**0.114*** **	**0.116*** **	**0.112*** **	**0.114*** **	**0.142*** **	**0.128*** **	**0.164*** **	**0.146*** **

NOTE: **BGM**- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

*- statistically significant at p<0.05, ** - p<0.01, ***- p<0.001, ^{ns}- non-significant based on 225 000 randomisations (after Bonferroni corrections)

For the structure analysis of the microsatellite dataset, a plot of the mean and standard deviation of the posterior probabilities, ***ln Pr (X/K)***, among runs for each value of ***K***, from 1–17 reached an asymptote at ***K***= 4, indicating that the true ***K*** value is possibly 4 (Figure 5a). The modal value for the distribution of ***ΔK*** (134.175) was at ***K*** = 4 as well supporting it to be the true number of clusters in the population studied (Figure 5b). By examining the measures of consistency (the pairwise ***G′*** values for each pair of runs) across multiple runs and the distribution of the standard deviations across these runs each revealed that at ***K*** = 4, the highest consistency (***G'***= 0.9985) and least deviation (standard deviation = 0.00041) was observed thus offering further support that the populations being studied are possibly clustered into four (Figure 5c).

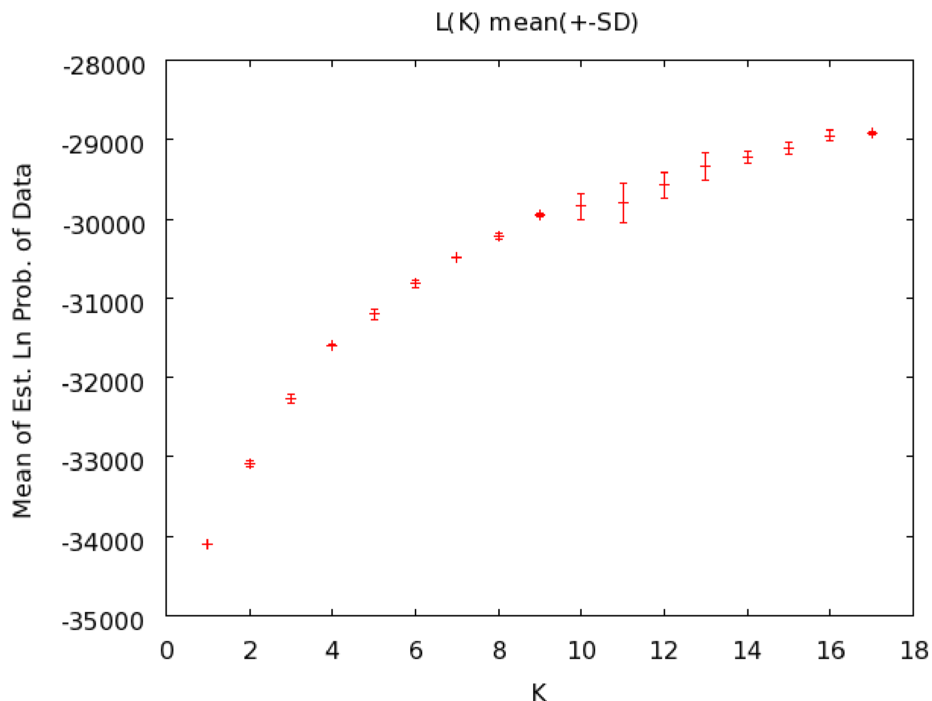


Figure 5(a) Plot of mean and standard deviation of the posterior probabilities, L (K), among runs for each value of *K*, 1–17.

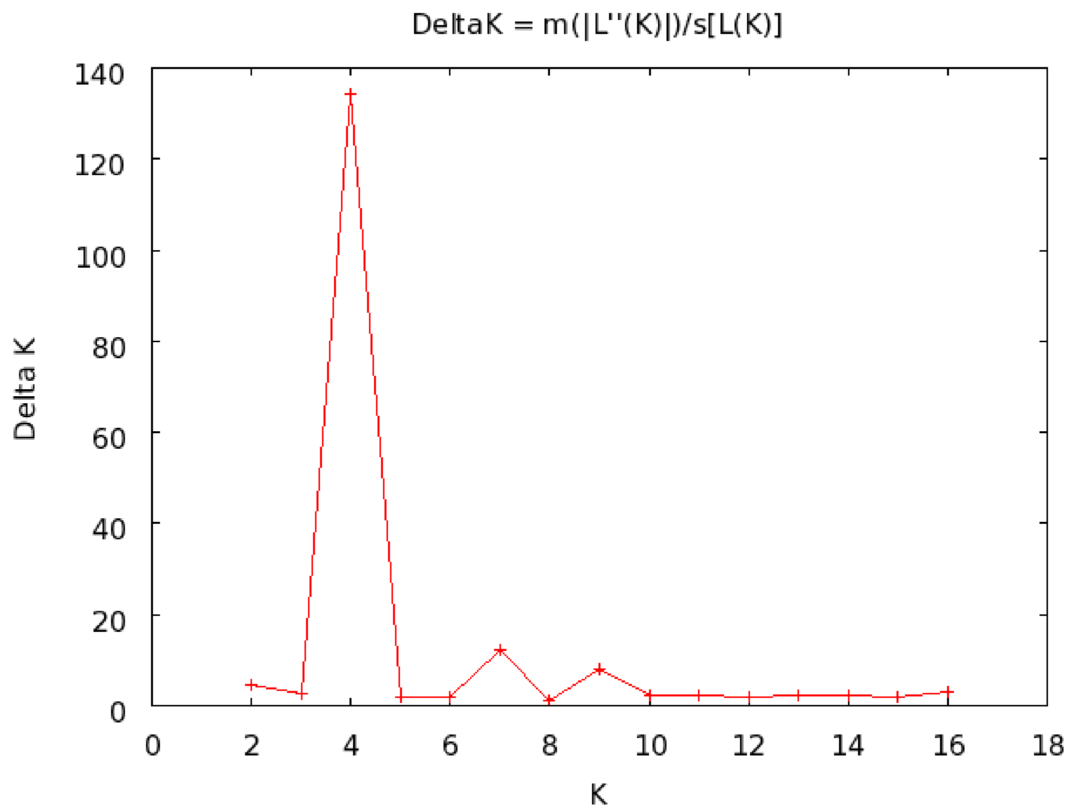


Figure 5(b) Plot of the modal value for the distribution of *ΔK* against *K*

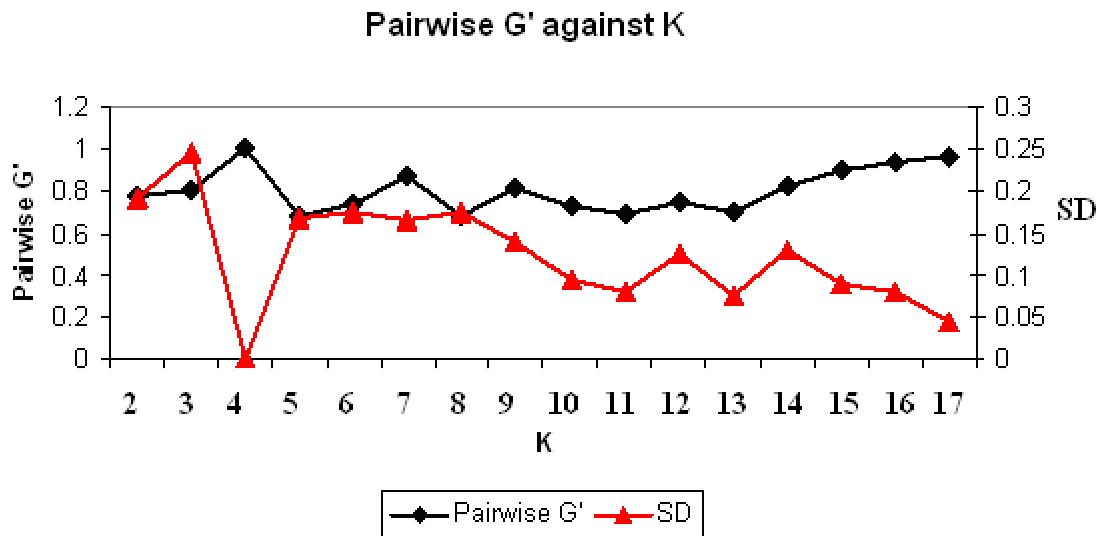


Figure 5(c) Plot of the pairwise similarity function *G'* and Standard Deviation (SD) against the number of clusters *K*, from *K*= 2 to 17

As shown in Table 11, the first partition put the populations into four clusters with the Kapiti populations (Red Maasai and Dorper) clustering with populations from Lokichoggio and West Pokot in Cluster I, the Bungoma and Olkiramatian populations (Blackhead and Red Maasai) constituting cluster II, while the Somali, Garissa, Lamu, Loitoktok, Moyale, and Mombasa Red Maasai populations made cluster III, and finally the Vipingo and Homabay populations were together in Cluster IV.

Table 11: Proportion of membership of each of the fifteen sheep populations in each of the four inferred clusters using the program STRUCTURE (at *K*=4)

Population	Cluster I[a]	Cluster II[a]	Cluster III[a]	Cluster IV[a]
BUNGOMA:	0.0446	**0.7797**	0.1188	0.0568
OKMBHS:	0.0280	**0.9276**	0.0194	0.0250
OKMRM:	0.0100	**0.9450**	0.0310	0.0140
VIPINGO:	0.0122	0.0070	0.0328	**0.9480**
HOMA BAY:	0.0220	0.0120	0.0140	**0.9520**
KAPITI DORPER:	**0.9652**	0.0080	0.0174	0.0094
KAPITI RED MAASAI:	**0.9098**	0.0212	0.0420	0.0270
LOKICHOGGIO:	**0.8528**	0.0900	0.0310	0.0262
WEST POKOT:	**0.6798**	0.0210	0.1908	0.1084
SOMALI:	0.0252	0.0120	**0.8896**	0.0732
GARISSA:	0.0470	0.1600	**0.7776**	0.0154
LAMU:	0.0242	0.0512	**0.8453**	0.0792
LOITOKTOK:	0.0270	0.1048	**0.8184**	0.0498
MOYALE:	0.0750	0.0358	**0.8418**	0.0474
REMA:	0.0522	0.0232	**0.6177**	0.3069

NB: [a]Estimates assumed admixture in the sampled genotypes. Contributions higher than 0.60 are in bold. For population acronyms, see Table 2.

The clusters obtained at K= 4 were further analysed to determine the number of subgroups in each cluster (Figure 6; Table 12a-d). Further genetic structure within cluster II was detected with a modal value for the distribution of ΔK found at K = 3, corresponding to the Bungoma, Olkiramatian Blackhead Somali and Red Maasai

populations (Figure 6). In cluster IV, with a modal value for the distribution of $\Delta\boldsymbol{K}$ found at $K = 2$ the Vipingo and Homa bay populations sub-clustered separately from each other (Figure 6). The Kapiti Dorper, Red Maasai Lokichoggio and West Pokot populations sub-clustered separately as shown in Figure 6, cluster I when a modal value for the distribution of $\Delta\boldsymbol{K}$ was at $K = 3$ (Figure 6). From the results (Table 12, Figure 6, cluster I) it looks like there was Dorper contamination in the Kapiti Red Maasai population. With a modal value for the distribution of $\Delta\boldsymbol{K}$ at K=6, cluster III separated into six sub-clusters comprising the Somali, Garissa, Lamu, Loitoktok, Moyale and Mombasa Red Maasai populations (Figure 6). Evidently, there was contamination of the Loitoktok and Mombasa Red Maasai population with individuals from the other population sub-clusters majorly comprising the Blackhead Somali sheep (Figure 6).

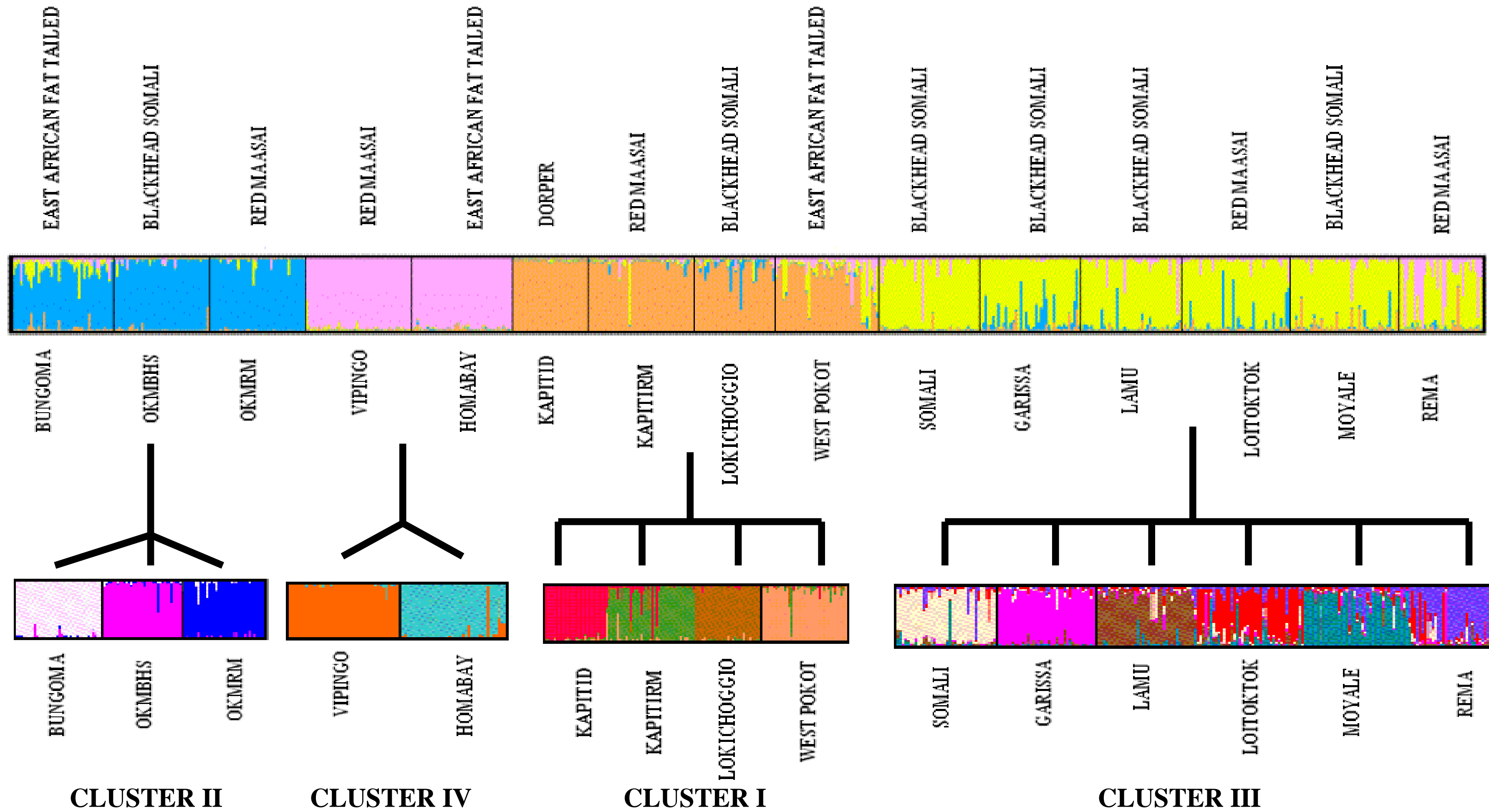


Figure 6. Population partitioning of the sheep populations as suggested by STRUCTURE based on 15 microsatellite markers using individual Q matrices. Junctions show where the data was split into *K* populations and re-run on the sub-data. Black lines separate the individuals of different populations. Population names are indicated below and phenotypic breed identities above the diagram. For population acronyms see Table 2.

Table 12(a-d): Proportion of membership of the re-analysed sheep population clusters

a) Cluster I at *K*=4

Population	**Sub-cluster1[a]**	**Sub-cluster2[a]**	**Sub-cluster3[a]**	**Sub-cluster4[a]**
KAPITI DORPER:	0.0340	0.0196	0.0080	**0.9384**
KAPITI RED MAASAI:	0.0448	**0.8219**	0.0290	0.1043
LOKICHOGGIO:	0.0190	0.0140	**0.9500**	0.0170
WEST POKOT:	**0.9090**	0.0540	0.0194	0.0176

b) Cluster II at *K*=3

Population	**Sub-cluster1[a]**	**Sub-cluster2[a]**	**Sub-cluster3[a]**
BUNGOMA:	0.0262	0.0110	**0.9628**
OLKIRAMATIAN BLACKHEAD SOMALI:	**0.9402**	0.0378	0.0220
OLKIRAMATIAN RED MAASAI:	0.0208	**0.9442**	0.0350

c) Cluster III at *K*=6

Population	**Sub-cluster1[a]**	**Sub-cluster2[a]**	**Sub-cluster3[a]**	**Sub-cluster4[a]**	**Sub-cluster5[a]**	**Sub-cluster6[a]**
SOMALI:	0.0534	0.0270	0.0160	**0.7731**	0.0488	0.0816
GARISSA:	0.0220	0.0314	**0.8554**	0.0356	0.0266	0.0290
LAMU:	0.0562	**0.7732**	0.0384	0.0584	0.0330	0.0408
LOITOKTOK:	0.1022	0.0484	0.0430	0.0614	**0.6484**	0.0966
MOYALE:	**0.7229**	0.0396	0.0358	0.0408	0.0574	0.1034
REMA:	0.0200	0.0320	0.0472	0.0600	0.1622	**0.6786**

d) Cluster IV at *K*=2

Population	**Sub-cluster1[a]**	**Sub-cluster2[a]**
VIPINGO:	**0.9768**	0.0232
HOMA BAY:	0.0818	**0.9182**

NB: [a]Estimates assumed admixture in the sampled genotypes. Contributions higher than 0.50 are in bold

As shown in Figure 7, Independent Component Analysis clustered populations in a manner similar to the clusters obtained using STRUCTURE. The first three components (IC1, IC2 and IC3) accounted for 0.3791, 0.3192 and 0.3017 of the variation respectively.

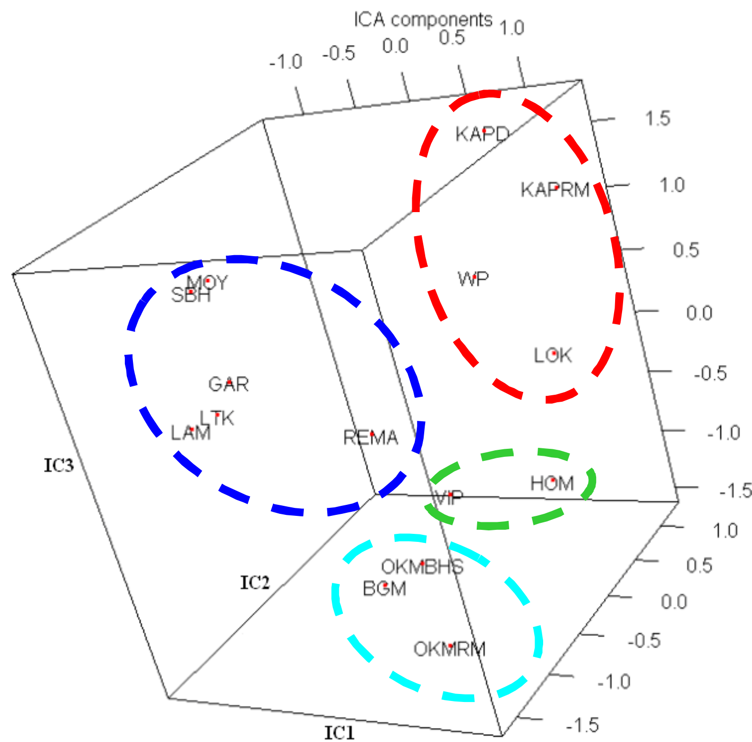


Figure 7: The Independent Component Analysis plot for the fifteen sheep populations

4.3 Hardy Weinberg Equillibrium

As shown in Table 13, 86 tests (38.2 %) out of 225 possible tests were in Hardy Weinberg disequilibrium, these are more than what would be expected by chance alone. All the loci showed significant deviations from Hardy Weinberg Equilibrium but none of them deviated in all the populations due to heterozygote deficiency (Table 13).

ILSTS005 had the highest number of deviations (12) and *SRCRSP9* had the least deviations (3). Garissa and Mombasa Red Maasai populations recorded the highest number of loci showing deviations (8), while the Kapiti Dorper population had the least number of loci (2) showing deviations (Table 13). With the alternative hypothesis $\boldsymbol{H_1}$ as Heterozygotes excess, five loci (*MCM42*, *OARHH47*, *DYMS1*, *OARVH72* and *SRCRSP9*) showed significant deviations from Hardy Weinberg Equillibrium, they all had a single deviation (Table 13). The Kapiti Red Maasai population had the highest number of loci (2) showing significant deviations while the Bungoma, Kapiti Dorper and Kijipwa populations each had one loci deviating significantly (Table 13).

Table 13: Results of the Fishers exact test on Hardy-Weinberg Equillibrium for all breeds and loci

Locus/Popln	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	REMA	WP	HOM
BM8125	ns	$*^{1}$	ns	ns	ns	ns	ns	$*^{1}$	ns	$**^{1}$	$***^{1}$	ns	ns	ns	ns
DYMS1	ns	ns	ns	ns	$*^{2}$	$**^{1}$	ns	$*^{1}$	ns	ns	ns	$*^{1}$	$*^{1}$	ns	ns
HSC	ns	ns	ns	ns	ns	$*^{1}$	ns	ns	ns	ns	ns	ns	$*^{1}$	$*^{1}$	ns
HUJ616	ns	ns	ns	ns	$*^{1}$	$*^{2}$	$*^{1}$	ns	ns	ns	$*^{1}$	ns	ns	$***^{1}$	ns
ILSTS005	$**^{1}$	$*^{1}$	$*^{1}$	ns	$***^{1}$	$**^{1}$	ns	ns	$***^{1}$	$**^{1}$	$*^{1}$	$**^{1}$	$***^{1}$	$***^{1}$	$***^{1}$
MAF209	ns	$*^{1}$	$***^{1}$	ns	ns	ns	ns	$*^{1}$	ns	ns	ns	ns	$*^{1}$	$*^{1}$	$*^{1}$
MCM42	ns	ns	$*^{1}$	ns	$**^{2}$	ns	ns	$**^{1}$	ns	ns	ns	ns	$*^{1}$	$*^{1}$	ns
OARFCB11	$**^{1}$	ns	ns	ns	ns	ns	ns	ns	$*^{1}$	ns	ns	$**^{1}$	$*^{1}$	ns	ns
OARFCB20	$*^{1}$	ns	$**^{1}$	ns	ns	ns	ns	ns	ns	ns	ns	$*^{1}$	ns	ns	$*^{1}$
OARFCB226	ns	ns	$*^{1}$	ns	ns	ns	ns	ns	ns	$*^{1}$	$**^{1}$	$***^{1}$	ns	$**^{1}$	$*^{1}$
OARHH47	$*^{2}$	ns	$***^{1}$	ns	ns	$*^{1}$	$*^{1}$	$*^{1}$	ns	$*^{1}$	ns	ns	$*^{1}$	ns	ns
OARJMP29	$**^{1}$	ns	$*^{1}$	ns	ns	ns	ns	ns	$*^{1}$	$*^{1}$	$**^{1}$	$*^{1}$	$*^{1}$	ns	ns
OARVH72	ns	ns	$*^{1}$	$*^{1}$	$***^{1}$	$*^{2}$	$*^{1}$	ns	$**^{1}$	ns	ns	$*^{1}$	ns	ns	ns
SRCRSP9	ns	ns	ns	$*^{2}$	ns	ns	ns	ns	ns	$*^{1}$	ns	ns	ns	ns	$*^{1}$
TGLA53	$*^{1}$	ns	ns	$*^{1}$	ns	$***^{1}$	$***^{1}$	$*^{1}$	$***^{1}$	$**^{1}$	ns	ns	ns	$**^{1}$	ns
Overall	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns

NOTE: **BGM**- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

Statistical significance: *- $p<0.05$, **- $p<0.01$, ***- $p<0.001$ ^{ns} – none significant (after Bonferroni corrections).

[1]- Heterozygote deficit

[2] - Heterozygote excess

4.4 Nei's Genetic Distances

To show the genetic relationships between the populations, Nei's chord distances (D_A) were computed (Table 14). The highest chord distance when all the markers were considered was 0.852, this occurred between the Olkiramatian Red Maasai and the Kapiti Dorper populations whereas the least (0.143) occurred between the Mombasa Red Maasai and the Loitoktok populations (Table 14). The Kapiti Dorper and Red Maasai populations were separated by a chord distance of 0.285, whereas the Kapiti Red Maasai and the Olkiramatian Red Maasai populations were separated by a chord distance of 0.755 (Table 14).

The Kapiti Dorper and Red Maasai populations were separated by chord distances of 0.506 and 0.361 respectively with the Kijipwa Red Maasai. The chord distance between the Somali and Loitoktok populations was 0.161 (Table 14). The Olkiramatian and Vipingo Red Maasai populations showed a chord distance of 0.741 (Table 14). The chord distances observed between the Lokichoggio as well as Olkiramatian Red Maasai populations and all the other populations studied were the largest (Table 14).

4.5 Phylogenetic relationships

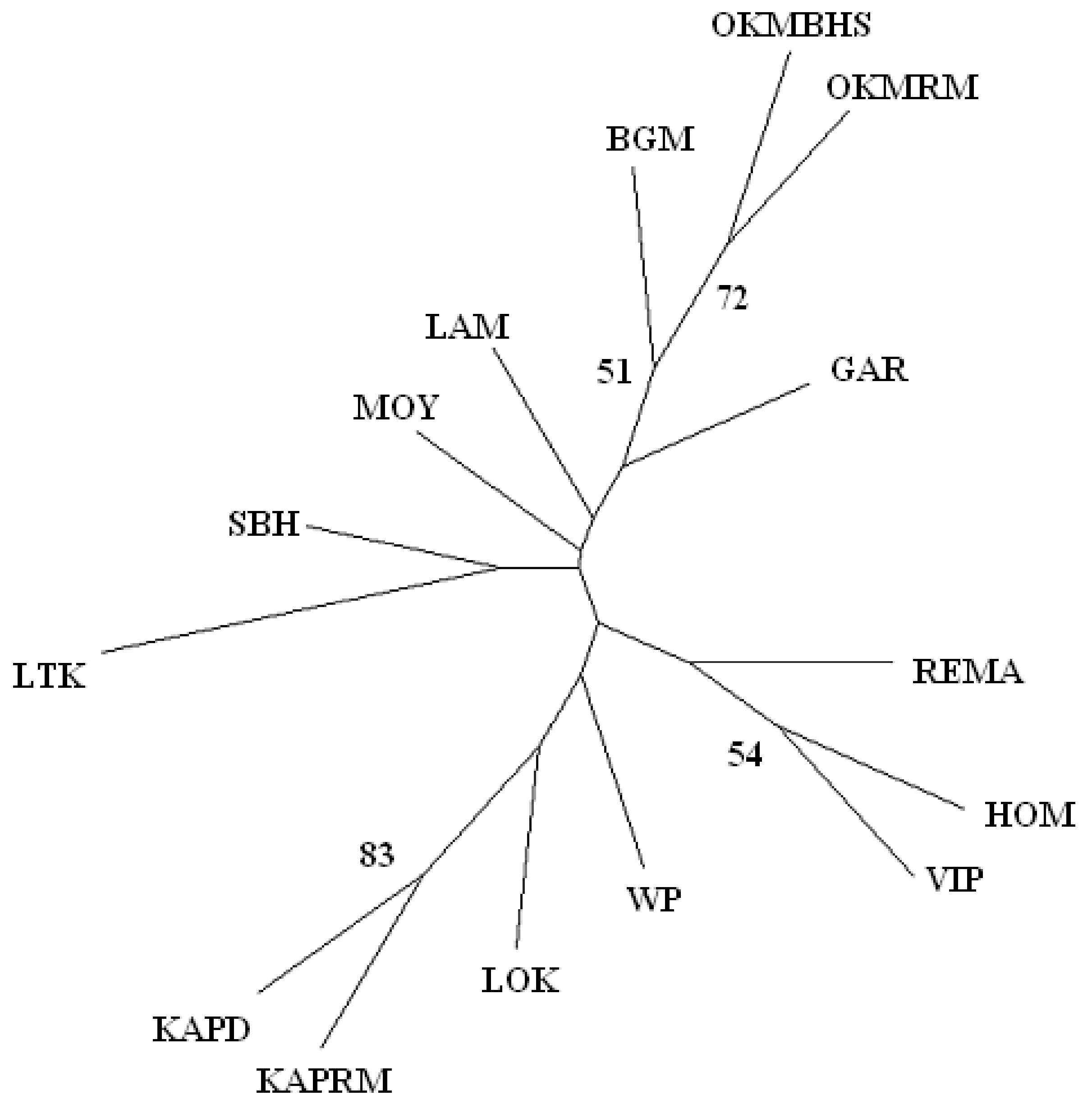
The un-rooted Neighbour Joining phylogenetic tree shown in Figure 8 was built using D_A distance matrices. The populations mainly clustered based on their geographical locations and population identity as well. The only surprising exceptions were the Bungoma population which clustered more closely with the Olkiramatian populations and the two Kapiti populations (Dorper and Red Maasai) which also

appeared more to be together (Figure 8). However, the bootstrap support across the phylogenetic tree was low with only Olkiramatian Red Maasai-Olkiramatian Blackhead populations, Bungoma-Olkiramatian populations, Kijipwa Red Maasai-Homabay populations and Kapiti Dorper- Kapiti Red Maasai groupings showing more than 50 % support (Figure 8). Such low bootstrap values could possibly signify the instability of the topology observed in the phylogenetic tree.

Table 14: Pairwise population matrix of Nei's chord distances (D_A) for the fifteen Kenyan sheep populations studied

POPLN	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	REMA	WP	HOM
BGM	0.000														
SBH	0.378	0.000													
GAR	0.319	0.232	0.000												
KAPD	0.557	0.419	0.419	0.000											
KAPRM	0.405	0.305	0.402	0.285	0.000										
VIP	0.339	0.316	0.348	0.506	0.361	0.000									
LAM	0.270	0.206	0.209	0.521	0.391	0.297	0.000								
LOK	0.577	0.465	0.665	0.618	0.497	0.704	0.655	0.000							
LTK	0.192	0.161	0.214	0.449	0.331	0.278	0.179	0.543	0.000						
MOY	0.243	0.193	0.224	0.396	0.334	0.368	0.175	0.509	0.152	0.000					
OKMBHS	0.323	0.357	0.302	0.617	0.521	0.620	0.357	0.536	0.333	0.296	0.000				
OKMRM	0.386	0.551	0.389	0.852	0.755	0.741	0.521	0.796	0.439	0.533	0.315	0.000			
REMA	0.255	0.177	0.226	0.456	0.281	0.196	0.251	0.513	0.143	0.226	0.419	0.533	0.000		
WP	0.274	0.254	0.327	0.378	0.234	0.313	0.358	0.538	0.208	0.266	0.473	0.625	0.206	0.000	
HOM	0.418	0.446	0.513	0.608	0.426	0.276	0.504	0.617	0.363	0.439	0.630	0.651	0.346	0.293	0.000

NOTE: **BGM**- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.



NOTE:BGM- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**- Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

Figure 8: An unrooted neighbour joining phylogenetic tree showing the relationship among the fifteen Kenyan sheep populations studied (only values showing >50 % bootstrap support are reported).

4.6 Linkage Disequilibrium

Appendix II shows the Linkage Disequilibrium (LD) chi-square results obtained for all the sheep populations using the Fisher's method. As shown three pairs of loci (*HUJ616* and *OARHH47*, *OARFCB226* and *OARHH47*, *OARFCB11* and *SRCRSP9*) had their *p*-values highly significant thus evidence for linkage disequilibrium.

4.7 Analysis of Molecular Variance (AMOVA)

Of the total genetic variation shown by Analysis of Molecular Variance (Table 15), 83.3 % is found within populations as opposed to 16.7 % which is among populations.

Source of Variation	Degrees of freedom	Sum of Squares	Mean Squares	Estimated Variation	Percentage variation
Among Populations	14	1489.365	106.383	2.431	16.7
Among individuals within populations	567	6875.362	12.126	12.126	83.3
Total	**581**	**8364.727**		**14.557**	**100**

Table 15: Analysis of Molecular Variance

4.8 Genetic Bottleneck

The results of tests for heterozygosities excess and deficit under the two phase mutation model (TPM) at $p<0.001$, provided evidence for a recent genetic bottleneck in the Kapiti Dorper population (Table 16). The estimations were based on 10 000 simulation replicates. Under the TPM, Kapiti Dorper and Lokichoggio populations had the highest heterozygote deficit of 0.99997 and 0.99377 respectively (Table 16). The Olkiramatian Red Maasai population had the least heterozygote deficit of 0.07571 (Table 16).

Population	Sample Size	One Tail P-value (TPM) H.deficit	H.excess
Bungoma	40	0.10388	0.90619
Somali Blackhead	40	0.55481	0.46704
Garissa	40	0.91559	0.09381
Kapiti Dorper	30	0.99997	0.00005***
Kapiti Red Maasai	42	0.97232	0.03186
Vipingo	42	0.88535	0.12619
Lamu	40	0.17957	0.83487
Lokichoggio	32	0.99377	0.00754
Loitoktok	43	0.70026	0.31934
Moyale	43	0.87381	0.13843
Olkiramatian Blackhead Somali	38	0.64014	0.38077
Olkiramatian Red Maasai	38	0.07571	0.93231
Mombasa Red Maasai	33	0.97937	0.02396
West Pokot	41	0.26224	0.75565
Homa Bay	40	0.84860	0.16513

***- statistically significant at p<0.001 based on 10 000 replications (after Bonferroni corrections)

Table 16: The Wilcoxon test for genetic bottlenecks

Table 17: Locus by population table of estimated null allele frequencies

Locus	Populations													
	BGM	SBH	GAR	KAPD	KAPRM	VIP	LAM	LOK	LTK	MOY	OKMBHS	OKMRM	MOMBRM	WP	HOM
BM8125	0.006	0.110	0.000	0.000	0.000	0.000	0.042	0.609	0.023	0.206	0.205	0.000	0.039	0.003	0.059
DYMS1	0.039	0.033	0.000	0.042	0.000	0.305	0.159	0.131	0.010	0.000	0.178	0.095	0.147	0.011	0.425
HSC	0.038	0.000	0.009	0.000	0.057	0.083	0.090	0.050	0.052	0.000	0.068	0.148	0.131	0.055	0.000
HUJ616	0.037	0.018	0.088	0.000	0.097	0.000	0.038	0.062	0.039	0.061	0.089	0.000	0.067	0.217	0.134
ILSTS005	0.229	0.130	0.188	0.018	0.285	0.288	0.024	0.031	0.234	0.197	0.132	0.357	0.212	0.296	0.275
MAF209	0.045	0.075	0.473	0.000	0.032	0.000	0.038	0.099	0.045	0.063	0.562	0.653	0.131	0.097	0.099
MCM42	0.000	0.070	0.111	0.184	0.000	0.020	0.007	0.389	0.055	0.017	0.071	0.011	0.094	0.113	0.030
OARFCB11	0.160	0.049	0.048	0.134	0.000	0.090	0.000	0.017	0.081	0.127	0.057	0.153	0.126	0.000	0.084
OARFCB20	0.075	0.006	0.310	0.029	0.046	0.000	0.000	0.108	0.000	0.033	0.005	0.974	0.000	0.013	0.088
OARFCB226	0.000	0.042	0.144	0.000	0.000	0.000	0.000	0.022	0.002	0.047	0.143	0.274	0.000	0.138	0.130
OARHH47	0.000	0.093	0.184	0.000	0.102	0.337	0.084	0.107	0.000	0.119	0.010	0.000	0.086	0.011	0.045
OARJMP29	0.139	0.000	0.111	0.131	0.022	0.073	0.036	0.000	0.090	0.030	0.181	0.067	0.118	0.014	0.000
OARVH72	0.040	0.000	0.118	0.101	0.185	0.000	0.058	0.206	0.161	0.000	0.038	0.064	0.008	0.080	0.038
SRCRSP9	0.000	0.034	0.000	0.000	0.033	0.000	0.098	0.302	0.023	0.082	0.038	0.022	0.007	0.000	0.054
TGLA53	0.196	0.211	0.159	0.086	0.104	0.293	0.310	0.187	0.200	0.264	0.059	0.178	0.016	0.162	0.066

NOTE: **BGM**- Bungoma, **SBH**-Somali Blackhead, **GAR**- Garissa, **KAPD**- Kapiti Dorper, **KAPRM**- Kapiti Red Maasai, **VIP**-Vipingo, **LAM**-Lamu, **LOK**- Lokichoggio, **LTK**- Loitoktok, **MOY**- Moyale, **OKMBHS**- Olkiramatian Blackhead Somali, **OKMRM**- Olkiramatian Red Maasai, **REMA**- Mombasa Red Maasai, **WP**- West Pokot, **HOM**- Homa Bay.

CHAPTER FIVE

5.1 DISCUSSION

Evolution is the outcome of population-level processes that transform genetic variation within species into genetic differences among species in time and space. Two central goals of evolutionary biology are to describe both the branching order of the history of life (phylogeny) and to describe the evolutionary forces (selective and non-selective) that explain why species differ from one another both in time and space.

Which evolutionary forces then could have led to the results obtained in this study?

5.1.1 Genetic Diversity

After genotyping the fifteen sheep populations across fifteen microsatellite loci, the mean number of alleles (MNA), expected and observed heterozygosities obtained were 7.70, 0.723 and 0.648 respectively, an indication of a high genetic diversity base among the Kenyan sheep populations studied. These values are comparable to the ones reported by Muigai *et al.,* (2003) MNA, H_E and H_O of 7.24, 0.74 and 0.69 respectively for the Sub-Saharan sheep as well as Lidga *et al.,* (2009) with MNA, H_E and H_O as 8.34, 0.74 and 0.696 respectively for Greek sheep breeds.

According to Buchanan *et al*., (1994), the number of alleles observed at a locus indicates the level of genetic variability at that locus and this has a direct impact on breed differentiation within a species. Nei (1987) indicates that MNA is an appropriate measure of genetic variation in comparison to heterozygosity since it is expected to be larger when the extent of polymorphism is higher, while the heterozygosity is hardly

affected by the low-frequency alleles. The mean allele number (allele diversity) therefore, provides a reasonable indicator of the levels of variability present within a breed assuming that the population is in mutation-drift equilibrium (MacHugh *et al.,* 1998). The allele diversity measure (7.70) reflected substantial level of genetic variability in the Kenyan sheep studied. Kapiti Red Maasai appeared to be most variable population as it harbored the highest number of alleles (140) whereas the Olkiramatian Red Maasai populations had the least number of alleles with 100 (Table 4).

It is plausible that the observed differences in the genetic variability between the nucleus herds' and farmers' sheep populations could be due to it that in the nucleus herds, mating between genetically related animals is highly regulated, as opposed to the farmers flocks where closely related individuals often mate. Usually the animals are grazed and housed separately based on their sex and are only brought together during the breeding and or mating period. This is quite unlike in the farmers' situations where the absence of clear pedigree records limits their ability to control mating between related animals. Since most of the farmers graze all the animals (both male and female) on the same field and no mating protective devices such as aprons are used, and the animals are also kept in the same 'bomas', thus the chances of related animals mating is expected to be high thereby reducing genetic variability observed in these populations.

The Olkiramatian Red Maasai population had the least diversity among the populations owned by farmers. This could be because very few farmers keep Red Maasai sheep in Olkiramatian (phenotypically speaking i.e. just by looking at the coat colour) and the number kept has also reduced. Given the importance of these sheep in

the Maasai culture some farmers still want to keep them, the farmers therefore have to share breeding rams between homesteads since they want to ensure the Red Maasai phenotype is maintained. This explains the high f_{IS} value in Table 10. The relatively higher genetic diversity observed in the Olkiramatian Blackhead Somali population could be due to the fact that since these animals are not native in this area and are brought in from different places with the origin mainly in northern Kenya to the market centres in Kajiado. Thus the probability that they share a recent or close ancestry is rather remote. The low genetic variability in the Bungoma and West Pokot populations could be attributed to the sharing of rams since many homesteads kept just a few animals. This therefore means that the number of rams kept will also be few thus the chances of sharing among homesteads increase.

The Kijipwa population, a nucleus herd owned and maintained by the Lafarge Ecosystems in Mombasa, had lower diversity compared to the Kapiti populations (a nucleus herd owned and maintained by the International Livestock Research Institute). The pedigree records of the Kijipwa sheep population are not meticulously maintained thus, mating between related animals cannot be avoided hence lowering the genetic variability resident in the herd. The main purpose for keeping and maintaining this herd is for the animals to graze in the trees, thus the chances for new animals to be brought into the flock are rather remote. The implication for this is that the animals held in the flock at any time are the ones which are likely to be ‘parents’ for the future animals thus increasing the chances of identity by descent of the alleles observed in the population. This can be supported by the high value of f_{IS} in Table 10. The high variability observed

in the Lokichoggio and Moyale populations could possibly be a result of crossbreeding with other populations from the neighbouring countries. These are mainly Sudan and Ethiopia, respectively, as they are found on the border between which the farmers move with their animals.

5.1.2 Genetic Bottleneck

Populations deviating from mutation–drift equilibrium, as indicated by Cornuet and Luikart (1996), exhibiting a significant heterozygosity excess for selectively neutral markers would be considered as having experienced a recent genetic bottleneck. This was observed in the Kapiti Dorper population (Table 16). The animals used to create this population were obtained from several source populations (thus forming a much wider base population than can be maintained on a single farm conveniently) which are genetically differentiated, it is thus expected that the resultant subpopulation will have an excess of observed heterozygotes with respect to expected heterozygotes in the Hardy-Weinberg Equillibrium (HWE).

The mixing of individuals from various populations with different allelic frequencies even though the source populations were in HWE will more often than not lead to such deviations. As one obtains only a few animals from one population, the chances for bringing along all the rare alleles present in that population are remote. The animals that are used to establish this Dorper herd are likely to be less than their effective population sizes as well. However, on mating the animals obtained from the many populations with different genetic backgrounds, heterozygosity will be high as

opposed to the mean number of alleles which will be reduced. Thus an imbalance between heterozygosity and allele number (Luikart *et al.,* 1998) is created which is evidence for a genetic bottleneck.

5.1.3 Genetic Distances and Phylogenetic Relationship

The genetic distances between the respective sheep populations as indicated in (Table 14) were highly varied. The highest population pairwise distances were between the Kapiti Dorper and the Lokichoggio populations with the rest of the sheep populations. The high distance observed between the Kapiti Dorper and the other populations was rather expected. This is because Dorper being an exotic sheep breed, it is genetically distantly related to other indigenous populations. This can imply that majority of the populations from where the samples were obtained do not have 'a lot' of the Dorper introgression. The only exception surprisingly was the close relationship found between the two Kapiti Populations- Dorper and Red Maasai. This being a nucleus herd, one would have expected to find the two animal breeds to be rather genetically distinct. The relationship observed could be due to it that, the animals used to establish the herd particularly the Red Maasai sheep were obtained from farmers and other nucleus herds thus a possibility of them not being pure to start with.

Based on the admixture results (Figure 6 in cluster I) the Kapiti Red Maasai population appeared to be the most affected. This could be because most of the Dorper sheep are likely to have been imported unlike the Red Maasai which are likely to have been obtained locally from farmers thus increasing the chances of genetic admixture

with other exotic breeds (Dorper in this case). Some Red Maasai rams were obtained from the Mombasa Red Maasai flock, realizing that this flock was not pure but had some Somali Blackhead and Dorper blood as well could possibly explain these results as well. Kwallah (2007) found out that the Olmagogo nucleus herd in Naivasha had 8 % Dorper in their genetic constitution. This finding could as well explain the observed Dorper proportion in the Kapiti Red Maasai population since some breeding rams to help establish the Kapiti flocks were obtained from Olmagogo. One could also implicate the herd management with the genetic relationship of the Kapiti populations, this could not be ruled out completely although it is very unlikely since any laxity on their side could be expected to affect both breeds in question. Allele scoring in the laboratory or sample spillover could be implicated as well. But since the observation is not limited to a few markers, this is unlikely though not impossible. A more important observation is that the Kapiti Red Maasai does not represent the whole Kenyan Red Maasai population and one should consider expanding the genetic base using other still existing Red Maasai populations. This observed admixture though undesirable should not be cursed in its entirety but the main focus therefore should be to limit the possible impact in future.

The high genetic distances observed between the Lokichoggio population and the other sheep populations can be attributed to the physical geographical barrier, given the long distance between Lokichogio and the other sites from which other sheep populations were sampled. This could therefore imply that this population has a lot of influence from the Sudan populations as a result of the common markets they share or even the raiding of animals across the Kenya-Sudan border. The close phylogenetic

relationship observed between the Kijipwa, Mombasa and Homa Bay populations as shown in Figure 8 was rather not surprising. This occurrence could be due to the identity by descent of the animals found in these populations. The Mombasa Red Maasai samples were obtained from a nucleus herd in Mombasa from which the animals making up the Vipingo herd were developed. The Homa Bay sheep belong to the East African fat-tailed sheep from which the Red Maasai was developed thus the close relationship between these populations.

The close relationships between the Garissa, Moyale, Lamu and Somali Blackhead populations as shown by the small genetic distances is in line with their geographical locations and pastoral-nomadic way of life of the occupants of these places. This can suggest the presence of common markets amongst them or even sharing of pastures and watering points thus enhancing gene flow between these populations. The frequent droughts occurring in most of these areas usually lead to massive deaths of animals and thus restocking is done mostly using animals bought from the immediate neighborhood of the affected areas. Quite surprisingly though, was the close relationship observed between the Somali Blackhead and the Loitoktok populations, this is an indication of the high gene flow from Somali into Kajiado, the home district of the Red Maasai sheep in Kenya. Equally, the relationship obtained in the Olkiramatian population pair; the Red Maasai and Blackhead shows the rate of indiscriminate crossbreeding going on in Kajiado between these two populations. This could be as a result of the animal market in Dagoretti Nairobi, where animals from Northern Kenya and Somali are brought in and the farmers from Kajiado buy them. These animals are

then let to breed with the sheep in Kajiado either knowingly or unknowingly by the farmers, this can be in an effort to increase the body size of the Red Maasai sheep originally kept. This is a really worrying trend as the unplanned crossbreeding will eventually lead to the dilution of the desired Red Maasai traits for sustainable agriculture like the resistance to *Haemonchus contortus*, ability to walk for long distance in search of water and pasture among others.

5.1.4 Population Genetic Differentiation

The probability that two alleles in an individual both descended from a single allele in an ancestor (that is, that they are ‘identical by descent’) is termed the inbreeding coefficient. Inbreeding depression which is the reduction in the fitness of offsprings produced by consanguineous mating (mating between two closely related individuals), can affect most fitness components causing reduced viability, lower fecundity, increased sterility, slower development and increased susceptibility to environmental stress (Frankham, 1995a; Keller and Waller, 2002), and consequently significantly decrease individual fitness thus markedly increasing the risk of extinction.

The average within population inbreeding was 0.109 for the fifteen sheep populations suggesting a rather high level of inbreeding since most of the populations had open breeding structures. The farmers’ flocks are relatively isolated so that the local parental individuals contribute majority of the next generation and the other farmer or village uses different animals as parents for the next generation as well. As shown the nucleus herds particularly Kapiti displayed the least inbreeding coefficients as compared

to the farmers' flocks. The high estimated inbreeding level in the Vipingo nucleus herd was most likely due to the poor pedigree recording thus increasing the chances for consanguineous matings.

According to Balloux and Lugon-Moulin (2002), differentiation estimators still remain the most commonly used parameters for defining population structures and assign individuals to source populations. Two parameters of population differentiation, θ_{ST} and its analogue G_{ST} (the coefficient of gene differentiation) were used to estimate the degree of population differentiation. All the populations studied showed highly significant differentiation between them (Table 9). All G_{ST} values obtained were below 0.2 which is a clear indication that all markers had low discriminatory power between populations. θ_{ST} values (Table 9) ranged between 0.026 and 0.127, implying some level of structuring within the subpopulations. Similar results were reported by Muigai (2003) and Kwallah (2007).

As shown in Table 8, G'_{ST} and θ_{ST} values indicate that most of the genetic variation (89.8 %) is within populations. The pairwise between-population differentiation test indicated that most populations were significantly different from each other with an overall θ_{ST} of 10.1 %. This is supported by the AMOVA analysis (Table 15), which shows that up to 83.3 % of genetic variation is within populations. These results are similar with what was reported by Muigai *et al.,* (2003) for the sub Saharan African sheep, Lawson *et al.,* (2007), Kusza *et al.,* (2008) and Tapio (2005b) all in European sheep but slightly higher than those obtained in studies of sheep using molecular markers like Bozzi *et al.,* (2009) in Italian sheep, Sodhi *et al.,* (2006) in

Indian sheep, and in Muigai *et al.,* (2009) in Kenyan sheep. Such differences are probably due to two main reasons: the different number and type of markers analyzed and the different geographic origin of the samples analysed. STRUCTURE (Figure 6), Independent Component Analysis (Figure 7) and Phylogenetic analysis (Figure 8) results suggested that the sheep populations can be separated into four clusters. The populations can be given as those with the highest Dorper genotype composition in cluster I (i.e. Kapiti Dorper, Kapiti Red Maasai, West Pokot and Lokichoggio), those with the highest Fat-tailed genotype composition in cluster IV (i.e. Vipingo and Homa Bay), sheep with the highest composition of Blackhead Somali genotype in cluster III (i.e. Somali Blackhead, Garissa, Lamu, Loitoktok, Moyale, and Mombasa Red Maasai) and cluster II comprising the Bungoma and Olkiramatian Sheep populations (East African Fat-tailed, Blackhead Somali and Red Maasai).

Upon further analysis of the four clusters in Figure 6, more sub-structuring became evident. Cluster I gave four sub-clusters with Kapiti Red Maasai having more than 10 % of Dorper genotype with the rest displaying approximately 2 % of the same. This finding supports the idea that the Red Maasai sheep used to establish the Kapiti nucleus herd had Dorper introgression as well. The West Pokot sheep were only 5.4 % Red Maasai, as we had indicated that this is a population likely influenced by the genotypes of the sheep from Uganda since it is found close to the border. The Lokichoggio population had the least Dorper influence, Red Maasai or even the sheep from West Pokot (keeping in mind that West Pokot was the closest sampling site to Lokichoggio than any other). This site (Lokichoggio) was as well close to Sudan than to

any other Kenyan site thus increasing the possibility of there being gene flow from Sudan into the Lokichoggio sheep.

Cluster II comprising the Bungoma and Olkiramatian populations (Blackhead Somali and Red Maasai) revealed that these three populations are not closely related to each other. The probable explanation of the observed close clustering could be that these populations are relatively far from all the other populations in this study. The consistently high pairwise θ_{ST} estimates between the Olkiramatian Red Maasai and all other populations in this study could be a reflection of different allele frequencies rather than unique alleles in this population.

Cluster III further sub-clustered to give Somali, Garissa, Moyale, Lamu, Loitoktok and Mombasa Red Maasai populations. The surprising inclusions in this cluster were the Loitoktok and Mombasa Red Maasai populations. Loitoktok population found in the heart of Maasai land was expected to be a Red Maasai population but its close relationship with Blackhead sheep is evidence to the extent to which the Red Maasai sheep in Kajiado have been introgressed with other sheep genotypes considered to be exotic to Kenya. This population had the highest contribution made by the Somali Blackhead population than even to the other Blackhead populations and as a result had between 4 % and 10 % of such exotic genotype in it. The Mombasa Red Maasai population as well had between 2 % and 6 % membership coefficient from Blackhead populations. This could as well explain why the Red Maasai in Kapiti are not pure since there are rams that came from this nucleus herd to help establish the flock at Kapiti. Of

the Blackhead populations, Garissa had the least 'foreign' genotype (not from Blackhead) composition of 2-3 %.

Cluster IV subdivided into the Vipingo and Homa Bay populations. As expected Homa Bay an outbreeding population had 8 % of the Vipingo genotype but the Vipingo only had 2 % of the Homa Bay genotype since the population has been having a closed breeding system with no pedigree records thus leading to inbreeding. In summary therefore, geographical proximity and interbreeding determine the population structure of the sheep populations studied.

5.1.5 Hardy-Weinberg Equillibrium

A significant deviation from HWE ($p < 0.05$) was observed in most of the loci within the populations analyzed (Table 13). Potential causes for such deviations in a given population include: (i) technical artifacts, such as the occurrence of null alleles, stuttering or large allele drop-out, (ii) the Wahlund effect, (iii) the selection of specific alleles and (iv) inbreeding (Hoarau *et al.*, 2002; Pereira *et al.*, 2009).

Table 17, shows the estimates of the occurrence of null alleles in this study. The occurrence of null alleles is a common problem in the study of microsatellites, and may be explained by the low efficiency of the primer hybridization used to amplify some loci, due to point mutation in one or more annealing sites of these primers (Dakin and Avise, 2004), besides the possible differential amplification of alleles with different sizes (Wattier *et al.*, 1998). The frequencies of the null alleles in this study vary from one population to another but some occur with a frequency higher than 20 % which

according to Dakin and Avise, 2004; Chapuis and Estoup, 2007 could partly explain the observed deviations from HWE.

Some populations as shown in Table 5, had the observed heterozygosity less than the expected heterozygosity which, according to Lawson *et al.,* 2007 and Peter *et al.,* 2007, suggests some degree of local inbreeding, presence of null alleles or population subdivision in each of the sampling regions. Weir and Cockerham's f_{IS} (0.109; Table 8), a measure of the deviation of the genotype proportions from the Hardy–Weinberg equilibrium at the population level, points to the presence of inbreeding and/or a Wahlund effect, as positive values suggest an excess of observed homozygotes (Wright, 1965). Also, at all loci (apart from SRCRSP9) the f_{IS} values showed significant ($p<0.05$) deviations from panmixia (Table 8).

The high value of Weir and Cockerham's F_{IT} (0.199) indicates a homozygote excess at the entire population level, as well. The possibility of population subdivision leading to the observed heterozygote deficiency is remote. This is because apart from the Kijipwa, Mombasa Red Maasai and Kapiti populations where one can expect that the animals might have been bought from different subpopulations, (so as to minimize pedigree relationship as the herds are established), most of the other sampled populations were farmer populations where animals were either born in the flock or bought from a nearby market place or neighbour. However, as shown in Table 18, the occurrence of null alleles in these markers is at high frequencies hence could also account for this heterozygote deficit. It can therefore be concluded that inbreeding and the occurrence of null alleles best explain the HWE departure in this study.

Six loci (*ILSTS005*, *TGLA53*, *OARVH72*, *OARJMP29*, *OARHH47* and *OARFCB226*) were found to be significantly associated with heterozygote deficiency in almost all the studied populations. There having been no selective association described for these markers, it is likely that this deficit could be a result of the farmers maintaining functional mutations through phenotypic selection. According to Charon (2004), *DYMS1* shows genetic linkage with *DYA* gene which is a possible candidate gene for resistance to *Haemonchus contortus* in sheep. It could therefore be expected that if selection was to be in force then, *DYMS1* (due to its linkage with *DYA* gene) should significantly contribute to the deviations from HWE in majority of the populations studied which actually is not the case here. Due to this, it might therefore be germane to exclude the possibility of selection contributing to the observed heterozygote deficit in this study.

5.1.6 Linkage Disequilibrium

The extent and range of Linkage Disequilibrium (LD) of two loci in a population, according to Feng-Xing *et al.,* (2007), is jointly affected by evolutionary forces (such as random drift, natural selection, mutation, and line origin), molecular forces such as historical recombination events, and the population's breeding history such as historical effective population sizes, intensity and direction of artificial selection, population admixture, and mating patterns.

In the current study however, random drift, inbreeding and population substructure (Wahlund effect) are very unlikely to have contributed to the disequilibrium observed. If one considers inbreeding as a possible causative factor, then all loci should

be affected equally by this phenomenon (as most populations showed the presence of inbreeding, Table 10) this is not the case here since only three locus pairs were in linkage disequilibrium. The only population that is likely to be affected by the Wahlund effect is the Kapiti population since the animals used to breed were initially obtained from various populations to minimize close genetic relationships. This being the case therefore, it is least expected that the effect from this population could be strong enough to affect the whole study population. The effect of random drift becomes more dramatic usually in cases of smaller effective population sizes which is least expected of the populations studied. It is therefore very tenuous to consider these factors as having led to the observed disequilibrium.

Occurrences of genetic admixture, mutations and Recombination events among loci could possibly account for the linkage disequilibrium observed in the sheep populations studied. The genetic admixture of the sheep populations seen from the STRUCTURE results (Figure 6) creates new LD among loci that are in no previous LD in all parental populations and alter the extent of LD for loci that are in LD in the parental populations. Microsatellite markers used in this study are known to undergo high mutation rates thus yielding new mutants. When such new mutations occur in a finite population, LD is created and the degree is dependent on the frequency of the allele that is haplotyped with the new mutation. As the copies of the mutant allele accumulate, the LD between this locus and other loci depend on recombinant rate, random drift, population admixture, and selection.

Alleles at neighboring loci tend to be inherited together and remain closely associated even in a segregating population. However, recombination (a genetic process by which the combinations of alleles observed at different loci in two parental individuals become shuffled in offspring individuals) usually reduces the possibility that such neighbouring loci will remain linked. In a large population under no selection, the estimate of LD decays at rate of 1- θ under random mating, where θ is the recombination fraction. In these populations studied, recombination effects could have possibly resulted in the observed LD due to the shuffling of the loci found in the parental population to create new combinations in the offsprings which might lead to novel LD associations or severing already existing LD associations.

5.2 Implications for the sake of conserving the Indigenous sheep studied

Diversity among farm animals within and among countries is of major interest to the scientific community as it is a significant resource for livestock development and for responding to the changing needs and production requirements. With increasing world population, there is concern that the growing demands for animal products are eroding these genetic resources especially in developing countries, where most of the diversity is found. In recognition of this concern, many efforts have begun to characterize animals in developing countries to provide a foundation for developing sustainable genetic improvement approaches and for the sake of conservation as well.

In this study for instance, the microsatellite markers used are neutral markers which are recommended for the conservation aspects but they do not give any indication on the functional diversity which is equally crucial for decision making on what to conserve. These markers are good for deciphering the general pattern in terms of population history but in case of selection then deviations from the general pattern at either single and or multiple genes will become evident. In the event that the selection is weak, then the microsatellites will be informative but if the selection is weak then no deviation *per se* might be observed. It is therefore prudent that markers found in coding regions are used so as to give more insight into the traits related largely to productivity, survival and reproductive traits, to which could be added disease resistances and consumer preferences which are needed for conservation are understood.

Having shown that the indigenous sheep in Kenya are composed of a diverse genetic pool threatened with inbreeding and admixture with exotic breeds, there is urgent need to set up conservation measures urgently. Conserving FAnGR in general presents analogies with the more general question of preserving biological diversity. In both cases, owing to the limited resources which can be devoted to conservation, the central question is ‘what to protect’. The choices are difficult and an operational framework is needed. Since the financial and other resources that need to be invested into conservation are always limited, these resources have to be spent in the most efficient and cost effective way. However, considering the genetic differentiation observed among the sheep populations studied, preservation of any one population will not protect all the variation in the species. Therefore, several populations of the

indigenous sheep throughout the entire range should be considered for conservation, thus there is need to develop a framework to allow rational decision-making in conservation programmes with regard to the question: which populations need be considered for conservation? Setting priorities for the conservation of the sheep populations studied will require a process that enables the identification of which populations contribute most to genetic diversity and have the greatest potential to contribute efficiently to present and future utilization and further development of that diversity. Additional criteria, such as cultural or heritage values of a breed, also affects priorities for conservation.

There is need to study the available trait diversity in these sheep populations. Trait diversity is based on heritable phenotypic differences among these populations. When the populations are put under comparable environmental conditions, trait diversity is necessarily indicative of underlying functional genetic diversity. For this reason, those that possess unique or distinctive trait combinations should be given high priority for conservation, because their unique phenotypic characteristics necessarily reflect unique underlying genetic combinations. Trait diversity expressed at the level of complex quantitative traits such as disease resistance, milk production or growth rate is generally given higher priority in conservation decisions than trait diversity associated with simply inherited traits such as coat or plumage colour, horn shape or body type. These simply inherited traits can be changed rapidly in response to owner preferences, whereas differences in complex quantitative traits generally involve larger numbers of genes,

take longer to change, and therefore have greater potential to reflect underlying genetic diversity.

Molecular measures of genetic diversity reflect differences in evolutionary history, but provide only indirect indications of genetic diversity in functional or potentially functional regions of the DNA. The populations that appear closely related based on allelic frequencies at neutral loci may nonetheless differ importantly at functional loci as a result of divergent selection histories. For this reason, trait diversity generally warrants first consideration in choosing candidates for conservation. However, phenotypically similar breeds may evolve as a result of different genetic mechanisms, and measures of molecular genetic diversity can aid the identification of populations that are superficially similar but genetically distinct. Conservation of genetically unique populations is, likewise, justified because these breeds are more likely to exhibit functional genetic diversity for traits previously unmeasured or unexpressed, but that may be of future importance in new markets, with exposure to new diseases, or under different production conditions.

Historical information or evidence of long term genetic isolation of the sheep populations studied can be used in the absence of information on trait or molecular genetic diversity, but can also be misleading. Population genetics theory shows that very low levels of movement of animals between seemingly isolated populations can effectively prevent meaningful genetic differentiation. Thus, populations with a history of genetic isolation are important candidates for careful trait and molecular genetic characterization, but final decisions on genetic uniqueness are better made using more

objective tools. It should be recognized, however, that livestock populations developed as a result of cultural preferences in isolated rural communities may be an important part of community identity and heritage. Conservation of such populations may merit consideration as part of broader community development efforts, regardless of their predicted value as a unique global genetic resource.

5.3 CONCLUSIONS

i. The mean number of alleles and heterozygosity levels observed in the sheep studied signify the high genetic diversity base that these genetic resources are.

ii. The level of genetic introgression of the indigenous sheep particularly in the heart of Maasai land, Loitoktok and Olkiramatian with the Somali Blackhead populations and genotypes from neighbouring countries is evident and worrying.

iii. The populations studied were found to be genetically differentiated based on geographical proximity of the populations and interbreeding.

iv. Most of the sheep populations were found to be undergoing some level of inbreeding which puts them at risk since it compromises their variability thus evolution.

v. The Red Maasai population from Kapiti nucleus herd had the highest Dorper introgression in all the populations studied while the Vipingo nucleus flock had a relatively high inbreeding rate a likely consequence for the lack of properly maintained pedigree records for the animals kept at the nucleus herd.

5.4 RECOMMENDATIONS

i. A detailed study of phenotypic similarities, management practices by the farmers, performance evaluations as well as the use of functional markers needs to be done so as to provide additional information for rational decision making on the conservation and genetic improvement of the sheep populations.

ii. Conservation efforts both *in situ* and *ex situ* need to be put in place urgently to curb this precious genetic resource from genetic dilution and erosion. The most genetically diverse population based on **MNA**, $\boldsymbol{H_E}$ and $\boldsymbol{H_O}$ need be conserved. However, considering the genetic differentiation observed among the sheep populations studied, preservation of any one population will not protect all the variation in the species. Therefore several populations of the indigenous sheep throughout the entire range should be considered for conservation. As part of this effort, there is need to help smallholder farmers and pastoralist communities restock the Red Maasai sheep due to their adaptability.

iii. Proper and sustainable breeding programmes need to be developed so as to help reduce the inbreeding observed in the populations.

iv. More intense sampling needs be done in Kenya particularly in the central, Western and the Northern part so as to give a complete genetic diversity picture of the Kenyan sheep.

CHAPTER SIX: REFERENCES

Aitman, T.J. Dong, R. Vyse, T.J. Norsworthy, P.J. Johnson, M.D. Smith, J. Manglon, J. Roberton-Lowe C., Marshall A.J., and Petretto E., 2006. Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **16***:*851–855.

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., Hayden, M.J., Howes, N., Sharp, P.J., Vaughan, P., Rathmell, B., Huttner, E. and Kilian, A. 2006. Diversity array technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor Appl Genet* **113**:1409-1420.

Amer, P.R., Mpofu, N. and Bondoc, O., 1998. Definition of breeding objectives for sustainable production systems. *In: Proceedings of the Sixth World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Armidale, NSW, Australia*, **28**: 97–104.

Amos, W. and Harwood, J. 1998. Factors affecting levels of genetic diversity in natural populations. *Philosophical Transaction of the Royal Society B*. **353**:177-86.

Amos, B. Schlötterer, C. and Tautz, D. 1993. Social structure of pilot whales revealed by analytical DNA profiling. *Science*, **260**:670–672.

Anderson, E.C. and Garza, J.C. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference, *Genetics* **172**:2567–2582.

Arranz, J.J., Bayon, Y. and Primitivo, F.S. 2001. Genetic variation at microsatellite loci in Spanish sheep. *Small Ruminant Research*. **39**:3-10.

Avise, J.C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, NewYork.

Baker, R.L., Nagda, S., Rodriquez-Zas, S.L., Southey, B.R., Audho, J.O., Aduda, E.O., and Thorpe, W. 2003. Resistance and resilience to gastro-intestinal nematode parasites and relationships with productivity of Red Maasai, Dorper and Red Maasai x Dorper crossbred lambs in the sub-humid tropics. Journal of Animal Science **76**:272-278.

Baker, R. L., Mugambi, J. M., Audho, J. O., Carles, A. B., and Thorpe, W. 2002. Comparison of Red Maasai and Dorper sheep for resistance to gastro-intestinal nematode parasites, productivity and efficiency in a humid and semi-arid environment in Kenya. *In "Proceedings of the 7th World Congress of Genetics Applied to Livestock Production*", pp.639-642, Montpellier, France, 19-23 August 2002.

Baker, R.L., Mwamachi D.M., Audho, J.O., Aduda, E.O. and Thorpe W. 1999. Genetic resistance to gastro-intestinal nematode parasites in Red Maasai, Dorper and Red Maasai X Dorper ewes in the sub-humid tropics. Animal Science (UK). **69**:335-344.

Baker, R.L. and Rege, J.E.O. 1994. Genetic resistance to diseases and other stresses in improvement of ruminant livestock in the tropics, *Proceedings of the Fifth World Congress on Genetics Applied to Livestock Production, vol. 20* University of Guelph, Ontario, Canada, 7–12 August 1994 pp. 405–412.

Balloux, F. and Lugon-Moulin, N. 2002. "The estimation of population differentiation with microsatellite markers". *Molecular Ecology* **11**: 155-165.

Barker, J.S.F. 2001. Conservation and management of genetic diversity: a domestic animal perspective, *Can. J. For. Res.* **31**:588–595.

Baumung, R., Simianer, H. and Hoffmann, I. 2004. Genetic diversity studies in farm animals - a survey. *Journal of Animal Breeding and Genetics* **121**: 361–73.

Beardmore, J. 1983. Extinction, survival, and genetic variation. Pages 125–151 in C. M. Schonewald-Cox, S. M. Chambers, F. MacBryde, and L. Thomas, editors. Genetics and conservation: a reference for managing wild animal and plant populations. Benjamin/Cummings, Menlo Park, California.

Bennewitz, J. and Meuwissen, T.H. 2005. A novel method for the estimation of the relative importance of breeds in order to conserve the total genetic variance. *Genet. Sel. Evol.***37**:315–337.

Bichara, M. Pinet, I. Schumacher, S. and Fuchs, R. 2000. Mechanisms of dinucleotide repeat instability in *Escherichia coli*. *Genetics* **154**:533-542.

Blench, R. 2003. The conservation of agrobiodiversity: ethics and economics. *Paper given at the 2nd HENRY WALLACE inter-American scientific conference*, San Jose, 19th – 21st Mar 2003.

Bökönyi, S. 1976. Development of early stock rearing in the Near East. *Nature* **264**:19–23.

Bonin, A. Bellemain, E. Eidesen, P.B. Pompanon, F. Brochmann, C. and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. Mol.Ecol., **13**:3261–3273.

Bosman, H.G. Moll, H.A.J. and Udo, H.M.J. 1997. Measuring and interpreting the benefits of goat keeping in tropical farm systems. Agric. Syst. **53**: 349– 372.

Bostein, D. White, R.L. Skolnick, M. and Davis, R.W. 1980. Construction of genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* **32**:314–331.

Bouzat, J.L. Cheng, H.H. Lewin, H.A. Westemeier, R.I. Brawn, J.D. and Paige, K.N. 1998a. Genetic evaluation of a demographic bottleneck in the Greater Prairie Chicken. Conservation Biology **12**:836-849.

Bowling, A.T. Eggleston-Stott, M.L. Byrns, G. Clark, R.S. Dileanis, S. and Wictum, E. 1997. Validation of microsatellite markers for routine horse parentage testing. *Anim Genet* **28**:247–252.

Bozzi, R. Degl'Innocenti, P. Rivera Diaz, L. Nardi, A. Crovetti, C. Sargentini, A. Giorgetti A. 2009. Genetic characterization and breed assignment in five Italian sheep breeds using microsatellite markers Small Ruminant Research **85**:50–57.

Buchanan, F.C. Galloway, S.M. and Crawford, A.M. 1994. "Ovine microsatellites at the *OARFCB5, OARFCB19, OARFCB20, OARFCB48, OARFCB129* and *OARFCB226* loci" *Animal Genetics* **25**:60.

Buchanan, F.C. Adama, L.J. LittleJohn, R. P. Maddox, J.F. and Crawford A.M. 1994. Determination of evolutionary relationships among sheep breeds using microsatellites.*Genomics* **22**:397-403.

Buckland, P. R. 2003. Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Ann Med*; **35**:308-315.

Carlson, J.E. Tulsieram, L. K. Glaubitz, J. C. Luk, V. Kauffeldt, C. and Rutledge, R. 1991. Segregation of random amplified DNA markers in f l progeny of conifers. *Theor Appl Genet* **83**:194-200.

Cavalli-Sforza, L.L. and Edwards, A.W.F. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* **21**: 550–570.

Chapuis, M. P. and Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol* **24**:621–631.

Charon, K. M. 2004. "Genes controlling resistance to gastrointestinal nematodes in ruminants." *Animal Science Papers and Reports* **22**: 135-139.

Charters,Y. M. Robertson, A. Wilkison, M. J. and Ramsay, G. 1996. PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. *oleifera*) using 5'-anchored simple sequence repeats. *Theor. Appl. Genet.*, **92**:442–447.

Chiara Sabatti and Neil Risch 2002. Homozygosity and Linkage Disequilibrium *Genet* **160**: 1707-1719.

Cipriani, G. Lot, G. Huang, W.G. Matarazzo, M.T. Peterlunger, E. and Testolin, R. 1999. AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: Isolation, characterization and cross species amplification in *Prunus*. *Theoretical and Applied Genetics* **99**:65-72.

Clark, J.D. and Williams, M. A. J. 1978. Recent archaeological research in southeastern Ethiopia (1974-1975: Some preliminary results.*Annalsd'Ethiopie* **11**: 19-44.

Cloete, S.W.P. Snyman, M.A. and Herselman, M.J. 2000. Productive performance of Dorper sheep. *Small Ruminant Research*. **36**(2):119-136.

Clutton-Brock, J. 1999. A natural history of domesticated mammals. Cambridge, UK: Cambridge University Press.

Conrad, D. F. Andrews, T. D. Carter, N. P. Hurles, M. E. and Pritchard, J. K. 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet*. **38**:75-81.

Cornuet, J.M. and Luikart, G. 1997. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**:2001-2014.

Cornuet, J.M. and Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**:2001–2014.

Curtain, C.C. 1971. On the origin of domesticated sheep. *Antiquity* **45**: 303-304.

DAGRIS, 2006. Domestic Animal Genetic Resources Information System (DAGRIS) Addis Ababa, Ethiopia, International Livestock Research Institute.

Dain, A.R. 1970. Separating the sheep and the goats on the basis of their chromosomes. *Nature* **228**: 560-561.

Dakin, E. E. and Avise, J.C. 2004. Microsatellite null alleles in parentage analysis. *Heredity* **93**:504–509.

de Haan, C. Steinfeld, H. and Blackburn, H. 1996. Livestock and the Environment: Finding the Balance, WRENmedia, Suffolk, UK 115 pp.

Delgado, C. Rosegrant, M. Steinfeld, H. Ehui, S. and Courbois, C. 1999. Livestock to 2020: The Next Food Revolution. 2020 Vision Discussion Paper No. 28, International Food Policy Research Institute, Washington, D.C, 39 pp.

Devendra, C. Morton, J.F. and Rischkovsky, B. 2005. Livestock systems. In: Owen, E. Kitalyi, A. and Jayasuria, M.C.N. Editors, *Livestock and Wealth Creation*, Nottingham University Press, United Kingdom, pp. 29–52.

Devendra, C. 2002. Potential productivity from small ruminants and contribution to improved livelihoods in developing countries. *In*: "Proceedings of the Thirty Ninth Reuniao Anual, Sociedade Brasilia de Zootechnia," Brasilia, Brazil.

Di Rienzo, A, Peterson, A.C. Garza, J.C. Valdes, A.M. Slatkin, M. and Freimer, N.B. 1994. Mutational processes of simple sequence repeat loci in human populations. Proceeding of National Academy of Sciences **91**:3166-3170.

Dieringer, D. and Schlötterer, C. 2002. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* **3**:167–169.

Diez-Tascon, C. Littlejohn, R.P. Almeida, P.A.R. and Crawford, A.M. 2000. Genetic variation within the Merino sheep breed: Analysis of closely related populations using microsatellites. *Animal Genetics* **31**: 243-251.

Dorward, A. Anderson, S. and Clark, S. 2001. Asset functions and livelihood strategies: a framework for pro-poor analysis, policy and practice. European Association of Agricultural Economics. Seminar on Livelihoods and Rural Poverty, September 2001.

Earl, D.A. 2009. Structure Harvester v0.3, from website: http://users.soe.ucsc.edu/~dearl/software/struct_harvest/

Edwards, C. Connellan, J. J. Wallace, P. F. Park, S. D. E. McCormick, F. M. Olsaker, I. Eythórsdó ttir, E. MacHugh, D. E. Bailey, J. F. and Bradley, D. G. 2003. Feasibility and utility of microsatellite markers in archaeological cattle remains from a Viking Age settlement in Dublin. *Animal Genetics*, **34**:410–416.

Eichler, E.E. 2006. Widening the spectrum of human genetic variation. *Nat. Genet*.**38**: 9–11.

El Mousadik and Petit 1996a. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L. Skeels)] endemic to Morocco, *Theor. Appl. Genet.* **92**:832–836.

Epstein, H. 1971. *The Origin of the Domestic Animals of Africa.* Africana Publishing Corporation: New York, USA.

Epstein, H. 1954. The fat-tailed sheep of east Africa. The East African Agricultural Journal **20**:109–17.

Erlich, H.A. 1989. PCR technology: principles and applications for DNA amplification. Stockton Press, New York.

Erschadi, S. Haberer, G. Schöniger, M. and Torres-Ruiz, R.A. 2000. Estimating genetic diversity of Arabidopsis thaliana ecotypes with amplified fragment length polymorphisms (AFLP). *Theor App Genet* **100**: 633–640.

Estoup, A. Jarne, P. and Cornuet, J.M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**:1591- 1604.

Evanno, G. Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.

Excoffier, L. Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance Inferred from metric distances among DNA haplotypes, application to human mitochondrial DNA restriction data. *Genetics*, **131**:479-491.

Falconer, C.S. and Mackay, T.F.C. 1996. Introduction to Quantitative Genetics. Longman Group Ltd. Harlow.

Falush, D. Wirth, T. Linz, B. Pritchard, J.K. and Stephens, M. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* **299:** 1582–1585.

Fan, B. S. Yang, L. Liu, B. Yu, M. Zhao, S.H. and Li, K. 2003. Characterization of the genetic diversity on natural populations of Chinese miniature pig Breeds. *Animal Genetics*, **34**:465–476.

FAO, 2007. *The State of the World's Animal Genetic Resources for Food and Agriculture*, edited by Barbara Rischkowsky and Dafydd Pilling. Rome.

FAO, 2007. Report of the international technical conference on animal genetic resources for food and agriculture. Interlaken, Switzerland, 3–7 September 2007. http://www.fao.org/AG/againfo/programmes/en/genetics/angrvent2007.html.

FAO, 2004. DAD-IS. Domestic Animal Diversity Information Systems. Download from *http://www.fao.org/dad-is* at November 29, 2004.

FAO, 2000. World Watchlist for Domestic Animal Diversity. Rome. Italy, FAO.

FAO, 1998. Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Measurements of Domestic Animal Diversity (MoDAD): OriginalWorking Group Report.

FAO, 1997. The Global Strategy for the Management of Farm Animal Genetic Resources: a Call for Action. FAO Publishing, Rome. *Genetics* **121**: 361–73.

Felsenstein, J. 2004. PHYLIP: Phylogeny Inference Package, Version 3.6. University of Washington, Seattle, USA.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791.

Feng-Xing Du, Archie, C. and Michael, M. Lohuis 2007. Characterizing Linkage Disequilibrium in Pig Populations *Int. J. Biol. Sci.*, **3** (3):166-178.

Feuk, L. Carson, A.R. and Scherer, S.W. 2006. Structural variation in the human genome. *Nat Rev Genet* **7**(2):85-97.

Field, D. and Wills, C. 1998. Long polymorphic microsatellites in simple organisms. Proceeding of the Royal Society of London, Series B: *Biological Sciences* 263:209-215.

Field, D. and Wills, C. 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc. Natl. Acad. Sci*. U.S.A. **95**: 1647–1652.

Frankham, R. Ballou J.D. and Briscoe, D.A. 2002. Introduction to Conservation Genetics, Cambridge University Press, Cambridge.

Frankham, R. 1995. Conservation genetics. *Annual Review of Genetics* 29: 305–327.

Freeman, J.L. Perry, G.H. Feuk, L. Redon, R. McCarroll, S.A. Altshuler, D.M. Aburatani, H. Jones, K.W. Tyler-Smith, C. and Hurles, M.E. 2006. Copy number variation: New insights in genome diversity. *Genome Res.* **16**: 949–961.

Freeman, A. R. Meghan, C. M. MacHugh, D. E. Loftus, R. T. and Achukwi, M. D. 2004. Admixture and diversity in West African cattle populations. *Mol. Ecol.* **13**: 3477.

Fries, R. and Durstewitz, G. 2001. Digital DNA signatures for animal tagging. *Nat.Biotechnol*. **19**: 508.

Gatenby, R.M. 1986. Sheep Production in the Tropics and Sub-Tropics, Longman Inc., New York 351 pp.

Gathuka, Z.G. 1986. Management of small ruminant production systems in the farming areas of Kenya In The "Workshop on the Improvement of Small Ruminants in Eastern and Southern Africa" (Adeniji, K.O. (ed.); Kategile, J.A. (ed.) p. 195-205 Nairobi (Kenya) 18-22 Aug 1986.

Geist, V. 1971. "Mountain sheep" United Press, Chicago.

Gicheha, M.G. Kosgey, I.S. Bebe, B.O. and Kahi, A.K. 2006. Evaluation of the efficiency of alternative two-tier nucleus breeding systems designed to improve meat sheep in Kenya, *J. Anim Breed. Genet.* **123**:247–257.

Giles, R.E. Blanc, H. and Cann, H.M. 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, **77**:6715–6719.

GOK (Government of Kenya), 2003. Sheep and Goats Annual Report 2003. (Ministry of Livestock and fisheries, Department of Sheep and Goats, Nairobi)

Gonzalez, E. Kulkarni, H. Bolivar, H. Mangano, A. Sanchez, R. Catano, G. Nibbs, R.J. Freedman, B.J. Quinones, M.P. and Bamshed, M.J. 2005. *The influence of CCL3L1 gene—containing segmental duplications on HIV-1/AIDS susceptibility. Science* **307***:*1434–1440.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm. Updated from Goudet (1995).

Grandin, B.E. Bekure, S. and Nestel, P. 1991. Livestock transactions, food consumption and household budgets. In: Bekure, S. de Leeuw, P.N. Grandin, B.E. Neate, P.J.H. (eds), Maasai Herding: An Analysis of the Livestock Production System of Maasai Pastoralists in Eastern Kajiado District, Kenya , Vol. 4 . International Livestock Centre for Africa (ILCA) Systems Study, Addis Ababa, Ethiopia, pp . 103–114.

Hahn, E. 1896. Die Haustiere und ihre Beziehungen zur Wirtschaft des Menschen. Eine geographische Studie. Leipzig, Germany: Duncker and Humblot.

Hamrick, J. L. and Godt, M. J. 1990. Allozyme diversity in plant species. In Brown, A. H. D. Clegg, M. T. Kahler, A. L. and Weir, B. S. [eds.], Population genetics and

germplasm resources in crop improve- ment, 43-63. Sinauer Associates, Sunderland,MA.

Hanslik, S. Harr, B. Brem, G. and Schloterer, C. 2000. Microsatellite Analysis Reveals Substantial Genetic Differentiation Between Contemporary New World And Old World Holstein Fresian Populations. *Animal Genetics*.**31**: 31-38.

Hardy G. 1908. Mendelian proportions in a mixed population *Science* **28**:49-50.

Hartl, D.L. and Clark, A.G. 1997. Principles of Population Genetics (3rd edition) Sunderland, MA, Sinauer Associates, Inc.

Hedrick, P.W. 2001. Conservation genetics: where are we now? *TREE* **16**: 629–636.

Hentschel, C.C. 1982. Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S1 nuclease. *Nature* **295**:714-716.

Heutink, P. and Oostra, B. 2002. Gene finding in genetically isolated populations. *Hum Mol Genet*; **11**: 2507–15.

Hiendleder, S. Kaupe, B. Wassmuth, R. and Janke, A. 2002. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **269**: 893-904.

Hiendleder, S. Mainz, K. Plante, Y. and Lewaiski, H. 1998. Analysis of Mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: No evidence for contribution from Urial and Argali. *Journal of Heredity* **89**:113-120.

Hoarau, G. Rijnsdorp, A.D. Van der Veer, H.W. Stam, W.T. and Olsen, J.L. 2002. Population structure of plaice (*Pleuronectes platessa* L.) in Northern Europe*:* microsatellites revealed large scale spatial and temporal homogeneity. *Mol.Ecol.***11**:1165-1176.

Holst, P.J. 1999. Recording and on-farm evaluations and monitoring: breeding and selection. *Small Ruminant Research* **34**: 197-202.

Hongtrakul, V. Huestis, G.M. and Knapp S.J. 1997. Amplified fragment length polymorphism as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. *Theor Appl Genet* **95**:400– 407.

Hu, N. Wang, C. Hu, Y. Yang, H. H. Kong, L.H. Lu, N. Su, H. Wang, Q.H. Golstein, A.M. Buetow, K.H. Emmert-Buck, M.R. Taylor, P. R. and Lee, M.P. 2006. Genome-wide loss of heterozygosity and copy number alteration in esophageal squamous cell carcinoma using the Affymetirx GeneChip Mapping 10K array. *BMC Genomics* **7**(299):1-16.

Huang, J. Wei, W. Zhang, J. Liu, G. Bignell, G.R. Stratton, M.R. Futreal, P.A.Wooster, R. Jones, K.W. and Shapero, M.H. 2004. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics* **1**: 287–299.

Huson, D. H. and Bryant, D. 2006. Application of Phylogenetic Networks in Evolutionary Studies, *Mol. Biol. Evol.* **23**(2):254-267.

Inoue, A. Yoshida, N. Omoto, Y. Oguchi, S. Yamori, T. Kiyama, R. and Hayashi, S. 2002. Development of cDNA microarray for expression profiling of estrogen-responsive genes. *Journal of Molecular Endocrinology* **29:**175–192.

Isagi, Y. and Suhandono, S. 1997. PCR primers amplifying microsatellite loci of *Quercus myrsinifolia* Blume and their conservation between oak species. *Molecular Ecology* **6**:897-899.

Jaccoud, D. Peng, K. Feinstein, D. and Kilian, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.*[online], **29**:4 e25.

Jaitner, J. Sowe, J. Secka-Njie, E. and Dempfle, L. 2001. Ownership pattern and management practices of small ruminants in The Gambia – implications for a breeding programme. *Small Ruminant Research* **40**:101-108.

Jakobsson M., and Rosenberg, N.A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**:1801-1806.

Jarne, P. and Lagoda, P. J. L. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**:424-429.

Jobling, S.A. Jarman, C. Holmberg, N. Blake, C. and Verhoeyen, M.E. 2003. Immunomodulation of enzyme function in plants by single-domain antibody fragments. *Nat. Biotechnol.* **21**:77–80.

Jobling. S. Reynolds, T. White, R. Parker M.G. and Sumpter, J.P. 1995. A variety of environmentally persistent chemicals, including some phthalate plasticizers, are weakly estrogenic. *Environmental Health Perspectives* **103**:582-587.

Jongmans, M.C. Admiraal, R.J. van der Donk, K.P. Vissers, L.E. Baas, A.F. Kapusta, L. van Hagen, J.M. Donnai, D. de Ravel, T.J. Veltman, J.A. Geurts van Kessel, A. De Vries, B.B. Brunner, H.G. Hoefsloot L.H. and van Ravenswaaij, C.M. 2006. CHARGE

syndrome: the phenotypic spectrum of mutations in the CHD7 gene. *J Med Genet* **43**:306–314.

Karp, A. and Edwards, K. J. 1997. Molecular techniques in the analysis of the extent and distribution of genetic diversity. In: Ayad, W.G. Hodgkin, T. Jaradat, A. Rao, V.R. (1997) Molecular genetic techniques for plant genetic resources. Report of an IPGRI workshop, 9–11 October 1995, Rome, Italy. International Plant Genetic Resources Institute, Rome, Italy, pp 11–22.

Keller, L.F. and Waller, D. M. 2002. Inbreeding effects in wild populations. *Trends Ecol. Evol.***17**:230–241.

Kennedy, G. 2003. Large scale genotyping of complex DNA. *Nat. Biotechnol.*, **21**:1233–1237.

Kilian, A. Huttner, E. Wenzl, P. Jaccoud, D. Carling, J. Caig, V. Evers, M. Heller-Uszynska, Cayla, C. Patarapuwadol, S. Xia, L. Yang, S. and Thomson, B. 2005. The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In: Tuberosa R, Phillips RL, Gale M (eds) Proceedings of the international congress "In the wake of the double helix: from the green revolution to the gene revolution", 27–31 May, 2003, Avenue Media, Bologna, Italy, pp 443–461.

Kinyamario, J.I. and Ekeya, W.N. 2001. Mechanisms of drought management by African pastoralists. In: Proceedings of the Animal Production Society of Kenya (APSK) Symposium, 7-8 March, 2001, Egerton University, Njoro, Kenya, pp. 117-122.

Kliebenstein, D. Lambrix, V. Reichelt, M. Gershenzon J. and Mitchell-Olds, T. 2001. Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in Arabidopsis thaliana. *Plant Cell* **13**: 681–693.

Kolodner, R.D. and Marsischky, G.T. (1999) Eukaryotic DNA mismatch repair. *Current Opinion Genetics* **9**:89-96.

Kosgey, I.S. Rowlands, G. J. van Arendonk, J. A. M. and Baker, R. L. 2008. Small ruminant production in smallholder and pastoral/ extensive farming systems in Kenya. *Small Ruminant Research* **77**: 11-24.

Kosgey, I.S. and Okeyo, A.M. 2007. Genetic improvements of small ruminants in low-input, smallholder production systems: Technical and infrastructural issues. In: Notter, D.R. Barker, R.L. Cockett, N. (Eds.), The Outlook for Quantitative and Molecular Genetic Applications in Improving Sheep and Goats. *Small Ruminant Res*. **70**:76–88.

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

Kosgey, I. S. Rowlands, G. J. van Arendonk, J. A. M. and Baker, R. L. 2006b. Small ruminant production in smallholder and pastoral/ extensive farming systems in Kenya. *Small Ruminant Research*.

Kosgey, I.S. 2004. Breeding Objectives and Breeding Strategies for Small Ruminants in the Tropics. **Ph.D. Thesis**, Wageningen University, The Netherlands, 272 pp.

Krawczak, M. (1999) Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis*, **20**: 1676–1681.

Kruglyak, L. 1997. The use of a genetic map of biallelic markers in linkage studies. *Nature Genet.* **17**: 21–24.

Kwallah, A.B. 2007. Characterization of Indigenous fat-tailed sheep in some parts of Kenya: A Microsatellite Approach. **MSc Thesis**, Biochemistry Department, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya.

Lande, R. 1988. Genetics and demography in biological conservation. *Science* **241**:1455–1460.

Lawson, L.J. Handley, K. Byrne, F. Santucci, S. Townsend, M. Taylor and G.M. Hewitt, 2007. Genetic structure of European sheep breeds, *Heredity* pp. 1–12.

Lebbie, S.H.B. and Ramsay, K. 1999. A perspective on conservation and management of small ruminant genetic resources in sub-Saharan Africa. *Small Ruminant Research* **34**:231- 247.

Ledin, I. 1997. Sheep and goats under African conditions, *IRDC Curr.* **13**–14 , pp. 12–15.

Levinson, G. and Gutman, G. A. 1987. Slipped strand mispairing: A major mechanism for DNA sequence evolution. *Molecular and Biology Evolution* **4**:203-221.

Lezar, S. Myburg, A. A. Berger, D. K. Wingfield, M.J. and Wingfield, B.D. 2004. Development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. *Theor Appl Genet* **109**:1329–1336.

Ligdaa, Ch. J. Altarayrah, A. Georgoudis, and the ECONOGENE Consortium, 2009. Genetic analysis of Greek sheep breeds using microsatellite markers for setting conservation priorities. *Small Ruminant Research* **83**:42–48.

Lindley, H. Broae, A. Damoglou, A.P. Darskus, R.L. Elleman, T.C. Gillespie, J.M. and Moore, C.H. 1971. The high-sulphur protein fraction of keratins. *Applied Polymer Symposium* **18**: 21-35.

Linnæus, C. 1758. Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata. - pp. [1-4], 1-824. Holmiæ. (Laurentii Salvii).

Luikart, G. Allendorf, F.W. Cornuet, J.-M. and Sherwin,W. B. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *J. Hered.* **89**: 238–247.

Lupski, J. R. and Stankiewicz, P. 2005. "Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes." *PLoS Genet* **1**(6): e49.

MacHugh, D. E. Loftus, R. T. Cunningham, P. and Bradley, D. G. 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim. Genet.* **29**:333–340.

MacHugh, D.E. Shriver, M.D. Loftus, R.T. Cunningham, P. and Bradley, D.G. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**:1071–86.

Marklund, S. Ellegren, H. Eriksson, S. Sandberg, K. and Andersson, L. 1994. Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Anim Genet* **25**:19–23.

Marshal, F. 2000. The origins and spread of domestic animals in East Africa. In: The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography (Ed. by R.M. Blench & K.C. MacDonald), pp. 191–218. University College London Press, London.

Marth, G. Yeh, R. Minton, M. Donaldson, R. Li, Q. Duan, S. Davenport, R. Miller, R.D. and Kwok, P.Y. 2001. Single nucleotide polymorphisms in the public domain: how useful are they? *Nat. Genet.*, **27**:371–372.

Martinez-Arias, R. Calafell, F. Mateu, E. Comas, D. Andres, A. and Bertranpetit, J. 2001. Sequence variability of a human pseudogene. *Genome Res.***11**:1071–1085.

Mason, I.L. 1996. A World Dictionary of Livestock Breeds, Types and Varieties. Fourth Edition. CAB International. Wallingford. UK. pp. 252.

Mason, I.L. 1988. A World Dictionary of Livestock Breeds Types and Varieties. CAB International, 348pp.

Mason, I.L. and Maule, J.P. 1960. *The Indigenous Livestock of Eastern and Southern Africa* (Tech.Comm.N° 14, Commonw.Bur.Anim.Breed. Genet.). Commonwealth Agricultural Bureaux: Farnham Royal, UK.

McCarroll, S.A. and Altshuler, D.M. 2007. Copy-number variation and association studies of human disease. *Nat. Genet.*, **39**:S37-S42.

McCarroll, S.A. Hadnott, T.N. Perry, G.H. Sabeti, P.C. Zody, M.C. Barrett, J.C. Dallaire, S. Gabriel, S.B. Lee, C. and Daly, M.J. 2006. Common deletion polymorphisms in the human genome. *Nat. Genet.* **38**: 86–92.

Meadows, J. R. S. 2005. Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. Journal of Heredity, **96**(5) 494-501.

Mills, L. S. and Smouse, P. E. 1994. Demographic consequences of inbreeding in remnant populations. *American Naturalist* **144**:412–431.

Milne, C. 2000. The history of the Dorper sheep. *Small Ruminant Research*. **36**(2):99-102.

Miyata, T. 1982. Evolutionary changes and functional constraint in DNA sequences. pp.233-266, *in* Kimura, M. ed. Molecular evolution, protein polymorphism and the neutral theory. Japan Scientific, Tokyo.

MLFD 2003. Ministry of Livestock and Fisheries Development, Annual Report. MLFD, Nairobi, Kenya.

Mueller, J.C. 2004 "Linkage disequilibrium for different scales and applications", *Briefings in Bioinformatics* **5**(4):355-364.

Mueller, U. G. and Wolfenbarger, L. L. 1999. AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**: 389-394.

Muigai, A.W.T. Okeyo, A.M. Kwallah, A.K. Mburu, D. and Hanotte O. 2009. Characterization of sheep populations of Kenya using microsatellite markers: Implications for conservation and management of indigenous sheep populations. *South African Journal of Animal Science*, **39** (Supplement 1).

Muigai, A.W.T. 2003. Characterization and conservation of indigenous animal genetic resources: Genetic diversity and relationships of fat-tailed and thin-tailed sheep of Africa. **PhD Thesis**, Biochemistry Department, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya.

Mwacharo, J.M. Otieno, C.J. Okeyo, A.M. and Aman, R.A. 2002. Characteristics of indigenous fat-tailed and fat-rumped hair sheep in Kenya: diversity in blood proteins. *Tropical Animal Health and Production* **34**:515-524.

Nadler, C. F. Korobitsina, K.V. Hoffmann, R.F. and Vorontsov, N.N. 1973. Cytogenetic differentiation, Geographic distribution and Domestication in Palearctic sheep (*Ovis*) *Zeitschrift fûr Säugetier-kunde* **38**:109-125.

Naess, M.W. Lhagyal, D. Yangzom, D. Mathiesen, P. Fox, J.L. and Bardsen, B.J. 2004. Nomadic pastoralism in the Aru Basin of Tibet's Chang Tang, *Rangifer* **15**:39–46.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press: New York.

Nei, M, Tajima, F, and Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J Molec Evol* **19**: 153–170.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.* **41**: 225-233.

Nei, M. 1973. Analysis of genetic diversity in subdivided populations. Proceedings of the National Academy of Sciences **70**:3321-3323.

Nei, M. 1973. The theory and estimation of genetic distances. *In Genetic structure of Populations*. N.E Morten (ed.), University Press of Hawaii, Honolulu, 45-54.

Nei, M. 1972. Genetic distance between populations. *The American Naturalist*, **106**:283-292.

Olivier, J.J. Moyo, Montaldo, S. Thorpe, H.H. Valle, W. Zarate, A. Trivedi, K.R. 2002. Integrating genetic improvement into livestock development in medium- to low-input production systems. In: Proceedings of the Seventh World Congress on Genetics Applied to Livestock Production, vol. 33, 19-23 August, 2002, Montpellier, France, pp. 365-372.

Ota, T. 1993. Dispan: Genetic Distance and Phylogenetic Analysis. Pennsylvania, Pennsylvania State University.

Palacios, C. Kresovich, S. and González-Candelas, F. 1999. A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). *Mol. Ecol*., **8**:645–657.

Parker, H. G. Kukekova, A. V. Akey, D. Goldstein, T. O. and Kirkness, E. F. 2007. Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res.* **17**: 1562–1571.

Peakall R. and P. Smouse 2006. GenAlEx 6. Genetic Analysis in Excel. Population Genetic and Software for Teaching and Research.

Pelant, R.K. Chandra, B. Pu, J.B. Lohani, M. Suknaphasawat, N. and Xu, G. 1999. Small ruminants in development: the Heifer Project International experience in Asia, *Small Rumin. Res.* **34**: 249–257.

Pereira, L.H.G. Foresti, F. and Oliveria, C. 2009. Genetic structure of the migratory catfish *Pseudoplatystoma corruscans* (Siluriformes, Pimelodidae) suggests homing behavior.*Ecol Freshw Fish* **18**:215-225.

Peter, C.M. Bruford, T. Perez, S. Dalamitra, G. and Hewitt, G. and Erhardt, G. 2007. The genetic diversity and subdivision of 57 European and middle-eastern sheep breeds, *Anim. Genet.* **38**:37–44.

Peters, K.J. 1988. The importance of small ruminants in rural development. *Animal Research and Development* **28**:115-125.

Plaschke, J., Ganal, M.W. and Roder, M.S. 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.*, **91**:1001–1007.

Pritchard, J.K. Wen, X. and Falush, D. 2007. Documentation for *structure* software: Version 2.2. http://pritch.bsd.uchicago.edu/software/structure22/

Pritchard, J.K. Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**:945-959.

Raymond, M. and Rousset, F. 1997. GENEPOP, Version 3.1b: an updated version of GENEPOP (Version 1.2), Population genetics software for exact tests and ecumenism. *Journal of Heredity*, **86**:248-249.

Redon, R. Ishikawa, S. Fitch, K.R. Feuk, L. Perry, G.H. Andrews, T.D. Fiegler, H. Shapero, M.H. Carson, A.R. and Chen, W. 2006. Global variation in copy number in the human genome. *Nature* **444**(7118): 444-454.

Rege, J.E.O. and Gibson, J.P. 2003. Animal genetic resources and economic development: issues in relation to economic valuation. *Ecol. Econ*., **45**:319–330.

Rege, J. E. O. Yapi-Gnaore, C. V. and Tawah, C. L. 1996. The indigenous domestic ruminant genetic resources of Africa. *In* "2nd All Africa Conference on Animal Agriculture" 1st – 4th, April 1996, Pretoria, South Africa.

Rege, J.E.O. 1994. Indigenous African small ruminants: a case for characterisation and improvement. In: Lebbie, S.B.H., Rey, B., Irungu, E.K. (Eds.), Small ruminant research and development in Africa, Proceedings of the Second Biennial Conference of the African Small Ruminant Research Network, 7-11 December, 1992, AICC, Arusha, Tanzania, pp. 205-211.

Rice, W.R. 1989. Analyzing tables of statistical tests. *Evolution*, **43**:223-225.

Risch, H. Marrett, L.D. Jain, M. and Howe, G. R. 1996. Difference in risk factors for epithelial ovarian cancer by histologic type. Results of a case-control study. *Am J Epidemiol* **136**:1184–1203.

Roder, M.S. Korzun,V. Wendehake, K. Plaschke, J. Tixier, M.H. Leroy, P. and Ganal, M.W. 1998. A microsatellite map of wheat. *Genetics*, **149**: 2007–2023.

Roeleveld, A.C.W. 1996. The diagnostic phase in research on livestock systems. In: Roeleveld, A.C.W. and van den Broek, A. Editors, *Focusing Livestock Systems Research*, Royal Tropical Institute, Amsterdam, The Netherlands pp. 14–28.

Rosenberg, N. A. 2004. *Distruct*: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**: 137-138.

Rovelet-Lecrux, A. Hannequin, D. Raux, G. Le Meur, N. Laquerrie`re, A. and Vital, A. 2006. APP locus duplication causes autosomal dominant earlyonset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet* **38**: 24–6.

Ryder, M.L. 1984. Sheep. In: Evolution of Domestic Animals (Ed. by I.L. Mason), pp.63–85. Longman, London.

Ryder, M.L. 1983. *Sheep and man*. Duckworth, London.

Ryder, M.L. and Stephenson S.K. 1968. Wool Growth. In: Academic Press, London and New York, pp. 805.

Sachidanandam, R. Weissman, D. Schmidt, S.C. Kakol, J.M. Stein, L.D. Marth, G. Sherry, S. Mullikin, J.C. Mortimore, B.J. and Willey, D.L. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928–933.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.

Sambrook, J. Fritsch, E. F. and Maniatis, T. 1989. *Molecular cloning: a laboratory manual.* Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Scarpa, R. Ruto, E. S. K. Kristjanson, P. Radeny, M. Drucker, A. and Rege, J. E. O. 2003a. Valuing indigenous cattle breeds in Kenya: an empirical comparison of stated and revealed preference value estimates. *Ecological Economics* **45**:409-426.

Scherer, S.W. Lee, C. Birney, E. Altshuler, D.M. Eichler, E.E. Carter, N.P. Hurles, M.E. and Feuk, L. 2007. Challenges and standards in integrating surveys of structural variation. *Nat. Genet.* **39**:S7–S15.

Scherf, B.D. (Ed.). 2000. World Watch List for domestic animal diversity. Food and Agricultural Organization of the United Nations, Rome, Italy, 3rd edition.

Schlötterer, C. 2004. The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics* **5**:63–9.

Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromossoma* 109:365-371.

Schlötterer, C. 2000. Microsatellite analysis indicates genetic differentiation of the neo-sex chromosomes in *Drosophila americana americana*. *Heredity* **85**: 610–616.

Schlötterer, C. and Tautz, D. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Research **20**:211-215.

Schlötterer, C. Amos, B. and Tautz, D. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature **354**:63-65.

Schmidtt, J. and Ulbright, F. 1968. Die Chromosomen verchiedener Caprini Simpson,1945. *Z Saugetierkunde* **33**: 180-186.

Semagn, K. Bjornstad, A. and Ndjiondjop, M.N. 2006. An overview of molecular marker methods for plants. *African J. Biotech.,* **5**(25): 2540-2568.

Shaw-Smith, C. Redon, R. Rickman, L. Rio, M. Willatt, L. Fiegler, H. Firth, H. Sanlaville, D. Winter, R. Colleaux, L. Bobrow, M. and Carter, N. P. 2004. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J. Med. Genet.* **41**: 241–8.

Shields, G.F. Schmiechen, A.M. Frazier, B.L. Redd, A. Voevoda, M.I. Reed, J.K. and Ward, R.H. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am J Hum Genet*. Sep;**53** (3):549–562.

Sia, E.A. Butler, C.A. Dominska, M. Greenwell, P. Fox, T.D. and Petes, T. D. 2000. Analysis of microsatellite mutations in the mitochondrial DNA of *Saccharomyces cerevisiae*. Proceedings of the National Academy of Sciences **97**:250-255.

Sia, E.A. Kokoska, R.J. Dominska, M. Greenwell, P. and Petes, T. D. 1997. Microsatellite instability in yeast: Dependence on repeat unit size and DNA mismatch repair genes. *Molecular and Cellular Biology* **17**:2851-2858.

Singleton, A.B. Farrer, M. Johnson, J. Singleton, A. Hague, S. and Kachergus, J. 2003. Alpha synuclein locus triplication causes Parkinson’s disease. *Science* **302**: 841.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462.

Sodhi, S. Mukesh, M. and Bathia, S. 2006. Characterizing Nali and Chokla sheep differentiation with micro satellite markers. Small Rumin. Res. **65**:185–192.

Spencer, C.C. Neigel, J.E. and Leberg, P.L. 2000. Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks, *Mol. Ecol.* **9**:1517–1528.

Strand, M. Prolla, T.A. Liskay, R.M. and Petes, T. D. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**:274-276.

Streisinger, G. and Owen, J.E. 1985. Mechanisms of spontaneous and induced frame shift mutation in bacteriophage T4. Genetics **109**:633-659.

Szilvia Kusza , Istv´an Nagy, Zsuzsanna Sasv´ari , Anik´o St´agel , T´ımea N´emeth, Andr´as Moln´ar , Kristaq Kume, Zsuzsanna B˝osze , Andr´as J´avor , S´andor Kukovics 2008. Genetic diversity and population structure of Tsigai and Zackel type of sheep breeds in the Central, Eastern and Southern European regions. *Small Ruminant Research* **78**:13–23.

Takezaki, N, and Nei, M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**: 389–399.

Tapio, I. Tapio, M. Grislis, Z. Holm, L.E. Jeppsson. S. Kantanen, J. Miceikiene, I. Olsaker, I. Viinalass, H. and Eythorsdottir, E. 2005a. Unfolding of population structure in Baltic sheep breeds using microsatellite analysis. *Heredity* **94**:448–56.

Tapio, M. Tapio, I. Grislis, Z. Holm, L.E. Jeppson, S. Kantanen, J. Miceikiene, I. Olsaker, I. Viinalass, H. Eythorsdottir, E. 2005b. Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Mol. Ecol.* **14**:3951–3963.

Tautz, D. and Renz, M. 1989. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* **12**: 4127–4138.

The Microsatellite Toolkit [http://animalgenomics.ucd.ie/sdepark/ms-toolkit/]

Thomas, D. L. 1991. Hair sheep genetic resources of the Americas. In: S. Wildeus (Ed.) Proc. Hair Sheep Res. Symp. p 3. Univ. Virgin Islands, St. Croix.

Toro M. A. and Caballero, A. 2005. Characterisation and conservation of genetic diversity in subdivided populations Phil. Trans. Roy. Soc. Series B, **360**: 1367-1378.

Torroni, A. Lott, M.T. Cabell, M.F. Chen, Y.S. Lavergne, L. Wallace, D.C. 1994 mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the Dloop region. *Am J Hum Genet* **55**:760-776.

Torroni, A. Schurr, T.G. Cabell, M.F. Brown, M.D. Neel, J.V. Larsen, M. and Smith, D. G. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* **53**:563-590.

Tóth, G. Gáspari, Z. and Jurka, J. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* **10**:967-981.

Trevor, R. 1991. Wilson Bartridge Partners Umberleigh North Devon, UK. Small ruminant production and the small ruminant genetic resource in tropical Africa. Food and Agriculture Organization of the United Nations. FAO Animal Production and Health Paper Rome, 1991.

Tuzun, E. Sharp, A.J. Bailey, J.A. Kaul, R. Morrison, V.A. Pertz, L.M. Haugen, E. Hayden, H. Albertson, D. and Pinkel, D. 2005. Fine-scale structural variation of the human genome. *Nat. Genet.* **37:** 727–732.

Uerpmann, H.P. 1979. Probleme der Neolithisierung des Mittelmeerraums. Tübinger Atlas des Vorderen Orients, B. 28. Wiesbaden, Germany: Ludwig Reichert.

Upton, M. 1985. Returns from small ruminant production in South West Nigeria. *Agricultural Systems* **17**:65-83.

Vignal, A. Milan, D. San Cristobal, M. and Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Gen Sel Evol **34**:275–305.

Vigne, J.D. Carre`re, I. and Guilaine, J. 2003. Unstable status of early domestic ungulates in the Near East: the example of Shillourokambos (Cyprus, IX-VIIIth millennia Cal. B.C.). *Bull Corr Hell Suppl* **43**:239–51.

Vos, P. Hogers, R. Bleeker, M. Reijans, M. Van De Lee, T. Hornes, M. Frijters, A. Pot, J. Peleman, J. Kuiper, M. and Zabeau, M. 1995. AFLP: a new technique for DNA Fingerprinting. *Nucleic Acids Res* **23**: 4407-4414.

Wallace, D. Ye, C.J. Neckelmanng, S. N. Singh,K . Webster, A. and Reenberg, B. D.G 1987. Sequence analysis of cDNAs for the human and bovine ATP synthase β subunit: mitochondrial genes sustain seventeen times more mutations. *Curr. Genet.* **12**: 81-90.

Wanyangu, S.W. Mwendia, C.M.T. Bain, R.K. Mugambi, J.M. and Stevenson, P. 1993. *Production potentials of the Red Maasai and Black-headed Somali sheep under pastoral management conditions without trypanosome and helminth control.* Small Ruminant Collaborative Research Support Programme, Nairobi (Kenya). pp. 167–180.

Waters-Bayer, A., and Bayer, W. 1992. The role of Livestock in the rural economy. Nomadic Peoples 31.

Wattier, R. Engel, C.R. Saumitou-Laprade, P. and Valero, M. 1998. Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*,**7**:1569–1573.

Weber, J.L. and May, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**:388–396.

Weinberg, W. 1908. On the demonstration of heredity in man *Naturkunde in Wurttemberg, Stuttgart* **64**:368-382.

Weir, B.S. and Cockerham, C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**:1358-69.

Weising, K. Atkinson, R.G. and Gardner, R.C. 1995. Genomic fingerprinting by microsatellite primed PCR: a critical evaluation. *PCR Meth Applications* **4**: 249-255.

Wenzl, P. Carling, J. Kudrna, D. Jaccoud, D. Huttner, E. Kleinhofs, A. and Kilian, A. 2004. Diversity arrays technology (DArT) for whole-genome profiling of barley. Proc. Natl. Acad. Sci. U.S.A. **101**: 9915–9920.

White, G. and Powell, W. 1997. Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): An endangered tropical hardwood species. *Molecular Ecology* **6**:851-860.

Williams, J.G.K. Kubelik, A.R.K. Livak, J.L. Rafalski, J.A. and Tingey, S.V. 1990. DNA polymorphisms amplified by random primers are useful as genetic markers. *Nucleic Acids Res*. **18**: 6531–6535.

Wilson, R.T. 1991. Small Ruminant Production and the Small Ruminant Genetic Resources in Tropical Africa. Food and Agriculture Organization (FAO) of the United Nations Animal Production and Health Paper, 88, Rome, Italy, 231 pp.

Wittenberg, A.H. Van Der Lee, T. Cayla, C. Kilian, A. Visser, R.G. and Schouten, H. J. 2005. Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Mol Gen Genet* **274**:30-39.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**: 395-420.

Wright, S. 1951. The genetical structure of populations. *Annual Eugenics* **15**:323-354.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97-159.

Xia, X. McClelland, M. and Wang, Y. 2005. WebArray: An online platform for microarray data analysis. *BMC Bioinform* **6**:306.

Yuriy Mileyko, Richard, Joh, I. and Joshua, S. Weitz 2008. Small-scale copy number variation and largescale changes in gene expression. *Proc. Natl. Acad. Sci.* U. S. A.**105**:16659–16664.

Zabeau, M. and Vos, P. 1993. Selective restriction fragment amplification: a general method for DNA fingerprinting. Publication 0 534 858 A1, bulletin 93/13. European Patent Office, Munich, Germany.

Zeuner, F.E. 1963. "A History of Domesticated Animal" Hutchinson, London.

Zogopoulos, G. Ha, K.C. Naqib, F. Moore, S. Kim, H. Montpetit, A. Robidoux, F. Laflamme, P. Cotterchio, M. and Greenwood, C. 2007. Germ-line DNA copy number variation frequencies in a large North American population. *Hum. Genet.* **122**:345-353.

Zucchi, M.I. Brondani, R.V. Pinheiro, J.B. Brondani, C. and Vencovsky, R. 2002. Transferability of microsatellite markers from *Eucalyptus* spp. to *Eugenia dysenterica* (Myrtaceae family). *Molecular Ecology Notes* **2**:512-514.

6.2 APPENDICES

6.2.1 APPENDIX I:

Protocol for Nucleic acid measurement

1. Turn on the computer and the ND-1000 at the mains plug.
2. Start the Nanodrop software by double-clicking the desktop icon or by selecting the following path: ***Start→Programs→NanoDrop→ND-1000[version]***

   (The applications module will appear) Click Nucleic Acid button to select Nucleic acid measurement.
3. Clean the measurement pedestal by very gently loading 1µl of water onto the lower pedestal (avoid touching it with the pipette tip).
4. Gently lower the sampling arm into the 'down' position and then click OK. The spectrophotometer will initialize and be ready for use. Raise the sampling arm and wipe out the water with a soft laboratory wipe tissue paper.
5. **BLANKING:** Load a blank sample (the buffer, solvent, or carrier liquid used with your samples) onto the lower measurement pedestal and lower the sampling arm into the 'down' position and click on Blank button or F3 key. Wait for a few seconds.
6. Wipe the Blanking buffer from both pedestal surfaces using a laboratory wipe.
7. Analyze a fresh replicate of the blanking solution as though it were a sample. This is done using the Measure button (F1). The result should be a spectrum with a relatively flat baseline.

8. Wipe the blank from both measurement pedestals surfaces and repeat the process until the spectrum is flat.
9. Wipe the blank from both measurement pedestals surfaces using a laboratory wipe.
10. Pipette 1μl of sample onto the lower measurement pedestal and lower the sampling arm into the ‘down’ position.
11. Click the Measure button (or press F1 key) to initiate the measurement sequence for all samples (non-blanks). The entire measurement cycle takes about 10 seconds.
12. When the measurement is complete, open the sampling arm and wipe the sample from both the upper and lower pedestals using a soft laboratory wipe to prevent sample carry-over to successive measurements.
13. Save your outputs in a folder.

6.2.2 APPENDIX II:

Genotypic linkage disequilibrium *p*-values for each locus pair across all populations using the Fisher's method

Markov chain parameters
Dememorisation : 10000
Batches : 1000
Iterations per batch : 5000

Locus pair		Chi^2	df	*p*-Value
BM8125	& DYMS1	21.643802	30	0.866684
BM8125	& HSC	26.461742	30	0.651370
DYMS1	& HSC	22.773780	30	0.824444
BM8125	& HUJ616	30.428181	30	0.443889
DYMS1	& HUJ616	35.387131	30	0.228827
HSC	& HUJ616	22.544351	30	0.833519
BM8125	& ILSTS005	30.125238	30	0.459253
DYMS1	& ILSTS005	31.105465	30	0.410235
HSC	& ILSTS005	26.049511	30	0.672602
HUJ616	& ILSTS005	19.927123	30	0.918425
BM8125	& MAF209	17.213425	30	0.969926
DYMS1	& MAF209	23.524522	30	0.793097
HSC	& MAF209	39.238895	30	0.120466
HUJ616	& MAF209	35.959818	30	0.209396
ILSTS005	& MAF209	18.847163	30	0.943150
BM8125	& MCM42	24.089948	30	0.767940
DYMS1	& MCM42	27.344691	30	0.605120
HSC	& MCM42	25.769805	30	0.686827
HUJ616	& MCM42	47.302217	30	0.023244
ILSTS005	& MCM42	48.534026	30	0.017524
MAF209	& MCM42	36.834235	30	0.182024
BM8125	& OARFCB11	31.986149	30	0.368172
DYMS1	& OARFCB11	29.668311	30	0.482732
HSC	& OARFCB11	9.928277	30	0.999790
HUJ616	& OARFCB11	23.012304	30	0.814754
ILSTS005	& OARFCB11	34.096693	30	0.276984
MAF209	& OARFCB11	40.279194	30	0.099569
MCM42	& OARFCB11	22.405728	30	0.838881
BM8125	& OARFCB20	14.217942	30	0.993466
DYMS1	& OARFCB20	17.722175	30	0.962861

HSC	& OARFCB20	8.420246	30	0.999965
HUJ616	& OARFCB20	13.592191	30	0.995589
ILSTS005	& OARFCB20	16.159181	30	0.981355
MAF209	& OARFCB20	35.355947	30	0.229920
MCM42	& OARFCB20	33.955756	30	0.282605
OARFCB11	& OARFCB20	34.394256	30	0.265346
BM8125	& OARFCB226	17.721287	30	0.962874
DYMS1	& OARFCB226	50.082364	30	0.012160
HSC	& OARFCB226	23.645708	30	0.787812
HUJ616	& OARFCB226	46.531758	30	0.027629
ILSTS005	& OARFCB226	12.747438	30	0.997531
MAF209	& OARFCB226	29.916578	30	0.469932
MCM42	& OARFCB226	27.831667	30	0.579357
OARFCB11	& OARFCB226	18.476438	30	0.950299
OARFCB20	& OARFCB226	17.967300	30	0.959061
BM8125	& OARHH47	40.249378	30	0.100124
DYMS1	& OARHH47	45.113050	30	0.037677
HSC	& OARHH47	21.824218	30	0.860361
HUJ616	**& OARHH47**	**Infinity**	**30**	**Highly sign.**
ILSTS005	& OARHH47	21.917670	30	0.857022
MAF209	& OARHH47	19.701133	30	0.924092
MCM42	& OARHH47	16.214488	30	0.980854
OARFCB11	& OARHH47	19.524056	30	0.928347
OARFCB20	& OARHH47	25.329415	30	0.708870
OARFCB226	**& OARHH47**	**Infinity**	**30**	**Highly sign.**
BM8125	& OARJMP29	40.048196	30	0.103934
DYMS1	& OARJMP29	22.781127	30	0.824150
HSC	& OARJMP29	29.608330	30	0.485838
HUJ616	& OARJMP29	24.587244	30	0.744832
ILSTS005	& OARJMP29	32.643209	30	0.338233
MAF209	& OARJMP29	36.401558	30	0.195223
MCM42	& OARJMP29	39.339148	30	0.118311
OARFCB11	& OARJMP29	31.635693	30	0.384660
OARFCB20	& OARJMP29	33.361815	30	0.307057
OARFCB226	& OARJMP29	23.628244	30	0.788577
OARHH47	& OARJMP29	25.753756	30	0.687639
BM8125	& OARVH72	15.543643	30	0.986286
DYMS1	& OARVH72	26.779003	30	0.634851
HSC	& OARVH72	32.849661	30	0.329104
HUJ616	& OARVH72	26.899868	30	0.628524
ILSTS005	& OARVH72	32.691013	30	0.336107
MAF209	& OARVH72	50.379835	30	0.011321
MCM42	& OARVH72	50.759305	30	0.010328
OARFCB11	& OARVH72	31.504340	30	0.390927

OARFCB20	& OARVH72	27.421673	30	0.601054
OARFCB226	& OARVH72	23.021841	30	0.814361
OARHH47	& OARVH72	24.535987	30	0.747253
OARJMP29	& OARVH72	31.780562	30	0.377802
BM8125	& SRCRSP9	60.142193	30	0.000885
DYMS1	& SRCRSP9	16.892200	30	0.973845
HSC	& SRCRSP9	25.067540	30	0.721743
HUJ616	& SRCRSP9	28.856260	30	0.525157
ILSTS005	& SRCRSP9	31.065015	30	0.412216
MAF209	& SRCRSP9	29.618457	30	0.485313
MCM42	& SRCRSP9	39.492824	30	0.115067
OARFCB11	**& SRCRSP9**	**Infinity**	**30**	**Highly sign.**
OARFCB20	& SRCRSP9	32.360059	30	0.350974
OARFCB226	& SRCRSP9	27.591314	30	0.592084
OARHH47	& SRCRSP9	20.131328	30	0.913077
OARJMP29	& SRCRSP9	23.348640	30	0.800660
OARVH72	& SRCRSP9	28.206809	30	0.559478
BM8125	& TGLA53	25.296705	30	0.710488
DYMS1	& TGLA53	24.888227	30	0.730445
HSC	& TGLA53	30.859307	30	0.422347
HUJ616	& TGLA53	32.236625	30	0.356605
ILSTS005	& TGLA53	19.292576	30	0.933669
MAF209	& TGLA53	20.217722	30	0.910749
MCM42	& TGLA53	16.961916	30	0.973029
OARFCB11	& TGLA53	27.275030	30	0.608795
OARFCB20	& TGLA53	28.938493	30	0.520830
OARFCB226	& TGLA53	17.853572	30	0.960857
OARHH47	& TGLA53	27.123815	30	0.616762
OARJMP29	& TGLA53	38.850542	30	0.129112
OARVH72	& TGLA53	20.679798	30	0.897633
SRCRSP9	& TGLA53	31.114145	30	0.409811

NB: Loci pairs showing evidence of Linkage disequilibrium are indicated in bold