# GENETIC CHARACTERIZATION OF INDIGENOUS GOAT POPULATIONS OF ETHIOPIA USING MICROSATELLITE DNA MARKERS

THESIS SUBMITTED TO THE
NATIONAL DAIRY RESEARCH INSTITUTE
(DEEMED UNIVERSITY)
KARNAL (HARYANA)
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

**DOCTOR OF PHILOSOPHY**
**IN**
**ANIMAL GENETICS & BREEDING**

**By**

TESFAYE ALEMU TUCHO

**DIVISION OF DAIRY CATTLE BREEDING**
**NATIONAL DAIRY RESEARCH INSTITUTE**
**(ICAR)**
**KARNAL – 132 001 (HARYANA) INDIA**

**2004**

**REGN. No.1062k01**

# GENETIC CHARACTERIZATION OF INDIGENOUS GOAT POPULATIONS OF ETHIOPIA USING MICROSATELLITE DNA MARKERS

By

**TESFAYE ALEMU TUCHO**

A THESIS SUBMITTED TO THE
NATIONAL DAIRY RESEARCH INSTITUTE
(DEEMED UNIVERSITY)
KARNAL (HARYANA)
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

**DOCTOR OF PHILOSOPHY**
IN
**ANIMAL GENETICS & BREEDING**

**Approved by:**

| | | |
|---|---|---|
| **(  )** | **(OLIVIER HANOTTE)** | **( B.R. YADAV )** |
| **EXTERNAL EXAMINER** | **CO-GUIDE (ADVISOR)** | **MAJOR ADVISOR & CHAIRMAN (GUIDE)** |

**Members, Advisory Committee**

1. **Dr. V.K. Batish** _____
   Principal Scientist, DM Division
2. **Dr. A.K. Chakravarty** _____
   Senior Scientist, DCB Division
3. **Dr. A. K. Gupta** _____
   Scientist, DCB Division
4. **Dr. T.K. Mohany** _____
   Senior Scientist, DCB Division
5. **Dr. Surjeet Singh** _____
   Senior Scientist, Animal Biochemistry Division

# CERTIFICATE

This is to certify that the thesis entitled, **"GENETIC CHARACTERIZATION OF INDIGENOUS GOAT POPULATIONS OF ETHIOPIA USING MICROSATELLITE DNA MARKERS"** submitted by **Mr. TESFAYE ALEMU TUCHO** towards the partial fulfilment of the award of the degree of **DOCTOR OF PHILOSOPHY** in **ANIMAL GENETICS & BREEDING** of the **NATIONAL DAIRY RESEARCH INSTITUTE (DEEMED UNIVERSITY)**, Karnal (Haryana), India, is a bonafide research work carried out by him under our supervision, and no part of the thesis has been submitted for any other degree or diploma.

**(OLIVIER HANOTTE)**
**CO-GUIDE**
**(ADVISOR)**

Dated:_____

**( B.R. YADAV )**
**MAJOR ADVISOR & CHAIRMAN**
**(GUIDE)**

Dated:_____

# ACKNOWLEDGEMENTS

‘Then Samuel took a stone and set it up between Mizpah and Shen, and called its name ***EBENEZER***, saying ‘***’Thus far the Lord has helped us’’*** (I Samuel 7:12). I give thanks and all the glory to the Almighty God for seeing me, not only through four years of relentless effort to acquire acknowledge, but also a frantic race against time and resources. Thus far the Lord has brought me! I am grateful!!

The work presented in this thesis would not have been accomplished without the support and assistance of many people. It is impossible to adequately acknowledge everyone who contributed to this work, in one way or the other, within the space provided here. Should I miss your name please do not consider it as an act of ingratitude. I consider this achievement yours too and I am grateful to all and sundry whose contribution(s) have enabled this work to be concluded fruitfully.

I am grateful to my supervisors Dr. B. R. Yadav and Olivier Hanotte, for giving me an opportunity to work with them, for their firm support, for patiently teaching me many things I know today about characterization of Indigenous Animal Genetic Resources. I would also like to thank Dr Edward Rege for accepting me as a graduate fellow at ILRI.

I would like to acknowledge Dr. Han Jianlin for assisting me with the data analysis and for reading drafts of my thesis. I am indebted to Russ Kruska and Michael Arunga for providing me with the geographic distance matrix for the Ethiopian goat populations. Special thanks to Joel Ochieng, who “got me started” and trained me on use of the various lab equipment for my work as well as providing me with invaluable technical support. It has been a great pleasure to work with the members of the Animal Genetics Resources group.

I am also thankful to the members of my Advisory Committee: Dr. V.K. Batish, Dr. A.K. Chakravarty, Dr. A.K. Gupta, Dr. T.K. Mohany and Dr. Surjeet Singh, for their suggestions and support during my study.

To all my colleagues in Lab. 7; Paul Wafula, Moses Ogugo, Daniel Mwangi, Nemwel Nyamweya, John Wambugu, Samuel Njomo, Joel Mwakaya, Joseph Nthale, Henrie Gathuo, Sarah Waita, Sarah Nyongesa, Gerold Kierstein, Sonya Kierstein, Donald Rugira Kugonza, Batsuuri Lkhagua, Yeon Heun-Seong and Chang Yeon Cho, Suzan Musembi and Richard Kang’ethe. Thank you all so much for all your contributions into this thesis. Special thanks also to Manish Thakur and Zenu who “got

me started'' and trained me while I was at NDRI Livestock Genome Analysis laboratory at the beginning.

I would like to acknowledge all the individuals who participated in collecting the goat samples for this project, especially Dr. Workneh Ayalew, Dr. Markos Tibo, Ato Girma Abebe, Ato Michal Temesgen, and Ato Eshetu Zerihun.

To my colleagues and fellow PhD-students in the lab, especially David Odongo, Xuebin Qi, David Menge, Joseph Ng'ang'a, Lucy Muthoni and Mujibi Fidalis (Msc student), thanks all for your scientific, social and friendly discussions and ideas that we shared.

I would like to acknowledge the contributions made by the following organizations: The Ethiopian Agricultural Research Organization (EARO) and Oromiya Agricultural Research Institute (OARI) for the financial assistance and study leave; International Livestock Research Institute (ILRI) and the Department for International Development (DFID) UK, for financial support, and all the field staff from the Ministry of Agriculture, who allowed the field team easy access to the site of interest.

To my beloved wife, Meaza and our children, Elim (daughter), Yeabkal, Yididiya and Kebron (Sons), your prayer, sacrifice and support has been with me to the end, and I would like to thank you so much and say God bless you abundantly, I love you all from the bottom of my heart. For the tremendous family prayer and support I would like to thank my Dad (Alemu Tucho), brothers, Hailu and Tsegaye, sisters, Asnakech (with her family), Meseret, Tayech and Tseganesh, my mother-in-law W/ro Ejigayehu and Father-in-law, Ato Workagegnehu, and all my brothers and sisters-in-law.

Dated November, 2004 (**Tesfaye Alemu Tucho**)

# CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

## APPENDICES

# ABSTRACT

Indigenous goats are an important resource for African farmers, providing meat, milk, manure, fibres and hides, and satisfying various cultural and religious functions. In Africa, the survival of many indigenous goat populations is threatened by diseases, adverse climatic conditions, civil strife, pressure of economic development, abandonment of traditional farming practices, and more importantly through crossbreeding or replacement with animals from the developed world.

Indigenous Ethiopian goat genetic resources have been classified phenotypically into 11 types. These are Abergalle, Arsi-Bale, Afar, Central Highland, Gumez, Hararghe Highland, Keffa, Long-eared Somali, North-West Highland, Short-eared Somali, and Woyto-Guji. In the present investigations two main studies were carried out. The first study dealt with the genetic diversity within and between the Ethiopian goats. The second study dealt with the genetic diversity and relationship of the Ethiopian goat in relation to goat populations from other countries of African and outside African continents. For the first study, the 11 indigenous Ethiopian populations were sampled. In the second study, five populations from other African countries [(Boran-Galla and Small East African from Kenya), WAD from Guinea Bissau (West Africa), Tswana from Botswana and Baladie from Egypt] and four populations from outside African continent (Italian Alpine from Italy, Ardi from Saudi Arabia, Hair from Turkey and Zalaajinistiin from Mongolia) were sampled. All the populations were assessed for genetic diversity using 15 microsatellite loci. A wide range of statistical analyses were performed on the data to answer questions on population genetics diversity, relationship and differentiation.
The genetic diversities observed in both the Ethiopian and non-Ethiopian populations were similar. The analysis of molecular variance (AMOVA) indicated a greater proportion of the genetic diversity within the Ethiopian populations than between populations. Genetic distances were observed to be relatively small for Ethiopian populations. The high correlations observed between geographical distances and genetic distances for Ethiopian populations supported this result. On the other hand, relatively large genetic distances were observed between Ethiopian and Non-Ethiopian populations with exception of Kenyan populations. The results have shown that more than 60% of the goat populations showed significant inbreeding effect, which leads to loss of genetic diversity. Phylogenetic trees, population structure and principal component analyses showed that all the Ethiopian populations are genetically distinct from all the reference populations except with two Kenyan populations. However, they also indicated that all the Ethiopian goats populations are very closely related to each other. Our results indicate that the 11 Ethiopian populations can be grouped as eight distinct genetic entities: Arsi-Bale, Gumez, Keffa, Woyto-Guji, Abergalle, Afar, Highland goats (previously separated as Central and North West Highland) and the goats from the previously known Hararghe, Southeastern Bale and Southern Sidamo provinces (Hararghe Highland, Short-eared Somali and Long-eared Somali goats). Weitzman analysis indicates that about 75% of the total genetic diversity of the Ethiopian goats is present in four breeds: Afar, Abergalle, Gumez and Keffa with marginal loss of diversity of 24.32%, 19.22%, 16.59% and 12.99%, respectively.

Genetic differentiation tests showed that all the Ethiopian and non-Ethiopian populations were well differentiated. Individuals from Ethiopian populations were assigned to their source populations with a low degree of accuracy indicating their genetic similarity. However, the proportion of correct assignment for the non-African population was higher than that for the Ethiopian populations. Results support the theory of two separate entry points for the goats into the African continent: the Isthmus of Suez for the North African and the horn of Africa for the Eastern African goats. Ethiopia hosts a large indigenous goat genetic resource adapted to the complexity agricultural production systems of the country; it represents a unique resource that has potential for further and future genetic improvement of its productivity.

# CHAPTER ONE

## 1.0. INTRODUCTION

Most food production systems depend heavily on utilisation of locally adapted animal species. These animal genetic resources consist of species that are of agricultural, cultural and economic importance to man. Different cultures and regions of the world bear importance to different domestic animals. However, the common agricultural species kept in most regions include sheep, goats, cattle, horses, pigs, buffaloes and chicken (FAO, 1994; FAO, 1999). In sub-Saharan Africa, indigenous breeds of sheep and goats are very important, in fact more important than cattle for the small-scale farmers since they are easier to acquire and to maintain. These serve as a secure form of investment, a means of income, source of manure and for various religious and ceremonial functions (Chenyambuga, 2002). These animals often provide the only practical means of utilising vast areas of natural grasslands in the areas where crop production is uneconomical (Rege *et al*., 2002). Adapted to the local environment, the African sheep and goat represent a unique genetic resource for the farmer. However, the need for increased economic gains has led to crossbreeding these indigenous breeds with imported exotic breeds or directly replacing the indigenous genotypes (Wollny, 2003). To compound the problem, indigenous breeds have not been characterised in many African countries and it would be tragic to lose these unique genetic resources that are the result of centuries of human and natural selection (Leak *et al*., 2002). The rate of erosion of indigenous animal genetic resources therefore threatens the prospect of improving the livelihood of present and future human generations. This is especially so when one considers that the demand for animal products is expected to double over the next 20 years as a consequence of urbanisation, population growth and increased income (FAO, 1999). Developing countries host a vast majority of indigenous farm animal genetic resources (over 90 percent) that unfortunately are not being improved to respond to global food security needs (FAO, 1999). FAO estimates that 30 percent of livestock breeds are at risk of extinction and about six breeds are lost per month. Moreover, more than half of these breeds are found in developing countries (FAO, 2000). However, sustainable conservation and utilization of these indigenous goat resources cannot begin until their genetic diversity is understood and quantified. Most indigenous goat breeds, despite being best suited

for the local conditions, are not yet genetically classified and cannot thus be maximally exploited. These indigenous genetic resources therefore need to be characterized before the conservation and utilization programs can begin.

Ethiopia has the largest livestock population among all the countries in Africa. Ethiopia with its great variation in agro-ecological zones represents a potential reservoir of sheep and goat diversity. They inhabit a wider range of environments extending from a tropical to cool temperate climates. The sheep and goat population of Ethiopia, though rather variably estimated, ranks high both in Africa and in the world. According to FAO (1994), there are approximately 21.7 and 16.7 million sheep and goats, respectively. Sheep and goats are distributed in all agro-ecological zones of the country although the majority of the sheep population is concentrated in the highlands. The majority of the goat population is found in large flocks in the arid and semi-arid lowlands where pastoralists in the south, east, and west keep them for milk production, for slaughter and for sale. Goats in the highlands are widely distributed in the mixed crop-livestock production systems with very small flock size as a means of cash earnings and meat. Although there are severe environmental constraints to increase goat productivity, there is considerable potential for improved goat production in the country, where goat milk, meat, and skin are valued commodities.

Ethiopia has been long back recognized as a source of genetic diversity in plants and animals. It appears that Ethiopia has served as a gateway of domestic animal genetic materials from Asia to Africa, and its diverse ecology leads to further diversification and development of different genotypes. It is believed that the first goats reached Egypt by 5000 BC and subsequently spread south and west of the African continent. Remains of a small goat have been found in a site in Sudan thought to date before 3300 BC and the lop-eared Nubian type has been found in sites dating prior to 2500 BC (Mason, 1984). Therefore it is likely that the first wave of goats entered Ethiopia from the North of Africa sometime between 2000 and 3000 BC.

The identification, description, classification and naming of livestock types are the first step in the assessment of a country's livestock resources. The physical description of types can only dimly reveal the genetic relationships between individuals, though it is the first step in classifying a diverse population into relatively homogeneous sub-populations.

Work done in 1990s on physical description and management system at Alemaya University of Agriculture together with FARM – Africa (DGDP - Dairy Goat

Development Project) has shown the presence of 14 goat types in Ethiopia and Eritrea (Workneh, 1992; Alemayehu, 1993; Nigatu, 1994). This survey was the very first step in breed characterization of Ethiopian goats through differentiating a heterogeneous population into relatively homogeneous sub-populations, which may then be named as breeds, types or subtypes. However, it remains unclear how different at the genetic level are these 14 goat types. The availability of new molecular techniques for large-scale genetic characterization is providing unique opportunities to tackle the above issues. These techniques are able to reveal genetic distances between populations allowing a better understanding of how these populations evolved.

In view of the above, it would be ideal to characterize at the molecular level the Ethiopian goats in order to determine their genetic diversity and describe the relationship among the different populations. The information provided will allow formulation of conservation programs, maximizing diversity conserved and it may guide breeding improvement activities for meat and milk production.

The primary measure should be to determine the magnitude of genetic differentiation and relationship among the populations. One way of doing this is to determine genetic distances between pairs of breeds/populations. The computation of genetic distances requires estimation of allele frequency data using either protein markers or DNA markers. However, techniques for the determination of polymorphism of protein products lack the power to resolve the differences between closely related breeds since a great deal of genetic variation remains undetected by these methods (Meghen *et al*., 1994; Dowling *et al*., 1996).

In recent years, a range of innovations in molecular genetics has been developed for the study of genetic variation and evolution of populations using DNA genotyping information. The most utilized DNA marker for population genetics of livestock is microsatellite. Microsatellite markers, also called short tandem repeats (STRs) or simple sequence repeats (SSRs), are a relatively new class of genetic marker. Over a few years they have become a tool of choice to address population genetics and demographic questions (Sunnucks, 2000). The application of microsatellite markers is currently considered to be useful in the analysis of genetic diversity as they are numerous, randomly distributed in the genome, highly polymorphic, and they show codominant mode of inheritance (Ellegren, 1993). They allow the study of genetic diversity and differentiation of closely related populations. Microsatellite, as genetic markers, have been applied successfully in the study of genetic variation of livestock

including between and among European and African cattle breeds (Machugh *et al*., 1997; Okoma *et al*., 1998), between Swiss goat breeds (Saitbekova *et al*., 1999), between Indian goat breeds (Ganai and Yadav, 2001) and between Chinese indigenous goat populations (Li *et al*., 2002). Microsatellite DNA variation has also been employed to assess genetic distances between sheep breeds (Buchana *et al*., 1994; Arranz *et al*., 1998). Unfortunately, for goats, throughout the world very limited work has been done to determine genetic variation among breeds/populations and in fact, information on microsatellite polymorphism of Ethiopian goats is completely lacking.

Therefore, to utilize the Ethiopian goat genetic resources effectively, it is necessary to characterize Ethiopian indigenous goat populations genetically. Such characterization would provide a comprehensive database of genetic variation among the goat populations in Ethiopia. It would provide furthermore information as to which of the populations represent homogenous populations and which of them are genetically distinct. It would also provide information on the risk status of the goat populations. The information generated would contribute to the understanding of the evolutionary history of goats in Ethiopia. It will also contribute to the conservation and management of Ethiopian goat genetic resources.

The overall goal of this project is to obtain detailed information on the genetic diversity and differentiation of the indigenous goat populations of Ethiopia with the following specific objectives:

1. To quantify genetic diversity within and among Ethiopian goat populations.
2. To clarify the evolutionary genetic relationships among the Ethiopian goat populations.
3. To clarify the origin and relationships of Ethiopian goat populations through the study of non-Ethiopian breeds of references.
4. To compare the morphological classification of the Ethiopian goat populations with the one based on molecular genetic characterization.

# CHAPTER TWO

## 2.0 LITERATURE REVIEW

### 2.1. Classification of goats

Taxonomically the domestic goats (*Capra hircus*) belong to the order *Artiodactyla*, suborder *Ruminantia*, family *Bovidae* and tribe *Caprini*. The tribe *Caprini* includes different members: - The goat-like sheep or sheep like goat: *Ammotragus lervia* (Barbary sheep) and *Pseudois nayaur* (Blue sheep), the goat like antelope: *Oreamnos americanus* (Mountain goat), and the sheep and goats. **Table 1** shows the main fifferences between goats and sheep.

Table 1. **The main differences between goats and sheep**

| Goats | Sheep |
|---|---|
| 60 chromosomes | 54 chromosomes |
| Presence of beard in most | Absence of beard |
| Presence of face and caudal glands | Absence of face and caudal glands |
| Absence of foot glands | Presence of foot glands |
| Tail held up | Tail hanging down |

(http://capra.iespana.es/capra/ingles/origen/origen.htm).

Goats are divided into two genera – *Capra* and *Hemitragus*. The later is the tahr, which has three species (Ellerman and Morrison-Scott 1966): - *H. jemlahicus*, found in the Himalayan region from Kashmir to Sikkim (northern India); *H. jayakari*, found in Oman; and *H. hylocrius*, found in Nilgiri Hills and adjacent ranges in southern India.

The genus *Capra* is divided into eight species (Alados 1985b; *Ansell in* Meester and Setzer 1977, Corbet 1978) (http://www.press.jhu.edu/books/walkers_mammals_of_the_world/artiodactyla/artiodactyla.bovidae.capra.html):

*(i)* *C. aegagrus* (bezoar or wild goat), found in mountains from Asia Minor to Afghanistan and Pakistan, Oman, Crete, Aegean Islands;

*(ii)* *C. hircus* (domestic goat), found worldwide in association with people, feral in many / some areas;

*(iii)* *C. ibex* (ibex), found in European Alps, Palestine, Sinai, Arabian Peninsula, mountainous region from Afghanistan and northern India to Lake Baikal, Egypt and Sudan east of the Nile, northern Ethiopia;

*(iv)* *C. walie* (Walia ibex), Simien Mountains of north-central Ethiopia;

*(v)* *C. caucasica* (west Caucasian tur), western Caucasus;

*(vi)* *C. cylindricornis* (east Caucasian tur), eastern Caucasus;

*(vii)* *C. pyrenaica* (Spanish ibex), Spain and formerly southern France and Portugal;

*(viii)* *C. falconeri* (markhor), mountains of southern Uzbek and Tadzhik (formerly Soviet Socialist Republic), Afghanistan, northern and central Pakistan, and Kashmir (India).

Various authors have allegedly suggested that all these different species or subgenera have contributed to the domesticated stock (Epstein, 1971). According to Mason (1984), the bezoar is the direct ancestor of the domestic goat and the markhor is said to have contributed to some breeds of Central Asia (Mason, 1981a). However, for Mason (1984) it is less likely that the markhor has influenced some breeds. The ibexes and turs have not been domesticated.

**2.2. Domestication of goats**

Domestic animals have played a key role in human history. Despite their importance, the origin of most domestic species remains poorly understood. The goat is the earliest domestic animal and probably the first ruminant livestock, after the wolf was domesticated (Hole, 1996; Uerpmann, 1996). There are two reasons for this: Firstly, the wild goat was reported to be present in the regions of southwest Asia during the time when agriculture was developing. Secondly, the goat is an extremely hardy animal, hence, could have withstood the rigours of being reduced to the state of domestication better than other ruminants. Goats were domesticated at first for meat. However, as a dairy animal, the goat is regarded as the oldest, even older than cattle (Devendra and Burns, 1983).

Domestic goats (*Capra hircus*) might have played a central role in the Neolithic agricultural revolution and the spread of human civilizations around the globe (Porter, 1996; Pringle, 1998a). The evidence for time and place for goat domestication is based on the identification of bones from archaeological sites. Domestication of goats is considered to have occurred in the mountainous area of western Asia between the 7$^{th}$

and 9th millennium B.C. (Epstein, 1971; Devendra and Burns, 1983). The origin of domestic goats remain uncertain and controversial, however, archaeological evidence suggests that they were probably first domesticated in the Fertile Crescent region of the Near East possibly first in the Zagros Mountains area ≈10,000 years ago (Zeder and Hesse, 2000). However it has been suggested that the goats could have also been domesticated outside the Zagros Mountains range. Some studies hint for a second domestication area in Sindhu valley civilization in northwest part of India (now in Pakistan) from which the cashmere breeds would have originated (Porter, 1996). Others suggest that at least two wild species of *Capra (*Shackleton, 1997*)* could have contributed to the gene pool of domestic goats (Clutton-Brock, 1981).

Takada *et al*. (1997), studied the diversity of the cytochrome b gene of mitochondrial DNA (mtDNA) in Asian goats, and concluded that the strongest candidate for a matriarch ancestor of domestic goats is the bezoar or pasang (*Capra aegagrus*). Furthermore, phylogeographic analysis of mtDNA revealed three highly divergent lineages (estimated divergence > 200,000 years ago) suggesting three separate maternal origins of domestic goats (Luikart *et al.*, 2001). Sultana *et al.* (2003) studied the mtDNA diversity of Pakistani goats, and reported four distinct mtDNA-lineages termed as A, B and C (previously reported) and a new lineage D. The estimated divergence times between the most recently evolved mtDNA-lineages A and D were from 260 483 to 371 052 year ago. They concluded that at least four different strains of wild *Capra* might have been the source of the modern domestic goats with the most likely wild ancestor being the bezoar (Harris 1961; Clutton-Brock 1981). Joshi *et al.* (2004) have undertaken the investigation of 363 goats belonging to 10 different breeds from different geographic regions of India using mtDNA sequence data from the hypervariable region I. They found evidence for population structure and new mitochondrial DNA in Indian goats and cannot reconcile the genetic diversity found within the major lineage with domestication starting 10,000 years ago from a single mtDNA ancestor. Thus, they proposed a more complex origin for domestic goats.

The Bezoar (*Capra aegagrus*) is thought to be the progenitor of the domestic goat. *Bezoar* means counter poison, as in the savage animals it is possible to find a concrete in the stomach which believed to be an antidote. The main characteristic of the Bezoar is the long saber or scimitar shaped horns.

It is believed that goats were probably present at Cayönü in Southeast Anatolia about 7000 BC. They were present at Jarmo in Iraqi Kurdistan where the site is dated

from about 6500 BC to the end of the 6th millennium BC and at Sialk on the Iranian plateau as early as the 6th millennium BC. Domestic goats with scimitar horns like the bezoar spread over a large part of the old World during the Neolithic period. They were represented in vase paintings from ancient Mesopotamia in the 4th millennium BC. The small short-eared sabre-horned generalized type was occasionally also illustrated in the Mesopotamia tomb painting of the same period. It is believed that by the 5th millennium BC goats had reached Egypt and by about 3500 BC goats with spiral or corkscrew horns entered Egypt from the Middle East. From Egypt the goats moved to the South and West part of the continent. The dwarf goat is recorded at Shaheinab near Khartoum in the Sudan from 3300 BC while first evidence for the screw-horned lop-eared type is found in Nubia and Libya at site dated around 2650 BC. The goats of sub-Saharan Africa are predominantly small, short-eared, short-haired and short-horned. Goats are noted to appear in East Africa by 4000 – 3500 years BP (Marshall, 2000) and in southern Africa by 4th to 7th centuries AD (Smith, 2000).

**2.3 Goat genetic resource of Ethiopia**

Indigenous breeds constitute well over 95 percent of small ruminant populations in Africa (Rege, 1992). These are well adapted to the environment and the ravages of various kinds such as drought, famine and civil wars that continually plague the continent. Their adaptive features enable them to effectively cope with the stressful nature of marginal lands (Chenyambuga, 2002). The population of goats in sub-Saharan Africa is estimated to be 163 million (Rege *et al.*, 2002). Migration to a new habitat and consequently the effect of natural and artificial selection have led to the evolution of breeds and types of goat, which differ in appearance and performance. Around 90 'breeds' of African goats have been recognized using criteria as geographic distributions, ecotypes or communities-tribe ownership (Rege, 1992). They presumably derive from goats that spread south from Egypt at an early date. Generally, goats of sub-Saharan Africa are divided into three major types following their morphology; the long lop-eared type in north east and southern Africa, the small short-eared type dominant in eastern Africa and the dwarf short-eared type of West Africa. Intermediates morphological types are numerous.

The majority of Ethiopian goat population is found in large flocks in the arid and semi-arid lowlands where pastoralists in the South, East, and West keep them for

milk and meat production and for sale. Goats in the highlands are widely distributed in the crop-livestock production systems with very small flock sizes as a means of cash earnings and meat. Despite the huge resource potential, production and export opportunities, goat production in Ethiopia is relatively undeveloped. Although there are severe environmental constraints to increase goat productivity, there is considerable potential for goat production in the country, where goat milk, meat, and skin are valued commodities.

Information compiled on physical description and management system revealed that there are 14 goat types in Ethiopia and Eritrea (Farm-Africa, 1996). Out of these eleven are found in today Ethiopia. However, goats in Africa have traditionally been divided into three main families – the Dwarf goats of West and Central Africa, the Savannah goats of sub-Saharan Africa and the Nubian type goat of Northern Africa (Epstein, 1971; Wilson, 1991). Using a set of morphological characters and multivariate statistical analysis the goat of Ethiopia and Eritrea have been classified into four major families: - the Somali family (Short-eared Somali, Long-eared Somali, and Hararghe Highland), the Nubian family (Nubian and Barka), the small Rift valley family (Abergalle, Worre, Afar, Arsi-Bale, and Woyto-Guji) and finally the more heterogeneous Small East African family (Western Highland, Keffa, Central Highland and West Lowland). The names given to the goat types of Ethiopia reflect mainly their geographical locations and to some extent also their ethnic affiliation. These are:

**i) Afar**

The Afar goat is also known as *Adal* or *Danakil*. It is descendant of the rift valley goat thought to have entered Ethiopia from Yemen and Saudi Arabia. It is kept almost exclusively by the Afar-ethnic group in the Ethiopian and Eritrean rift valley, the Danakil depression; Gewane and northern and western Haraghe (see Fig. 1). These goats are well adapted to arid environments, infrequent watering (every 3-4 days) and can travel long distances. Where it is possible flocks may move to highland areas during the dry season. They have concave facial profile, narrow face, prick-eared, leggy, long thin upward-pointing horns, and patchy coat colour. The coat is very fine and short with variable colours and colour patterns with 48% of individuals white, 25% light brown, 27% black and flecks and patches are also common. Afar pastoralists identify their goats by branding the ears or hindquarters, or by notching the ears. Afar goats are extensively milked (*hadore hana*) for food, medicine and sale. Udder size is

relatively small with a circumference of 21 cm and length of 17 cm. Meat may be consumed fresh or air-dried (*teru*). Kids less than one month old may be sold as a delicacy known as *bekel* (male) and *bekelo* (female). Fresh blood is consumed to treat malaria and bullet wounds (Farm-Africa, 1996).


Figure 1: **Afar female goat**

**ii) Abergale**

Abergale goats are descendents of the rift valley goat type from south-west Asia. Their localized distribution is along the Tekeze River in southern Tigray, Northern Wollo and Eastern Gonder (see Fig. 1). The Agew and Tigray ethnic groups keep the goats. Abergalle goat milk is made into butter for home consumption and sale. Goat milk (*seba*) butter (*t'esmi*) is made and used for cooking and as a cosmetic (*likai*). Milk is also made into yogurt (*rugeo*) and cottage cheese (*ajebo*). Skins are processed locally and used as apron, as container for grain, water, butter or honey, a small mat or shoulder cape, pages of religious books (*brana*), sandals, belts and edges of baskets. The manure is widely used as a soil fertilizer in Tigray (Farm-Africa, 1996).


Figure 2. **Abergalle male goat**

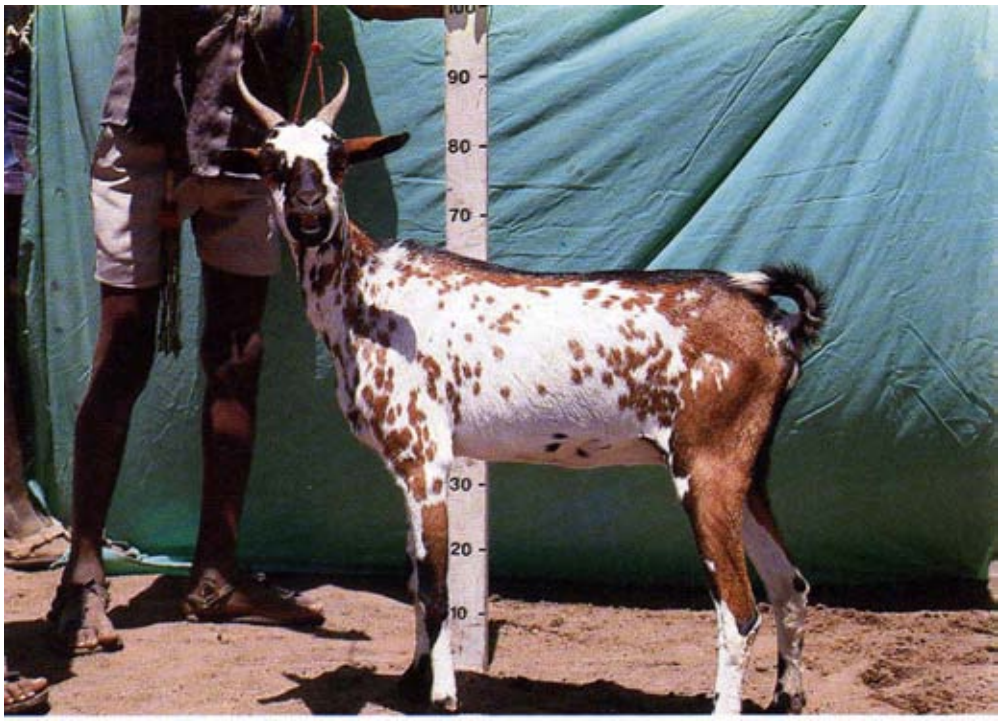


Figure 3. **Abergalle female goat**

**iii) Arsi-Bale**

The Arsi-Bale goat is also part of the rift valley family (previously known as *Galla-Sidamo* in the literature). They are distributed throughout Arsi and Bale regions,

up to an altitude of 4000 meters. They are also found in the higher altitudes of Sidamo and western Hararghe. They occupy all the agro-pastoral lowlands within the rift valley, from Lake Abaya in the South to south Shoa in the North (Figure 1. ). The Arsi-Bale goat is a relatively tall goat with a predominantly straight facial profile (98%). Males have curved (47%) and straight (41%) horns mainly pointing backwards (58%) with some pointed straight upward (28%). The coat colour varies between any of seven colours (white, black, brown, fawn, gray, roan and red), with the first three being most dominant. The ears are fairly long. About 10% have a semi-pendulous ear form, the rest being erect or horizontal. About half of the females and 75% of the males in Sidamo are bearded, but in the agro-pastoral areas beards on females are less frequent (38.7%). The goats in Sidamo appear to attain a greater number of average kidding for a given dentition class than other goat types in the region. Arsi-Bale goats are extensively milked. Among the Sidamo people goat milk is a highly valued source of nourishment for children. The demand for goat milk is so high that the smallest amount of milk available is diluted with some water to make sure that as many children as possible drink some. Goat meat is widely eaten, mostly during social and festive occasions. This goat type is more prolific than other goat types in the region and is potentially a desirable meat producer in the highlands. The skins of the hairy Arsi-Bale goats are highly valued as saddle covers (Farm-Africa, 1996).

Figure 4. **Arsi-Bale male goat**

**iv) Woyto-Guji**

Woyto-Guji goats are also known as *Woyto*, *Guji* or *Konso*. They are part of the rift valley family of goats. They are distributed in north and south Omo, southern Sidamo and parts of Walayita (see Fig. 1). These goats are mainly kept by pastoral ethnic groups (Tsemay, Malie, Hamer, Bena, Dasenatch, Bumie and Guji) and by a few agricultural groups (Konso and Gardula). More notably this goat type inhabits those areas in Sidamo known to be endemic with trypanosomiasis, especially the Gelo valley to the south of Lake Abaya and the western Ghenale catchment's area. They are Brown, black or red colour with a shiny, smooth coat and small head with a straight or concave facial profile. Coat colours are often marked with black or brown stripe along the back, on the underside or on the front of the legs. They are medium-sized goats with a mainly straight (89%) to concave (11%) facial profile. Straight horns occur in 71% of the males, curved in 26%, with polled goats forming 3% of the population. Horns mainly point backwards (75%) or upwards (21%) and in a few cases laterally (2%). The average number of kids born per breeding female is 4.3. Goats may be milked once or twice a day for one to six months irrespective of the season. Milk is sometimes made into butter and used in traditional healing treatment. The Woyto-Guji goats are known to be good for meat. Almost all pastoral groups in south Omo use goat skins to make traditional leather garments for women (Farm-Africa, 1996).

Figure 5. **Woyto-Guji Make goat**


Figure 6. **Woyto-Guji female goat**

**v) Hararghe Highland**

The Hararghe Highland goats are most likely derived from the Somali goat type. They are distributed in the highlands of east and west Hararghe (see Fig. 1). They are

small, white, brown or black, commonly polled with a straight (60%) or concave (40%) facial profile. Horned goats have short hair and straight (32%) or concave horns (29%). Beard is present on 72% of males but there are no ruffs. Wattles are present on 14% of goats. The average number of kids born per breeding female is 2.2. They are commonly milked (*anan ra'e*) for home consumption and sale in some markets. Goat meat is widely eaten and preferred to sheep meat. Skins are used for beds and prayer mat. They are also made into blacksmith's bellows (*buufa*) (Farm-Africa, 1996).


Figure 7. **Hararghe Highland female got**

**vi) Short-eared Somali**

The Short-eared Somali goats are also known as *Denghie*r or *Deghiye*r. They are probably related to the Arab goats in Somalia, which were introduced directly from Arabia. They are distributed in the northern and eastern parts of Ogaden i.e. Jijiga, Degeh Bur and Werder (see Fig. 1) where they are maintained by the Isaaq and Mijertein Somali clans, and Dire Dawa, Issa and Gurgura communities. They are medium sized; mainly have white, short hair. The Short-eared Somali goat is smaller than the Long-eared Somali goat type. It has a straight facial profile and males have straight (46%) and upward pointing (46%) horns. Females have generally curved horns (50%), most of which point upwards (55%), but 27% are oriented backwards and 12% are lateral pointing. Polled goats are found in 5% of males and 7% of females. The

average number of kids born per breeding female is 2.4. Somali goats are widely milked and milk is consumed fresh and also made into butter. The Somalis use butter for both food and medical purposes. Meat may be eaten fresh or preserved by cutting into slices, frying it in butter or animal fat and keeping it in a container (Farm-Africa, 1996).


Figure 8. **Short-eared Somali male goat**

Figure 9. **Short-eared Somali female goat**

**vii) Long-eared Somali**

Long-eared Somali goats are also known as *Large-white Somali*, *Degheir*, *Galla*, *Digodi*, and *Melebo*. They are probably related to the descent of the Arab goats in Somalia introduced directly from Arabia. Long-eared Somali goats are distributed throughout the Ogaden, lowlands of Bale, Borana and southern Sidamo (see Fig. 1). They are kept by the Hawia, Ogaden, Rare Bare, Digodi clans of the Somali ethnic group and the Boran, Gabra and Geri ethnic groups. The Long-eared Somali is a large white goat with a predominantly straight facial profile. Horns are mainly curved (in 41% males and in 46% females), and pointed backwards in 38% of males and upwards in 48% of females. Horns in some 13% of both sexes have lateral orientation. The ear is horizontal (in 46 % of females and in 66.8% of males) or semi-pendulous (in 22% of females and in 16% of males). Vestigial ears, which are not found in other goat types, occurred concentrated in one site, Dollo, in Borana. The average number of kids born per breeding female is 3.2. Somali goats are extensively milked by the Somali and Boran pastoralists. Goat meat is preferred to mutton in most areas where the Long-eared Somali goat is kept. Meat may be eaten fresh or preserved by cutting into slices, frying it in butter or animal fat and keeping it in a container (Farm-Africa, 1996).


Figure 10. **Long-eared Somali female goat**

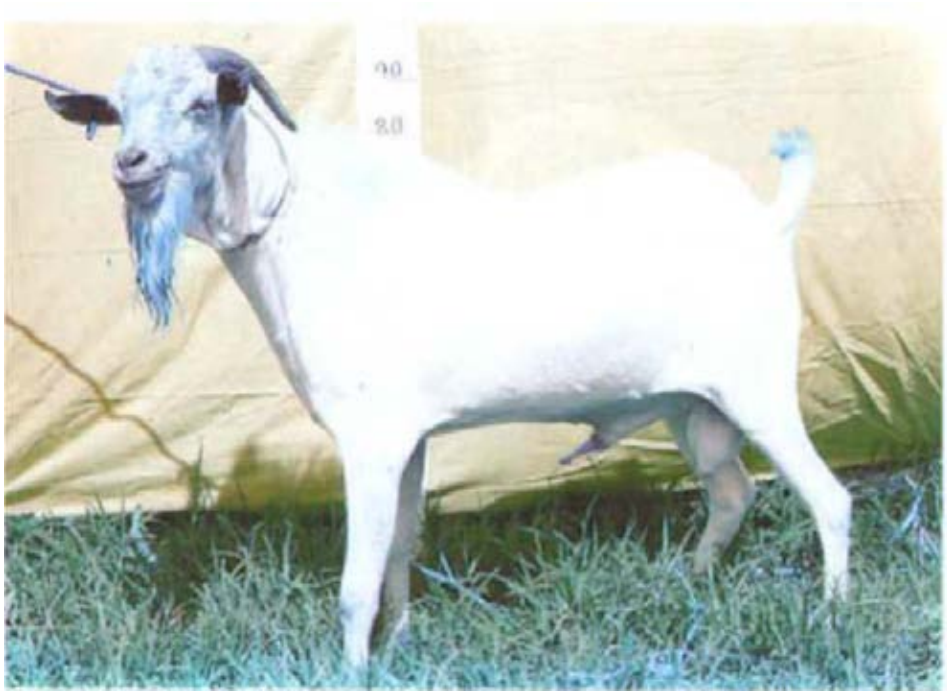

Figure 11. **Long-eared Somali male goat**

**viii) Central Highland**

Central Highland goats are also known as *brown goats*. They are highland type most likely derived from mixing of types in the past. They are distributed in the central

highlands west of the rift valley escarpment in central Tigray, Wollo, Gonder and Shoa (see Fig. 1). They are owned by settled farmers of the Tigray, and Amhara ethnic groups. Central Highland goats are medium sized, broad-faced with thick horns and reddish-brown colour. The Central Highland goat has a predominantly straight (71%) facial profile; 29% of the goats have a concave profile. Virtually all males have horns, 82% being straight and pointing backwards, 13% curved and 5% spiral. The coat type is short and smooth with 51% plain colour, 42% patchy and 7% spotted. The predominant colour is red-brown (41%), the remaining goats being split between black, white and gray. Most males (82%) have beard and ruff (99%). Wattles are present on 6% of males. The average number of kids born per breeding female is 2.9. Goats are expensively milked in Tigray. The skins from central highland goats are an important export product. The best quality skin 'Bati Genuine' is made from skins of brown/fawn coloured goats in the Bati area of Wollo (Farm-Africa, 1996).


Figure 12. **Central Highland male goat**

**ix) Western Highland (North West Highland)**

The Western Highland goats are also known as *Agew* locally. They are most likely derived from mixing of types in the past, distributed in the Highlands of south Gonder, Gojam, Wellega, and western Shoa (see **Fig. 19a**). They are relatively tall with

a concave facial profile (100%). It has a relatively coarse long coat (82%); 12% of the animals have hair on their thighs. The coat colour is mainly plain (51%), 42% have a patchy colour pattern, and 7% are spotted. The main colours are white (42%) and fawn (42%), and combinations of these colors. There are relatively a high proportion of polled goats in the western Highland population (14%). The horned goats have straight (76%) horns directed backwards (73%). A ruff is present on 99% of the males, and beard on 84%. Wattles are present on 12% of the goats. An unusual high proportion of polled and hermaphrodite goats were found around Lake Tana. The goats showed hermaphroditism based on examination of external genitalia. There is an established relationship between hermaphroditism and polledness in goats (Gall, 1981). The average number of kids born per breeding female is 3.6 (Farm-Africa, 1996).


Figure 13. **Western Highland male goat**


Figure 14. **Western Highland female goat**

**x) Western Lowland (Gumez)**

Western Lowland goats are also known as *Shankela* and *Gumez* locally. They are derived from mixing of types in the past. They are distributed in the western lowlands bordering Sudan in Gojam (Metekel), Wellega (Assosa) and Illubabor (Gambela) (see Fig. 1). The Western Lowland goat is a relatively short goat with a straight facial profile (100%). It has a predominantly short smooth coat (81%), with 16% having a relatively coarse coat. The main colours are white (42%) and fawn (38%), with some black (9%) and gray (11%), occurring mainly in patches (73%). Most male goats have straight horns (85%) oriented backwards (77%). There are 12% polled males in the population. A ruff is present in 96% of all males and beard in 70% of males. Wattles are present in 12% of all male goats. The average number of kids born per breeding female is 3.5. Western Lowland goats are remarkably prolific with 56% giving single births, 41% twin births and 3% triplets. Quadruplets were also reported to occur. Pastoral and agro-pastoral groups in the area of distribution use goat milk extensively. Goat meat is widely eaten. The horn is used as a musical instrument (*zoombara*) in Assosa (Farm-Africa, 1996).


Figure 15. **Western Lowland male goat**


Figure 16. **Western Lowland female goat**

**xi) Keffa**

Keffa goats are derived from mixing of types in the past. They are distributed in the highlands and lowlands of Keffa, and parts of south Shoa, Kambata and Hadiya

(see **Fig. 19a**). The Keffa goat is relatively short with a predominantly straight facial profile (92%). Most males (83%) have straight horns pointing backwards (80%). A small proportion, 14% have curved horns. The incidence of polledness is low at 3%. Keffa goats have short pricked ears. Most Keffa goats have a coarse (38%) to hairy (27%) coat type; some 16% have hair on their thighs. Plain colors predominate (52%), with some patchy colour patterns (45%). The main colours are black (30%) or brown (31%). Among males, 88% have beards, 97% have ruffs. Wattles are present in 12% of all goats. The average number of kids born per breeding female is 3.1. The pastoral ethnic group, the Surmas, drink the milk and blood of goats. Goat meat is widely eaten. The horn is used as a container for igniting a mixture of explosive minerals (Farm-Africa, 1996).


Figure 17. **Keffa male goat**


Figure 18. **Keffa female goat**

## 2.4 Conservation of domestic animals diversity

### 2.4.1 Meaning and importance of domestic animals diversity

Domestic animal diversity has been defined by Hammond (1993) as the spectrum of genetic variation existing among species, breeds and individuals of all animal species, which have been domesticated, and their immediate wild relatives. This genetic variation has been developed during the millions of years of evolution to form and stabilize each species. Globally, there are about 40 species of domestic livestock (FAO, 1995), which have been domesticated by humans over the last 12,000 years. In the process of domestication, however, separate and genetically unique breeds and strains have developed within each species as a result of human development and occupation of new areas over the planet. According to FAO (1995), there are about 4,000 to 5,000 breeds and strains of domestic animals in the world, of which 300 are breeds of goats (Devendra and Burns, 1983). These breeds and strains are referred to as the global animal genetic resources and the genetic variation both between and within the breeds are described as the diversity within the species of domestic animals. The selection process, both environmental- and human-directed, has resulted in much of the diversity existing between the breeds. Differences among the breeds have been created

by reproductive isolation, often imposed by human through pursuit of different breeding objectives and physical separation for various length of time. As a consequence of physical separation, each breed/strain has been adapted to particular ecological condition to suit the local climate and the requirements of the community. Thus, in the conservation of domesticated species, it is the diversity between breeds, rather than the variation between species, which is of crucial importance (Hammond, 1993; Barker, 1994).

Intra-specific diversity has got both within-population and between-population genetic components. Genetic variation at the population level consists of the differences in the types of alleles present and their frequencies across all members of a population considered together. Genetic variation within population is caused by change of allele frequencies over time due to selection, random genetic drift and gene flow (immigration from or emigration to other population). Genetic variation within population is important as it is related to heterozygosity, which is known to enhance fitness-related characteristics (Allendorf and Leary, 1986). Variability between populations arises both from random processes (founder effects, demographic bottlenecks, genetic drift and mutations) and from local selection imposed by environment and humans (Hartl and Clark, 1997). Variability between populations is also the result of adaptation of populations to their local conditions and it is important since locally adapted populations may have particular genes or gene combinations critical for viability in their local environment. For conservation of animal diversity, it is therefore, important that both within and between populations variation be maintained. This will enable better utilization of animals for production of food and agriculture to meet the current need of human kind while maintaining the potential to meet the range of future possible changes and needs (Hammond, 1993; FAO, 1995). This is because genetic diversity is the basis for a species evolutionary flexibility and responsiveness to environmental changes.

The main reason to advocate conservation of genetic diversity is that the improvement of domestic animals in order to meet human needs, is dependent on genetic variation both within- and between-breeds/strains. Loss of genetic diversity will likely decrease the ability of animals to respond to environmental change and will result in loss of genetic information potentially useful for breeding improvement (Primack, 1993; Hunter, 1996). The earth's climate varies widely and is continuously changing. It is the diversity of species and breeds that will allow the environment to be

utilized efficiently throughout the years both in temperate areas and in the tropics. The urge of conservation, particularly in developing countries, is strengthened by the predictions of the scale of human population growth. It has been predicted that, human needs for food will double over the next two generations with the demand for animal products increasing more rapidly than that of plants (Hammond, 1993; Hammond, 1994; Degado *et al.,* 1999). With increasing population pressure, the quantity of food and other products must increase. In developed countries consumer emphasis on product quality is increasing. Changing the production level and product quality require different types and varieties. In addition, changes in food and agriculture production will require new combinations of plant and animal species and lines (e.g. breeds and varieties) and changes of management and production strategies. Sustainability in this new production system will require different genetic types (FAO, 1995).

The second reason to advocate conservation of domestic animal diversity is that, we can reasonably expect that large differences observed today between geographic regions in human needs for food and agriculture and in production capability will persist in the future. According to Hammond (1994) almost three quarters of the world agriculture will remain at the low- to medium-input levels where animal production environments incorporate combinations of stress such as feed shortage, drought and heat. Breeds in these environments are reservoirs of genetic information responsible for the adaptations necessary for production under adverse environment. Although they have been overlooked, the needs for those, which are able to produce in hard conditions, is making and will make in future important contribution to feeding humans (Simm, 1998).

A third reason for conservation of domestic animal diversity is that sustainable crossbreeding schemes require at least two viable purebred populations and sometimes more. Trends towards fewer breeds will reduce the opportunities for improvement of productivity through crossbreeding.

The fourth reason is that the wide variety of livestock breeds and strains available today is part of our cultural heritage and it deserves protection. At the breed level, unique genetic differences exist between the breeds of each domestic species, for example in goats, the Nubians have high milk yield, the Small East African are hardy animals and the West African Dwarf goats are trypanotolerant. Thus, genetic diversity at the level of species can be exploited to satisfy human needs for varieties.

### 2.4.2 Loss of Genetic Diversity in Domestic Animals

Reduction in genetic diversity has been expressed primarily in terms of loss of breeds and strains. It is postulated that the current rate of extinction of species, breeds and strains is greater now than at any time in the past (Hammond, 1993). It is estimated that at least 30% to 40% of all animal genetic resources are currently at high risk of extinction (Hammond, 1994). However, adequate records do not exist to enable reliable estimates of either loss rates of the breeds or of domestic animal diversity itself. The existing data on the number of endangered breeds are likely to be underestimates of the magnitude of the problem. The loss of genetic diversity is occurring both within populations and among populations.

### 2.4.3 Loss of Diversity within Population

The factors that diminish genetic diversity within populations are genetic bottlenecks, random genetic drift, inbreeding and human activities. All the four factors are functions of population size (Lacy, 1987; Lande and Barrowclaugh, 1987). These described as follows:

#### 2.4.3.1 Genetic Bottlenecks

A demographic bottleneck occurs when a large population experiences a severe, temporary reduction in size due to environmental or demographic events. For example, natural catastrophes, which occur at unpredictable intervals including events, like drought, disease outbreak and war. These events may kill a certain percentage of a population and therefore reduce the effective population size. The result is that the genetic variability of all subsequent generations is contained in the few individuals that survive the bottleneck and reproduce. Hence, some genetic diversity is lost in the process. The magnitude of the loss in diversity depends on the size of the bottleneck and the growth rate of the population afterward (Hunter, 1996). Another demographic event which may lead to bottleneck effect is the founder event. A founder event occurs when a few individuals of a population establish a new population. The genetic constitution of the new population depends up on the genetics of the founder animals. The genetic diversity of the original larger populations is reduced because the sample of

genes in the few founder animals is not likely to be representative of the original gene pool.

Generally passing through a genetic bottleneck can create two problems (Carson, 1983): a loss of certain alleles, especially rare alleles, if no individuals possessing those alleles survive, and a reduction in the amount of variation in genetically determined characteristics due to the presence of fewer alleles and decline in heterozygosity. Baker and Moeed (1987) have shown that the founder populations of several species of introduced birds have fewer alleles and lower heterozygosity than the source populations. Bottlenecks have been shown to reduce allozyme heterozygosity. Leberg (1992) suggested that extremely low levels of allozyme heterozygosity in broad geographical surveys imply the occurrence of one or more recent severe bottlenecks. The overall effect of bottlenecks is the decline in fitness of the individuals in the population.

#### 2.4.3.2 Random genetic drift

Genetic drift is a random change in gene frequencies in small populations, attributable to sampling error. In small populations each generation retains just a proportion of the gene pool of the previous generation. Subsequently after many generations of random genetic drift, small populations will usually result in loss of variability as a consequence of fixation of some alleles on loci and loss of other alleles.

#### 2.4.3.3 Inbreeding depression

Inbreeding results from the mating of two closely related individuals. However, the degree of relationship may vary. "Closebreeding" refers to the mating of very close relatives such as sibling to sibling or parent to offspring. Its probability of occurrence increases in small populations if mating occurs at random. Inbreeding allows the rare, harmful recessive alleles to become expressed in the homozygous form, with resulting harmful effects on the offspring (Selander, 1983; Ralls *et al.*, 1988; and Charlesworth, 1987; Lomker and Simon, 1994) such as reduction in fertility, fecundity, offspring size, growth and survival, and physical deformities. Since inbreeding depresses reproductive fitness, it is assumed to increase the risk of extinction. This presumption is supported by correlation between extinction and inbreeding on laboratory and domestic animals

(Soulé, 1980). However, gene flow from outside populations is beneficial in avoiding inbreeding and the erosion of genetic diversity (Miller and Waits, 2003).

#### 2.4.3.4 Human activities

Extinction of species, breeds and strains/varieties through human activities represents the greatest threat to genetic diversity. The major threats to genetic diversity that result from human activity are habitat destruction and degradation, pollution, introduction of exotic species, and over-exploitation (Frankham, 1994). These threats are all caused by an ever-increasing use of the world's natural resources due to expanding human population and development of market economy. The growth of towns, factories and mines in developing countries creates a cash market for livestock products. Consequently, the traditional farmers who formerly kept animals for their own needs begin to supply the cash market. Genetic variation is being lost as farmers in developing countries abandon their local breeds in favour of high-yielding breeds for commercial production. Even if human activities do not directly eliminate a breed or strain, loss of genetic variation is taking place as the number of individuals in populations is reduced. In the long run the population size of a breed or strain may become so small that the bred/strain is no longer viable and may eventually go extinct.

### 2.4.4 Loss of Diversity among Populations

Loss of genetic diversity among-populations occurs when historically divergent and isolated populations experience an artificially high rate of gene flow from other populations. In domestic animals this is typically happening, as traditional farmers mix and interbreed their populations by moving from one place to another when they set up new settlements. It also occurs through crossbreeding by increasingly promoting universal use of very few superior breeds per species to upgrade local breeds. The result is that the uniqueness of formerly distinct breeds/strains is diminished or lost. Moreover, there are concerns that, if there exist co-adapted groups of genes within each subpopulation, merging may result in the break-up of these and the loss of the particular adaptedness of each subpopulation to its environment (Templeton, 1986). Lacy (1987) has shown that the effect of gene exchange between subpopulations is to increase the variance within groups, decrease the variance between groups and decrease the total

variance. According to Hammond (1994) about 50 % of the total variance for several quantitative traits in cattle, sheep and pig breeds is at the between-breed level, the remainder being common to all breeds. Hence, a move to one breed would eliminate half of the total variation. Loss of genotypic variance is irreversible as is the loss of alleles from a population. This is because it is impossible to sort out the alleles into groups of genotypes resembling the original subpopulations (Nelson and Soulé, 1987).

### 2.4.5 Approach to conservation of animal genetic diversity

Conservation of genetic resources has become a major issue of public interest in the last decade or so. This is due to increase awareness that the genetic diversity is the basis for a species evolutionary flexibility and responsiveness to environmental changes and that loss of genetic diversity will restrict the opportunities available to mould domestic animals to meet unpredictable future changes and requirements. Hence, genetic diversity needs to be preserved. The loss of variation within breeds is continually balanced by the introduction of new variation through new mutations; however, it is unwise to rely on new mutations as they are rare events and generally have deleterious effects on reproductive fitness (Frankham, 1994). Moreover, the genetic variation present as differences among breeds cannot be regenerated. Each breed or strain is a result of mutation, genetic drift as well as separate adaptation and evolution, often over many centuries following different selection pressures imposed by climate, endemic parasites and diseases, nutrition and human selection. Each breed thus comprises a unique set of genes, which must be conserved.

Conservation of domestic animal diversity has been defined as the total sum of all operations involved in the management of animal genetic resources so that the pool of genetic diversity is maintained over time (Hammond, 1993). It encompasses management of human activities in such a way that animal genetic resources are best utilized and developed to meet immediate and short term human needs for future generations. Most attention in conservation of animal diversity has been directed toward rare breeds. However, in management of animal genetic resources, the fundamental problem is not the distinction between the breeds that are endangered and those that are not, but between those that are perceived to have little or no current utility

and those which do have current utility or seem likely to have an immediate future use (Barker, 1994).

The critical issues in conservation of domestic animal diversity according to Notter *et al.* (1994) includes: (i) defining utilization strategies for available breeds to optimize production efficiency and sustainable utilization throughout the global range of environments and production systems; (ii) livestock improvement strategies that make appropriate use of global domestic animal diversity; (iii) development of optimum strategies for assessment and preservation of the genetic diversity, which exists in breeds that have currently little or no value in commercial production. FAO (1995) has recommended the following strategies for effective management of domestic animal diversity at global level and for each species:

- Identifying and listing all breeds.
- Describing and characterizing breeds in order to understand their unique qualities and potential contributions and to understand which breeds have the potential to make the greatest variety of future contributions.
- Monitoring the population statistics for each breed and regularly reporting to the world the populations currently at risk of extinction.
- Sorting adequate samples of as many breeds as possible, generally in the form of frozen semen, ova, and embryos, to enable the future regeneration of lost populations of animals.

An overview of the historical background of animal genetic resources and the reasons for concern about their future, both in the developing and developed world, has been presented by Barker (2002). He concluded that the primary focus in the conservation of domestic animal diversity is on conservation of breeds.

In general, two methods are used for conservation of genetic resources to support future livestock improvement: *in-situ* conservation and *ex-situ* conservation.

#### 2.4.5.1 In-situ conservation

This involves maintaining breeds in their farm environment. *In-situ* conservation is primarily the active breeding of animal populations for food and agriculture, such that diversity is best utilized in the short term and maintained for the longer term (Hammond, 1994). Operations pertaining to *In-situ* conservation include performance recording and development of breeding programmes. It also includes

ecosystem management for sustainable production of food and agriculture. It has been suggested that the number of animals required maintaining a breed in western countries ranges from 150 to 1500 breeding females, depending on the species and reproductive rate. However, in developing countries the number should not decline below 5000 breeding females (Hodges, 1990). The advantages of this approach are that the animals are still being utilized, the performance characteristics can be properly recorded and evaluated, and the breeds have the opportunity to evolve. The disadvantages are that selection and genetic drift may result in unfavorable genetic changes if the population is small. There is a risk of increasing inbreeding and hence homozygosity, which is associated with reduced fitness. The animals are at risk from disease and other natural disasters. Also, they are likely to be less productive and so more costly to maintain.

#### 2.4.5.2 Ex-situ conservation

In the context of conservation of domestic animal diversity, *ex-situ* conservation means storage. Generally three approaches are used for *Ex-situ* conservation.

##### 2.4.5.2.1 Maintaining breeds in farm parks or other animal collections

It involves the breeding animals of a sample of a breed outside its normal production environment or habitat. Many of the pros and cons of this approach are similar to the *in-situ* conservation method. However there is potentially more control over management of the population. It has been suggested that 50 individuals should be the minimum number necessary to maintain genetic variability (Franklin, 1980). With such population only 1 % of the genetic variability will be lost per generation. This figure is on the safe side. Indeed it has been suggested that animal stocks can be maintained with a loss of 2–3% of the variability per generation (Primack, 1993). According to Notter *et al.* (1994) populations can be maintained at an inbreeding rate of 1% per generation corresponding to an effective population size (*N*e) of 50%, which can be achieved with 14 males and 100 females under random mating.

##### 2.4.5.2.2 Creating a conservation herd (gene pool)

This involves crossing several rare breeds together, then breeding them to maintain genetic variability. It is an effective way of conserving genetic variation for two or three breeds. Maintenance of genetic diversity is almost better served by

pooling five breeds in a conservation herd with *Ne* of 250 (Notter *et al.,* 1994). However there is a greater risk of loosing useful genes when more populations are combined. The disadvantage of this approach is that, although useful genes may be conserved, the identity of individual breeds is lost.

**2.4.5.2.3 Cryopreservation**

This involves frozen storage of rare breeds in the form of living semen, ova, embryos or tissues, which can be used to regenerate animals. Cryopreservation of semen and embryos is a powerful tool for preservation of genetic diversity (Moore, 1985). In the situation of a critical threat with high probability that a breed will become extinct, preservation of genetic material of individual animals in the form of germ cells and embryos is necessary to ensure that an adequate genetic pool is retained for future improvement programs. Breeding technologies as artificial insemination (AI) and embryo transfer (ET) may provide support for this approach. The use of frozen semen in conservation program is particularly feasible where a tradition of use of AI is already strong. Collection of semen of endangered local breeds should take place as part of the AI program. According to Smith (1984) semen from 25 unrelated sires would provide an exceptionally good sample of the genetic diversity of a breed. Also Notter *et al*. (1994) recommended cryopreservation of semen from 5 breeds with 10 sires from each breeds and argued that it will provide greater preserved diversity than frozen semen of 25 sires of each of two breeds. For loci with multiple alleles, the probability of retaining all alleles from the rare breed is higher for the conservation herd supported with frozen semen than for a conservation herd maintained by random mating alone (Lomker and Simon, 1994), however the ability to rapidly recreate living purebreds is a key issue in germplasm utilization. If frozen semen is used as the only method of conservation, then several generations of backcrossing are needed to reinstate the breed concerned. In contrast, breeds can be reinstated rapidly from frozen embryos. Embryo preservation is particularly important when the reproductive rate is low or the generation interval is long. Furthermore, maternal cytoplasmic genes (e.g. mtDNA) cannot be preserved from semen, but can be effectively conserved by maintenance of even a modest number of female lineages. However, for all species, the likely cost per haploid genotype maintained or per animal sampled is much less for semen than for embryos (Notter *et al*., 1994). In general, embryo preservation should be attempted

only when ET is already in use for livestock improvement or when a breed has both clear potential value and a high likelihood of extinction.

Frozen storage of semen and embryos are relatively expensive, but it has the advantage that after the initial investment, semen and embryos provide much greater flexibility for finding ways of enhancing the reproduction of individuals, which are refractory to natural breeding conditions (Moore, 1985). In addition, the breed being conserved is free from unintended genetic changes (Simm, 1998). During storage frozen genetic material is at less risk from disease and natural disasters than live animals, but obviously at risk from technological failures. Splitting the material collected from a particular breed or population and storing it in/at different locations can minimize these risks. The disadvantages are that reproductive technologies are not uniformly successful or presently available for all individuals or all species and the expertise is not always available in the places where it is needed most. Also the breeds conserved in this way are not able to adapt to changes in the production environment or new disease challenges. Semen or embryos frozen now may as well be unable to meet future requirements for improvement.

## 2.5 Molecular techniques for the assessment of animal genetic diversity

### 2.5.1 Rationale for molecular genetic approach

Biological variation is of wide occurrence in nature. Morphological variations in type, traits and production characteristics have been known and characterized within and between species for many centuries. Phenotypic characterization has traditionally been used in the description of breeds. The characters vary from external features (horns, ears, coat colour), physical body measurements (heart girth, height, body length), production traits (body weight, milk yield, wool production, carcass traits), reproductive traits (fertility, age at first kidding, kidding interval), to survival traits (mortality, resistance to diseases, tolerance to heat and poor management and feeding) (Aboagye, 1992; Aboagye *et al.*, 1994). The description of breeds with respect to coat colour, horns, physical body measurements and productive traits is based up on observation of the phenotype. An organism's phenotype is principally a manifestation of its genotype. Phenotypic characters have the advantages of being easily seen and measured and usually much lower costs are incurred during phenotypic characterization

compared to genetic characterization (Minelli, 1993). For these reasons, phenotypic characters have been used extensively for characterization and identification of breeds. One of the problems associated with phenotypic characterization is the difficulty to combining different measures in order to provide a useful tool for the description of a breed. Most phenotypic characters are polygenetically inherited and many of them are influenced by environmental effects and with sometimes strong genotype-environment interaction. Furthermore, phenotypic characters are affected by natural selection. Therefore, due to the influence of different environmental conditions and different selection pressure on different kinds of characters, interbreed phenotypic comparison is unlikely to give meaningful results.

Earlier work to study genetic variability of individual animals and populations employed screening of protein variants by gel electrophoresis. Polymorphism in gene products such as enzymes, blood group systems and leucocytes antigens have been used for investigating genetic diversity at the molecular level. Using the above protein markers, numerous studies have estimated genetic variability, gene flow and phylogenetic relationships among populations (Barbancho *et al.*, 1984; Nguyen, 1990; Pepin and Nguyen, 1994). However, these techniques lack the power to resolve the differences between closely related breeds (Meghen *et al.*, 1994; Dowling *et al.*, 1996) since a great deal of genetic variation remains undetectable by using protein markers. Moreover, the genotype frequencies estimated from protein markers may be influenced by natural selection among alleles (Alexandrino *et al.*, 1983; Pemberton *et al.*, 1988; Mopper *et al.*, 1991) making it difficult to interpret inter-population comparisons.

The goal of breed characterization is to determine the genetic diversity within and between breeds. The near ultimate description of an animal should be a description of the sequence of nucleotides that comprise its genome. Describing differences and similarities in the DNA of two or more populations can provide the measure of relative genetic distances of such populations from each other. The analysis of DNA has several significant advantages over protein markers for the study of molecular population genetics and systematics (Dowling *et al.*, 1996): (i) the genotype rather than the phenotype is assayed; (ii) the expression of DNA markers is not influenced by development or by environmental factors; (iii) one or more sequences appropriate to a problem can be selected on the basis of evolutionary rate or mode of inheritance; (iv) the methods are for most part general to any type of DNA and, (v) DNA can be prepared from small amounts of tissues and is relatively stable, even in non-

cryogenetically stored tissues. The last attribute means that genetic information on rare or endangered species can be obtained without destructive sampling (Taberlet *et al.*, 1997) and it is possible to analyze DNA from extinct populations or species (Taylor *et al.*, 1994). More recently, molecular data from DNA markers have received particular attention in the study of population variability because of their possible use to determine the chronology of evolutionary events using neutral DNA markers.

### 2.5.2 Molecular techniques in the study of animal genetic diversity

DNA markers (molecular markers) are the most important tools added to the armory of the scientists to characterize, distinguish, or relate individuals within or between breeds, strains, or species. The advances of the last two decades in techniques on DNA polymorphism and subsequent data analysis have greatly increased the ability to understand the genetic relationship among species at the molecular level. DNA genetic markers have now been used for the molecular genetic characterization and genetic diversity studies in numerous species.

The DNA polymorphism to be used widely for genome characterization and analysis was restriction fragment length polymorphism (RFLP) (Botstein *et al*., 1980). Large amount of data have been generated since the first demonstration of RFLP (Grodzicker *et al.*, 1974) and RFLP have been used as markers in human genetics (Botstein *et al.*, 1980), in genetic improvement of plant (Tanksley *et al.*, 1989) and animals (Soller and Beckmann, 1982). The introduction of PCR (Saiki *et al*., 1988) in conjunction with the constantly increasing availability of DNA sequence data also represents a milestone in this endeavor. A variety of different molecular techniques are now being used in various laboratories for the study of inter-and intra-specific genetic variation at the DNA level. The most widely used techniques are restriction fragment length polymorphism (Botstein *et al*., 1980) of nuclear DNA (Jeffreys and Morton, 1987) and mitochondrial DNA (Loftus *et al.*, 1994), minisatellites probing (Jeffreys *et al*., 1985; Jeffreys and Morton, 1987), microsatellites PCR amplification (Litt and Luty, 1989; Weber and May, 1989), random amplified polymorphic DNA (RAPD) (Welsh and McClelland, 1990; Williams *et al*., 1990), amplified fragment length polymorphism (AFLP) (Vos *et al.*, 1995) and sequencing of mitochondrial DNA fragments (Cunningham *et al*., 1994). These techniques differ in the type of data they generate, in

the way that they resolve genetic variations and in the taxonomic levels at which they may be most appropriate.

#### 2.5.2.1 Restriction fragment length polymorphism (RFLP)

Polymorphism at the DNA sequence level is often detected as a variation in length of restriction fragments produced by cutting DNA with restriction enzyme (Botstein *et al.,* 1980). Such polymorphism is referred as RFLP and the first genetic map of the human genome was produced using this technique in the year 1980. RFLP is the result of differences in the sequence of nucleotides in the DNA of individual animal. These differences are the result of mutations in the genome occurring over time and are detected as variations (polymorphism) in length of restriction fragments. In RFLP, DNA is digested with restriction enzymes that cut DNA at particular sites within a specific recognition sequence, typically of 4-6 base pairs (bp) long (Dowling *et al*., 1996). Since each enzyme cleaves DNA at a characteristic recognition sequence, the complete digestion of a particular DNA produces an array of fragments. These fragments are separated by gel electrophoresis and blotted onto a filter and then probes are hybridized to the DNA. Variations in fragment lengths between individuals, populations or species, are results of mutation, which either alter restriction sites or lead to insertions/deletions of cleavage sites. Such polymorphism in a specific gene locus can be used to distinguish animal species, populations and individuals. The advantages are that RFLPs give highly reproducible band patterns and are co-dominant markers, hence, heterozygotes are distinguishable. The limitations of RFLPs are that, it is labor intensive and relatively expensive, as a good supply of probes is needed and the blotting and hybridization steps are time-consuming, and difficult to automate. Also the extent of detectable polymorphism may be limited for intra-species comparison and sufficient quantities (e.g. 10 µg per digestion) of good quality DNA are required. Because of the last limitation, RFLPs are not applicable when very limited amount of source material or preserved tissues are available.

#### 2.5.2.2 Mitochondrial DNA (mtDNA) analysis

Mitochondrial DNA (mtDNA) sequence differences among individuals or populations can be assayed indirectly through restriction site analysis (Loftus *et al*.,

1994) or directly through DNA sequencing (Cunningham *et al*., 1994; Bradley *et al*., 1996; Kikkawa *et al*., 1997). The former is often less expensive and allows rapid screening of large samples. The latter provides often more information as all changes at the nucleotide level will be revealed. In restriction site analysis, isolated mtDNA is digested with a restriction enzyme and the resulting fragments separated by gel electrophoresis. Genotypes are comprised of the composite restriction fragment patterns following digestion with 10-20 restriction enzymes. The pattern of restriction site loss or gain among genotypes for a panel of restriction enzymes can be used to estimate genetic differentiation and evolutionary history of populations or species (Avise, 1994). In sequencing, mtDNA genes are directly sequenced using universal or species specific primers that allow amplification of sequences as much as several thousand base pairs in length. Since different genes in the mtDNA genome evolved at different rates, rapidly evolving genes can be analyzed to determine the relationship of recently diverged populations, whereas slowly evolving genes can be used to answer systematic questions involving distantly related species.

#### 2.5.2.3 Random amplified polymorphic DNA (RAPD)

RAPD markers were developed in the early 1990s (Williams *et al*., 1990). The RAPD polymorphisms are based on mismatches in primer binding sites or insertion/deletion events and therefore usually result in the presence or absence of an amplified product from a single locus. RAPD involves the use of a single and short arbitrary primer which can be synthesized or are available commercially. Using of such primer in PCR amplification results in several discrete DNA band products. Each product is derived from a region of the genome that contains two short segments in inverted orientation, on opposite strands that are complementary to the primer and sufficiently close together for amplification to work. More particularly this technique utilizes random 10-base oligonucleotide as primer to amplify discrete fragments of genomic DNA through PCR (Williams *et al*., 1990) i.e. in contrast to using two specific primers as used in a conventional PCR for amplification of a specific region in DNA. RAPD employs a single primer (10 mer) to amplify DNA of any genome region that happens to be flanked in opposite orientation within 5000 bp of each other. During the annealing step the primer anneals to the template at two sites on complementary strand of DNA template. If these priming sites are within an amplifiable range of each other, a

discrete DNA product is formed through the amplification. In RAPD, the amplification products are separated on agarose gels in the presence of ethidium bromide and the resultant bands are visualized under ultraviolet light (Welsh and McClelland, 1990; Williams *et al*., 1990, 1993). The presence or absence of bands can be scored and the data converted into similarity matrices for calculation of genetic distances (Gwakisa *et al*., 1994).

RAPD method has opened a new area of genetic analysis due to its relative simplicity and speed. It is being applied widely in livestock animals for various purposes such as breed characterization (Gwakisa *et al*., 1994; Kantanen *et al*., 1995; Teale *et al*., 1995; Yadav *et al.,* 2001), species identification, parentage testing, and pedigree analysis (Chung *et al.,* 1998), population studies (Parejo *et al.,* 1998), and detection of genetic variations (Kantanen *et al*., 1995). Luo *et al*. (2001) applied RAPD technique to evaluate genetic relatedness and diversity among five goat breeds and confirmed that RAPD marker analysis can be used to determine genetic diversity and relationship among and within goat breeds (http://www2.luresext.edu/goats/library/Luo2001-1.html).

The advantages of RAPD are that there is no requirement for DNA probes or sequence information for the design of primers, the procedure is quick, simple and can be automated, and very small amount of DNA (e.g. 10 ng per reaction) is required, moreover this technique does not require any restriction enzyme and radioactivity, and the technique is cheaper (Williams *et al*., 1993; Kantanen *et al*., 1995). Disadvantages of RAPD are however many. First, RAPD methods are very sensitive to the PCR reaction conditions, DNA quality and PCR temperature profiles (Karp *et al*., 1998). It is absolutely critical to maintain strictly constant PCR conditions in order to achieve reproducible results. Second, the markers are dominant, thus heterozygotes cannot be detected. Third, in the absence of pedigree analysis, the identity of individual bands in the multi-band profiles is not known and there can be uncertainty in assigning markers to specific loci. This makes it difficult to use RAPD in inter-population or inter-specific comparisons (Dowling *et al*., 1996). Fourth, single bands on the gel can sometimes be comprised of several co-migrating amplification products.

#### 2.5.2.4 Amplified fragment length polymorphism (AFLP)

The AFLP method is a more recent PCR based technique (Vos *et al*., 1995), which is essentially intermediate between RFLPs and RAPDs. In AFLPs, the first step

involves restriction digestion of the genomic DNA and the oligonucleotide adapters are ligated to the ends of the restricted fragments. The second step involves selective amplification of sets of restriction fragments. Either a pre-selection step is performed using magnetic beads followed by a round of selective PCR or two selective rounds of PCR amplification are applied (Vos *et al*., 1995). In the last step, the amplified products are separated by polyacrylamide gel electrophoresis and can be visualized using radioactive or fluorescent labeling.

Compared to RAPDs, AFLP results are much more easily reproducible (Vos *et al*., 1995). Compared with RFLPs, AFLPs are faster, less labour intensive and provide more information (Powell *et al*., 1996). Because of their large genome coverage (on average they give 100 bands per gel compared with 20 for RAPDs), AFLPs are particularly good for mapping, and also fingerprinting and genetic distances can be calculated between genotypes. However, AFLPs share many of the limitations, with respect to band homologies and identities, as RAPDs, for example as RAPD bands, AFLP are codominant markers.

#### 2.5.2.5 Variable number of tandem repeats (VNTRs)

Throughout the genome of individuals there are many regions comprised of tandemly repeated simple sequences. These repeat sequences vary in number and hence in length and are generally called variable number of tandem repeats (VNTRs), although the terms minisatellites and microsatellite are used depending on the size of the sequence of base pairs which is repeated. Jeffreys *et al*. (1986) first reported the term minisatellites. Minisatellites region are characterized by randomly repeating oligonucleotide units ranging from 10 to 60 bp in length (Haberfeld *et al*., 1991, Armour and Jeffreys, 1992) whereas in microsatellite the repeated sequences are only 2 to 5 bp long (Tauz, 1989; Weber and May, 1989; Queller *et al*., 1993). Although variation in both systems is the result of change in copy number, the underlying mechanism of mutation and the chromosomal locations differ. Microsatellite loci are randomly distributed and subject to replication slippage (Weber and May, 1989; Schlötterer and Tautz, 1992) while minisatellites loci tend to be concentrated near telomeres and their length vary following intra-molecular or intra-allelic recombination and gene conversion mechanisms (Jeffreys *et al*., 1994).

Minisatellites are assayed by digestion of DNA with restriction enzymes which do not cleave within the tandem repeats and Southern blot transfer and hybridization using minisatellites sequences or synthetic oligonucleotide probes. Two strategies can be applied. In one approach, the mixture of restriction fragments can be separated by size on agarose gel in the presence of an electric field and visualized by hybridization with a radioactively labeled probe based on the core sequence of the repeat (Jeffreys *et al*., 1985). The result is complex patterns of bands that are usually unique to an individual. The complex band patterns are widely known as DNA fingerprint. DNA fingerprint has become an extremely powerful tool for testing parentage (Burke, 1989), relatedness of individuals as well as overall population variability (Wayne *et al*., 1991). A major advantage of this method is that many probes can be applied across a wide spectrum of plants and animals. An apparent problem with DNA fingerprinting is that it is a multilocus approach, which reveals multiple fragments some of which segregating together. Such linkage disequilibrium between bands can only be detected with the analysis of pedigrees. Moreover, assigning specific fragments to a particular locus is difficult and hence, identifying alleles and determining genotypes is not possible in the absence of pedigree analysis. Last but not least co-migrating fragments might not be homologous. This will lead to inflated band sharing coefficients between non-relatives (Lynch, 1988). With second approach, specific minisatellites loci are studied using single locus probe rather than multilocus ones (Nakamura *et al*., 1987). The advantage of this method is that alleles can be assigned to specific loci and genotypes can be identified. However, cloning of minisatellites loci is a relatively complex process. This method requires more efforts than multilocus fingerprinting and the variation at individual loci tends to be taxonomically restricted (Hanotte *et al*., 1991).

Microsatellite loci are usually examined one at a time via PCR (several loci can be examined simultaneously by multiplexing). The PCR amplification products can be visualized using radioactive or florescence methods. Under radioactive methods, PCR amplified microsatellites can be detected either by direct incorporation of a single labelled dioxynucleotide triphosphate (dNTP) like ($\alpha$-$^{32}$P) dCTP during thermal cycling (internal labeling) or a single 5' ($\gamma$-$^{32}$P) dATP end-labelled primer in the PCR mix (end labelling). The products are resolved on acrylamide gels, fixed, dried and autoradiographed. Under the fluorescence methods, fluorescently labeled dNTPs are used for internal labeling in the PCR. Alternatively and more commonly one of the

PCR primers will be fluorescently labeled. The PCR products, fluorescently labeled, are separated on polyacrylamide gel and detected when excited to fluorescence by a laser and analyzed by computer software. Although the initial cost is high, this system offers considerable advantage in terms of size accuracy and enhances throughout. An advantage of microsatellite over minisatellites is their smaller size and single loci can be amplified using PCR. Thus, there is a great potential for studying allele frequencies at single hypervariable loci in contemporary and historic populations. Moreover, co-dominant genotypes can be scored and exact allele sizes can be determined. The technique has great potential for speed and accuracy once the appropriate PCR primers are known and a great deal of polymorphism can be determined. The use of microsatellites over RFLPs and RAPDs is that their genetic basis of variability is readily apparent. Unique primers amplify a genomic region including a well-defined repeat structure that is responsible for the observed variation. Compared to minisatellites microsatellites are more conveniently spaced throughout the genome while minisatellites tend to be found more frequently near the ends of chromosomes. Also most minisatellites alleles are too long to be amplified easily by PCR. Microsatellites typically are less than 400-500 bp long, consisting of 10- 30 copies of a repeat that is four bp in length and so they are more amenable to PCR typing. The disadvantage of microsatellite is that, the work required to develop primers for each new species examined and that only a few allelic states are possible, hence, increasing the chance of parallel evolution of a particular sequence repeat. However, with the increasing number of microsatellite maps for economically important species and due to the fact that primers developed for one particular species can be sometimes applied across a wide range of related taxa (Moore *et al*., 1991). Cloning of microsatellites for development of primers for new species will be unnecessary in near future.

### 2.5.3 Application of Microsatellite markers for the analysis of population

#### 2.5.3.1 Microsatellite as genetic markers for population studies

Microsatellites usually called Simple Tandem Repeats (STRs), are highly polymorphic class of genetic markers (Weber and May, 1989), consist of short sequence repeat motifs of tandemly repeated di, tri, tetra or penta nucleotide sequences that occur at large number of loci throughout the eukaryotic genome (Hamada *et al*.,

1984, Tauz, 1989). The dinucleotide repeats cytosine – adenine (CA) or Guanine – thymine (GT) on the other strand is the most common in mammalian genomes (Meghen *et al*., 1994). Polymorphism of microsatellite markers takes the form of variation in the number of repeats at any given locus. Microsatellites with more than ten dinucleotide repeats tend to be highly informative (Weber, 1990). For instance, in haploid human genome there are at least 35,000 CA repeats. That is they occur every 100,000 bp (Weber, 1990). Tri- and tetra-nucleotides have been shown to occur at a frequency of 1 every 300 – 500 kb on X-chromosome (Edwards *et al*., 1991). Due to exceptionally high rate of mutation, the majority of microsatellite loci are highly polymorphic in most mammalian species (Jeffreys *et al*., 1988; Weber, 1990). Alleles at microsatellite loci conform generally to Hardy-Weinberg principle and segregate in a Mendelian fashion. Most microsatellite loci are selectively neutral. This makes them compatible with the assumption of the neutral theory of population genetics (Kimura and Crow, 1964). Microsatellite polymorphism has advantages over other techniques in that it is a very reliable, highly accurate and repeatable method and has the potential of being standardized all over the world and the other potentially very large advantage of microsatellite in population genetics is the fact that primers developed for a particular species can be sometimes applicable across a wide range of related taxa. For example, some primers developed for cattle can be used for sheep, goats, yak, etc.

Microsatellite markers are currently the markers of choice for a wide range of molecular genetic studies such as establishing genetic linkage maps (Kappes *et al*., 1997), analysis of mating system and population structures (Bruford and Wayne, 1993; Queller *et al*., 1993; Schlötterer and Pemberton, 1994) and reconstruction of phylogenetic relationships among populations (Bowcock *et al*., 1994; Roy *et al.*, 1994, Forbes *et al*., 1995; Takezaki and Nei, 1996; MacHugh *et al.*, 1997). Their desirable characteristics are that: microsatellite loci are found in large numbers and are relatively evenly distributed throughout mammalian genomes, Mendelian co-dominant inheritance and relative ease of scoring through polymerase chain reaction. Microsatellites are conserved across related species. Moreover microsatellite variation is independent of age, sex and environmental changes and hence can be detected at the early stage of development.

The majority of microsatellites are found in non-coding regions of genome. However Morin *et al.* (1994) also reported the presence of microsatellite in protein coding region exhibiting regulatory role in gene expression and trinucleotide repeats

have been used for linkage analysis in association with disease susceptibility genes in human (Richards and Sutherland, 1994). Microsatellites have been found conserved between closely related species (Stallings *et al*., 1991) allowing primer sets to be used across species (Moore *et al*., 1991). They have been extensively used for linkage mapping in diverse organisms from human to mosquitoes (Weissenbach *et al.*, 1992; Crawford *et al.*, 1995; Vaiman *et al.*, 1996; Barendse *et al.*, 1997) and their use has enabled the identification of quantitative trait loci in major livestock species (Georges *et al.*, 1993). Microsatellites have proven useful in the analysis of paternity and kinship (Queller *et al*., 1993) and statistical calculation of probability at the individual (Edward *et al.*, 1992) and the population levels (Paetkau *et al.*, 1995b).

Microsatellite variation has been used to study the amount of hybridization between closely related species (Gottelli *et al*., 1994; Roy *et al*., 1994). Spatial distribution of alleles has been used to study local genes flow and population sub-structure (Gottelli *et al*., 1994; Allen *et al.*, 1995). Effective population sizes and inbreeding have been estimated from microsatellite data (Edwards *et al*., 1992). Microsatellites have been increasingly used for the study of genetic variation between and within animal populations. They have been successfully applied in the study of genetic variation in vertebrates (Kashi *et al.*, 1990a, b) and more particularly in livestock species such as poultry, sheep, goats, buffaloes and bovines (Haberfeld *et al*., 1991; Ganai and Yadav, 2001, Arora R *et al.*, 2004, Li and Valentini 2004). Microsatellite DNA variation has been employed to assess genetic distances between strains of poultry (Kuhnlein *et al.*, 1989) and within and among European and African cattle (MacHugh *et al.*, 1994; MacHugh *et al*., 1997; Okomo *et al.*, 1998) and sheep breeds (Buchanan *et al.*, 1994; Arranz *et al.*, 1998). Microsatellite markers are well suited for forensic applications; population genetics studies etc. and consequently, have become the basis of genomic analysis in rodents, pigs, dogs, cattle and goats (Barendse *et al.*, 1994; Ellegren, 1993; Roy *et al*., 1994; Serikawa *et al.*, 1992; Ganai and Yadav, 2001).

#### 2.5.3.2 Estimation of genetic distances from microsatellite data

One of the most important problems in the conservation of genetic diversity in domestic animals is how to choose appropriate breed or population for conservation, among the many still available in role to maximize genetic diversity conserved. One

way could be to compute the genetic distance between pairs of breeds, and to choose breeds that show the widest genetic differences among them (Barker, 1994). Genetic distances are used either for estimating divergent time or for construction of phylogenies, which in turn can be used to decide which breeds should be kept (Nei and Takezaki, 1994). Many genetic distances have been developed, of which a few remain in regular use (Nei, 1987). Each of these genetic distances has unique evolutionary and statistical properties, and evolutionary relationships inferred from each genetic distance can be quite different. The merits and demerits of each measure have been investigated (e.g. Takezaki and Nei, 1996).

Many of the earlier measures of genetic distances are geometric distances, which are based explicitly on geometric representations of gene frequencies rather than on any particular population genetic models. These distance measures obey the axioms of Euclidean geometry. The most commonly used of geometric distance is Cavalli-Sforza and Edwards's (1967) chord distance ($D_C$). In this distance measure, populations are represented as points on a multidimensional hypersphere with dimension equal to the number of alleles. $D_C$ is the chord distance between two points representing the gene frequencies in two different populations. This distance incorporates an angular transformation of gene frequencies and the principle of triangle inequality is fulfilled. The transformation has the effect of standardizing the distance with respect to random drift, so that the rate of increase in genetic distance under drift is nearly independent of the initial gene frequencies. However, little attention is paid to the relationship between genetic distance and evolutionary change of populations. The problem of geometric distance is that the absolute values of these measures do not have any particular biological meaning (Nei *et al.*, 1983), and only the relative values are important for finding the genetic relationship among populations.

Another widely used measure of genetic distance, which is close to $D_{C,}$ is the Nei *et al.*'s (1983) angular distance ($D_A$). This is a modified Cavalli-Sforza and Edwards distance measure. It is based on genetic drift model and is more efficient in determining the true topology of phylogenetic trees especially for closely related populations. It is the most widely used distance measure since it reportedly increase more slowly with time and maintains a linear relationship for longer periods of time (Nei, 1984).

The most commonly used genetic distances are based on population genetic models and are supposed to be linearly proportional to time: either to the average time

to a common ancestor (coalescence time) of pairs of genes or to the time since two populations diverged from one another. There are two most important genetic models, which have been proposed to explain the mode of mutation as cause of evolutionary divergence between copies of homologous genes since the time of their common ancestor. The models are infinite alleles model (IAM) and stepwise mutational model (SMM). The IAM model assumes that every new mutation gives rise to an allele that does not already exist in the population while the SMM assumes that mutations increase or decrease allele size by single unit. IAM seems to apply to classical genetic markers such as protein and blood polymorphisms (Nei, 1987) and SMM apply to microsatellite loci (Goldstein *et al.*, 1995*a*) since alleles at microsatellites are thought to evolve by a stepwise mutation process. However, Edwards *et al*. (1992) and Valdes *et al.* (1993) have shown that the pattern of mutation for microsatellite loci follow a kind of a stepwise mutation model which is close to the IAM of mutation when a relatively short evolutionary time is considered. Thus microsatellite data can be analyzed in a manner similar to protein polymorphisms. The most widely used distance measure based on IAM is Nei (1972) standard genetic distance ($D_S$). $D_S$ is intended to measure the number of codon substitutions per locus that have occurred after divergence between a pair of populations and is expected to increase linearly with time (Nei and Takezaki, 1994). $D_S$ assumes that the rate of gene substitution per locus is uniform across loci and lineages. This assumption is violated under microsatellite loci. According to Farris (1981) $D_S$ is not appropriate for making a phylogenetic tree because it is not a metric measurement and does not obey the triangle inequalities. Another widely used distance measure under IAM is Rogers' (1972) distance ($D_R$). This measure has the virtues of simplicity and it satisfies the principle of triangle inequality. However, $D_S$ is better than $D_R$ with respect to the linear relationship with time. Both $D_S$ and $D_R$ have the undesirable property of being influenced by within-taxon heterozygosity. That is the distance between two taxa that are fixed for alternate alleles exceeds that between two taxa in which one or both are heteroallelic but have no alleles in common.

Although a large number of genetic distances could conceivably be applied to microsatellite, estimators of population parameters based on the IAM are unlikely to apply to microsatellite (Goldstein *et al.*, 1995*a*). This is because the majority of mutations at microsatellite loci are stepwise in nature, changing allelic sizes up or down by one or very few number of repeats. Thus, distance measures, which apply to microsatellites generally, assume the SMM model. Three distances have been

developed specifically for application to microsatellite evolution assuming the SMM model: Goldstein *et al*.'s (1995*a*) genetic distance $(\delta\mu)^2$, Goldstein *et al*.'s (1995b) and Slatkin (1995) average square distance (ASD) and Shriver *et al.* (1995) stepwise distance (DSW). The expected values of $(\delta\mu)^2$ and ASD increase linearly with time under the SMM model, but DSW seems to be non-linear with time. Although these distances apply to microsatellite, their efficiencies for obtaining the correct topology of the phylogenetic tree are generally low (Takezaki and Nei, 1996). This is because ASD, $(\delta\mu)^2$ and DSW have large sampling errors due to their dependence on the variation within population and on the sensitivity to fluctuation in effective population size (Takezaki and Nei, 1996).

Other widely used measures of genetic distances are various estimators of population subdivision ($F_{ST}$) such as $G_{ST}$ (Nei, 1973; Weir and Cockerham, 1984) and $R_{ST}$ (Slatkin, 1995).

Different measures of genetic distances are appropriate for different purposes. Nei and Takezaki (1994) and Forbes *et al.* (1995) have reported that classical distances based on the IAM model and the distances based upon multidimensional geometric considerations are more effective in obtaining the correct tree topology than the SMM-based genetic distances. Nei and Takezaki (1994) and Takezaki and Nei (1996) have shown that $D_C$ and $D_A$ are more efficient in obtaining the correct tree topology than other distances whereas $D_S$ and $(\delta\mu)^2$ are more appropriate for estimating evolutionary time than other distances.

#### 2.5.3.3 Microsatellite genetic distances in Livestock

Moazami-Goudarzi *et al*. (1997) used 17 microsatellite loci and 13 biochemical markers for studying the genetic variability between 10 cattle breeds. They used Nei's (1972) genetic distance and chord distance ($D_C$) (Cavalli-Sforza and Edwards, 1967) for studying the genetic relationships between breeds. Yadav *et al*. (1998) reported the use of 22 bovine microsatellite primer pairs in 36 animals of Murrah breed of buffalo. All the 22 primer pairs amplified a product, and 14 of which were polymorphic. Pandey (1999) also reported the conservation of cattle microsatellites in two breeds of Indian buffalo.

Loftus *et al.* (1999) carried out a microsatellite survey on 8 humpless cattle breeds from the Near East, 3 from Europe, 1 from West Africa and 2 zebu breeds from India with 20 microsatellite loci. They reported a large genetic divergence between the non-hybrid taurine and zebu groups. Mommens *et al*. (1999) studied variation between an African and five European taurine using data on 31 microsatellites loci. They calculated genetic distance and constructed Phylogenetic tree using neighbor joining (NJ) method. They reported American bison as an out-group. Wilson and Strobeck (1999) have shown genetic relatedness among wood and plains bison populations by using 11 microsatellite loci. Yang *et al.* (1999) studied the genetic relationship among five indigenous Chinese goat breeds with six microsatellite markers. A neighbour-joining tree was constructed using Nei's standard genetic distance. In this tree, Neimonggol and Liaoning were grouped together, then with Taihang; while Tibetan and Matou each had their own branch.

Hanslik *et al.* (2000) studied the introgression of new world Holstein-Friesian into European Holstein-Friesian populations. A tree of individuals based on the proportion of shared alleles was produced; a clear distinction between old world and new world Holstein-Friesian populations was found. Ritz *et al*. (2000) studied the phylogenetic analysis of the tribe bovine using 20 microsatellite loci. They inferred that the genetic distances used for the study: $(\delta\mu)^2$ and $D_C$ are both considered being appropriate for phylogenetic studies among species and subgenus. They concluded that *Bos*, *Poephagus* and *Bibos* each emerged as a subgenus within the genus *Bos*. On the other hand, *Bison*, *Bubalus* and *Syncerus* each formed a separate discrete genus. A suitable large panel of polymorphic markers for each species, especially those that are less common and less studied. Cross-species utilization of microsatellite loci not only save time and efforts in the laboratory, but also enable the construction of comparative maps between related species.

MacHugh *et al*. (1998) studied genetic demarcation among the breed of seven European cattle using 20 microsatellite markers. They reported correct breed designation with accuracy approaching 100% using data from a panel of 10 microsatellite loci. Martin *et al*. (1999) studied genetic diversity in six Spanish native cattle breeds using 30 microsatellites markers. They calculated allele frequencies and genetic distance for the characterization of breeds and construction of dendrogram respectively. They found closest population those representing Austrian breeds, the most divergent being Menorquina and fighting bull.

## 2.6 Methods for the reconstruction of phylogenetic tree

Phylogenetic trees are graphical representations of nodes (taxonomic units) and branches (pathways connecting nodes) that summarize the evolutionary relationships among organisms (Avise, 1994). Reconstruction of a phylogenic tree is considered as statistical inference of a true phylogeny, which is unknown (Nei, 1996). A large number of methods have been described for phylogenetic tree building. In principle there are two ways of building trees, either by clustering methods or by exhaustive search methods. These methods have the advantage of being ease to perform, resulting in very fast computer programs (Saitou and Imanishi, 1989). The clustering methods follow a set of steps (an algorithm) and leading to a tree. The biggest limitation is that cluster methods do not provide a ranking criterion for evaluating trees. Exhaustive search methods use optimality criteria for choosing the best tree among the set of all possible trees. This criterion is used to assign to each tree a score or a rank, which is a function of the relationship between the tree and data. Exhaustive Search methods have the advantage that they provide a ranking criterion under which all trees can be evaluated (Page and Holmes, 1998). However, they are computationally demanding, resulting in computer programs, which are much slower to perform and cannot give exact results when there are more than 12 taxa (Nei, 1996).

The methods of phylogenetic inference currently used in molecular phylogenetic can be classified in to three major groups: distance methods, parsimony methods and likelihood methods.

### 2.6.1 Distance methods

In distance methods, an evolutionary distance is computed for all pairs of populations and a phylogenetic tree is constructed from the pairwise distances. The most commonly used methods are the ‘unweighted pair group method with arithmetic averages’ (UPGMA) of Sneath and Sokal (1973) and the Neighbour-joining (NJ) of Saitou and Nei (1987).

The UPGMA has gained popularity mostly because of its simplicity and because of its speed (though many other distance methods are as fast). The UPGMA uses real (uncorrected) distance values and a sequential clustering algorithm. This method of tree construction is not sensitive to differences in branch length or unequal rates of evolution. The UPGMA is a clustering method, which operates as follows: A distance

matrix is scanned for the smallest distance element and the corresponding operation taxonomic units (OTUs) are joined. This distance element in the matrix is then discarded. The matrix is again scanned for the smallest remaining distance element and the process is repeated until all OTUs are finished. In each cycle of the clustering procedure, OTUs or previously formed clusters are grouped according to the smallest mean distance between the populations involved. Each OTU contributes equally to the calculation of these mean distances. UPGMA produces a dendrogram, which is rooted implicitly at the point where the last clusters join. The major assumption of UPGMA clustering is the equal rate of evolution along all dendrogram branches (Nei, 1987). UPGMA is considered as a heuristic method for finding the least square ultra metric tree for a distance matrix. UPGMA method performs well in recovering correct tree topologies (Tateno *et al*., 1982; Nei *et al*., 1983). This is due to the fact that genetic distance estimates are subject to large stochastic error and the distance averaging aspects of UPGMA reduces the effects of this error (Nei, 1987).

The NJ is a clustering method, which attempts to correct the UPGMA method for its strong assumption (frequently invalid) that the same rate of evolution applies to each branch. Hence this method assumes unequal rates of molecular change among the branches (Nei, 1987) and yields an unrooted tree. In the NJ method the raw data are provided as a distance matrix and the construction of a tree begins by assuming that the populations are all related to one another by a star phylogeny, thus the initial tree is a star tree. The next step is to begin a procedure that groups certain populations together. In order to do this, first, branch lengths for tree are estimated by least square and the sum of the branch lengths for the entire tree is calculated. Then pairs of populations that give the minimum sum of branch lengths are considered as neighbours and are joined and considered as single entity. At each step of the analysis a transformed distance matrix is constructed in which the separation between each pair of nodes is adjusted on the basis of their average divergence from all other nodes. The tree is constructed by linking the least distance pair of nodes as defined by the modified matrix. When two nodes are linked, their common ancestral node is added to the tree and the terminal nodes with their respective branches are removed from the tree. This pruning process converts the newly added common ancestor into terminal node on a tree of a reduced size. The process of considering all possible pairs is repeated and completed when two nodes remain, separated by a single branch. According to Nei (1996), NJ is a heuristic method for estimating the minimum evolution tree. NJ

produces a reasonably good phylogenetic tree when the extent of population differences is large and a large number of loci are examined (Rzhetsky and Nei, 1992). The efficiency of the NJ method in obtaining the true tree is due to the fact that in each step of population clustering the principle of minimum evolution is applied, and according to Nei (1996) the repeated application of this principle reduces the effects of sampling errors in topology construction.

### 2.6.2 Maximum parsimony method

Parsimony implies that simpler hypotheses are preferable to more complicated ones. Maximum parsimony (MP) is a character-based method that infers a phylogenetic tree by minimizing the total number of evolutionary steps required to explain a given set of data, or in other words by minimizing the total tree length. Maximum Parsimony methods fall under optimality criteria methods. Under MP the best tree is the one that requires the smallest number of evolutionary changes to explain the observed differences among OTUs. A number of related methods have been described for inferring phylogenies by parsimony approach, the most common employed are Fitch and Wagner parsimony (Fitch, 1971; Farris, 1970), Camin-Sokal parsimony (Farris, 1971), Dollo parsimony (Farris, 1981) and generalized parsimony (Swofford and Olsen, 1990). These methods differ in their underlying evolutionary assumptions but they are united by the goal to reconstruct the evolution of OTUs on a tree based on the constraint of invoking the fewest possible evolutionary changes (Page and Holmes, 1998). Among the advantages of parsimony methods are that the approach apparently makes few assumptions about the evolutionary process and it is based on implicit assumption about evolution that evolutionary changes are rare (Swofford *et al*., 1996). The rarity of evolutionary changes implies that the tree that minimizes the change is likely to be the best estimate of the actual phylogeny (Pages and Holmes, 1998). According to Swofford *et al*. (1996), if there are no multiple nucleotide substitutions at each site, MP is expected to generate the correct topology as long as enough parsimony-informative sites are examined. The main disadvantage of MP is that under some models of evolution it is not consistent. Indeed, when the true tree has a special type of topology and branch lengths, MP may generate an incorrect topology even if an infinite number of loci are examined (Nei, 1996). Takezaki and Nei (1996) have shown that this can happen even if the rate of nucleotide substitution is

constant for all evolutionary lineages. Another disadvantage of MP is that in parsimony analysis it is difficult to treat the phylogenetic inference in a statistical framework because there is no natural way to compute the means and variances of minimum number of substitutions obtained by the parsimony procedure. However, under certain circumstances, MP is quite efficient in obtaining the correct topology and it is the only method that can take care of insertions and deletions of nucleotides, which may give important phylogenetic information (Nei, 1996).

### 2.6.3 Maximum likelihood

Maximum Likelihood (ML) evaluates a hypothesis about evolutionary history in terms of the probability that the proposed model and the hypothesized history would give rise to the observed data set. The supposition is that a history with a higher probability of reaching the observed state is preferred to a history with a lower probability. The method searches for the tree with the highest probability or likelihood. The data are observed gene frequencies or nucleotide sequences; the unknowns are the branching order and branch lengths of the tree (Felsenstein, 1981). In ML approach, a model of the evolutionary process that accounts for the conversion of one sequence or allele into another must first be specified. The next step is to calculate the probability that the observed data would have been generated by a particular tree topology under that model. The tree that makes the observed data the most probable evolutionary outcome (i.e. the tree that gives the highest likelihood) is chosen as the ML estimate of the phylogeny. Maximum Likelihood is an appealing method of inference as it can incorporate explicit models of evolution and also allows statistical test of evolutionary hypothesis. Hasegawa *et al.,* 1991 reported that the maximum likelihood method is superior to distance methods in the efficiency to produce the correct true topology particularly when the evolutionary rate differs among lineages. There advantages of maximum likelihood methods over other methods are: (i) they have often lower variance than other methods (i.e. it is frequently the estimation method least affected by sampling error); (ii) they tend to be robust to many violations of the assumptions in the evolutionary model; (iii) even with very short sequences they tend to outperform alternative methods such as parsimony or distance methods; (iv) the method is statistically well founded; (v)they evaluate different tree topologies, and they use all the sequence information. The objection to likelihood methods is that the dependence on a

model raises the question of which model to use as the pattern of nucleotide substitution varies from site to site and with evolutionary time (Tateno *et al*., 1982). The biggest problem with ML is a technical difficulty for computing the likelihood itself. This is computationally time-consuming, and recently it has been shown that more than one maximal likelihood values may exist for a given tree (Nei, 1996); making it difficult to guarantee that the likelihood value for that tree is actually maximal.

## 2.7. Statistical analysis

### 2.7.1. Hardy-Weinberg equilibrium (HWE)

In a large random mating population with no selection, mutation or migration, the allelic or gene frequencies and the genotype frequencies remain constant from one generation to the next. Thus a population with constant gene and genotype frequencies is said to be in Hardy-Weinberg equilibrium (HWE) (Falconer and Mackay, 1996). Factors that affect the HWE include:

*Non-random mating*: Non-random mating can occur in cases where related individuals have a greater probability of mating with each other than with other members of the population and where individuals that are geographically close are more likely to mate with each other than those that are not geographically close.

*Mutation*: This is the process that produces a gene or chromosome that is different from the wild type. A mutation occurs when a DNA gene is damaged or changed in such a way as to alter the genetic message carried by that gene. Mutation brings about genetic variation in a population by producing novel variants of genes.

*Migration*: The movements of genes caused by individuals moving, including new individuals entering (immigration) or leaving (emmigration) a population, introducing or removing genetic material and thereby changing allele frequencies. http://www.modernhumanorigins.com/m.thml). Migration of genes into a population results in an increase in a population's genetic variation and the migration of genes out of a population may result in a reduction of the genetic variation.

*Selection*: A natural process resulting in the evolution of organisms' best adapted to the environment (http://www.cogsci.princeton.edu/cgi-bin/webwn?stage=1&word =selection). Only the individuals that are better adapted to the environment or

able to mate successfully could pass their genes on to the next generation. Selection generally results in a reduction of genetic variation in a population.

*Random genetic drift* i.e. the random sampling of alleles at every generation: Random genetic drift is a stochastic process (by definition). Random genetic drift is the change in gene frequencies due to chance or sampling effects. It is very dependent upon population size. Genetic drift is fundamentally the result of a finite population size. The effect of genetic drift is infinitely proportional to the population size. Thus when the population size is small, e.g. due to strong bottleneck effects in the past, there are greater changes in gene frequency under genetic drift at every generation (Cavalli-Sforza *et al*., 1994). Therefore, the smaller the population size, the greater the chances of sampling errors occurring.

In the study of many populations, it is important to determine whether the loci and the populations genotyped were in HWE and whether there were any significant deviations from the HWE.

### 2.7.2. Linkage disequilibrium (LD)

Linkage disequilibrium is often termed ‘allelic association’ and is a measure of the degree of association between two alleles in a population. Measures of linkage disequilibrium (such as D) quantify how frequently two alleles are found on the same chromosome in a certain population (http://dblab8.csie.ncu.edu.tw/Glossary.htm). A population is said to be at linkage disequilibrium at a set of loci if the alleles are not randomly assorted in the next generation, but are inherited together as a unit. Linkage disequilibrium can be generated by finite population size, random genetic drift, mutation, selection, non-random mating, and migration (http://darwin.cwru.edu/~conti/notes/linkagedisequilibrium.htm). It can have positive or negative values, the tendency for two 'alleles' to be present on the same chromosome (positive LD) or not to segregate together (negative LD). LD is decreased by recombination. Thus, it decreases every generation of random mating unless there are some processes opposing the approach to linkage 'equilibrium'.

### 2.7.3. Genetic diversity

Genetic diversity refers to the variation at the level of individual genes (polymorphism), and provides a mechanism for populations to adapt to their ever-changing environment. The more variation the better the chance that at least some of the individuals will have an allelic variant that is suited for the new environment, and will produce offspring with the variant that will in turn reproduce and continue the population into subsequent generations (http://genetics.nbii.gov/GeneticDiversity.html).

The mean number of alleles (MNA) detected in each population and the expected heterozygosities are good indicators of the genetic polymorphism within the populations. The MNA is the average number of alleles observed in a population, while the expected heterozygosities are the proportion of heterozygote expected in a population. Numbers of alleles per locus per population are obtained by direct counting. Generally, the MNA is dependent on the sample size because of the presence of unique alleles that occur in low frequencies in populations and also because the number of observed alleles tend to increase with increases with population size (Nei, 1987). Therefore the comparison of the MNA between samples of different sizes may not be meaningful unless the sample sizes are more or less the same (Nei, 1987; Hart and Clark, 1989).

### 2.7.4. Genetic relationship between the populations

#### 2.7.4.1. Genetic distance

Genetic distance is a measurement of genetic relatedness of samples of populations (whereas genetic diversity represents diversity within a population). The estimate is based on the number of allelic substitutions per locus that have occurred during the separate evolution of two populations. This difference measured between two populations provides a good estimate of how divergent they are genetically (Avise, 1994). Thus when the genetic distance is large, the genetic similarity is low and the time they diverged from each other is greater, while when genetic distance is small the genetic similarity is high and the time they diverged from each other is smaller (Avise, 1994). Cavalli-Sforza and Edwards (1964; 1967) were the first to use gene frequency data to classify populations. They used genetic distance data to construct a phylogenetic tree to infer the history of various human populations.

One of the common measures of genetic distance in use today is Nei's standard genetic distance ($D_S$) (Nei, 1972, 1978). The $D_S$ distance is based on the infinite alleles model (IAM) and is said to be more appropriate for estimating evolutionary time. $D_S$ increases linearly with time from zero to infinity and will have a value proportional to the mutation rate (Kalinowski, 2002). However, Nei *et al*. (1983) noted that the modified Cavalli-Sforza and Edwards distance measure ($D_A$) usually known, as angular distance is more efficient in recovering the true topology of an evolutionary tree of closely related populations. The $D_A$ distance has proven to be useful for reconstructing phylogenies (Takezaki and Nei, 1996). $D_A$ has also been reported to increase more slowly with time and to maintain a linear relationship for longer periods of time (Nei *et al*., 1983).

**2.7.4.2. Construction of phylogeny**

The neighbor-joining (NJ) methodology (Saitou and Nei, 1987) implemented in the DISPAN program was used to construct the phylogenetic tree of population relationships. The NJ is known to be more efficient than many other methods in obtaining the correct tree (Rzhetsky and Nei, 1992). The neighbor-joining method does not assume an equal rate of evolution between lineages unlike the unweighted pair group-method with arithmetic mean (UPGMA), another method commonly used to construct phylogenetic trees (Nei, Tajima, and Tateno, 1983).

The DISPAN program (Ota, 1993) was used to generate pair-wise lower triangle matrices of genetic distances between the populations. A neighbor-joining and UPGMA trees were then constructed from these matrices using $D_S$ and $D_A$ (Ota, 1993). The reliability of the constructed trees was examined by a bootstrap test with 1000 replicates re-sampling of loci with replacement (Felsenstein, 1985).

**2.7.4.3. Principal component analysis (PCA)**

Construction of phylogeny and calculation of genetic distances have the disadvantage of not being able to detect the effects of past bottleneck effects on populations and any admixture that has taken place between populations (Cavalli-Sforza *et al*., 1994). An alternative method taking into account the admixture is

multivariate analysis e.g. principal component analysis. PCA can be used to reduce a large number of variables to a smaller number of variables without losing too much information (Alt, 1990; Cavalli-Sforza *et al.*, 1994). PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components*. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. http://www.fon.hum.uva.nl/praat/manual/Principal_component_analysis.html.

PCA simply describes the variation of the multivariate data in terms of some linear combinations of the variables, in this case the observed microsatellite allele frequencies, which are resolved into smaller reference vectors. The new variables are uncorrelated with each other. The coefficients defining the linear transformation are chosen so as to minimize the variation of the transformed data measure along each new co-ordinate axis. PCA was performed using the MVSP program (version 3.12d) from (http://www.kovcomp.com/mvsp/downl.html).

#### 2.7.4.4. Genetic differentiation of the populations

The populations of most, if not all, species show some levels of genetic structure, which may be due to a variety of non-mutually exclusive agents. Environmental barriers, historical processes and life histories (e.g. mating system) may all, to some extent, shape the genetic structure of populations (e.g. Donnelly and Townson, 2000; Gerlach and Musolf, 2000; Palsson, 2000; Tiedemann *et al.*, 2000). The understanding of the genetic structure of a population or differentiation between populations is of interest to population geneticists because it reflects the number of alleles exchanged between populations which influence the genetic composition of individuals within these populations (Balloux and Lugon-Moulin, 2002). Gene flow between populations determines the effects of selection and genetic drift, generates new polymorphisms and increases the local effective population size (Balloux and Lugon-Moulin, 2002). The simplest parameters for assessing diversity between breeds using microsatellite data are the genetic differentiation or fixation indices. Several estimators have been proposed (e.g. $F_{ST}$, $G_{ST,}$ $\theta$), the most widely used being $F_{ST}$ (Weir and Basten, 1990), which measures the degree of genetic differentiation among

subpopulations through the calculation of the standardized variances in allele frequencies among populations.

#### 2.7.4.4.1. Standardized variance in allele frequencies among populations ($F_{ST}$)

$F_{ST}$ measures the effect of population subdivision, which is the reduction in heterozygosity in a subpopulation due to genetic drift. $F_{ST}$ is the most inclusive measure of population substructure and is most useful for examining the overall genetic divergence among subpopulations. It is also called co-ancestry coefficient ($\theta$) [Weir & Cockerham, 1984] or 'fixation index' and is defined as correlation of gametes within subpopulations relative to gametes drawn at random from the entire population (Subpopulation within the Total population). It is calculated using the subpopulation (average) heterozygosity and total population expected heterozygosity. $F_{ST}$ is always positive; it ranges between 0 = panmixis (no subdivision, random mating occurring, no genetic divergence within the population) and 1 = complete isolation (extreme subdivision). $F_{ST}$ values below 0.05 indicate negligible genetic differentiation whereas > 0.25 means very great genetic differentiation within the population analysed. $F_{ST}$ is usually calculated for different genes, and then averaged across all loci, and all populations ($F_{ST} = (H_T - H_S)/H_S$). $F_{ST}$ can also be used to estimate gene flow: $Nm = 0.25\ (1-F_{ST})/F_{ST}$. This highly versatile parameter is even used as a genetic distance measure between two populations instead of a fixation index among many populations (see Weir, Genetic Data Analysis II, 1996, and Kalinowski, 2002).

$F_{ST}$ estimates were calculated using the program FSTAT 2.9.3 (Goudet, 2001; http://www.unil.ch/izea/software/fstat.html) and GENEPOP version 3.4 (Raymond and Rousset, 1995).

#### 2.7.4.4.2 The coefficient of gene differentiation ($G_{ST}$)

The coefficient of gene differentiation ($G_{ST}$) developed by Nei (1973) is an extension of the (Nei, 1972) genetic distance theory between a pair of populations. $G_{ST}$ can be computed directly from allele frequencies in terms of expected heterozygosities within and between populations. Unlike the $F_{ST}$, the estimation of heterozygosities in $G_{ST}$ relies only on allele frequencies (Nei, 1987). This method offers several advantages because it is not affected by the number of alleles at the locus and neither is affected by the evolutionary forces such as mutation, selection and migration, which may be taking place in the organism. $G_{ST}$ can be defined as:

$G_{ST} = D_{ST}/H_T$

Where $D_{ST}$ is average gene diversity between and within populations and $H_T$ is the expected total heterozygosity. The $G_{ST}$ estimates in this study were calculated using the DISPAN program (Ota, 1993).

#### 2.7.4.5 Analysis of molecular variance (AMOVA)

Population differentiation and the distribution of genetic variation can also be determined by AMOVA (Excoffier *et al*., 1992). AMOVA is built on the classical analysis of variance. However, the data analyzed with AMOVA uses allele frequency data. The AMOVA analysis was done using the software Arlequin version 2.000 from (http://lgb.unige.ch/arlequin/).

#### 2.7.4.6 Assignment of individuals to populations

In population genetics, individuals in a sample have to be classified into specific populations or breeds using a set of phenotypic criteria. Population assignment methods assign individuals to populations using a clustering method. Here individuals that are genetically similar are assigned to the same cluster. There are broadly two types of clustering methods that could be used; distance based and model based methods.

Distance based methods first calculate a pair-wise distance matrix (for e.g. Nei's distance or Cavalli-Sforza and Edwards Chord distance), then use them to assign the individual to the closest population. These methods are very easy to apply but have the disadvantage of being too dependent on the distance measure and it is difficult to asses the confidence level that should be assigned to clusters obtained (Pritchard *et al*., 2000).

Model based methods include the maximum likelihood based methods, which include the frequency and the Bayesian methods. The frequency method, which was first described by Paetkau *et al*. (1995a, b) first computes the allele frequencies in the populations, then computes the likelihood of the individual's genotype occurring in each population and then assigns an individual to the population in which the individual's genotype is most likely to occur. This method assumes that the populations are in HWE and there is no linkage disequilibrium. However, the major drawback of this method is that when there are null or rare alleles in a population, the corresponding

allele frequency is zero and hence a likelihood of zero leading to the elimination of the population from the analysis.

The Bayesian method (Rannala and Mountain, 1997) is similar to the frequency method; here the sample population frequencies are used to derive the probability density of the population allele frequencies. In this method, the problem of dealing with null allele is eliminated.

Individual population assignments were done using two programs WHICHRUN 4.1.0 (Banks and Eichert, 2000) and GeneClass (Cornuet *et al*., 1999). Using the maximum likelihood approach that incorporated jackknife iterations, the WHICHRUN 4.1.0 program used allele frequency distributions of the multilocus genotypic data to assign individuals to a population using the method of Paetkau *et al*. (1995a, b). Here individuals are sampled one at a time and the allele frequency recalculated in the absence of each genotype before the individual is assigned to the most likely source population. The program then computes a ratio between the likelihood of the most likely populations and the second most likely population. The allocation stringency was of a log of odds (LOD) ratio greater than two or $P < 0.001$.

Population assignments were also done using the GeneClass program, which allows for different estimation procedures. The probability that the individual belongs to a population (*P*-value) was calculated by simulating 10,000 genotypes using Monte Carlo sampling of allele frequencies observed in a population. The observed frequency of the tested individual's genotype was computed in the population under investigation. An individual was considered correctly assigned to a population when it was excluded from all other populations with a high statistical significance of $P < 0.001$.

### 2 7.4.7 Inference of population structure

Admixture analysis at both the population and individual level and individual specific assignment to the populations was performed using the program STRUCTURE version 2.0 (Pritchard *et al.*, 2000), which implements the Pritchard, Stephens and Donnelly (2000) method, a model based clustering algorithm that infers population structure using multilocus genotypes. This Bayesian method assumes Hardy-Weinberg as well as linkage disequilibrium between loci within each population. The method places individuals in K populations or clusters and individuals can have membership in

multiple clusters, if their genotypes indicate that they are admixed. These membership coefficients add up to one across the various clusters for each individual or population. Five runs were made at each four Ks, where K is the number of inferred clusters and a burn-in period of 100,000 steps as well as a Markov Chain Monte Carlo (MCMC) of 500,000 being applied. The program DISTRUCT offers various options for displaying these results, where an individual's estimated membership coefficient are displayed as various segments partitioned into different colors, depending on the value of K chosen.

# CHAPTER THREE

## 3.0. Materials and methods

### 3.1. Populations/Breeds and sampling strategy

#### 3.1.1. Animals

The aim of this study was to quantify genetic diversity within and among Ethiopian goat populations/types and to clarify the genetic relationships between them. For this purpose, 11 indigenous goat populations/types of Ethiopia and nine reference breeds (four from other African countries viz. two from Kenya, one from Guinea Bissau and one from Botswana), and five from non-African countries (one breed from Turkey, Egypt, Italy, Mongolia and Saudi Arabia each) were included in this study. The sampled animals of Ethiopian goats were selected to represent the type/population as much as possible based on the available information of phenotypic characteristics and geographical distribution from previous survey (physical description and management systems) by Farm-Africa (FARM – Africa, 1996) (**Fig. 19a**). The finally identified animals for the study were sampled from distantly located villages and sampling was made 1-5 animals per flock. In order to avoid sampling of related individuals, farmers were asked about the source and familial relationships of the individuals. For all populations/types/breeds, approximately equal number of females and males were sampled. The breed, number of animals per breed, country of origin, and phenotypic characters/subtypes are shown in **Table 2**. For this study, seven Ethiopian goat types (Woyto-Guji, Long-Ear Somali, Hararghe Highland, Central Highland, Keffa, Short-Ear Somali and Abergalle) were sampled fresh from the field, while for the project also used DNA or blood samples or FTA® cards already available at ILRI for remaining four populations of Ethiopia and all the reference breeds.

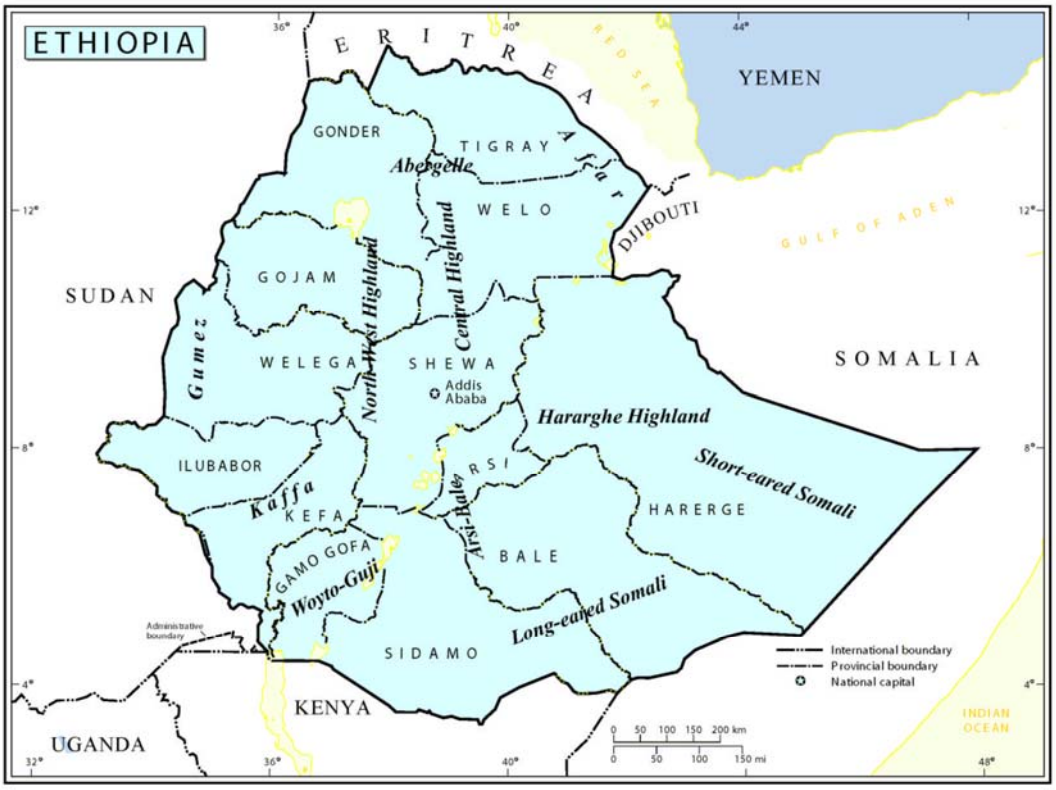


Figure 19a. **Map of Ethiopia showing distribution of goat populations/types sampled.**

### 3.1.2. Sampling protocol

Blood was collected from jugular vein by puncher using 10 ml capacity vacutainer tubes containing EDTA (100 µl neat). Every time a new needle was used for each animal to avoid cross contamination between samples. The blood was mixed with EDTA by gently inverting the tube about three to five times. Blood collected at the field site was brought to laboratory with all necessary precautions to avoid exposure to extreme temperature (hot or cold) and direct sunlight. In the laboratory blood samples were centrifuged at 2000 rpm for 15 – 20 minutes the same day evening. The plasma was discarded and the buffy-coat containing the peripheral blood lymphocytes (PBL) was transferred with a wide-mouth Pasteur pipette into a 2 ml cryotube. Then 1 ml of 8 M urea was added into the tube, the content of the tube mixed by gentle inversions five times. The tube properly labeled and stored in upright position at room temperature, ready for DNA extraction.

In the case of sampling using the FTA cards, the blood in the tube was placed on ice and transferred onto FTA classic cards (Whatman® Bioscience) on the same day. After drying, these cards were transferred into bags containing silica gel for storage and future use.

Table 2 **Populations/types used in the analysis of genetic diversity of Ethiopian goats and the reference Breeds/populations from outside Ethiopia.**

| Population/Breed | Number genotyped | Origin | Sub-type |
|---|---|---|---|
| Abergale | 46 | Ethiopia | Afar and Worre type |
| Arsi-Bale | 46 | Ethiopia | The Woyto-Guji goat is the closest relative |
| Afar | 44 | Ethiopia | Worre and Abergalle types |
| Central Highland | 40 | Ethiopia | Western Highland and Keffa goat types |
| Western Lowland (Gumez) | 43 | Ethiopia | Most closely related to the Central and Western Highland goats |
| Hararghe Highland | 40 | Ethiopia | Related to both the Short- and Long-eared Somali |
| Keffa | 38 | Ethiopia | Western Highland goat |
| Long-eared Somali | 48 | Ethiopia | Short-eared Somali |
| North-West Highland | 45 | Ethiopia | Central Highland and Keffa types |
| Short-eared Somali | 43 | Ethiopia | Long-eared Somali |
| Woyto-Guji | 46 | Ethiopia | Related to the Arsi-Bale |
| Boran Galla | 35 | Kenya | Short Ear Short Horn |
| Small East Africa | 39 | Kenya | Short Ear Short Horn |
| Guinea Bissau | 46 | Guinea Bissau | |
| Tswana | 29 | Botswana | Short Ear Short Horn |
| Alpine | 31 | Italy | |
| Ardi | 25 | Saudi Arabia | |
| Baladie | 22 | Egypt | |
| Hair | 31 | Turkey | |
| Zalaajinistiin | 39 | Mongolia | |

### 3.2. DNA extraction

DNA was extracted either from peripheral blood lymphocytes (PBL) or FTA® blood cards (Whatman® Bioscience). The extraction procedures applied are described in the following section.

#### 3.2.1. DNA extraction from peripheral blood lymphocytes (PBL)

Buffy-coat of each sample (200 µl) was put into 1.5 µl-eppendorf tube and 500 – 700 µl of digestion buffer was added and mixed by vortexing. Then 40 µl of proteinase-K (20mg/µl) + 100 µl of 10% SDS (Sodium Dodecyl Sulphate) solution was added and mixed and incubated for 3 hours at 50$^{o}$C. Then equal volume of a mixture of phenol, chloroform, and Isoamylalcohol (25:24:1) was added, mixed by vortexing and centrifuged at rpm (13,000 to 14,000 rpm) for 10 minutes and the supernatant transferred to a new eppendorf tube. To the supernatant, one volume of chloroform was added, mixed for 5 minutes by vortexing and centrifuged for 10 minutes at 13,000 to 14,000 rpm and the supernatant transferred to a new eppendorf tube. To the supernatant, two volumes of absolute ethanol were added to precipitate DNA and kept for 10 minutes and spin/centrifuged at 13,000 to 14,000 rpm for 3 minutes and the supernatant was discarded. Then the pellet was washed twice in 70% ethanol and air-dried or oven-dried at 36.8$^{o}$C for an hour and dissolved in 200 µl of sterile triple distilled water. The concentration of DNA was determined by spectrophotometer and then the DNA concentration was adjusted to 20 ng/µl by adding required amount of water.

#### 3.2.2. DNA extraction from FTA® blood cards (Whatman® Bioscience)

Around thirty disks (1 mm) were punched out of the FTA® blood cards for each sample, from within the region well-stained by the blood smear. These disks were put into an eppendorf tube. FTA purification reagent at the rate of 10 µl for each disk i.e. 300 µl was added. The tubes were rotated up and down for 30 minutes and the spent solution was poured out carefully. This process of rotation was repeated twice. Tris-EDTA buffer (500 µl) at pH 6.7 was added and the tubes rotated up and down for 30 minutes and the spent

solution was poured out carefully. This step was repeated once. The FTA® blood card disks were dried overnight in an oven at 37°C. Thus the disks were ready for use and kept at room temperature. One of the disks was used for PCR reaction.

### 3.3. Microsatellite markers, PCR conditions and genotyping

Eighteen microsatellite loci were chosen from the goat genome database maintained at INRA (Institut national de Rechrché Agronomique) (http://www.locus.jouy.inra.fr/) and from the sheep genome database maintained at Roslin Research Institute (http://www.ri.bbsrc.ac.uk/). The characteristics of the 18 microsatellites are shown in **Table 3**.

Table 3. **Autosomal microsatellite loci used in the study and their characteristics.**

| Locus | Chromosome | ABI (colour) | Primer sequences 5' - 3' | Annealing temperature $^{o}$C |
|---|---|---|---|---|
| BM 1818 | 23 | Tet | F-AGC TGG GAA TAT AAC CAA AGG<br>R-AGT GCT TTC AAG GTC CAT GC | 55 |
| BMC 1222 | 13q12 | Hex | F-CCA ATT TTG CAG ATA AGA AAA CA<br>R-CCT GAG TGT TCC TCC TGA GT | 55 |
| BMS 0357 | 12 Ovine | Hex | F-TCC AAA CAA GTC TTC TCT ATT TAC C<br>R-CCA AAT AAT TGC TGG TCA GG | 58 |
| BMS 1494 | 21 Bovine | Hex | F-TCT GGA GCT TGC AAA AGA CC<br>R-AAT GGA TGA CTC CTG GAT GG | 55 |
| ILSTS 0005 | 10 | Fam | F-GGAAGCAATGAAATCTATAGCC<br>R-TGTTCTGTGAGTTTGTAAGC | 55 |
| ILSTS 0011 | 14 | Hex | F-GCT TGC TAC ATG GAA AGT GC<br>R-CTA AAA TGC AGA GCC CTA CC | 55 |
| ILSTS 0044 | 1 Ovine | Tet | F-AGT CAC CCA AAA GTA ACT GG<br>R-ACA TGT TGT ATT CCA AGT GC | 57 |
| ILSTS 0087 | 6 Ovine | Tet | F-AGC AGA CAT GAT GAC TCA GC<br>R-CTG CCT CTT TTC TTG AGA GC | 58 |
| INRA 0005 | 12 | Fam | F-CAA TCT GCA TGA AGT ATA AAT AT<br>R-CTT CAG GCA TAC CCT ACA CC | 52 |
| INRA 0063 | 18q22 | Fam | F-ATT TGC ACA AGC TAA ATC TAA CC<br>R-AAA CCA CAG AAA TGC TTG GAA G | 55 |
| INRA 0132 | 23 | Hex | F-AAC ATT TCA GCT GAT GGT GGC<br>R-TTC TGT TTT GAG TGG TAA GCT G | 57 |
| MAF 0035 | 23 Ovine | Hex | F-TCA AGA ATT TTG GAG CAC AAT TCT GG<br>R-AGT TAC AAA TGC AAG CAT CAT ACC TG | 55 |
| MAF 0065 | 15 | Tet | F-AAA GGC CAG AGT ATG CAA TTA GGA G<br>R-CCA CTC CTC TGA GAA TAT AAC ATG | 50 |
| MAF 0209 | 17 | Fam | F-GAT CAC AAA AAG TTG GAT ACA ACC GTG G<br>R-TCA TGC ACT TAA GTA TGT AGG ATG CTG | 57 |
| OarAE 0129 | 7 | Fam | F-AAT CCA GTG TGT GAA AGA CTA ATC CAG<br>R-GTA GAT CAA GAT ATA GAA TAT TTT TCA ACA CC | 60 |
| OarFCB 0304 | 19 Ovine | Hex | F-CCC TAG GAG CTT TCA ATA AAG AAT CGG<br>R-CGC TGC TGT CAA CTG GGT CAG GG | 55 |
| SRCRSP 0003 | 10q36 | Tet | F-CGG GGA TCT GTT CTA TGA AC<br>R-TGA TTA GCT GGC TGA ATG TCC | 55 |
| SRCRSP 0007 | 6 | Fam | F-TCT CAG CAC CTT AAT TGC TCT<br>R-GTC AAC ACT CCA ATG GTG AG | 57 |

The 18 microsatellite loci were analyzed at ILRI, Nairobi, Kenya. The steps of PCR protocol applied was as follows: all PCR amplifications were performed in a total volume of 10 µl (except for some FTA® blood disks, where a total volume of 20 µl was used) on GeneAMP®PCR system 9700 (Applied Biosystem) thermocycler. Each PCR reaction contained 1.0 µl of 20 ng template DNA (one FTA card disk), 1.0 µl of PCR buffer (19.2 mM $MgCl_2$, 100 mM Tris-HCl (pH 8.3), 500 mM KCl, 0.01% gelatin, 0.25% Tween 20 and 0.25% Nonidet P40), 0.5 µl of 2.5 mM dNTP (i.e. mixture of dATP, dCTP, dGTP, and dTTP), 0.08 µl of 200 mM each primer (forward and reverse), 0.04 µl of Taq (*Thermus aquaticus*) DNA polymerase (Promega) and 7.3 µl of sterile triple distilled water. The PCR amplification for FTA® blood with a total volume of 20 µl reaction contains 1 disc, 2.0 µl of PCR buffer (19.2 mM MgCl2, 100 mM Tris-HCl (pH 8.3), 500 mM KCl, 0.01% gelatin and 0.25% Tween 20 and 0.25% Nonidet P40), 1.0 µl of 2.5 mM dNTP (i.e. mixture of dATP, dCTP, dGTP, and dTTP), 0.16 µl of 200 mM of each primer (forward and reverse), 0.08 µl of Taq (*Thermus aquaticus*) DNA polymerase (Promega) and 15.6 µl of sterile triple distilled water. All amplification conditions included an initial denaturing step of 5 min at 95°C, followed by 35 cycles of 30 seconds at 95°C, 1 minute at an appropriate primer annealing temperature (**Table 2**), and 1 minute at 72°C. Duration of final extensions was 10 minutes at 72°C. The PCR products were analyzed by electrophoresis on a 4.25% denaturing polyacrylamide gel using a 377 ABI automatic DNA sequencer (ABI PRISM 377, Applied Biosystems). For this purpose, the forward primer for each locus was labeled with a fluorescent dye, either FAM (blue) or TET (green) or HEX (yellow). Two to four PRC products in different size ranges and labeling of the forward primers were co-loaded in a single lane of the gel. In preparation for loading, the samples were usually prepared by mixing 0.5-1.0 µl of each PCR products and 1.8 µl of loading mix/buffer (made of GENESCAN™ TAMRA 350, loading dye and formamide in the ratio of 1:1:4 with a little modification). The mixture was heated to 95°C for 2 to 4 minutes and immediately put on ice and then 1.5-2.0 µl was loaded in each lane of the gel. The GeneScan™ 350-TAMRA served as an internal size standard in each lane. A control sample was included in addition to the GeneScan™ 350-TAMRA for further standardization of fragment sizes. Fragment lengths were identified using the GeneScan® software (version 3.1.2 Applied Biosystems), while allele sizes identifications were performed by GENOTYPER® software (version 2.0 Applied Biosystems) using the third order least squares method for size calling (**Fig. 19b – g**).

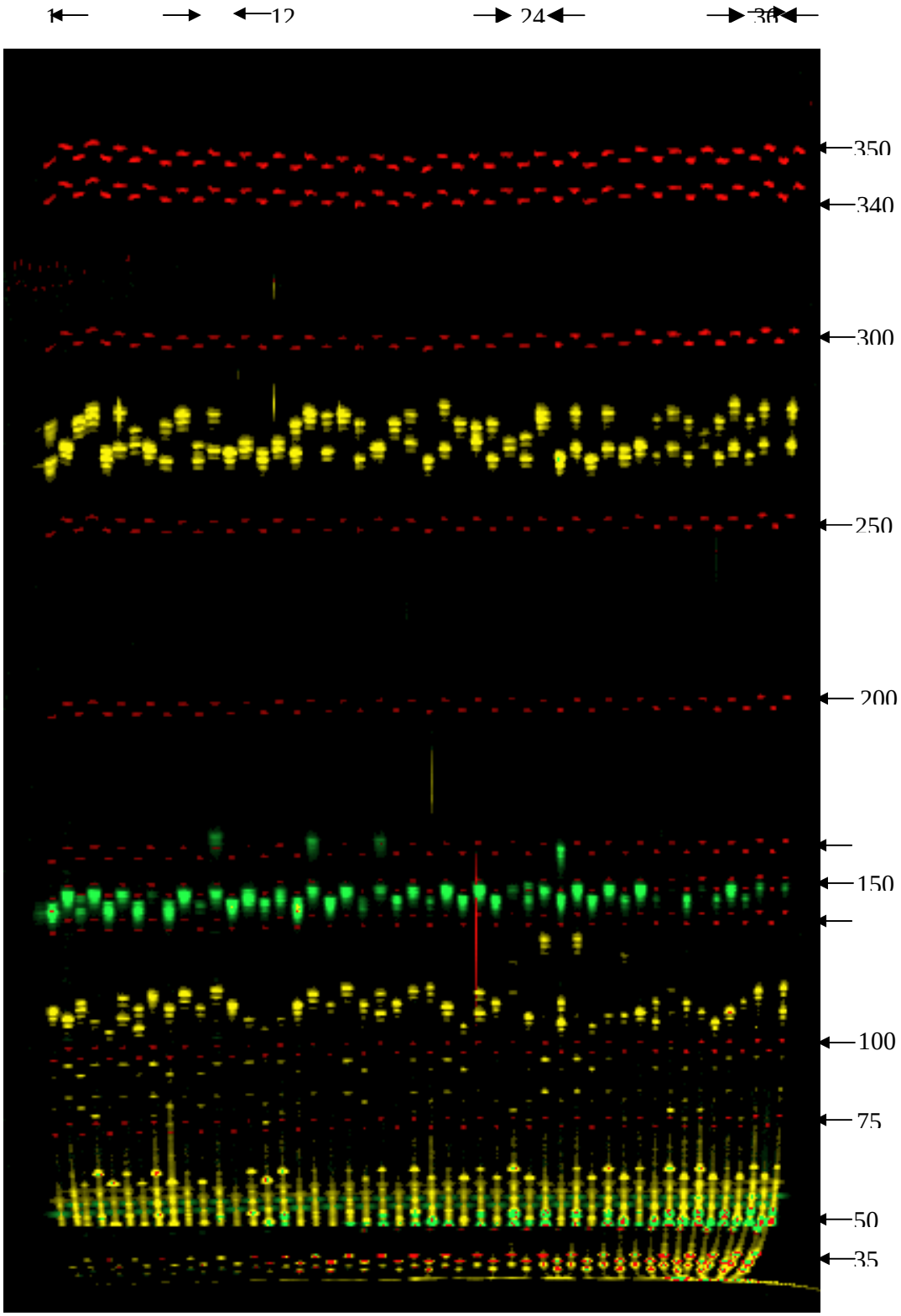


Figure 19b: Gel image of Abergalle breed shown by BMS357, ILSTS87 and BMC1222

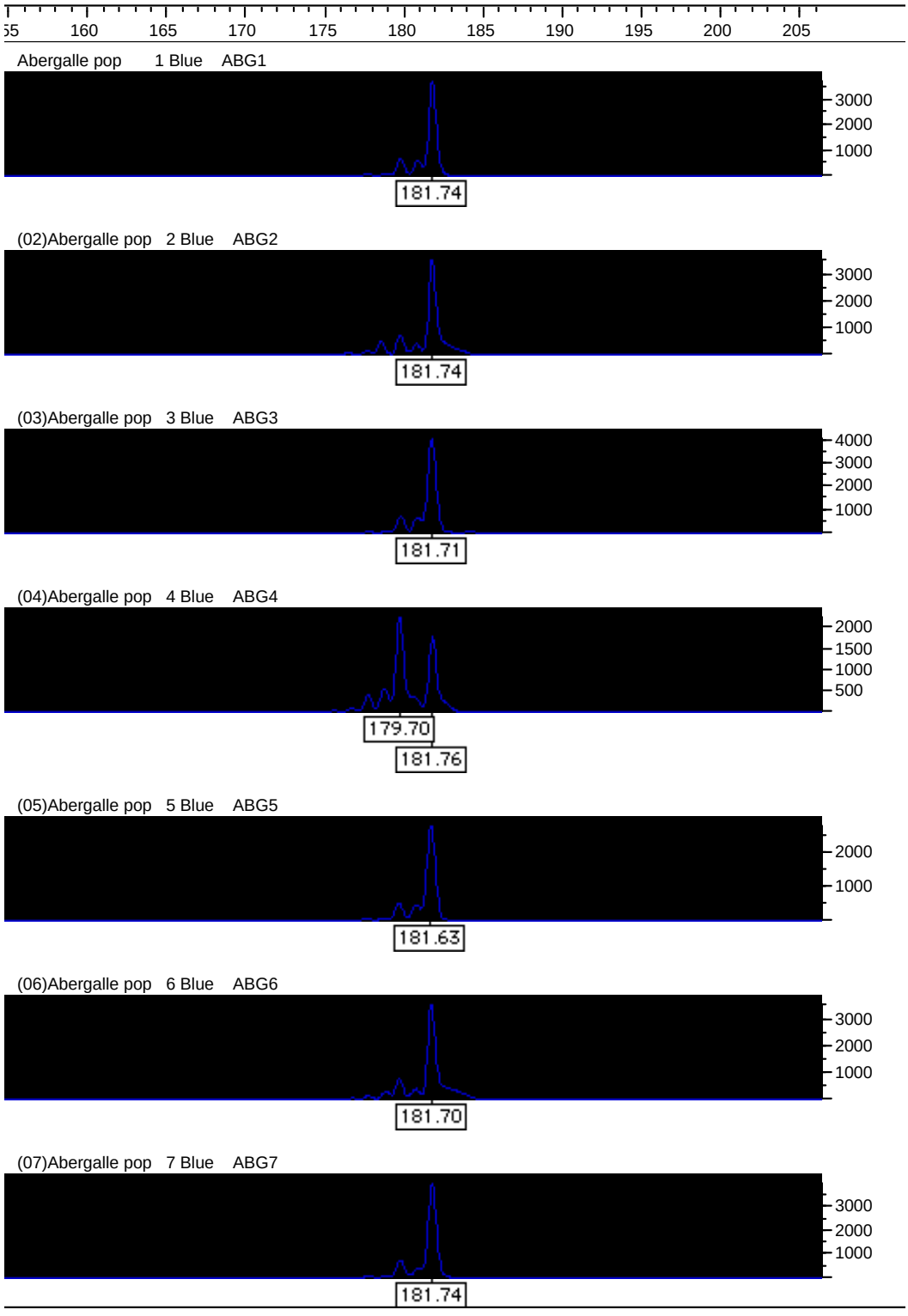


Figure 19c: **Allele scoring for Abergalle population using ILSTS05**

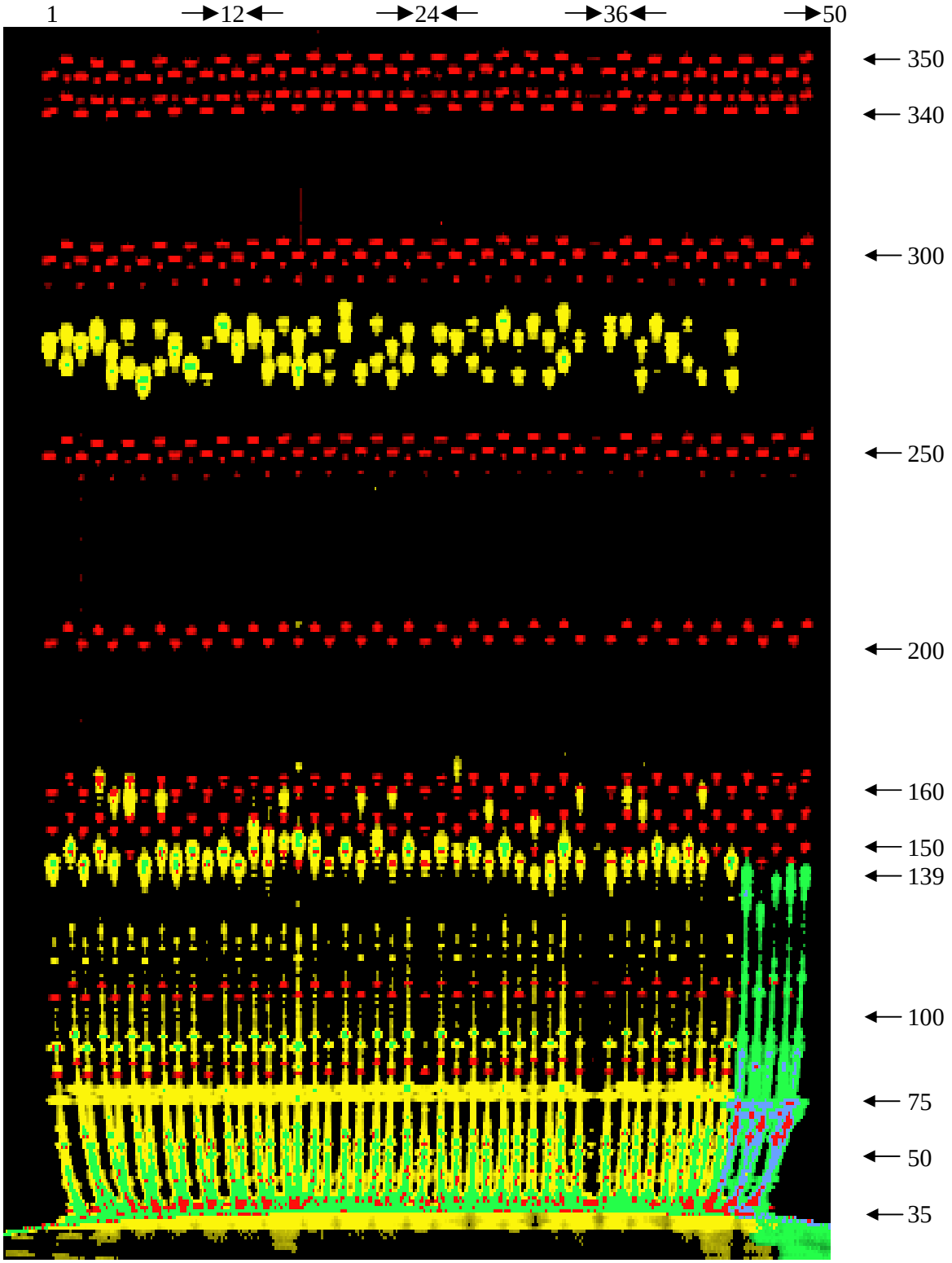


Figure 19d: **Gel image of Short-eared Somali population shown by OarFCB304, ILSTS11 and MAF65**

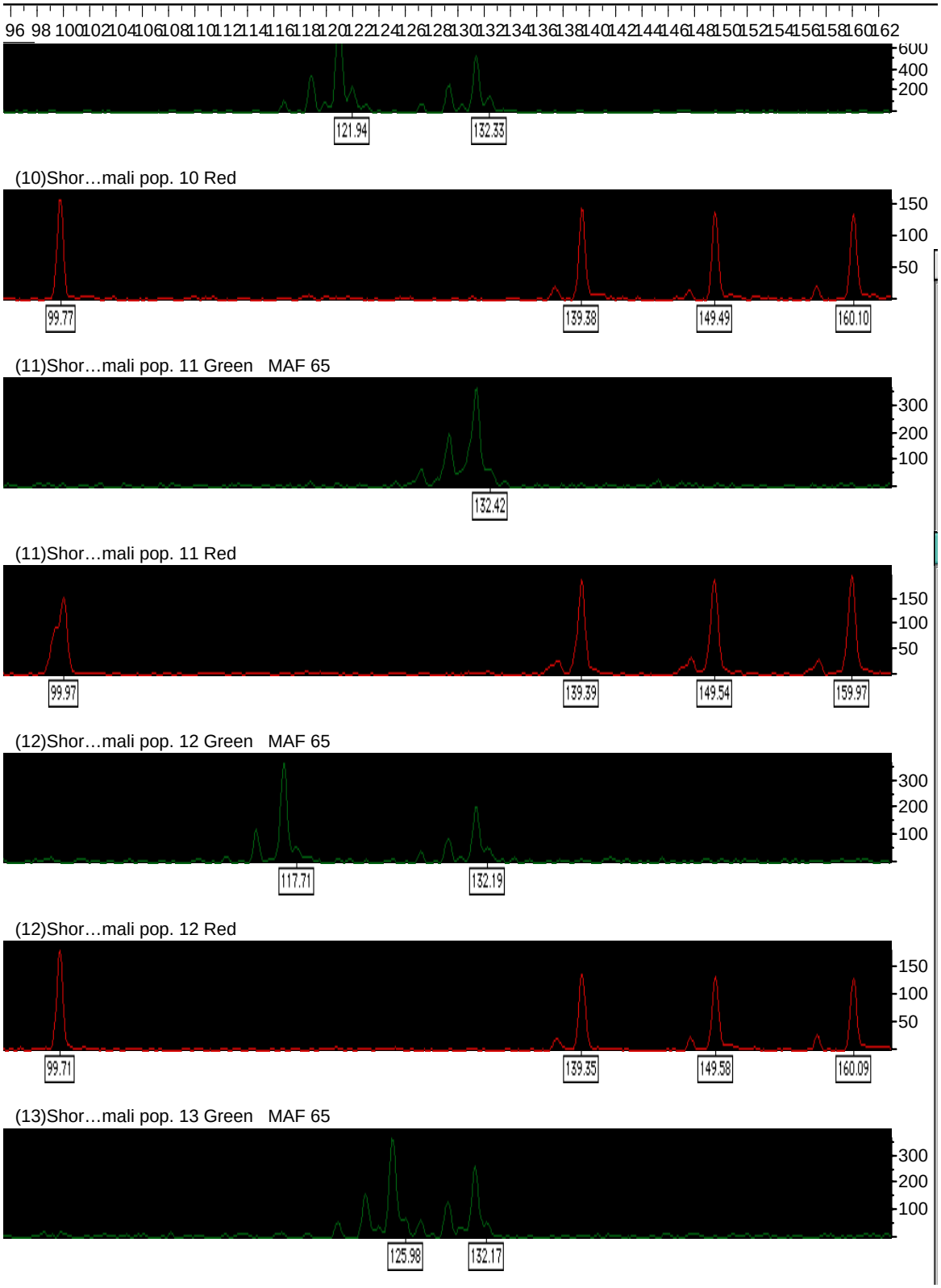


Figure 19e: **Allele scoring for Short-eared Somali population using MAF065 (Red color is internal standard)**.

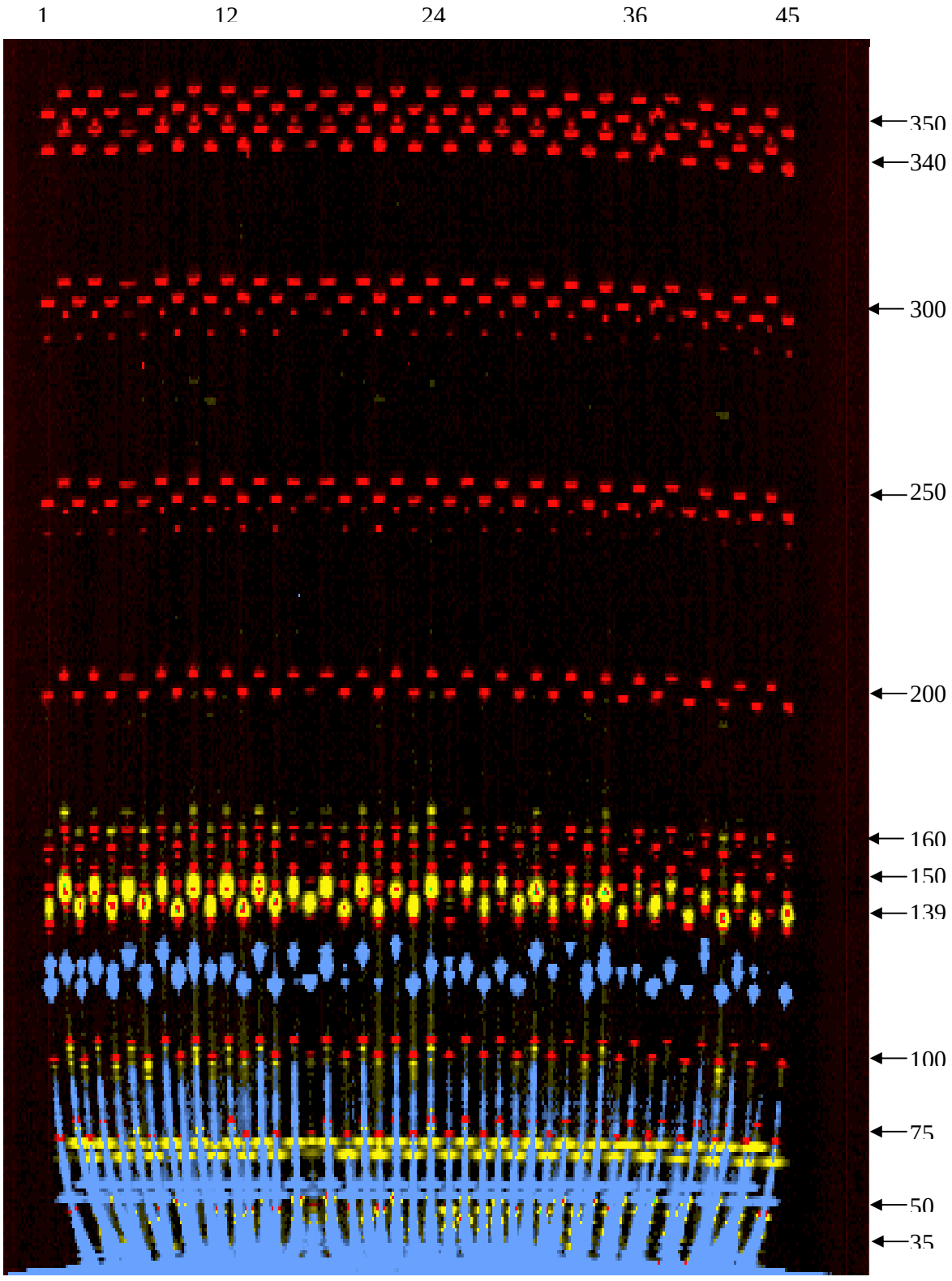


Figure 19f: **Gel image for Short-eared Somali population shown by INRA132-SRCRSP07**

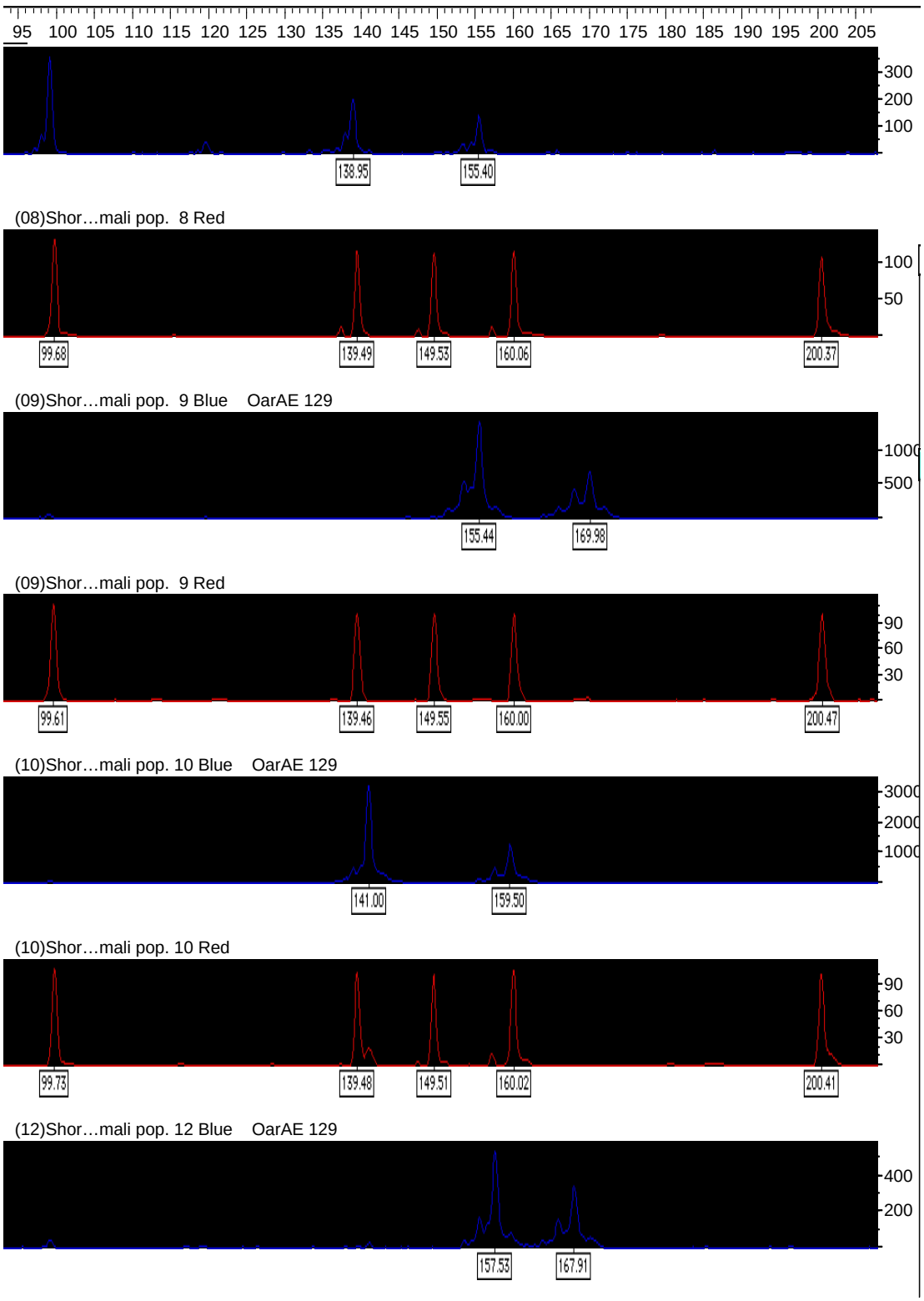


Figure 19g: **Allele scoring for Short-eared Somali population using Oarae129 (Red color is internal standard)**.

## 3.4. Data analysis

In this study, deviations of the locus/population combinations from HWE were determined using the program GENEPOP version 3.4 June 2003 (Raymond and Rousset, 1995). This involved the use of the Markov chain method to obtain unbiased P estimates of Fisher's exact test (Raymond and Rousset, 1995) for each locus/population combination through 30,000 iterations. In order to control type I error, i.e. accepting the hypothesis that the population are in HWE when it is not the case, a sequential Bonferroni correction (Rice, 1989) was performed.

Observed allele frequencies were computed using the excel macros (a program kindly provided by Stephen Park, (Spark@tcd.ie). The allele frequencies were used in the determination of both within- and between-population genetic variation.

Linkage disequilibrium analysis among pairs of loci in each population was also done using the program GENPOP version 3.4 June 2003 using Fisher's exact tests (Raymond and Rousset, 1995) with unbiased P values being derived by Markov chain method with 20,000 iterations per batch. A sequential Bonferroni correction (Rice, 1989) was performed to control type I error.

Genetic variation was determined as the mean number of alleles (MNA) per locus and average observed and expected heterozygosities. Average observed and expected heterozygosities were calculated using Microsatellite Analyzer (MSA) version 3.12 (Dieringer *et al*., 2003). The MNA and allele frequencies were calculated using the program MICROSATELLITE TOOLKIT version 3.1 (Park, 2001). To remove the sample bias, the MNA was calculated for all the individuals in a population and for random sample of 20 individuals, with 250 re-sampling of 20 individuals with replacement using the program ALLELE SAMPLER version 1.2 (Mburu *et al.,* 2003).

Both $D_S$ and $D_A$ genetic distance measures were computed using the DISPAN program (Ota, 1993). Phylogenetic trees were constructed from the DS and DA genetic distance matrices generated using the Neighbour Joining (NJ) and the Unpaired Group with

Arithmetic Mean (UPGMA) methods in the TREEVIEW program. The reliability of the constructed trees was examined by a bootstrap test with 1000 replicates re-sampling of loci with replacement (Felsenstein, 1985).

Principal component analysis, which simply describes the variation of the multivariate data in terms of some linear combinations of the variables was performed by Eigen analysis of covariance matrix calculated from allele frequencies data using the Multivariate Statistical Package (MVSP) program (version 3.12d) available from (http://www.kovcomp.com/mvsp/downl.html).

$F_{ST}$ estimates were calculated using the program FSTAT 2.9.3 Goudet, 2000 (http://www.unil.ch/izea/software/fstat.html) and GENEPOP version 3.4 (Raymond and Rousset, 1995), while the *GST* estimates were calculated using the DISPAN program (Ota, 1993). P-values were obtained after 55,000 permutations using standard Bonferroni corrections.

Analysis of molecular variance, AMOVA (Excoffier *et al*., 1992) implemented in the program ARLEQUIN version 2.000 (Schneider *et al*., 2000) available from (http://lgb.unige.ch/arlequin/) was performed to quantify further the extent of population differentiation and the distribution of genetic variation in the sample populations.

Individual population assignments were done using two programs WHICHRUN 4.1.0 (Banks and Eichert, 2000) and GeneClass (Cornuet *et al*., 1999).

Admixture analysis at both the population and individual level and individual specific assignment to the populations was performed using the program STRUCTURE version 2.0 (Pritchard, 2000), which implements the Pritchard, Stephens and Donnelly (2000) method, a model based clustering algorithm that infers population structure using multilocus genotypes. Five runs were made at each four Ks, where K is the number of inferred clusters and a burn-in period of 100,000 steps as well as a Markov Chain Monte Carlo (MCMC) of 500,000 being applied. The program DISTRUCT offers various options for displaying these results, where an individual's estimated membership coefficient are displayed as various segments partitioned into different colors, depending on the value of K

chosen. The program DISTRUCT offers various options for displaying these results, where an individual's estimated membership coefficient are displayed as various segments partitioned into different colors, depending on the value of K chosen.

The data was also tested for possible scoring errors due to stuttering or large allele drop-out and occurrence of null alleles using the program MICROCHECKER version 2.2 (Van-Oosterhout *et al.*' 2003), where 1,000 Monte Carlo bootstrap simulations were used to calculate probabilities with a 99% confidence interval.

CHAPTER FOUR

# 4.0. RESULTS

A number of analyses were performed using allelic variation obtained at 15 microsatellite loci. Allele frequencies at each locus were calculated for each population. Microsatellite variability such as the observed and expected numbers of alleles at each locus for each population, mean number of alleles per locus, observed heterozygosity and expected heterozygosity were calculated to quantify the genetic variation within the population. Statistics such as fixation indices or *F*-statistics were performed to describe the population structure and genetic differentiation among populations. Phylogenetic analysis and multivariate analysis such as principal component analysis were performed to infer the evolutionary relationship among populations. A number of computer programs were used to perform all the analyses.

## 4.1. Genetic diversity and relationship between Ethiopian and non-Ethiopian goat populations

### 4.1.1. Test for deviation from Hardy-Weinberg equilibrium (HWE)

Deviations from HWE proportions were determined for each of 360 locus-population combinations using Fisher's exact test for contingency tables, which was first applied by Haldane (1954). Exact tests are appropriate even when many rare alleles are present (Guo and Thompson, 1992; Chakraborty and Zhong, 1994), and should therefore be used for population genetic analysis of hypervariable markers such as microsatellite loci (Rousset and Raymond, 1995). Three distinct tests were performed for testing of deviation from HWE proportions under the same null hypothesis ($H_O$ = random union of gametes): probability test, heterozygosity deficiency and heterozygosity excess. An exact HW test (Haldane, 1954; Guo and Thompson, 1992) was applied for the probability test, while an exact score test (U-test) as described by Rousset and Raymond (1995) was applied to the heterozygosity deficiency and heterozygosity excess. The probability of the observed sample has been used to define the reject zone, and the *P*-value of the test has been defined as a sum of the exact probabilities of all tables (with the same allelic counts) with the same or lower probability than the observed table. Unbiased exact *P*-values were computed using

the complete enumeration algorithm for loci with two or three alleles (Louis and Dempster, 1987) or estimated by a Markov chain algorithm (Guo and Thompson, 1992) for loci with four or more alleles. The tests were implemented in the program GENEPOP version 3.4 (Raymond and Rousset, 1995) with the parameters as 30,000 steps of dememorization, 500 batches and 30,000 iterations per batch. Significance levels for multiple tests were adjusted with the Bonferroni corrections (Rice, 1989).

A total of 776 individuals from 20 different populations located in 9 countries (Ethiopian (11), Kenyan (2), Guinea Bissau (1), Mongolian (1), Saudi Arabian (1), Egypt (1), Italy (1), Botswana (1) and Turkey (1)) were genotyped using 18 microsatellite markers. After testing for linkage disequilibrium, null allele and HWE, 17, 18 and 14 populations were observed to have a significantly lower proportion of heterozygotes than that expected at HWE for BMC1222, BMS357 and SRCRSP07, respectively (**Table 4**). Consequently, these three markers were removed from further analysis.

Single locus exact test for HWE was conducted for the 20 populations and the remaining 15 microsatellite loci. Results indicated that 39 out of 300 i.e. 13% of the comparisons ($P < 0.05$; **Table 5**) deviated significantly from HWE, a figure higher than the 5% that would be expected from type I error alone. Application of sequential Bonferroni adjustment for multiple testing reduced this number of significant locus-population combinations to only three.

Out of the 15 loci, there was no locus that deviated from HWE for all the populations and also no population was observed to deviate from HWE for all the loci. However, as it is indicated in **Table 5**, there were some populations with several loci that deviate significantly from HWE. On one extreme, five loci found to deviate from HWE in the Hararghe Highland population, while on the other extreme, the Gumez and Boran-Galla didn't deviate from HWE at any locus. Arsi-Bale, Afar, Short-eared Somali, Woyto-Guji, Small-east Africa, Guinea Bissau and Mongolia deviated from HWE at only one locus. Most of these loci deviations from HWE corresponded to either heterozygote deficit or excess (**Table 5**).

Five out of twenty populations had HWE deviation at loci ILSTS44, INRA05, and INRA63, while four out of the twenty populations had HWE deviation at loci NRA132 and SRCRSP03.

Table 4. **Results of Fisher's exact test for Hardy-Weinberg equilibrium for 20 populations and 18 microsatellite markers.**

| Population | BM 1818 | BMC 1222 | BMS 1494 | BMS 357 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 | SRCRSP 07 | All loci |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | *1 | **1 | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| ABA | ns | **1 | ns | **1 | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | ns | ns | ** |
| AFA | ns | **1 | ns | **1 | ns | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| CEH | ns | **1 | ns | **1 | ns | ns | ** | ns | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | **1 | ** |
| GUM | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| HAH | ns | **1 | ns | **1 | ns | ns | *1 | ns | * | ns | *2 | ns | ns | *1 | ns | ns | **1 | ns | ** |
| KEF | ns | **1 | ns | **1 | ns | *1 | ns | ns | ns | ns | ns | ns | *1 | *1 | ns | ns | ns | *1 | ** |
| LES | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | ** | ns | ns | ns | ns | ** | ns | * | **1 | ** |
| NWH | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | **1 | ns | ns | **1 | ** |
| SES | *1 | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| WGJ | ns | **1 | ns | **1 | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| BGA | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| SEA | ns | **1 | ns | **1 | ns | ns | ** | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ** |
| GUB | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | ns | **1 | ns | ns | ns | ns | ns | **1 | ** |
| TSW | ns | ns | ns | ns | ns | ns | **1 | ns | ns | *1 | * | ns | ns | ns | *1 | ns | ns | *1 | ** |
| ALP | ns | ns | ns | ns | ns | ns | ns | ns | ** | ns | ns | ns | **1 | ns | ns | ns | *1 | *1 | * |
| ARD | ns | ns | **2 | **1 | ns | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | * |
| BAL | ns | **1 | ns | **1 | ns | ns | ns | *1 | * | ns | ns | ns | ns | ns | ns | ns | ns | ns | ** |
| HAI | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | *1 | *1 | ns | ** |
| MON | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | **1 | ns | ns | ns | ns | ns | ns | **1 | ** |
| All population | ns | **1 | ns | **1 | ns | ns | **1 | ns | * | **1 | **1 | ns | *1 | ns | ns | ns | *1 | **1 | |

ns = not significant; * = $P < 0.05$; ** = $P < 0.01$; 1 = heterozygote deficit; 2 = heterozygote excess.

ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran Galla, SEA – Small East African, GUB– Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia.

Table 5. **Results of Fisher's exact test for Hardy-Weinberg equilibrium for 20 populations and 15 microsatellite markers.**

| Population | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 | All loci |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | *1 | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | ns |
| ABA | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| AFA | ns | ns | ns | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| CEH | ns | ns | ns | ns | ** | ns | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | ** |
| GUM | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| HAH | ns | ns | ns | ns | * | ns | * | ns | *2 | ns | ns | *1 | ns | ns | **1 | ** |
| KEF | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | *1 | *1 | ns | ns | ns | ns |
| LES | ns | ns | ns | ns | ns | ns | ns | ** | ns | ns | ns | ns | ** | ns | * | ** |
| NWH | ns | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | **1 | ns | ns | ns |
| SES | *1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| WGJ | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| BGA | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| SEA | ns | ns | ns | ns | ** | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| GUB | ns | ns | ns | ns | ns | ns | ns | ns | ns | ** | ns | ns | ns | ns | ns | ns |
| TSW | ns | ns | ns | ns | **1 | ns | ns | *1 | * | ns | ns | ns | **1 | ns | ns | ** |
| ALP | ns | ns | ns | ns | ns | ns | ** | ns | ns | ns | ** | ns | ns | ns | *1 | ns |
| ARD | ns | **2 | ns | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns |
| BAL | ns | ns | ns | ns | ns | *1 | * | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| HAI | ns | ns | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | *1 | *1 | * |
| MON | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ns | ns | ns | ns | ns | ns | * |
| All population | ns | ns | ns | ns | **1 | ns | * | **1 | **1 | ns | *1 | ns | ns | ns | *1 | ns |

ns = not significant; * = $P < 0.05$; ** = $P < 0.01$; 1 = heterozygote deficit; 2 = heterozygote excess.

ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran Galla, SEA – Small East African, GUB– Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia.

### 4.1.2. Linkage disequilibrium

The linkage disequilibrium tests within each of the 20 populations were performed for the 105 pair-wise combinations of the 15 loci. Assuming a 5% type I error rate ($P < 0.05$), about four to five pairs of loci are expected to be in linkage disequilibrium. On average 6% of the locus pairs were in linkage disequilibrium within the populations. A large number of locus pairs that were in linkage disequilibrium ($P < 0.05$) were observed in the Arsi-Bale and Guinea Bissau (11 pairs), followed by the Long-eared Somali, Italian Alpine and Mongolia (nine pairs) populations, Boran-Galla and Short-eared Somali (eight pairs) populations, Afar, Small-East Africa and Woyto-Guji (seven pairs) populations, Central Highland (six pairs) population, Hair, Gumez and Keffa (five pairs) populations, Abergalle, Hararghe Highland and North west Highland (four pairs) populations, Ardi and Baladie (three pairs) populations, and Tswana population which had only one out of 105 pairs of loci in linkage disequilibrium (**Appendix 1**).

The global exact test for linkage disequilibrium across all 20 populations (105 pair-wise comparisons) yielded significant deviations in the following comparisons: BM1818/ILSTS11 ($P < 0.05$); BM1818/ILSTS44 ($P < 0.05$); BM1818/MAF209 ($P < 0.05$); BM1818/OarAE129 ($P < 0.05$); BM1494/ILSTS11 ($P < 0.05$), BM1494/ OarAE129 ($P < 0.05$); ILSTS05/SRCRSP03 ($P < 0.01$); ILSTS87/INRA132 ($P < 0.05$); MAF209/OarFCB304 ($P < 0.01$) (**Appendix 2**). Only one out of these nine tests remained significant following application of sequential Bonferroni corrections.

### 4.2.3. Inference of the recent bottleneck within populations

Many natural populations are suffering bottlenecks (i.e. severe reductions in size) as a result of habitat fragmentation and isolation. Understanding the effect of population bottlenecks on genetic variation has become increasingly important in population genetics, speciation theory, and conservation biology. In order to identify conservation objectives properly, it is important to identify populations that have lost genetic variability recently, as they may be more susceptible to demographic stochasticity (Lande, 1988; Mills and Smouse, 1994). Unfortunately, it is often difficult to determine if a population has recently experienced a bottleneck since historical population size and levels of genetic variation are

seldom known. Identifying recently bottlenecked populations (populations severely reduced in size) is important as bottlenecks can increase demographic stochasticity, rate of inbreeding, loss of genetic variation, and fixation of deleterious alleles and, thereby reduce adaptive potential and increase the probability of population extinction (Luikart and Cornuet, 1998). Cornuet and Luikart (1996) developed a statistical test (a sign test for heterozygosity excess) to detect recent historical bottlenecks using allele frequency data. The test requires no data on historical population size or levels of genetic variation; it requires only measurement of allele frequencies from 5 to 20 polymorphic loci in a sample of approximately 20 – 30 individuals. The test has reasonable statistical power when applied to allele frequency data sets generated by computer simulations (Cournet and Luikart, 1996).

In natural populations, allele number and heterozygosity at selectively neutral loci result from an equilibrium between mutation and genetic drift. The heterozygosity expected at a locus in an equilibrium population (*Heq*) can be calculated from the number of alleles observed and the sample size of individuals, assuming neutrality and mutation-drift equilibrium. In non-bottlenecked population defined as not has been recently bottlenecked and is therefore likely to be near mutation-drift equilibrium, the expected heterozygosity (*Heq*) will equal the measured Hardy-Weinberg equilibrium heterozygosity ($H_E$). However, if a population has suffered a recent bottleneck, the mutation-drift equilibrium is transiently disrupted and the heterozygosity measured at a locus ($H_E$) will exceed the heterozygosity (*Heq*) computed from the number of alleles' sampled (Luikart and Cornuet, 1998).

Bottlenecks generate 'heterozygosity excess' as alleles are generally lost faster than heterozygosity during a bottleneck. During bottleneck rare alleles are lost rapidly since they have little effect on heterozygosity (Hedrick *et al.,* 1986). Thus, many alleles can be lost without much reduction in heterozygosity in a bottlenecked population. The bottleneck-induced heterozygosity excess is transient and is likely to be detected only for a short time, approximately 0.02 – 4.0 *Ne* generations (*N*e is the bottleneck effective size), until a new equilibrium between mutation and drift is reached at the new *N*e (Cornuet and Luikart, 1996). Thus, only bottlenecks that have occurred in the recent past (less than 4 *N*e generations ago) are likely to be detectable by the sign test for heterozygosity excess. This

window of time is approximate and depends not only on *N*e but also on factors such as the mutation rate and mutation model of the loci sampled (Cornuet and Luikart, 1996).

In a non-bottlenecked, equilibrium population, approximately 50% of the loci sampled are expected to have a slight excess of heterozygosity (*H*e > *H*eq), and 50% will have a slight deficiency of heterozygosity (*H*e < *Heq*), resulting from genetic drift and sampling error. However, in recently bottlenecked populations, a majority of loci will exhibit an excess of heterozygosity (Luikart and Cornuet, 1998). In order to determine if a population exhibits a significant number of loci with heterozygosity excess, Piry *et al*. (2004) proposed three statistical tests: Sign test, a standardized differences test (Cornuet and Luikart, 1996; Luikart and Cornuet, 1998), and a Wilcoxon's signed rank test (Luikart *et al*., 1997). They also proposed a graphical descriptor of the shape of the allele frequency distribution ('mode-shift' indicator), which can differentiate between bottlenecked and stable populations (Luikart *et al*., 1998). The Wilcoxon's test (although this test is analogous to the sign test) is generally the most useful of all the tests because it is the most powerful (along with the standardized differences test) and robust (like the sign test) when used with few (< 20) polymorphic loci. In testing for bottlenecks, the null hypothesis of Wilcoxon's test is no significant heterozygosity excess (on average across loci). Thus the alternative hypothesis is significant heterozygosity excess (and thus evidence of a recent bottleneck). This is a one tailed test that requires at least four polymorphic loci to have many possibility of obtaining a significant ($P < 0.05$) test result. The sign test (Cornuet and Luikart, 1996) determines if a significant majority of loci in a population have a heterozygosity excess, and thus if a population appears to have been recently bottlenecked.

Test for heterozygosity excess should not be confused with test for Hardy-Weinberg proportions. Tests for Hardy-Weinberg proportions compare the observed proportion of heterozygosity ($H_O$) to the heterozygosity expected ($H_E$) when a population is in Hardy-Weinberg proportions. The test for Heterozygosity excess compares $H_E$ to the heterozygosity (*Heq*) expected at mutation drift equilibrium in a sample that has the same size and the same number of alleles as the sample used to measure heterozygosity expected ($H_E$).

Alternatively, a quantitative graphical method ('mode-shift' indicator) can be used for detecting a bottleneck-induced distortion of allele frequency distributions (Luikart *et al*.,

1998). This method involves comparing the distribution of allele frequencies observed in a population suspected to have been bottlenecked to the distribution expected in a non-bottlenecked population. The graphical method consists of grouping alleles from a sample of many polymorphic loci (at least five loci) in to each of 10 allele frequency classes and then plotting a frequency histogram. The graphical method concludes that a population has been recently bottlenecked if fewer alleles are found in the low frequency classes than in one or more intermediate frequency classes.

The recent bottleneck effect was inferred for each population using a Wilcoxon's signed rank test (Cornuet and Luikart, 1996), and the results are presented in **Table 6** and **Figure 20**. The calculations were implemented in the computer program BOTTLENECK version 1.2.02 (Cornuet and Luikart, 1996; http://www.montpellier.inra.fr/URLB/bottleneck) based on 10,000 simulation replicates assuming that microsatellite alleles are evolved under the stepwise mutation model. Generally, there was no recent bottleneck effect detected in any population for both Wilcoxon's signed rank test (**Table 6**) and quantitative graphical method (**Figure 20**). Interestingly, ten populations out of the total 20 populations showed a significant deficiency of heterozygosity. This suggests that these populations are not at mutation-drift equilibrium but instead have experienced a recent expansion in population size or perhaps a recent entry of rare alleles from genetically distinct immigrants.

Table 6. **The Wilcoxon test of recent bottleneck effect for each population assuming that all the microsatellite alleles are evolved under stepwise mutation model. The estimation is based on 10,000 simulation replicates.**

| Population | Sample size | One-tail *P*-value | |
|---|---|---|---|
| | | Heterozygote deficit | Heterozygote excess |
| Abergale | 46 | 0.0535 | 0.9527 |
| Arsi-Bale | 46 | 0.12619 | 0.88535 |
| Afar | 44 | 0.06769 | 0.93973 |
| Central Highland | 40 | 0.0535 | 0.9527 |
| Gumez | 43 | 0.00513 | 0.99582 |
| Hararghe Highland | 40 | 0.01767 | 0.98492 |
| Keffa | 38 | 0.03186 | 0.97232 |
| Long-eared Somali | 48 | 0.01279 | 0.98923 |
| North-West Highland | 45 | 0.12619 | 0.88535 |
| Short-eared Somali | 43 | 0.24435 | 0.77286 |
| Woyto-Guji | 46 | 0.11465 | 0.89612 |
| Boran-Galla | 35 | 0.01279 | 0.98923 |
| Small East Africa | 39 | 0.06027 | 0.94650 |
| Guinea Bissau | 46 | 0.04730 | 0.95837 |
| Tswana | 29 | 0.00903 | 0.99246 |
| Italian Alpine | 31 | 0.10388 | 0.90619 |
| Ardi | 25 | 0.00084 | 0.99934 |
| Baladie | 22 | 0.01767 | 0.98492 |
| Hair | 31 | 0.11465 | 0.89612 |
| Mongolia | 39 | 0.01077 | 0.99097 |

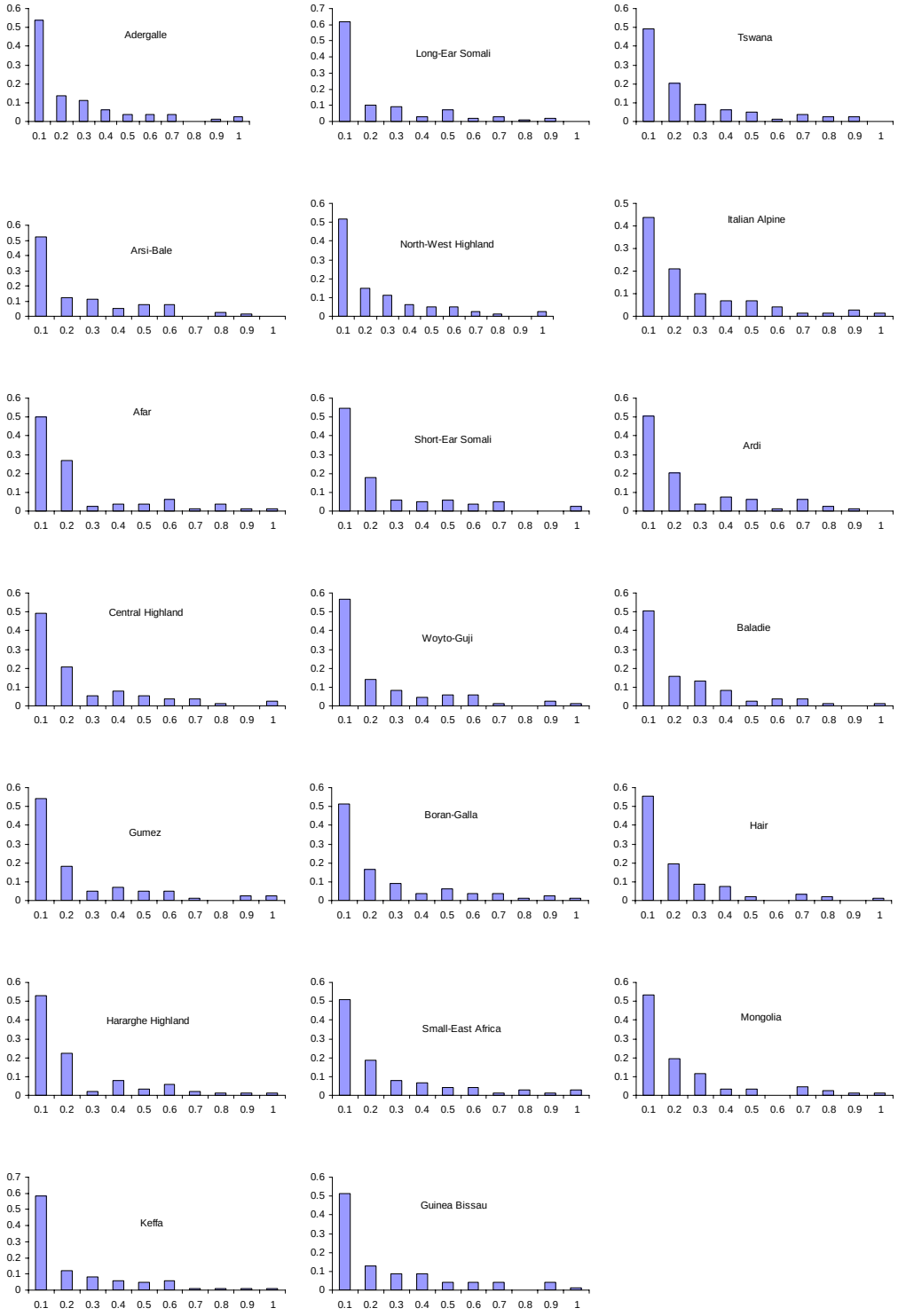


Figure 20. **Distributions of allele frequency from 15 microsatellite loci in 20 goat populations. The bars represent the proportion of alleles expected in each of 10 allele frequency classes calculated from 10,000 bottleneck simulation replicates.**

## 4.3. The extent and distribution of the genetic diversity within the 20 populations

### 4.3.1. Mean number of alleles and heterozygosities

#### 4.3.1.1. Analysis of allelic variation

Genetic data for a population can be expressed as gene or allele frequency. The portion of a particular allele out of all the alleles at a locus observed in all individuals of a population is called the allele frequency. It is a fundamental parameter in the study of population genetics, because the genetic change in a population is usually described by the changes in allele frequencies. Microsatellite are co-dominant markers and all alleles are identifiable, so the allele frequency can be obtained by simply calculating the proportion of a particular allele over all alleles at a specific locus detected in a population. The allelic frequencies were calculated using a computer program the EXCEL MICROSATELLITE TOOLKIT version 3.1 (available from http://oscar.gen.tcd.ie/~sdepark/ms-toolkit), and allelic frequencies for every locus and population are given in **Appendices 3** to **17**. The alleles observed at a locus varied in length from two to 60 base pairs. **Figure 21** illustrates the allelic ranges and their frequencies detected at each of 15 microsatellite loci in 776 goat individuals of 20 populations.

The number of alleles observed and expected under stepwise mutation model (SMM, Ohta and Kimura, 1973) and infinite allele model (IAM, Kimura and Crow, 1964) at each locus for each population were calculated using the computer program MICROSATELLITE ANALYSER (MSA) version 2.39 (Dieringer and Schlötterer, 2003). In the SMM, an allele at a locus is assumed to mutate to an allele either with gain or loss of a repeat with an equal probability, while the IAM assumes that new mutant alleles are always different from the existing ones in the populations.

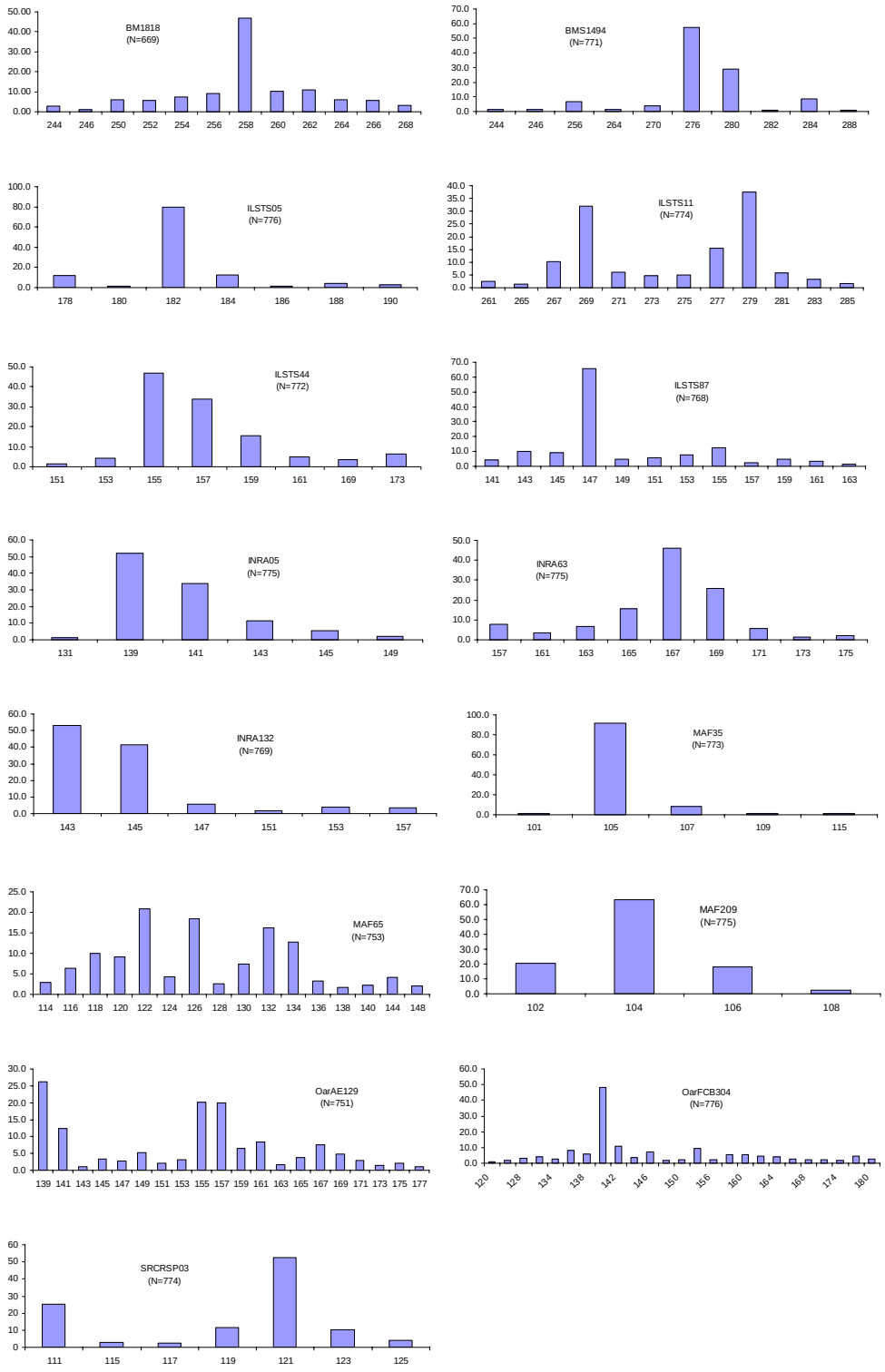


Figure 21. **Histogram showing the allele size range and their frequency detected in 776 goat individuals of 20 populations at 15 microsatellite loci.** The horizontal axis represents the allele size (base pairs) and the vertical axis their frequency in percentage.

The observed number of alleles, mean number of alleles expected under SMM ($MNA_{SMM}$) and IAM ($MNA_{IAM}$) are shown in **Table 7.** The number of observed alleles is highly dependent on sample size, and thus comparison of this quantity between different populations is not meaningful unless the sample size is more or less the same. The sample size dependence occurs as there are many low-frequency alleles in natural populations, and the number of observed alleles increases with the increase of the sample size. The mean number of alleles per locus (MNA) is the average number of alleles observed in a population over all loci genotyped. It is an indicator of genetic variability due to allelic diversity with the conditions that the populations under investigation are at mutation-drift equilibrium and that sample size is more or less the same for each population. To minimize the variation of MNA among populations affected by the sample size, we calculated the MNA for a random sample of 20 animals after 250 re-samplings with replacement for every population.

A total of 159 alleles were detected at 15 microsatellite loci in 776 goat individuals of 20 populations. Allele frequencies across all loci ranged from 1.04% to 98.91% (**Appendix 3** to **17**). Ten of these alleles were specific to Ethiopian populations, being observed in a minimum of two of the 11 Ethiopian populations analysed (**Table 8**). Private alleles meaning alleles that were unique to a single or observed in only one population were observed in the following Ethiopian populations: Arsi-Bale (one), Afar (one), Central Highland (two), Keffa (two), Long-eared Somali (two) and North-West Highland (two). However, the frequencies of these private alleles were less than 0.05 in all cases. In addition to these, nine alleles were specific to non-Ethiopian reference populations, being observed in a minimum of two of the nine non-Ethiopian reference populations analysed (**Table 8**). Private alleles, that were unique to a single population were observed in the following non-Ethiopian reference populations: Italian Alpine (one), Ardi (five), Boran-Galla (one), Small-East Africa (one), Guinea Bissau (one), Tswana (one), Baladie (two), Hair (two) and Mongolia (three). However, the frequencies of these private alleles were also less than 0.05 in all cases.

The most polymorphic marker was OarFCB304 with 25 alleles, while the least polymorphic marker is MAF209 with four alleles in the pooled population. The total

number of alleles sampled in a population across all loci ranged between 80 in Guinea Bissau population from West Africa and 116 in Long-eared Somali population from Ethiopia. The $MNA_{OBS}$ across all populations was 5.46 ± 0.06, varying between 4.67 ± 2.74 again in Guinea Bissau population and 6.47 ± 4.03 again in Long-eared Somali. The $MNA_{OBS}$ for uniformed sample size with 20 animals was 4.88 ± 0.08, varying between 4.19 ± 2.26 again in Guinea Bissau and 5.74 ± 2.89 in Hair populations. In general, the mean numbers of alleles were comparable among all the populations from the different geographic regions.

The expected MNA across all populations was 4.60 ± 0.08 and 7.42 ± 0.16 for SMM and IAM, respectively, ranging between 4.05 ± 1.59 in Guinea Bissau and 5.77 ± 2.70 in Hair for SMM, and between 6.39 ± 2.34 in Ardi and 9.29 ± 4.85 in Hair populations for IAM. The results indicate that the stepwise mutation model (SMM) is more appropriate for this particular dataset and should be preferred in the model-based algorithms such as the inference of the recent population bottleneck.

The expected heterozygosity ($H_E$) and observed heterozygosity ($H_O$) values averaged over loci showed an overall pattern similar to that observed for MNA per locus (**Table 9**). The highest value for $H_{E,}$ around 64% was observed within the Hair (Turkey) population, while the lowest value around 52% was again for Guinea Bissau. The highest value for $H_{O,}$ around 63% was observed within the Ardi population, while the lowest value around 52% was observed for Afar, Boran Galla and Tswana when all individuals in the populations are considered in both cases. In all cases average observed heterozygosities were lower than that expected under HWE except for Guinea Bissau and Ardi populations.

The null alleles were tested for every population at each locus using Brookfield null allele estimator 2 (Brookfield, 1996) which assumes that a proportion of the sample are not amplified because these are in fact null allele homozygotes. The test, implemented in the computer program Micro-Checker version 2.2.1 (available from http://www.microchecker.hull.ac.uk), were carried out with 1000 Monte Carlo bootstrap simulations over the observed data to calculate probabilities and 99% confidence interval for both homozygote frequencies and heterozygote size differences. Three loci (BMC1222, BMS357 and SRCRSP07) showed evidence for a null allele in 17, 18 and 14 populations,

respectively. Neither a single locus nor a single population showed evidence for null allele across all populations or across all loci.

Table 7. Observed number of alleles, mean number of alleles expected under SMM and IAM for each population and each locus.

| Pop | BM1818 | | | BMC1222 | | | BMS1494 | | | BMS357 | | | ILSTS05 | | | ILSTS11 | | | ILSTS44 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM |
| ABG | 8 | 5.2 | 9.3 | 8 | 4.2 | 6.7 | 4 | 3.7 | 5.6 | 8 | 5.4 | 9.8 | 3 | 1.6 | 1.7 | 7 | 5.2 | 9.2 | 4 | 4.7 | 8.1 |
| ABA | 9 | 6.5 | 12.2 | 6 | 4.4 | 7.3 | 4 | 3.7 | 5.6 | 7 | 5.8 | 10.2 | 3 | 2.7 | 3.4 | 6 | 5.0 | 8.8 | 4 | 4.4 | 7.1 |
| AFA | 10 | 3.9 | 5.8 | 7 | 5.6 | 10.1 | 4 | 3.2 | 4.3 | 9 | 6.8 | 12.7 | 5 | 2.9 | 3.8 | 5 | 4.4 | 7.2 | 4 | 5.0 | 8.6 |
| CEH | 9 | 5.2 | 9.1 | 5 | 4.8 | 8.1 | 5 | 3.7 | 5.3 | 7 | 6.2 | 11.0 | 2 | 1.5 | 1.5 | 6 | 4.7 | 7.7 | 4 | 5.0 | 8.6 |
| GUM | 8 | 6.4 | 11.8 | 5 | 4.5 | 7.4 | 4 | 2.5 | 2.9 | 9 | 7.6 | 13.8 | 3 | 1.8 | 2.0 | 5 | 4.4 | 7.0 | 5 | 4.5 | 7.5 |
| HAH | 8 | 4.4 | 7.1 | 5 | 4.8 | 7.9 | 4 | 3.7 | 5.4 | 9 | 8.0 | 14.6 | 2 | 2.1 | 2.3 | 6 | 5.0 | 8.5 | 4 | 5.1 | 8.7 |
| KEF | 7 | 3.6 | 5.2 | 4 | 2.9 | 3.7 | 5 | 4.2 | 6.5 | 9 | 4.8 | 8.0 | 2 | 2.5 | 3.0 | 7 | 5.0 | 8.6 | 4 | 4.8 | 8.0 |
| LES | 12 | 5.7 | 10.7 | 7 | 4.9 | 8.6 | 5 | 3.4 | 4.8 | 8 | 6.1 | 11.5 | 3 | 2.7 | 3.3 | 6 | 5.0 | 8.8 | 5 | 4.7 | 8.0 |
| NWH | 9 | 5.6 | 10.2 | 5 | 4.7 | 7.9 | 4 | 3.7 | 5.4 | 8 | 5.5 | 9.5 | 2 | 1.7 | 1.8 | 6 | 4.4 | 7.3 | 4 | 4.9 | 8.5 |
| SES | 10 | 5.8 | 10.4 | 6 | 6.0 | 10.9 | 3 | 3.5 | 5.1 | 9 | 7.2 | 13.1 | 3 | 1.8 | 2.0 | 7 | 5.2 | 9.2 | 4 | 5.2 | 9.0 |
| WGJ | 10 | 4.9 | 8.5 | 6 | 4.7 | 8.0 | 4 | 3.9 | 6.1 | 9 | 6.3 | 11.8 | 3 | 2.3 | 2.7 | 5 | 5.1 | 9.1 | 4 | 4.6 | 7.8 |
| BGA | 10 | 6.2 | 10.9 | 6 | 3.6 | 5.2 | 3 | 3.3 | 4.4 | 9 | 5.5 | 9.1 | 2 | 1.5 | 1.6 | 4 | 4.6 | 7.4 | 4 | 4.7 | 7.5 |
| SEA | 7 | 5.0 | 8.5 | 5 | 3.0 | 3.9 | 4 | 3.3 | 4.6 | 9 | 4.9 | 8.0 | 2 | 1.8 | 1.9 | 6 | 4.9 | 8.2 | 5 | 5.1 | 8.7 |
| GUB | 9 | 6.7 | 12.8 | 3 | 3.1 | 4.1 | 4 | 3.7 | 5.6 | 4 | 4.1 | 6.2 | 2 | 2.2 | 2.5 | 4 | 4.1 | 6.4 | 3 | 2.3 | 2.6 |
| TSW | 8 | 5.2 | 8.1 | 6 | 6.4 | 10.0 | 5 | 4.5 | 7.0 | 8 | 5.3 | 8.5 | 3 | 4.7 | 7.5 | 7 | 7.0 | 11.7 | 5 | 2.8 | 3.5 |
| ALP | 6 | 6.0 | 10.2 | 5 | 4.0 | 5.8 | 4 | 3.8 | 5.4 | 8 | 4.3 | 6.3 | 3 | 4.0 | 5.9 | 5 | 5.7 | 9.5 | 5 | 2.6 | 3.1 |
| ARD | 6 | 4.8 | 7.3 | 5 | 5.2 | 8.1 | 3 | 3.5 | 4.7 | 11 | 6.5 | 10.5 | 4 | 3.0 | 3.7 | 6 | 4.5 | 6.9 | 3 | 3.0 | 3.9 |
| BAL | 9 | 7.4 | 11.6 | 5 | 5.5 | 8.4 | 3 | 3.6 | 5.0 | 11 | 9.3 | 13.9 | 4 | 2.9 | 3.6 | 6 | 5.8 | 9.1 | 4 | 3.3 | 4.2 |
| HAI | 9 | 8.6 | 14.8 | 7 | 5.9 | 10.0 | 4 | 4.9 | 7.9 | 11 | 12.2 | 19.0 | 5 | 3.7 | 5.2 | 8 | 6.6 | 11.4 | 4 | 3.2 | 4.2 |
| MON | 8 | 8.7 | 15.8 | 6 | 5.6 | 9.9 | 4 | 3.5 | 4.9 | 11 | 8.4 | 15.1 | 5 | 4.0 | 6.1 | 6 | 5.7 | 10.2 | 5 | 2.4 | 2.9 |
| **Mean** | **8.6** | **5.8** | **10.0** | **5.6** | **4.7** | **7.6** | **4.0** | **3.7** | **5.3** | **8.7** | **6.5** | **11.1** | **3.1** | **2.6** | **3.3** | **5.9** | **5.1** | **8.6** | **4.2** | **4.1** | **6.5** |
| **SD** | **1.5** | **1.4** | **2.7** | **1.1** | **1.0** | **2.1** | **0.6** | **0.5** | **1.0** | **1.7** | **1.9** | **3.2** | **1.1** | **0.9** | **1.7** | **1.0** | **0.7** | **1.4** | **0.6** | **1.0** | **2.4** |

Cont.

| Pop | ILSTS87 | | | INRA05 | | | INRA63 | | | INRA132 | | | MAF035 | | | MAF065 | | | MAF209 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM |
| ABG | 6 | 2.4 | 2.9 | 4 | 3.8 | 5.7 | 5 | 4.5 | 7.5 | 3 | 3.5 | 5.1 | 2 | 1.9 | 2.0 | 8 | 6.8 | 13.0 | 3 | 4.0 | 6.1 |
| ABA | 6 | 3.4 | 4.9 | 4 | 4.4 | 7.3 | 5 | 4.1 | 6.4 | 3 | 3.7 | 5.6 | 3 | 2.1 | 2.4 | 9 | 9.6 | 18.3 | 3 | 4.4 | 7.3 |
| AFA | 5 | 2.5 | 3.0 | 4 | 3.8 | 5.7 | 4 | 4.5 | 7.4 | 3 | 3.6 | 5.3 | 2 | 1.6 | 1.7 | 9 | 9.1 | 17.3 | 3 | 3.3 | 4.6 |
| CEH | 5 | 3.4 | 4.7 | 3 | 3.9 | 5.9 | 5 | 4.0 | 6.0 | 3 | 3.5 | 5.0 | 2 | 1.6 | 1.6 | 8 | 9.0 | 16.3 | 3 | 4.0 | 6.2 |
| GUM | 6 | 2.6 | 3.2 | 4 | 4.0 | 6.2 | 5 | 4.7 | 7.9 | 3 | 4.4 | 7.1 | 2 | 1.6 | 1.7 | 11 | 7.8 | 14.3 | 3 | 3.4 | 4.9 |
| HAH | 6 | 3.0 | 3.8 | 4 | 4.4 | 7.0 | 6 | 4.3 | 7.0 | 3 | 3.7 | 5.4 | 3 | 1.6 | 1.6 | 11 | 10.7 | 19.4 | 4 | 3.9 | 5.9 |
| KEF | 5 | 2.8 | 3.5 | 4 | 4.5 | 7.4 | 8 | 5.2 | 8.9 | 3 | 3.6 | 5.2 | 2 | 1.5 | 1.5 | 9 | 8.4 | 15.3 | 4 | 4.4 | 7.0 |
| LES | 5 | 3.3 | 4.5 | 4 | 4.1 | 6.6 | 6 | 5.0 | 8.9 | 3 | 3.7 | 5.5 | 2 | 2.2 | 2.6 | 10 | 8.8 | 17.2 | 3 | 3.9 | 6.1 |
| NWH | 7 | 3.0 | 3.9 | 4 | 4.2 | 6.6 | 5 | 4.7 | 7.9 | 2 | 3.5 | 5.0 | 2 | 1.9 | 2.1 | 9 | 8.2 | 15.7 | 3 | 3.7 | 5.6 |
| SES | 7 | 3.5 | 4.9 | 4 | 4.2 | 6.6 | 5 | 5.0 | 8.6 | 2 | 3.6 | 5.1 | 2 | 1.7 | 1.8 | 9 | 9.2 | 17.5 | 3 | 3.4 | 4.7 |
| WGJ | 7 | 2.3 | 2.6 | 4 | 4.0 | 6.2 | 5 | 4.3 | 7.0 | 3 | 3.8 | 5.7 | 2 | 1.9 | 2.1 | 10 | 9.1 | 17.6 | 3 | 4.4 | 7.3 |
| BGA | 5 | 2.6 | 3.2 | 4 | 3.6 | 5.1 | 5 | 4.0 | 6.0 | 3 | 3.5 | 4.9 | 2 | 2.0 | 2.2 | 10 | 7.9 | 14.0 | 3 | 3.3 | 4.4 |
| SEA | 5 | 2.2 | 2.6 | 4 | 3.0 | 3.8 | 5 | 5.6 | 9.8 | 3 | 3.4 | 4.7 | 2 | 1.6 | 1.7 | 10 | 8.8 | 16.0 | 3 | 4.3 | 6.7 |
| GUB | 4 | 2.5 | 3.0 | 3 | 3.9 | 5.9 | 4 | 3.8 | 5.7 | 3 | 4.4 | 7.3 | 2 | 1.1 | 1.1 | 8 | 5.7 | 10.4 | 3 | 3.5 | 4.9 |
| TSW | 6 | 3.3 | 4.4 | 4 | 4.1 | 6.1 | 5 | 5.4 | 8.8 | 5 | 3.8 | 5.5 | 3 | 1.9 | 2.1 | 8 | 5.8 | 9.6 | 3 | 3.5 | 4.9 |
| ALP | 6 | 5.7 | 9.6 | 4 | 3.5 | 4.8 | 6 | 6.4 | 11.0 | 2 | 3.5 | 4.8 | 4 | 1.5 | 1.5 | 6 | 5.4 | 8.8 | 3 | 2.0 | 2.1 |
| ARD | 6 | 4.4 | 6.7 | 4 | 4.3 | 6.4 | 5 | 4.4 | 6.5 | 4 | 3.5 | 4.8 | 2 | 2.0 | 2.1 | 9 | 7.2 | 11.5 | 3 | 3.7 | 5.1 |
| BAL | 7 | 5.4 | 8.2 | 5 | 4.1 | 5.8 | 5 | 5.5 | 8.4 | 4 | 3.3 | 4.3 | 2 | 1.7 | 1.8 | 8 | 7.6 | 11.8 | 2 | 3.1 | 4.0 |
| HAI | 10 | 8.4 | 14.4 | 4 | 4.5 | 7.0 | 5 | 6.4 | 10.9 | 2 | 3.1 | 4.1 | 2 | 1.7 | 1.8 | 11 | 9.2 | 15.6 | 3 | 2.7 | 3.4 |
| MON | 8 | 5.8 | 10.2 | 3 | 3.0 | 4.0 | 6 | 5.2 | 8.9 | 4 | 3.4 | 4.7 | 2 | 1.7 | 1.8 | 8 | 5.7 | 9.9 | 2 | 2.6 | 3.1 |
| **Mean** | **6.1** | **3.6** | **5.2** | **3.9** | **4.0** | **6.0** | **5.3** | **4.8** | **8.0** | **3.1** | **3.6** | **5.2** | **2.3** | **1.7** | **1.9** | **9.1** | **8.0** | **14.5** | **3.0** | **3.6** | **5.2** |
| **SD** | **1.3** | **1.6** | **3.1** | **0.4** | **0.4** | **1.0** | **0.9** | **0.7** | **1.5** | **0.8** | **0.3** | **0.8** | **0.6** | **0.3** | **0.3** | **1.3** | **1.5** | **3.2** | **0.5** | **0.6** | **1.4** |

cont.

| Pop | OarAE129 | | | OarFCB304 | | | SRCRSP03 | | | SRCRSP07 | | | Total | Mean number of alleles per locus (SD) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | Obs | SMM | IAM | | OBS(all) | OBS(20) | SMM | IAM |
| ABG | 8 | 6.9 | 13.2 | 11 | 5.8 | 10.8 | 4 | 5.7 | 10.4 | 3 | 3.8 | 5.7 | 99 | 5.33 (2.55) | 4.65 (2.09) | 4.40 (1.50) | 7.37 (3.40) |
| ABA | 7 | 6.3 | 11.9 | 10 | 5.4 | 9.7 | 4 | 4.8 | 8.1 | 6 | 4.3 | 7.1 | 99 | 5.33 (2.41) | 4.87 (2.28) | 4.73 (1.64) | 7.97 (3.65) |
| AFA | 11 | 9 | 17.1 | 9 | 5.2 | 9.2 | 4 | 4.6 | 7.6 | 5 | 4.7 | 7.9 | 103 | 5.47 (2.83) | 4.77 (2.34) | 4.64 (1.99) | 7.74 (4.36) |
| CEH | 9 | 6.6 | 11.4 | 9 | 5.5 | 9.7 | 4 | 5.1 | 8.8 | 4 | 4.1 | 6.3 | 93 | 5.13 (2.53) | 4.65 (2.15) | 4.54 (1.75) | 7.39 (3.53) |
| GUM | 11 | 8.2 | 15.6 | 9 | 4.3 | 7 | 4 | 4.4 | 7.2 | 5 | 4 | 6.2 | 102 | 5.53 (2.90) | 4.71 (2.31) | 4.51 (1.91) | 7.42 (4.08) |
| HAH | 9 | 6.9 | 12.5 | 13 | 5.5 | 9.7 | 6 | 4.6 | 7.5 | 4 | 4 | 6 | 107 | 5.93 (3.13) | 5.1 (2.53) | 4.75 (2.11) | 7.80 (4.29) |
| KEF | 12 | 7.4 | 13.5 | 8 | 5 | 8.5 | 4 | 4.1 | 6.2 | 5 | 5.4 | 9.3 | 102 | 5.60 (2.82) | 5.05 (2.37) | 4.45 (1.65) | 7.17 (3.45) |
| LES | 14 | 8.3 | 16.2 | 14 | 4.2 | 6.7 | 5 | 4.5 | 7.5 | 4 | 4.3 | 7.1 | 116 | 6.47 (4.03) | 5.28 (2.90) | 4.72 (1.71) | 8.02 (3.92) |
| NWH | 12 | 7.6 | 14.6 | 8 | 5 | 8.8 | 4 | 4.9 | 8.5 | 4 | 4.5 | 7.5 | 98 | 5.40 (3.02) | 4.67 (2.26) | 4.54 (1.64) | 7.60 (3.62) |
| SES | 10 | 8.3 | 15.7 | 10 | 4.3 | 7 | 5 | 4.8 | 8.3 | 5 | 4.1 | 6.5 | 104 | 5.60 (3.00) | 4.9 (2.41) | 4.82 (1.98) | 8.13 (4.26) |
| WGJ | 11 | 8 | 15.4 | 10 | 5.6 | 10.1 | 4 | 4.2 | 6.6 | 4 | 4.2 | 6.7 | 104 | 5.67 (3.09) | 4.99 (2.57) | 4.64 (1.81) | 7.85 (4.04) |
| BGA | 10 | 8 | 13.6 | 8 | 4 | 6 | 5 | 4.7 | 7.6 | 4 | 3.8 | 5.5 | 97 | 5.20 (2.88) | 4.63 (2.48) | 4.26 (1.76) | 6.59 (3.48) |
| SEA | 7 | 6.4 | 11.4 | 8 | 5.1 | 8.8 | 4 | 4.4 | 7.1 | 4 | 4.1 | 6.3 | 93 | 5.00 (2.27) | 4.64 (2.10) | 4.26 (1.74) | 6.81 (3.63) |
| GUB | 7 | 5.6 | 10.1 | 11 | 7.4 | 14.3 | 3 | 4.2 | 6.8 | 3 | 4.6 | 7.8 | 80 | 4.67 (2.74) | 4.19 (2.26) | 4.05 (1.59) | 6.53 (3.54) |
| TSW | 6 | 3.9 | 5.6 | 7 | 5.5 | 9.1 | 4 | 2.5 | 3 | 4 | 3.1 | 4.1 | 97 | 5.27 (1.71) | 4.89 (1.63) | 4.39 (1.40) | 6.63 (2.71) |
| ALP | 7 | 5.4 | 8.9 | 6 | 5.1 | 8.4 | 4 | 3.2 | 4.3 | 3 | 4.7 | 7.5 | 87 | 4.73 (1.44) | 4.4 (1.46) | 4.25 (1.41) | 6.54 (2.83) |
| ARD | 8 | 5.4 | 8.6 | 12 | 3.8 | 5.3 | 4 | 4 | 5.8 | 5 | 4.7 | 7.3 | 100 | 5.27 (2.69) | 4.97 (2.41) | 4.32 (1.25) | 6.39 (2.34) |
| BAL | 8 | 6.3 | 9.8 | 12 | 8.8 | 13.4 | 4 | 3.9 | 5.5 | 4 | 4.1 | 5.9 | 103 | 5.53 (2.80) | 5.45 (2.75) | 5.08 (2.13) | 7.48 (3.58) |
| HAI | 9 | 7 | 12 | 11 | 7.3 | 12.5 | 5 | 3.9 | 5.7 | 5 | 4.6 | 7.3 | 115 | 6.13 (3.20) | 5.74 (2.89) | 5.77 (2.70) | 9.29 (4.85) |
| MON | 9 | 8.9 | 16.2 | 13 | 7.2 | 13.3 | 5 | 3.7 | 5.3 | 5 | 3 | 4 | 110 | 5.87 (2.95) | 5.09 (2.51) | 4.92 (2.25) | 8.12 (4.70) |
| **Mean** | **9.3** | **7** | **12.6** | **10** | **5.5** | **9.4** | **4.3** | **4.3** | **6.9** | **4.3** | **4.2** | **6.6** | **100.5** | **5.46** | **4.88** | **4.60** | **7.42** |
| **SD** | **2.1** | **1.3** | **3** | **2.2** | **1.3** | **2.5** | **0.7** | **0.7** | **1.7** | **0.8** | **0.5** | **1.3** | **8.5** | **0.09**$^{s.e.}$ | **0.08**$^{s.e.}$ | **0.08**$^{s.e.}$ | **0.16**$^{s.e.}$ |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Table 8. **List of specific alleles observed only in Ethiopian or reference populations.**

| Locus | Allele size | Type | Populations observed in |
|---|---|---|---|
| BM1818 | 246 | Ethiopian specific | Afar and Long-eared Somali |
| BMS1494 | 288 | Ethiopian specific | Keffa and Long-eared Somali |
| ILSTS05 | 186 | Ethiopian specific | Afar and Short-eared Somali |
| ILSTS87 | 163 | Ethiopian specific | ABG, North-West Highland and Woyto-Guji |
| INRA63 | 157 | Ethiopian specific | Gumez and Keffa |
| MAF209 | 108 | Ethiopian specific | Hararghe Highland and Keffa |
| OarAE129 | 145 | Ethiopian specific | Long-eared Somali and North West Highland |
| OarAE129 | 175 | Ethiopian specific | ABG, CEH, Gumez, Keffa, LES, NWH, SES and WGuji |
| OarFCB304 | 124 | Ethiopian specific | ABG, AFA, CEH, HAH and NWH |
| OarFCB304 | 174 | Ethiopian specific | Long-eared Somali and Woyto-Guji |
| ILSTS05 | 188 | Reference population specific | Hair and Mongolia |
| ILSTS05 | 190 | Reference population specific | Ardi, Baladie and Mongolia |
| ILSTS44 | 169 | Reference population specific | Tswana, Alpine, Ardi, Baladie, Hair and Mongolia |
| ILSTS87 | 141 | Reference population specific | Hair and Mongolia |
| ILSTS87 | 143 | Reference population specific | Ardi, Baladie, Hair and Mongolia |
| ILSTS87 | 153 | Reference population specific | SEA, GUB Tswana Alpine, Baladie, Hair and Mongolia |
| ILSTS87 | 157 | Reference population specific | Ardi and Hair |
| INRA132 | 157 | Reference population specific | Tswana, Baladie and Mongolia |
| MAF65 | 140 | Reference population specific | Guinea Bissau and Hair |

Table 9. **Expected heterozygosity (*H*E), observed heterozygosity (*H*O) and mean number of alleles (MNA) of all 20 populations.**

| Population | Sample size | MNA (All ind) | MNA (20 ind) | $H_E$ ± s.e. | $H_O$ ± s.e. |
|---|---|---|---|---|---|
| Abergalle | 46 | 5.33 | 4.65 | 0.57 ± 0.056 | 0.54 ± 0.053 |
| Arsi-Bale | 46 | 5.33 | 4.87 | 0.61 ± 0.043 | 0.58 ± 0.045 |
| Afar | 44 | 5.47 | 4.77 | 0.56 ± 0.051 | 0.52 ± 0.049 |
| Central Highland | 40 | 5.13 | 4.65 | 0.57 ± 0.056 | 0.53 ± 0.058 |
| Gumez | 43 | 5.53 | 4.71 | 0.55 ± 0.057 | 0.55 ± 0.054 |
| Hararghe Highland | 40 | 5.93 | 5.10 | 0.58 ± 0.052 | 0.55 ± 0.053 |
| Keffa | 38 | 5.60 | 5.05 | 0.58 ± 0.050 | 0.56 ± 0.047 |
| Long-eared Somali | 48 | 6.47 | 5.28 | 0.59 ± 0.044 | 0.56 ± 0.044 |
| North West Highland | 45 | 5.40 | 4.67 | 0.57 ± 0.053 | 0.54 ± 0.049 |
| Small-eared Somali | 43 | 5.60 | 4.90 | 0.58 ± 0.054 | 0.57 ± 0.055 |
| Woyto-Guji | 46 | 5.67 | 4.99 | 0.58 ± 0.052 | 0.54 ± 0.049 |
| **Ethiopian populations** | | **5.59** | **4.88** | **0.58 ± 0.015** | **0.55 ± 0.015** |
| Boran-Galla | 35 | 5.20 | 4.63 | 0.55 ± 0.055 | 0.52 ± 0.049 |
| Small East Africa | 39 | 5.00 | 4.64 | 0.55 ± 0.058 | 0.53 ± 0.055 |
| Guinea Bissau | 46 | 4.67 | 4.19 | 0.52 ± 0.058 | 0.53 ± 0.057 |
| Tswana | 29 | 5.27 | 4.89 | 0.58 ± 0.045 | 0.52 ± 0.047 |
| Italian Alpine | 31 | 4.73 | 4.40 | 0.56 ± 0.055 | 0.53 ±0.057 |
| Ardi | 25 | 5.27 | 4.97 | 0.58 ± 0.037 | 0.63 ±0.050 |
| Baladie | 22 | 5.53 | 5.45 | 0.62 ± 0.050 | 0.61 ± 0.055 |
| Hair | 31 | 6.13 | 5.74 | 0.64 ± 0.055 | 0.62 ± 0.061 |
| Mongolia | 39 | 5.87 | 5.09 | 0.58 ± 0.057 | 0.57 ± 0.061 |
| **Reference population** | | **5.30** | **4.89** | **0.58 ± 0.017** | **0.561 ± 0.018** |
| Mean of all populations | | 5.46 ± 0.158 | 4.88 ± 0.08 | 0.58 ± 0.011 | 0.56 ± 0.012 |

### 4.4. Genetic relationship between the populations

Two measures of genetic distances, Nei's standard genetic distance **(**$D_S$) and Nei's $D_A$ genetic distance were used to estimate the genetic relationship for each pair of populations.

#### 4.4.1. Nei's standard genetic distance ($D_S$)

The Nei's standard genetic distance ($D_S$) values for the 20 populations are shown in **Table 10**. The smallest $D_S$ values were between Kenyan Boran Galla and Ethiopian Short-eared Somali (0.000 ± 0.002). The next smallest $D_S$ values were observed between Ethiopian Short-eared Somali & Hararghe Highland and Hair (from Turkey) & Baladie (from Egypt) (0.001 ± 0.004 and 0.002 ± 0.005), respectively, while the Highest $D_S$ was observed between Mongolia and Guinea Bissau populations (0.427 ± 0.152).

#### 4.4.2. Nei's $D_A$ genetic distances

When the genetic distances of the populations in the study were determined using Nei's $D_A$ index of genetic distance, the smallest $D_A$ was observed between Ethiopian Short-eared Somali and Hararghe Highland (0.024), the next smallest $D_A$ was between Ethiopian North-West Highland & Central Highland (0.026), followed by the next smallest $D_A$ between Boran-Galla of Kenya and Short-eared Somali of Ethiopia (0.028), while the highest $D_A$ was observed between Mongolia and Guinea Bissau populations (0.310) as in the case for the Nei's $D_S$ genetic distance (**Table 11**).

Table 10. **Nei's standard genetic distance (*D*S) for 11 Ethiopian and 9 reference populations/breeds genotyped, using 15 microsatellite markers (*D*S below the diagonal, standard errors above the diagonal).**

| | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | 0.013 | 0.018 | 0.008 | 0.014 | 0.014 | 0.016 | 0.014 | 0.007 | 0.013 | 0.014 | 0.019 | 0.022 | 0.078 | 0.068 | 0.088 | 0.060 | 0.038 | 0.049 | 0.097 |
| ABA | 0.034 | - | 0.018 | 0.006 | 0.009 | 0.007 | 0.011 | 0.005 | 0.006 | 0.007 | 0.007 | 0.013 | 0.017 | 0.080 | 0.046 | 0.071 | 0.044 | 0.025 | 0.036 | 0.073 |
| AFA | 0.055 | 0.047 | - | 0.014 | 0.016 | 0.006 | 0.025 | 0.011 | 0.019 | 0.005 | 0.014 | 0.009 | 0.026 | 0.080 | 0.046 | 0.075 | 0.033 | 0.022 | 0.033 | 0.085 |
| CEH | 0.010 | 0.019 | 0.035 | - | 0.012 | 0.005 | 0.007 | 0.008 | 0.004 | 0.006 | 0.013 | 0.014 | 0.026 | 0.085 | 0.056 | 0.066 | 0.047 | 0.028 | 0.035 | 0.075 |
| GUM | 0.031 | 0.031 | 0.046 | 0.030 | - | 0.015 | 0.013 | 0.005 | 0.009 | 0.009 | 0.008 | 0.011 | 0.026 | 0.086 | 0.059 | 0.070 | 0.043 | 0.027 | 0.041 | 0.080 |
| HAH | 0.031 | 0.014 | 0.010 | 0.008 | 0.033 | - | 0.010 | 0.005 | 0.008 | 0.004 | 0.012 | 0.008 | 0.026 | 0.077 | 0.043 | 0.067 | 0.037 | 0.023 | 0.029 | 0.074 |
| KEF | 0.036 | 0.023 | 0.046 | 0.022 | 0.036 | 0.021 | - | 0.009 | 0.007 | 0.012 | 0.011 | 0.022 | 0.022 | 0.071 | 0.040 | 0.076 | 0.057 | 0.029 | 0.040 | 0.091 |
| LES | 0.037 | 0.012 | 0.026 | 0.026 | 0.019 | 0.011 | 0.030 | - | 0.006 | 0.003 | 0.007 | 0.005 | 0.014 | 0.081 | 0.040 | 0.061 | 0.037 | 0.019 | 0.023 | 0.066 |
| NWH | 0.014 | 0.011 | 0.035 | 0.004 | 0.022 | 0.012 | 0.012 | 0.016 | - | 0.007 | 0.008 | 0.015 | 0.019 | 0.085 | 0.046 | 0.068 | 0.050 | 0.025 | 0.032 | 0.074 |
| SES | 0.032 | 0.019 | 0.014 | 0.017 | 0.024 | 0.001 | 0.030 | 0.005 | 0.007 | - | 0.006 | 0.002 | 0.019 | 0.078 | 0.042 | 0.059 | 0.037 | 0.022 | 0.023 | 0.062 |
| WGJ | 0.034 | 0.012 | 0.033 | 0.036 | 0.021 | 0.020 | 0.022 | 0.011 | 0.019 | 0.015 | - | 0.006 | 0.008 | 0.063 | 0.042 | 0.080 | 0.046 | 0.025 | 0.035 | 0.090 |
| BGA | 0.049 | 0.034 | 0.020 | 0.041 | 0.031 | 0.015 | 0.052 | 0.011 | 0.025 | 0.000 | 0.013 | - | 0.014 | 0.073 | 0.048 | 0.067 | 0.044 | 0.025 | 0.028 | 0.070 |
| SEA | 0.057 | 0.055 | 0.055 | 0.065 | 0.058 | 0.053 | 0.056 | 0.037 | 0.045 | 0.044 | 0.020 | 0.034 | - | 0.058 | 0.053 | 0.086 | 0.059 | 0.041 | 0.047 | 0.093 |
| GUB | 0.220 | 0.211 | 0.201 | 0.222 | 0.202 | 0.200 | 0.211 | 0.210 | 0.224 | 0.198 | 0.163 | 0.178 | 0.159 | - | 0.124 | 0.138 | 0.091 | 0.088 | 0.102 | 0.152 |
| TSW | 0.212 | 0.176 | 0.179 | 0.193 | 0.206 | 0.152 | 0.164 | 0.169 | 0.167 | 0.152 | 0.163 | 0.177 | 0.203 | 0.358 | - | 0.059 | 0.073 | 0.043 | 0.033 | 0.072 |
| ALP | 0.232 | 0.196 | 0.190 | 0.196 | 0.184 | 0.165 | 0.205 | 0.154 | 0.173 | 0.140 | 0.189 | 0.159 | 0.220 | 0.353 | 0.122 | - | 0.049 | 0.043 | 0.031 | 0.039 |
| ARD | 0.178 | 0.133 | 0.106 | 0.128 | 0.137 | 0.102 | 0.158 | 0.106 | 0.141 | 0.109 | 0.152 | 0.132 | 0.190 | 0.251 | 0.206 | 0.139 | - | 0.036 | 0.026 | 0.069 |
| BAL | 0.095 | 0.079 | 0.080 | 0.077 | 0.080 | 0.063 | 0.084 | 0.062 | 0.068 | 0.057 | 0.078 | 0.076 | 0.122 | 0.235 | 0.108 | 0.106 | 0.079 | - | 0.005 | 0.034 |
| HAI | 0.148 | 0.123 | 0.115 | 0.119 | 0.143 | 0.096 | 0.138 | 0.096 | 0.109 | 0.081 | 0.129 | 0.109 | 0.170 | 0.272 | 0.084 | 0.070 | 0.082 | 0.002 | - | 0.032 |
| MON | 0.239 | 0.205 | 0.206 | 0.199 | 0.206 | 0.187 | 0.232 | 0.174 | 0.186 | 0.159 | 0.225 | 0.183 | 0.252 | 0.427 | 0.193 | 0.089 | 0.172 | 0.077 | 0.070 | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia.

Table 11. **Nei's *D*A genetic distance for 11 Ethiopian and 9 reference populations/breeds genotyped, using 15 microsatellite markers.**

| | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | | | | | | | | | | | | | | | | | | | |
| ABA | 0.049 | - | | | | | | | | | | | | | | | | | | |
| AFA | 0.061 | 0.060 | - | | | | | | | | | | | | | | | | | |
| CEH | 0.034 | 0.039 | 0.051 | - | | | | | | | | | | | | | | | | |
| GUM | 0.054 | 0.049 | 0.056 | 0.055 | - | | | | | | | | | | | | | | | |
| HAH | 0.048 | 0.040 | 0.038 | 0.040 | 0.049 | - | | | | | | | | | | | | | | |
| KEF | 0.055 | 0.050 | 0.065 | 0.051 | 0.054 | 0.051 | - | | | | | | | | | | | | | |
| LES | 0.059 | 0.041 | 0.052 | 0.058 | 0.050 | 0.041 | 0.057 | - | | | | | | | | | | | | |
| NWH | 0.034 | 0.044 | 0.051 | 0.026 | 0.048 | 0.044 | 0.044 | 0.052 | - | | | | | | | | | | | |
| SES | 0.048 | 0.037 | 0.033 | 0.038 | 0.052 | 0.024 | 0.051 | 0.033 | 0.033 | - | | | | | | | | | | |
| WGJ | 0.051 | 0.030 | 0.043 | 0.050 | 0.044 | 0.041 | 0.046 | 0.035 | 0.041 | 0.033 | - | | | | | | | | | |
| BGA | 0.066 | 0.047 | 0.043 | 0.058 | 0.049 | 0.038 | 0.065 | 0.037 | 0.053 | 0.028 | 0.035 | - | | | | | | | | |
| SEA | 0.069 | 0.066 | 0.060 | 0.067 | 0.074 | 0.059 | 0.065 | 0.050 | 0.058 | 0.053 | 0.042 | 0.055 | - | | | | | | | |
| GUB | 0.216 | 0.210 | 0.198 | 0.225 | 0.191 | 0.195 | 0.198 | 0.179 | 0.227 | 0.193 | 0.170 | 0.181 | 0.159 | - | | | | | | |
| TSW | 0.152 | 0.137 | 0.143 | 0.156 | 0.138 | 0.122 | 0.147 | 0.121 | 0.137 | 0.125 | 0.127 | 0.151 | 0.143 | 0.234 | - | | | | | |
| ALP | 0.178 | 0.165 | 0.161 | 0.175 | 0.141 | 0.145 | 0.169 | 0.139 | 0.157 | 0.143 | 0.152 | 0.154 | 0.169 | 0.246 | 0.120 | - | | | | |
| ARD | 0.171 | 0.144 | 0.123 | 0.164 | 0.133 | 0.128 | 0.178 | 0.128 | 0.164 | 0.137 | 0.144 | 0.140 | 0.175 | 0.222 | 0.180 | 0.158 | - | | | |
| BAL | 0.118 | 0.111 | 0.100 | 0.120 | 0.093 | 0.101 | 0.128 | 0.090 | 0.105 | 0.094 | 0.101 | 0.104 | 0.125 | 0.216 | 0.117 | 0.133 | 0.123 | - | | |
| HAI | 0.131 | 0.127 | 0.122 | 0.133 | 0.128 | 0.114 | 0.141 | 0.094 | 0.119 | 0.099 | 0.112 | 0.116 | 0.131 | 0.221 | 0.106 | 0.119 | 0.129 | 0.061 | - | |
| MON | 0.198 | 0.182 | 0.183 | 0.195 | 0.172 | 0.171 | 0.215 | 0.160 | 0.183 | 0.160 | 0.190 | 0.176 | 0.203 | 0.310 | 0.178 | 0.149 | 0.203 | 0.113 | 0.104 | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia.

### 4.4.3. Phylogenetic analysis

The phylogenetic relationships between the populations in the study were determined using $D_S$ and $D_A$ genetic distances and their UPGMA and neighbour joining trees are shown in **Fig. 22** to **25**. The $D_S$ neighbour joining phylogenetic tree (**Figure 22**) for all populations revealed three clusters. The first cluster comprised of six reference populations (Italian Alpine, Mongolia, Hair, Tswana, Baladie and Ardi), the second cluster was made up of Afar, Haraghe Highland, Short-eared Somali, Long-eared Somali, Gumez, Arsi-Bale, Keffa, North-west Highland, Abergalle, and Central Highland, and the third cluster was made up of Boran Galla, Small-east Africa (both from Kenya), Woyto-Guji (from Ethiopia), and Guinea Bissau (from West Africa). The bootstrap values (number of times a node was observed in 1000 replicates of resample loci) ranged from 3% to 76%. The $D_A$ neighbour joining phylogenetic tree (**Figure 23**) for all populations also revealed three, but more clear clusters. The first cluster was made up of Abergalle, North-west Highland, Central Highland, Keffa, Gumez, Arsi-Bale, Woyto-Guji, Boran Galla, Short-eared Somali, Afar, Haraghe Highland, Long-eared Somali and Small-East Africa, the second cluster comprised of six reference populations (Italian Alpine, Mongolia, Hair, Tswana, Baladie and Ardi), and the third cluster was made up of only Guinea Bissau. In general, the bootstrap values observed for the $D_A$ (**Figure 23)** were relatively lower (from 16% to 68%). The UPGMA phylogenetic tree from $D_S$ genetic distances (**Figure 24**) for all populations also revealed three clear clusters similar to the $D_A$ phylogenetic tree with the bootstrap values relatively higher than for both $D_S$ and $D_A$ neighbour joining trees (from 8% to 99%). The UPGMA phylogenetic tree from $D_A$ genetic distances (**Figure 25**) for all populations also revealed three clusters very similar to the $D_A$ neighbour joining phylogenetic tree, but clustering the Ardi population as an intermediate fourth cluster between Guinea Bissau and the other five reference populations from which it is subdivided; with the bootstrap values relatively higher than for all the other three trees (from 17% to 99%). In all cases, the Ethiopian goat populations cluster more closely together with the two Kenyan populations than to other reference populations, suggesting a common ancestry for them.

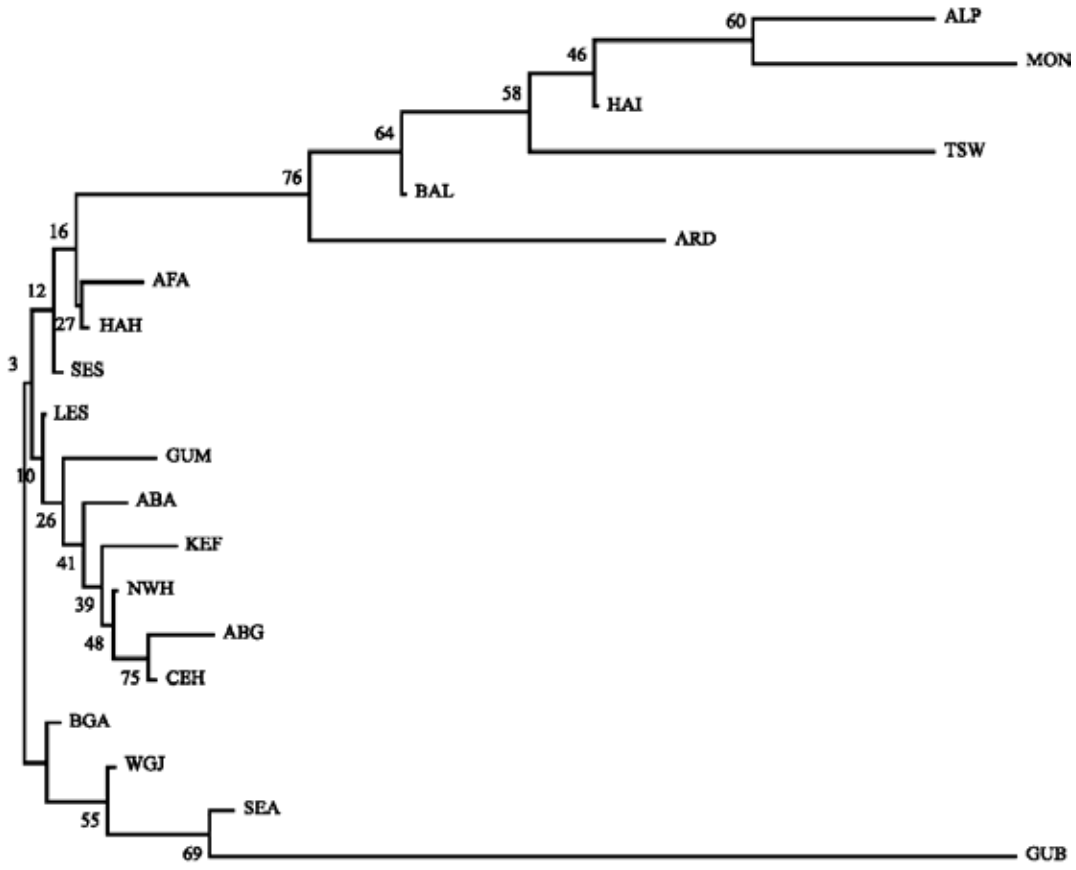


Figure 22. **Phylogenetic tree representing the genetic relationships between the 11 Ethiopian populations and 9 reference populations.** The NJ tree was constructed using the genetic distance *D*S calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci**.**

(ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

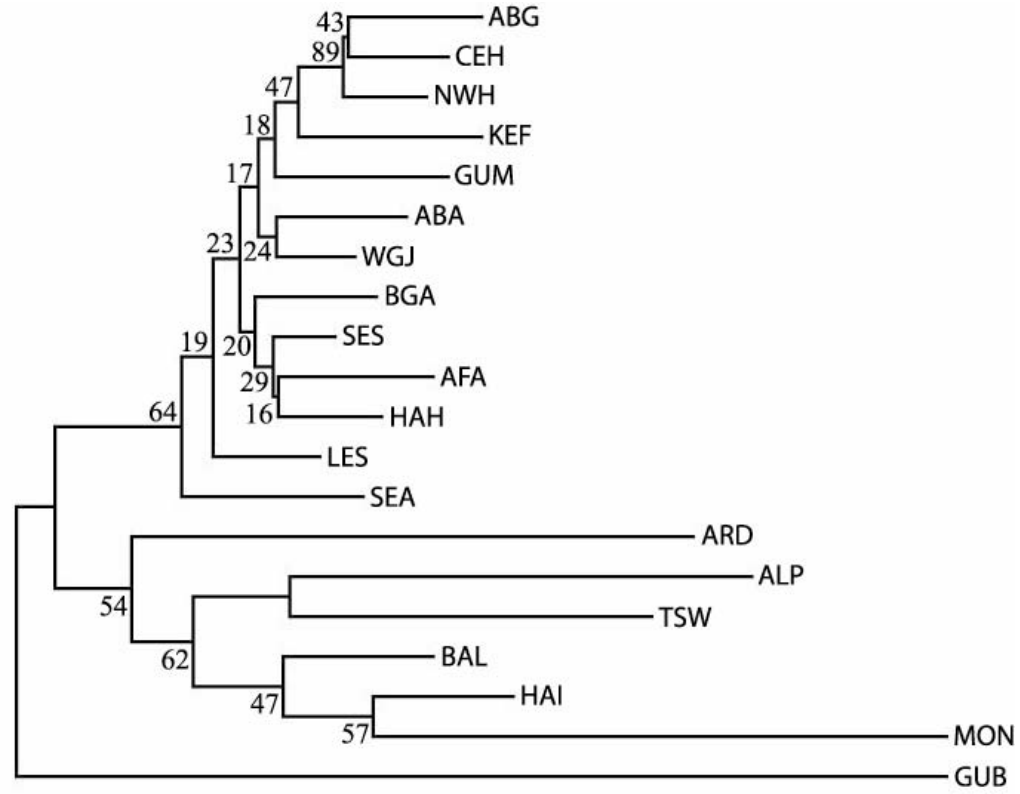


Figure 23. **Phylogenetic tree representing the genetic relationships between the 11 Ethiopian populations and 9 reference populations.** The NJ tree was constructed using the genetic distance *D*A calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

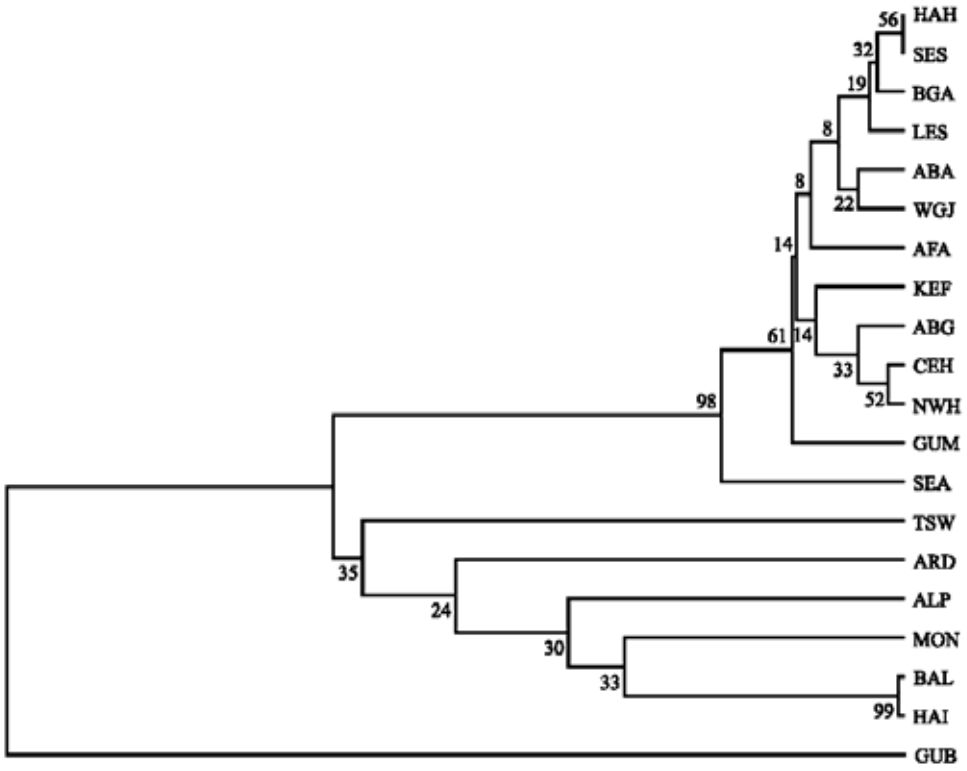


Figure 24. **Phylogenetic tree representing the genetic relationships between the 11 Ethiopian populations and 9 reference populations.** The UPGMA tree was constructed using the genetic distance DS calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci

(ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

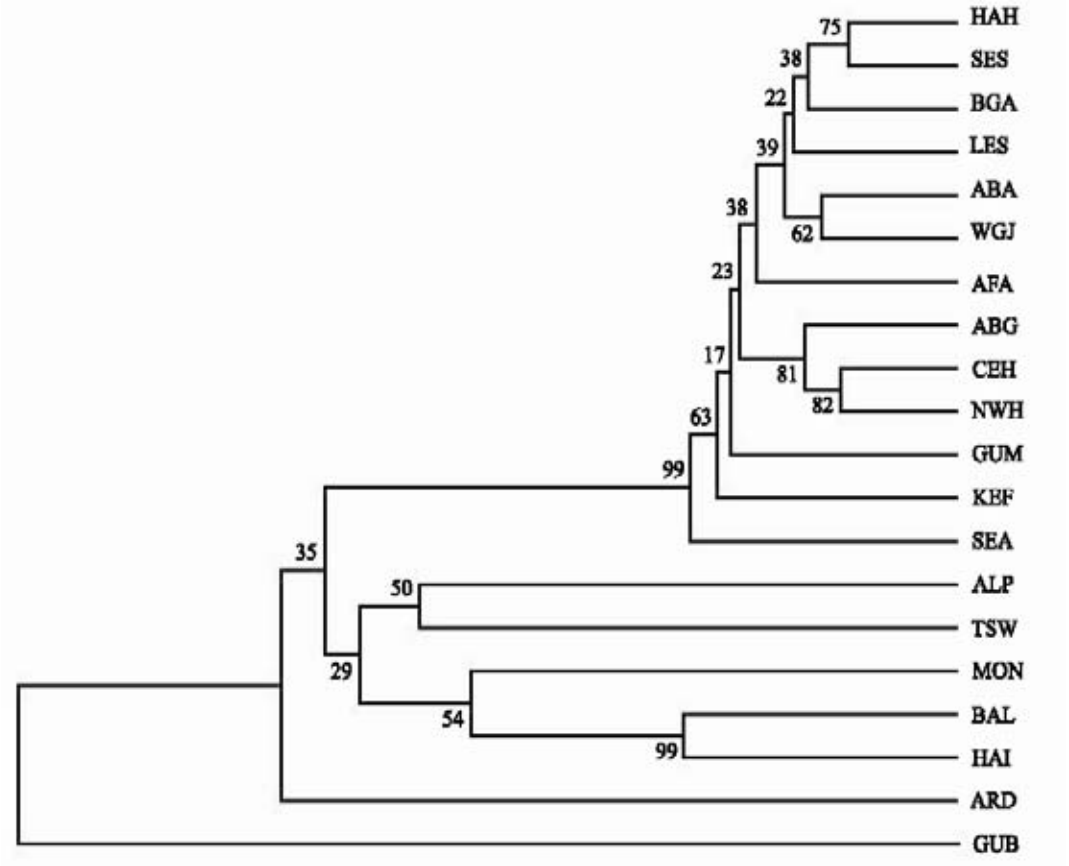


Figure 25. **Phylogenetic tree representing the genetic relationships between the 11 Ethiopian populations and 9 reference populations.** The UPGMA tree was constructed using the genetic distance *D*A calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

### 4.4.4. Principal component analysis (PCA)

The results of the PCA of the allele frequency data for the 11 Ethiopian and 9 other reference populations using 15 microsatellite markers are shown in **Table 12**. The variation accounted for by PC1, PC2 and PC3 was 31.81%, 18.04% and 10.38%, respectively, whish add up to 60.23%. The PC1 separated the 11 Ethiopian, 2 Kenyan

and 1 Guinea Bissau population from the other six (Mongolia, Italian Alpine, Hair, Tswana, Ardi and Baladie) populations on the other extreme. The second principal component (PC2) ranked Guinea Bissau, Ardi, Small-east Africa, Hair and Italian Alpine on one side and the remaining 10 Ethiopian, one Egyptian, one Botswana, one Kenyan and one Mongolian on the other extreme leaving Woyto-Guji population just in the center of the axis. The third principal component (PC3) ranked Ardi, Mongolia, Afar, Baladie, Central Highland, Gumez, Long-eared Somali, Hararghe Highland, Boran-Galla, Short-eared Somali, and Arsi-Bale on one side and Abergalle, Hair, North-West Highland, Guinea Bissau, Italian Alpine, Woyto-Guji, Keffa, Small-East Africa, and Tswana on the other extreme.

**Figure 26, 27** and **28** are two-dimensional plots of PC1 with PC2, PC1 with PC3 and PC2 with PC3 (in figure form), respectively, showing the same thing from the **Table 12**.

Table 12. **Ranked principal component (PC) values of allele frequencies from 20 populations using 15 microsatellite markers (Log e).**

| % variation accounted | 31.81 | | 18.04 | | 10.38 |
|---|---|---|---|---|---|
| Cumulative % | 31.81 | | 49.85 | | 60.23 |
| Breed | PC1 | Breed | PC2 | Breed | PC3 |
| MON | -0.179 | GUB | -0.257 | ARD | -0.118 |
| ALP | -0.166 | ARD | -0.059 | MON | -0.05 |
| HAI | -0.126 | SEA | -0.03 | AFA | -0.026 |
| TSW | -0.118 | HAI | -0.009 | BAL | -0.02 |
| ARD | -0.085 | ALP | -0.006 | CEH | -0.018 |
| BAL | -0.081 | WGJ | 0 | GUM | -0.017 |
| SES | 0.032 | BAL | 0.001 | LES | -0.013 |
| LES | 0.035 | BGA | 0.005 | HAH | -0.011 |
| HAH | 0.04 | AFA | 0.009 | BGA | -0.008 |
| Afar | 0.044 | TSW | 0.009 | SES | -0.008 |
| BGA | 0.046 | GUM | 0.02 | ABA | -0.001 |
| GUB | 0.052 | LES | 0.023 | ABG | 0.001 |
| ABA | 0.054 | SES | 0.026 | HAI | 0.001 |
| NWH | 0.054 | HAH | 0.027 | NWH | 0.003 |
| GUM | 0.057 | KEF | 0.031 | GUB | 0.016 |
| CEH | 0.06 | ABA | 0.034 | ALP | 0.023 |
| WGJ | 0.064 | ABG | 0.04 | WGJ | 0.025 |
| KEF | 0.065 | MON | 0.041 | KEF | 0.028 |
| SEA | 0.075 | CEH | 0.045 | SEA | 0.033 |
| ABG | 0.078 | NWH | 0.05 | TSW | 0.161 |

ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Short-Ear Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia.

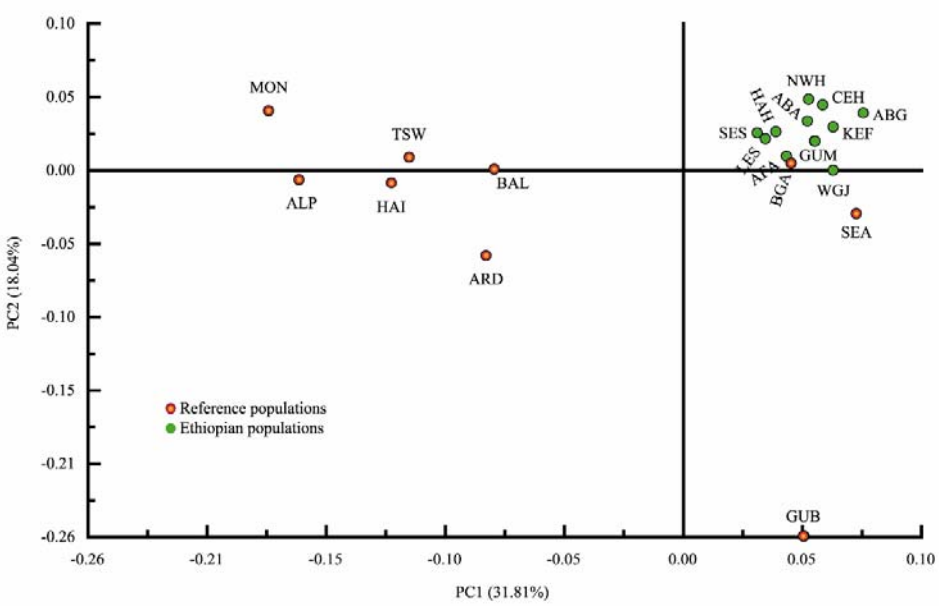


Figure 26. **Two-dimensional scatter plot representing the relationship between PC1 and PC2 using allele frequency data from 15 microsatellite loci typed in 20 populations.**

(ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

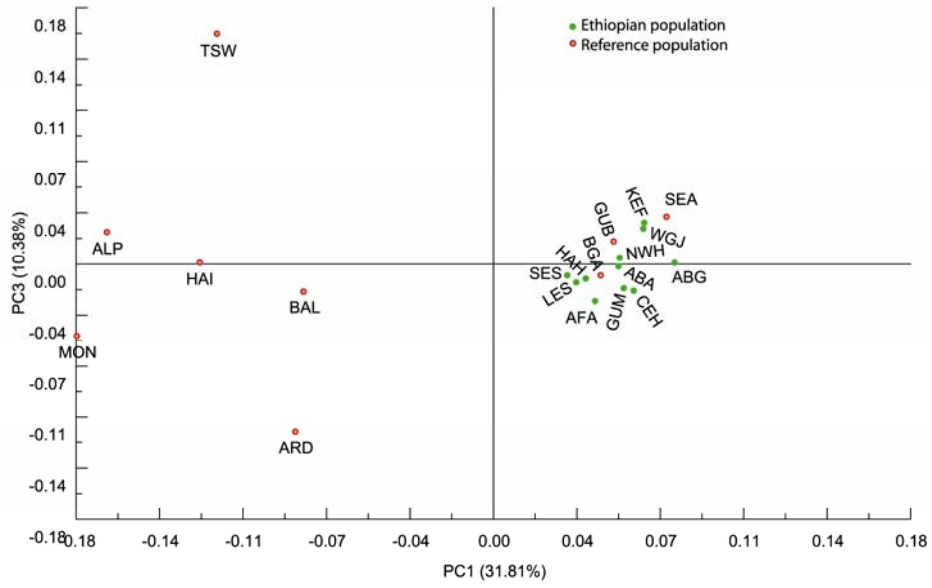


Figure 27. **Two-dimensional scatter plot representing the relationship between PC1 and PC3 using allele frequency data from 15 microsatellite loci typed in 20 populations.**

(ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

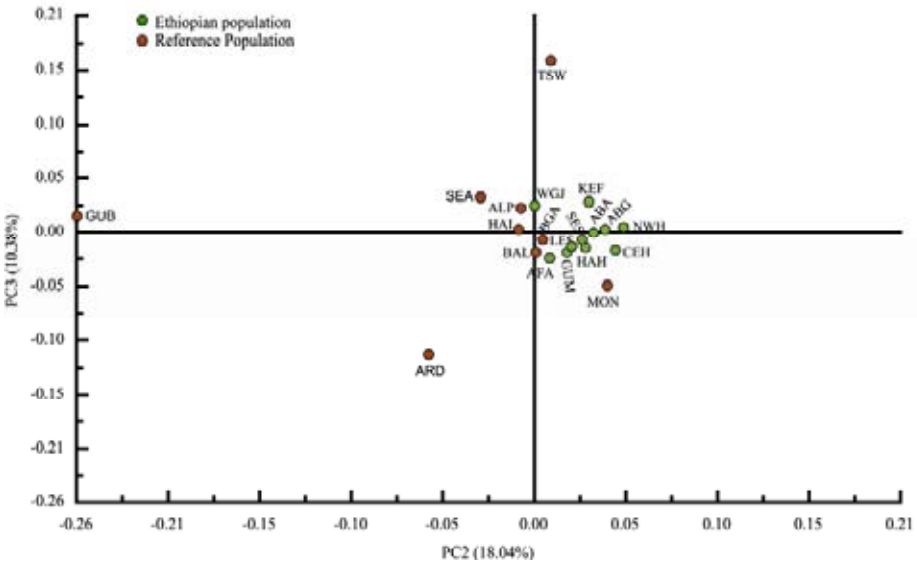


Figure 28. **Two-dimensional scatter plot representing the relationship between PC2 and PC3 using allele frequency data from 15 microsatellite loci typed in 20 populations.**

(ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

### 4.4.5. Inference of population structure

Pritchard *et al*. (2000) developed a model-based clustering approach to infer population structure and assign individuals to populations using multilocus genotype data, which identifies subgroups that have distinctive allele frequencies. This model-based algorithm detailed in Pritchard *et al*. (2000) and Falush *et al*. (2003), places individuals in to K clusters. K is a parameter that is chosen in advance, but that can be varied in independent structure runs. Individuals are given 'membership coefficient' for each cluster, such that the estimated membership coefficient of an individual sum to 1 across the K clusters. The matrix of membership coefficients, where the number of individuals is the number of rows and K is the number of columns, is referred to as the *individual Q-matrix.* For each population, membership coefficients for each cluster can

be averaged across individuals to create a *population Q-matrix.* The procedure was implemented using computer program STRUCTURE version 2.1 (available at http://pritch.bsd.uchicago.edu/software/structure2_1.html) with models assuming that individuals have mixed ancestry (admixture ancestry model; Pritchard *et al.*, 2000) and allele frequencies are correlated among populations (Falush *et al*., 2003) with default parameters. A burn-in period of 100,000 iterations of Markov chain Monte Carlo (MCMC) followed by 500,000 MCMC replications was used for running the structure program, and this parameter produced consistent results in five runs of each K. The estimated individual membership coefficients were graphically visualized using a computer program DISTRUCT (Rosenberg, 2004; available from http://www.cmb.usc.edu/~noahr/distruct.html). DISTRUCT is a program for the graphical display of population structure. Clusters are presented as colors, and individuals are depicted as bars partitioned into K colored segments that correspond to membership coefficient in K clusters.

**Figure 29** and **30** show the population structure, displayed with population Q-matrix and individual Q-matrix, respectively, inferred from allelic frequencies at 15 microsatellite loci. When K = 2 was applied, all the Ethiopian goat populations were clustered with the 2 Kenyan goat populations while all the remaining reference populations are clustered in the other cluster. When K = 3 was applied, all the Ethiopian goat populations and the 2 Kenyan goat populations remain the same while the West African Guinea Bissau was separated from the remaining reference populations forming the 3$^{rd}$ cluster. Then after K = 4 and K = 5 yielded no further population clustering. There was no clear population structure inferred from Ethiopian populations.

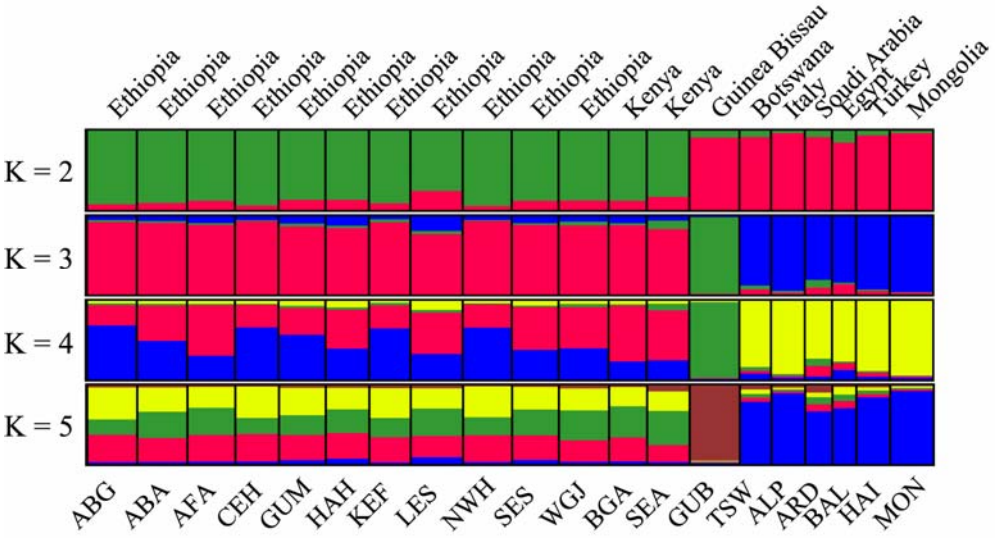


Figure 29. **Estimated population structure displayed with population Q-matrix. Populations were partitioned in to K-coloured segments, which represent population's estimated average membership fractions in the K clusters. The black lines separate the different populations used for the study. Populations are labeled below the figure with their country affiliations above the figure. Five runs at each K produced nearly similar population membership coefficients.**

**(**ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

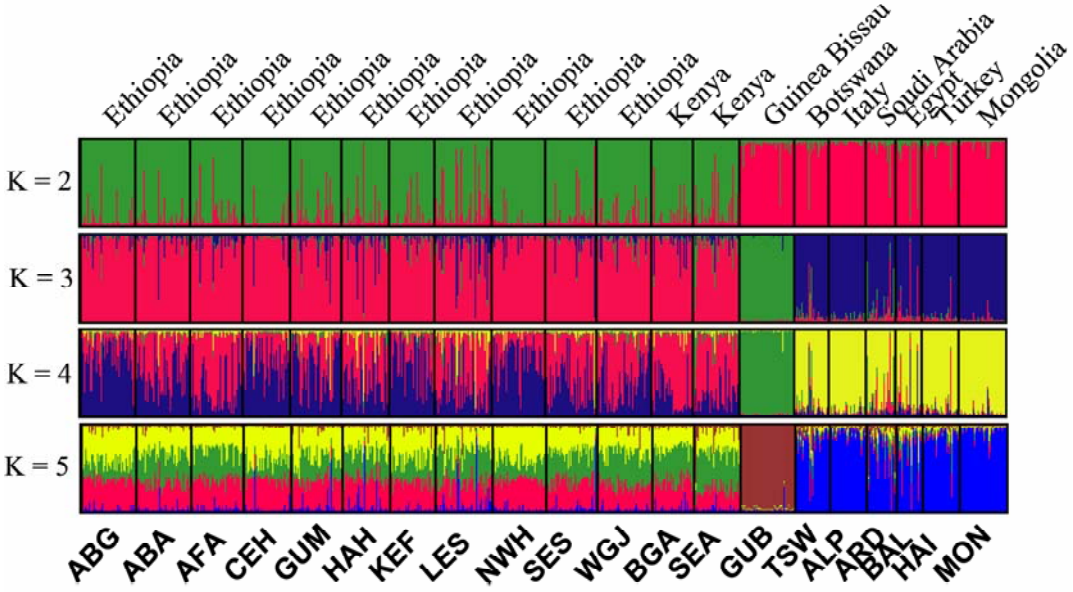

Figure 30. **Estimated population structure (graph) displayed with individual Q-matrix. Each individual was represented by a thin vertical line segment, which is partitioned in to K coloured segments, which represent individual's estimated average membership fractions in the K clusters. The black segments separate the individuals of different populations used for the study. Population names are below the figure with their country affiliations above the figure. Five runs at each K produced nearly similar individual membership coefficients.**

**(**ABG – Abergalle, ABA – Arsi-Bale, AFA - Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF– Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, BGA – Boran-Galla, SEA – Small East African, GUB – Guinea Bissau, TSW– Tswana, ALP – Italian Alpine, ARD - Ardi, BAL – Baladie, HAI - Hair, MON – Mongolia)

### 4.5. Genetic differentiation

#### 4.5.1. $G_{ST}$ and $F_{ST}$

The $G_{ST}$ estimates were slightly higher than the $F_{ST}$ values obtained for all the 15 loci (**Table 13**). The $G_{ST}$ values ranged from 0.014 for locus MAF35 to 0.162 for locus ILSTS87, while $F_{ST}$ ranged from 0.003 for locus MAF35 to 0.152 for locus ILSTS87. The overall $G_{ST}$ and $F_{ST}$ values indicate that 92.4% – 93.6% of the total genetic differentiation was due to variation between individuals within populations when all populations used in the study are considered together.

The $F_{ST}$ values and their level of significance for pairs of populations are shown in **Table 14**. The population pairs that were not genetically differentiated ($P > 0.05$) were the Afar/Hararghe Highland, Central Highland/Hararghe Highland, Hararghe Highland/Long-eared Somali, Central Highland/North West Highland, Afar/Short-eared Somali, Hararghe Highland/Short-eared Somali, Long-eared Somali/Short-eared Somali, North West Highland/Short-eared Somali, Arsi-Bale/Woyto-Guji, Long-eared Somali/Woyto-Guji, Hararghe Highland/Boran Galla, Long-eared Somali/Boran Galla, Short-eared Somali/Boran Galla, Woyto-Guji/ Boran Galla and Boran-Galla/Hair. The genetic differentiation observed among the remaining population pairs were highly significant ($P < 0.01$) although the $F_{ST}$ values were low except for two population pairs of Abergalle/Central Highland and Short-eared Somali/Woyto-Guji with significant levels at $P < 0.05$, indicating that there were significant genetic differentiation between these populations.

The $F_{ST}$ values were calculated for each locus to determine their ability to differentiate each pair of the populations. The efficiency of the 15 microsatellite loci to differentiate between the 190 population pair combinations is shown in **Table 15.** Out of the 15 loci, OarFCB304 was able to differentiate 85.26% of the population pair combinations (162 out of 190 population pair combinations) at $P < 0.05$ or lower, making it the most informative marker. The next informative markers were ILSTS11 and ILSTS87, which were able to differentiate 80% of the population pairs (152 out of 190 population pair combinations) at $P < 0.05$ or lower. In general, all the 15 loci were desirable at distinguishing between the population pairs, except for MAF35 which was able to differentiate only 13.16% of the population pairs (25 out of 190 population pair combinations) at $P < 0.05$ or lower, making it the least informative marker (**Appendix 18**). The genetic differentiation of the various goat populations studied is shown in the table below.

Table 13. **Measure of genetic differentiation in Ethiopian and reference goat populations.**

| Locus | $G_{ST}$ | $F_{ST}$ |
|---|---|---|
| BM1818 | 0.057 | 0.044 |
| BMS1494 | 0.102 | 0.101 |
| ILSTS05 | 0.099 | 0.085 |
| ILSTS11 | 0.076 | 0.066 |
| ILSTS44 | 0.133 | 0.121 |
| ILSTS87 | 0.162 | 0.152 |
| INRA05 | 0.085 | 0.074 |
| INRA63 | 0.050 | 0.033 |
| INRA132 | 0.052 | 0.036 |
| MAF35 | 0.014 | 0.003 |
| MAF65 | 0.057 | 0.042 |
| MAF209 | 0.066 | 0.053 |
| OarAE129 | 0.070 | 0.052 |
| OarFCB304 | 0.053 | 0.040 |
| SRCRSP03 | 0.061 | 0.050 |
| **Overall loci** | **0.076** | **0.064** |

Table 14. **Pair-wise tests of population differentiation between the 11 Ethiopian populations and 9 reference breeds. *F*ST values (below the diagonal) and their test of significance (above the diagonal) obtained after 190,000 permutations.**

| | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | ** | ** | * | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| ABA | 0.023 | - | ** | ** | ** | ** | ** | ** | ** | ** | ns | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| AFA | 0.04 | 0.033 | - | ** | ** | ns | ** | ** | ** | ns | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| CEH | 0.007 | 0.013 | 0.025 | - | ** | ns | ** | ** | ns | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| GUM | 0.024 | 0.023 | 0.034 | 0.022 | - | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| HAH | 0.022 | 0.009 | 0.007 | 0.005 | 0.024 | - | ** | ns | ** | ns | ** | ns | ** | ** | ** | ** | ** | ** | ** | ** |
| KEF | 0.025 | 0.015 | 0.033 | 0.015 | 0.027 | 0.014 | - | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| LES | 0.026 | 0.008 | 0.018 | 0.018 | 0.015 | 0.007 | 0.02 | - | ** | ns | ns | ns | ** | ** | ** | ** | ** | ** | ** | ** |
| NWH | 0.01 | 0.008 | 0.025 | 0.002 | 0.016 | 0.008 | 0.008 | 0.011 | - | ns | ** | ** | ** | ** | ** | ** | ** | ** | ** | ** |
| SES | 0.022 | 0.013 | 0.01 | 0.011 | 0.018 | -0.001 | 0.021 | 0.003 | 0.005 | - | * | ns | ** | ** | ** | ** | ** | ** | ** | ** |
| WGJ | 0.024 | 0.008 | 0.023 | 0.025 | 0.016 | 0.013 | 0.015 | 0.007 | 0.013 | 0.010 | - | ns | ** | ** | ** | ** | ** | ** | ** | ** |
| BGA | 0.036 | 0.025 | 0.015 | 0.03 | 0.024 | 0.011 | 0.038 | 0.008 | 0.019 | 0.000 | 0.009 | - | ** | ** | ** | ** | ** | ** | ** | ** |
| SEA | 0.041 | 0.038 | 0.04 | 0.046 | 0.044 | 0.038 | 0.04 | 0.027 | 0.033 | 0.032 | 0.014 | 0.026 | - | ** | ** | ** | ** | ** | ** | ** |
| GUB | 0.142 | 0.13 | 0.134 | 0.143 | 0.137 | 0.13 | 0.135 | 0.132 | 0.142 | 0.129 | 0.11 | 0.125 | 0.113 | - | ** | ** | ** | ** | ** | ** |
| TSW | 0.124 | 0.098 | 0.11 | 0.113 | 0.125 | 0.092 | 0.097 | 0.098 | 0.1 | 0.091 | 0.097 | 0.111 | 0.123 | 0.199 | - | ** | ** | ** | ** | ** |
| ALP | 0.137 | 0.112 | 0.119 | 0.119 | 0.118 | 0.102 | 0.121 | 0.094 | 0.107 | 0.088 | 0.114 | 0.105 | 0.135 | 0.202 | 0.078 | - | ** | ** | ** | ** |
| ARD | 0.109 | 0.079 | 0.072 | 0.082 | 0.092 | 0.066 | 0.096 | 0.067 | 0.089 | 0.070 | 0.093 | 0.089 | 0.119 | 0.158 | 0.119 | 0.09 | - | ** | ** | ** |
| BAL | 0.06 | 0.046 | 0.053 | 0.049 | 0.056 | 0.04 | 0.052 | 0.038 | 0.044 | 0.037 | 0.049 | 0.053 | 0.078 | 0.143 | 0.063 | 0.066 | 0.05 | - | ns | ** |
| HAI | 0.085 | 0.066 | 0.071 | 0.07 | 0.087 | 0.057 | 0.077 | 0.055 | 0.065 | 0.049 | 0.074 | 0.069 | 0.099 | 0.153 | 0.049 | 0.045 | 0.05 | 0.000 | - | ** |
| MON | 0.136 | 0.112 | 0.123 | 0.116 | 0.125 | 0.109 | 0.129 | 0.1 | 0.11 | 0.095 | 0.127 | 0.114 | 0.145 | 0.221 | 0.111 | 0.059 | 0.103 | 0.048 | 0.043 | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON – Mongolia.

Table 15. **Summary of genotypic differentiation tests for all the 190 pairs of population combinations (20 populations) for all the 15 loci.**

| Locus | Percentage of population pairs that could be distinguished by the marker, $P < 0.05$ |
|---|---|
| BM1818 | 63.68 |
| BMS1494 | 60.00 |
| ILSTS05 | 64.21 |
| ILSTS11 | 80.00 |
| ILSTS44 | 62.11 |
| ILSTS87 | 80.00 |
| INRA05 | 61.58 |
| INRA63 | 62.11 |
| INRA132 | 52.63 |
| MAF35 | 13.16 |
| MAF65 | 78.95 |
| MAF209 | 60.53 |
| OarAE129 | 73.16 |
| OarFCB304 | 85.26 |
| SRCRSP03 | 65.26 |
| Average for all population | 64.18 |

### 4.5.2. Analysis of molecular variance (AMOVA)

In order to understand the partitioning of the levels of genetic diversity of the 20 goat populations, an analysis of molecular variance (AMOVA) was conducted. The result of the AMOVA (**Table 16**) revealed that 93.62% of the total genetic diversity existed among the individuals within populations and only 6.38% of the total genetic diversity accounted for differences among populations, which is exactly in accordance with the $F_{ST}$ value.

Subsequently three types of nested analyses were then carried out. The first nested analysis was carried out to assess the level of partitioning of genetic diversity between three different geographic regions: East Africa, West Africa and all the others (North Africa, Europe, Asia and Middle East). The second nested analysis was carried out to assess the level of partitioning of genetic diversity between eight different geographic regions: Ethiopia, Kenya, Botswana, Egypt, Guinea Bissau, Turkey plus Italy (as Europe), Mongolia and Saudi Arabia. The third nested analysis was carried out to assess the level of partitioning of genetic diversity between nine different geographic regions: Ethiopia, Kenya, Botswana, Egypt, Guinea Bissau, Turkey, Italy, Mongolia and Saudi Arabia. The 20 populations were grouped into any of these regions depending on the country where they were sampled. The results of the nested analyses (**Table 16**) show that differences between the three different geographic regions accounted for 7.20% of the variation, while differences among populations within these groups accounted for only 2.82% of the total variation. The differences between the eight different geographic regions accounted for 7.07% of the total variation, while differences among populations within these groups accounted for only 1.78% of the variation. The differences between the nine different geographic regions and among populations within these groups accounted for 7.32% and 1.59% of the total variation.

Table 16. **Analyses of molecular variance (AMOVA) for 20 goat populations using genotype data from 15 microsatellite loci.**

| Source of variation | Degree of freedom | Sum of squares | Variance components | Percentage of variation |
|---|---|---|---|---|
| Among populations | 19 | 504.66 | 0.2883 | 6.38 |
| Within populations | 1532 | 6483.266 | 4.2319 | 93.62 |
| **Nested analysis one** | | | | |
| Among 3[A] groups | 2 | 258.765 | 0.33845 | 7.2 |
| Among populations within groups | 17 | 245.895 | 13245 | 2.82 |
| Within populations | 1532 | 6483.266 | 4.2319 | 89.99 |
| **Nested analysis two** | | | | |
| Among 8[B] groups | 7 | 370.761 | 0.32823 | 7.07 |
| Among populations within groups | 12 | 133.899 | 0.0826 | 1.78 |
| Within populations | 1532 | 6483.266 | 4.2319 | 91.15 |
| **Nested analysis three** | | | | |
| Among 9[C] groups | 8 | 388.14 | 0.33997 | 7.32 |
| Among populations within groups | 11 | 116.52 | 0.0741 | 1.59 |
| Within populations | 1532 | 6483.266 | 4.2319 | 91.09 |

[A] Three groups are: Ethiopian plus Kenyan, Guinea Bissau, and the other 6 populations together (Botswana, Egypt, Turkey, Italy, Mongolia and Saudi Arabia).

[B] Eight groups are: Ethiopia, Kenya, Botswana, Egypt, Guinea Bissau, Turkey plus Italy, Mongolia and Saudi Arabia.

[C] Nine groups are: Ethiopia, Kenya, Botswana, Egypt, Guinea Bissau, Turkey, Italy, Mongolia and Saudi Arabia.

### 4.6. Individual specific analysis

Individuals were assigned to source populations using the GeneClass2 program (Piry *et al.*, 2003) available at (http://www.montpellier.inra.fr/CBGP/softwares/), which allows different estimation procedures including the Bayesian (Rannala and Mountain, 1997), frequency (Paetkau *et al.*, 1995) as well as six distance methods; the Nei's $D_A$ (Nei *et al.*, 1983), Cavalli-Sforza and the Edwards chord distance (1967), the Nei Minimum distance (Nei, 1973), the Nei $D_S$ (Nei, 1972), the Jin and Chakraborty's (1993) $D_{AS}$ and the $\delta\mu^2$ distance method of Goldstein *et al.* (1995).

The efficiency of the different methods to assign individuals to their source population was tested and the results are shown in **Table 17**. The Bayesian method computed using GeneClass2 program gave relatively the best result with 51.07% of individuals being assigned to their correct source population. The frequency method and the maximum likelihood method of WHICHRUN also performed well with 49.66% and 49.37% of individuals being correctly assigned to their source populations, respectively. The other methods are relatively less accurate with Goldstein *et al.* (1995) $\delta\mu^2$ distance proving to be least accurate method with only 11.42% of the individuals being assigned to their correct source populations.

The percentage of correct assignment of individuals to their source populations using a single locus was very low ranged from 6.44% for INRA132 to 15.98% for OarAE129 and OarFCB304 (**Table 18**). However, when the six loci with the highest individuals i.e. BM1818, ILSTS11, ILSTS44, ILSTS87, OarAE129 and OarFCB304 (**Table 18**) were used together, the percentage of individuals correctly assigned was increased to 39.82%. Increasing the number of microsatellite loci to 15 resulted in correct assignment of 51.07% of the individuals.

The results of the population assignment test obtained using frequency data from the 20 populations are shown in **Table 19**. In the study 51.07% of the 776 individuals from the 20 populations (11 Ethiopian and 9 reference populations) were assigned to the corrected source population. The population with the highest level of correct assignment was the Guinea Bissau sampled in West Africa (Guinea Bissau) with 97.8% of the individuals being correctly assigned, followed by the Mongolian population with 94.9% of the individuals being correctly assigned to the source population. In general, all the

Ethiopian and Kenyan populations had low levels of correct assignment, Hararghe Highland population having the lowest level of assignment (5%), however, higher assignments of 54.5%, 74.2%, 82.8%, 88.0%, 93.5%, 94.9% and 97.8% were observed in seven reference populations namely Baladie, Hair, Tswana, Ardi, Italian Alpine, Mongolian and Guinea Bissau, respectively.

Table 17. **Percent of correct individual-population assignment from 15 microsatellite loci using different methods.**

| Methods | Software | % correct assignment |
|---|---|---|
| Bayesian | GeneClass | 51.07 |
| Frequency | GeneClass | 49.66 |
| Maximum likelihood | WHICHRUN | 49.37 |
| The Nei *et al*. (1983) $D_A$ | GeneClass | 48.38 |
| Cavalli-Sforza and the Edwards chord distance | GeneClass | 44.14 |
| Nei minimum distance | GeneClass | 41.40 |
| The Nei standard distance ($D_{S)}$ | GeneClass | 41.39 |
| The Chakraborty and Jin $D_{AS}$ | GeneClass | 39.37 |
| The Goldstein *et al*. (1995) $(d\mu)^2$ | GeneClass | 11.42 |

Table 18. **Percent of correct individual-population assignment (for 20 populations) estimated for each microsatellite locus using the Bayesian method.**

| Locus | % of correct assignment |
|---|---|
| OarAE129* | 15.98 |
| OarFCB304* | 15.98 |
| ILSTS11* | 15.08 |
| BM1818* | 14.82 |
| ILSTS44* | 14.56 |
| ILSTS87* | 14.43 |
| MAF65 | 13.66 |
| INRA05 | 12.63 |
| BMS1494 | 12.24 |
| SRCRSP03 | 11.68 |
| INRA63 | 11.08 |
| MAF209 | 10.44 |
| ILSTS05 | 8.76 |
| MAF35 | 7.47 |
| INRA132 | 6.44 |
| *When these six loci taken together | 39.82 |

Table 19. **Frequency of correct assignment of individuals to the 20 populations using genotype information from 15 microsatellite loci using the Bayesian method.**

| | No. | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | 46 | **45.7** | 10.9 | 0.0 | 6.5 | 0.0 | 2.2 | 4.3 | 4.3 | 15.2 | 2.2 | 8.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ABA | 46 | 6.5 | **37.0** | 2.2 | 4.3 | 0.0 | 4.3 | 6.5 | 8.7 | 0.0 | 13.0 | 8.7 | 4.3 | 4.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| AFA | 44 | 4.5 | 0.0 | **38.6** | 2.3 | 4.5 | 18.2 | 0.0 | 2.3 | 0.0 | 11.4 | 4.5 | 6.8 | 4.5 | 0.0 | 0.0 | 0.0 | 0.0 | 2.3 | 0.0 | 0.0 |
| CEH | 40 | 5.0 | 12.5 | 2.5 | **30.0** | 5.0 | 10.0 | 10.0 | 0.0 | 15.0 | 7.5 | 0.0 | 0.0 | 2.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| GUM | 43 | 2.3 | 2.3 | 0.0 | 4.7 | **48.8** | 2.3 | 7.0 | 2.3 | 9.3 | 0.0 | 9.3 | 11.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| HAH | 40 | 0.0 | 12.5 | 17.5 | 2.5 | 7.5 | **5.0** | 7.5 | 5.0 | 5.0 | 12.5 | 5.0 | 7.5 | 5.0 | 0.0 | 2.5 | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 |
| KEF | 38 | 2.6 | 0.0 | 2.6 | 2.6 | 10.5 | 2.6 | **44.7** | 5.3 | 10.5 | 5.3 | 7.9 | 0.0 | 5.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LES | 48 | 6.3 | 4.2 | 4.2 | 6.3 | 4.2 | 4.2 | 2.1 | **25.0** | 0.0 | 18.8 | 6.3 | 14.6 | 2.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.1 |
| NWH | 45 | 8.9 | 6.7 | 4.4 | 13.3 | 2.2 | 2.2 | 4.4 | 6.7 | **28.9** | 0.0 | 8.9 | 6.7 | 6.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| SES | 43 | 2.3 | 2.3 | 11.6 | 4.7 | 4.7 | 11.6 | 2.3 | 2.3 | 9.3 | **16.3** | 4.7 | 18.6 | 7.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.3 | 0.0 |
| WGJ | 46 | 2.2 | 8.7 | 4.3 | 0.0 | 2.2 | 4.3 | 8.7 | 6.5 | 8.7 | 6.5 | **30.4** | 8.7 | 6.5 | 0.0 | 2.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| BGA | 35 | 2.9 | 2.9 | 5.7 | 0.0 | 5.7 | 11.4 | 2.9 | 17.1 | 0.0 | 11.4 | 8.6 | **31.4** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| SEA | 39 | 5.1 | 2.6 | 7.7 | 0.0 | 0.0 | 5.1 | 0.0 | 0.0 | 0.0 | 10.3 | 10.3 | 2.6 | **53.8** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.6 | 0.0 |
| GUB | 46 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **97.8** | 0.0 | 0.0 | 0.0 | 0.0 | 2.2 | 0.0 |
| TSW | 29 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **82.8** | 3.4 | 0.0 | 6.9 | 6.9 | 0.0 |
| ALP | 31 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **93.5** | 0.0 | 6.5 | 0.0 | 0.0 |
| ARD | 25 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.0 | 0.0 | 0.0 | 0.0 | 4.0 | **88.0** | 4.0 | 0.0 | 0.0 |
| BAL | 22 | 0.0 | 0.0 | 0.0 | 0.0 | 4.5 | 0.0 | 0.0 | 4.5 | 0.0 | 0.0 | 0.0 | 4.5 | 0.0 | 0.0 | 0.0 | 4.5 | 0.0 | **54.5** | 27.3 | 0.0 |
| HAI | 31 | 0.0 | 0.0 | 3.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 22.6 | **74.2** | 0.0 |
| MON | 39 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.6 | 2.6 | **94.9** |

No. = number of animals. ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

## 4.7. Genetic diversity and relationship within and between Ethiopian goat populations

### 4.7.1. Hardy-Weinberg equilibrium (HWE)

A total of 479 individuals, from 11 Ethiopian populations were genotyped using 18 microsatellite markers. After testing for linkage disequilibrium, null allele and HWE three markers BMC1222, BMS357 and SRCRSP07 were found to deviate significantly out of Hardy-Weinberg equilibrium across 11, 11 and 9 populations, respectively (**Table 20**). These populations (eleven, eleven and nine) were observed to have significantly lower proportions of heterozygotes than that expected at HWE for BMC1222, BMS357 and SRCRSP07 respectively (**Table 20**). Consequently, these three markers were removed from further analysis.

Then single-locus exact test for HWE were conducted for the 11 population and remaining 15 microsatellite loci. Results indicated that 22 out of 165 or 13.33% of the comparisons ($P < 0.05$ or below; **Table 21**) deviated significantly from HWE, a figure higher than the 5% that would be expected from type I error alone. Application of sequential Bonferroni adjustment for multiple testing reduced the number of significant locus-population combinations to one.

Out of the 15 microsatellite loci used, no locus deviated from HWE for all the populations and also no population deviated from HWE for all the loci (**Table 21**). However, as it is indicated in **Table 21,** there were some populations with several loci deviate significantly from HWE. On one extreme, was found that five loci deviate from HWE in the Hararghe Highland population, while on the other extreme the Gumez population didn't deviate from HWE at any locus. Arsi-Bale, Afar, Short-eared Somali, and Woyto-Guji deviated from HWE at only one locus, while the remaining populations deviated from HWE at 2 to 3 loci. Most of these loci deviations from HWE corresponded to either heterozygote deficit or excess (**Table 21**).

Four out of eleven populations had HWE deviation at the locus INRA63, while three out of the eleven populations had HWE deviation at the loci ILSTS44 and INRA05.

Table 20. **Results of Fisher's exact test for Hardy-Weinberg equilibrium for all Ethiopian populations and 18 loci.**

| Population | BM 1818 | BMC 1222 | BMS 1494 | BMS 357 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 | SRCRSP 07 | All loci |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | *1 | **1 | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| ABA | ns | **1 | ns | **1 | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | ns | ns | ** |
| AFA | ns | **1 | ns | **1 | ns | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| CEH | ns | **1 | ns | **1 | ns | ns | ** | ns | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | **1 | ** |
| GUM | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| HAH | ns | **1 | ns | **1 | ns | ns | *1 | ns | * | ns | *2 | ns | ns | *1 | ns | ns | **1 | ns | ** |
| KEF | ns | **1 | ns | **1 | ns | *1 | ns | ns | ns | ns | ns | ns | *1 | *1 | ns | ns | ns | *1 | ** |
| LES | ns | *1 | ns | **1 | ns | ns | ns | ns | ns | ** | ns | ns | ns | ns | ** | ns | * | **1 | ** |
| NWH | ns | **1 | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | **1 | ns | ns | **1 | ** |
| SES | *1 | **1 | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| WGJ | ns | **1 | ns | **1 | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | ns | ns | **1 | ** |
| All population | ns | **1 | ns | **1 | ns | ns | **1 | ns | * | **1 | ns | ns | ns | ns | * | ns | ns | **1 | ** |

ns = not significant; * = $P < 0.05$; ** = $P < 0.01$; 1 = heterozygote deficit; 2 = heterozygote excess.
ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Table 21. **Results of Fisher's exact test for Hardy-Weinberg equilibrium for all Ethiopian populations and 15 microsatellite markers.**

| Population | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 | All loci |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | *1 | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | ns |
| ABA | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| AFA | ns | ns | ns | ns | **1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| CEH | ns | ns | ns | ns | ** | ns | ns | **1 | ns | ns | ns | ns | ns | *1 | ns | ** |
| GUM | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| HAH | ns | ns | ns | ns | * | ns | * | ns | *2 | ns | ns | *1 | ns | ns | **1 | ** |
| KEF | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | *1 | *1 | ns | ns | ns | ns |
| LES | ns | ns | ns | ns | ns | ns | ns | ** | ns | ns | ns | ns | ** | ns | * | ** |
| NWH | ns | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | **1 | ns | ns | ns |
| SES | *1 | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| WGJ | ns | ns | ns | ns | ns | ns | *1 | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| All population | ns | ns | ns | ns | **1 | ns | * | **1 | ns | ns | ns | ns | * | ns | ns | ** |

ns = not significant; * = $P < 0.05$; ** = $P < 0.01$; 1 = heterozygote deficit; 2 = heterozygote excess.
ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

### 4.7.2. Linkage disequilibrium

The linkage disequilibrium tests within each of the 11 populations were performed for the 105 pair-wise combinations of the 15 loci for each population. Assuming a 5% type I error rate ($P < 0.05$), about four to six pairs of loci are expected to be in linkage disequilibrium. On average 6% of the loci pairs were in linkage disequilibrium within the populations. As indicated in **Appendix 19,** large number of loci pairs that were in linkage disequilibrium ($P < 0.05$) were observed in the Arsi-Bale (11 pairs), followed by the Long-eared Somali (nine pairs), Short-eared Somali (eight pairs), Afar and Woyto-Guji (seven pairs each), Central Highland (six pairs), Gumez and Keffa (five pairs each), Abergalle, Hararghe Highland and North-West Highland populations (four pairs each).

As indicated in **Appendix 20 t**he exact test for genotypic disequilibrium across all 11 populations (105 pair-wise comparisons) yielded significant deviations in the following comparisons: BM1818/ILSTS11 ($P < 0.01$); BM1818/ILSTS44 ($P < 0.05$); BM1818/ MAF209 ($P < 0.01$); BM1818/OaeAE129 ($P < 0.01$); BM1494/OarAE129 ($P < 0.05$), BMS1494/OarFCB304 ($P < 0.05$); ILSTS11/INRA63 ($P < 0.05$); ILSTS87/INRA63 ($P < 0.05$); and MAF209/OarFCB304 ($P < 0.05$). These significant results observed are nine out of 105 pair-wise combinations (8.57%), a figure higher than the 5% that would be expected from type I error alone. However, none of these nine tests remained significant following application of sequential Bonferroni corrections.

### 4.7.3. Analysis of inbreeding coefficient within the Ethiopian goat populations

The inbreeding was estimated for each population using Weir and Cockerham's $F_{IS}$ estimator (Weir and Cockerham, 1984). The calculations and tests were implemented in the computer program FSTAT version 2.9.3 (Goudet, 2001). Significance levels were determined after 165,000 randomizations. The results of inbreeding coefficient ($F_{IS}$ estimator) are given in **Table 22**.

Table 22. **Inbreeding coefficients (*F*IS estimator) estimated from 15 microsatellite loci for Ethiopian goat populations.**

| Populations | $F_{IS}$ |
|---|---|
| Abergalle | $0.044^{*}$ |
| Arsi-Bale | $0.039^{ns}$ |
| Afar | $0.076^{**}$ |
| Central Highland | $0.077^{**}$ |
| Gumez | $-0.002^{ns}$ |
| Hararghe Highland | $0.051^{*}$ |
| Keffa | $0.03^{ns}$ |
| Long-eared Somali | $0.051^{*}$ |
| North-West Highland | $0.052^{*}$ |
| Short-eared Somali | $0.016^{ns}$ |
| Woyto-Guji | $0.059^{*}$ |

$^{ns}$ not significant ($P > 0.05$), $^{*}P < 0.05$, $^{**}P < 0.01$. Significance levels were determined after 165,000 randomizations.

Although the deficit (positive $F_{IS}$ value) was detected in 10 out of 11 populations, seven of these populations showed a significant inbreeding effect ($P < 0.05$ or below) (Abergalle, Afar, Central Highland, Hararghe Highland, Long-eared Somali, North-West Highland and Woyto-Guji populations). Only one population (Gumez) showed heterozygosity excess (negative $F_{IS}$ value), which doesn't show significant heterozygote excess ($P > 0.05$).

### 4.7.4. Analysis of Fixation Indices or *F*-statistics

The *F* statistics in population genetics has nothing to do with the *F* statistics evaluating differences in variances. Here *F* stands for fixation index, fixation being increased homozygosity resulting from inbreeding. Population subdivision results in the loss of genetic variation (measured by heterozygosity) within subpopulations due to being small populations and genetic drift acting within each one of them. This means that population subdivision would result in decreased heterozygosity relative to that expected heterozygosity under random mating as if the whole population was a single breeding unit. In the 1940s and 1950s, Sewall Wright developed three fixation indices to evaluate population subdivision: $F_{IS}$ (inter-individuals), $F_{ST}$ (subpopulations), $F_{IT}$ (total population).

$F_{IS}$ is a measure of the deviation of genotypic frequencies from panmixis frequencies in terms of heterozygous deficiency or excess. It is known as the inbreeding coefficient (*f*), which is conventionally defined as the probability that two alleles in an individual are identical by descent (autozygous). The technical description is the correlation of uniting gametes relative to gametes drawn at random from within a subpopulation (Individual within the Subpopulation) averaged over subpopulations. It is calculated in a single population as: $F_{IS} = 1 - (H_{OBS} / H_{EXP})$ [equal to $(H_{EXP} - H_{OBS} / H_{EXP})$ where $H_{OBS}$ is the observed heterozygosity and $H_{EXP}$ is the expected heterozygosity calculated on the assumption of random mating. It shows the degree to which heterozygosity is reduced below the expectation. The value of $F_{IS}$ ranges between -1 and +1. Negative $F_{IS}$ values indicate heterozygote excess (out-breeding) and positive values indicate heterozygote deficiency (inbreeding) compared with HWE expectations.

$F_{ST}$ measures the effect of population subdivision, which is the reduction in heterozygosity in a subpopulation due to genetic drift. $F_{ST}$ is the most inclusive measure of population substructure and is most useful for examining the overall genetic divergence among subpopulations. It is also called co-ancestry coefficient (*q*) [Weir & Cockerham, 1984] or 'Fixation index' and is defined as correlation of gametes within subpopulations relative to gametes drawn at random from the entire population (Subpopulation within the total population). It is calculated as using the subpopulation (average) heterozygosity and total population expected heterozygosity. $F_{ST}$ is always positive; it ranges between 0 = panmixis (no subdivision, random mating occurring, no genetic divergence within the

population) and 1 = complete isolation (extreme subdivision). $F_{ST}$ is usually calculated for different loci, and then averaged across all loci, and all populations.

$F_{IT}$ is rarely used. It is the overall inbreeding coefficient ($F$) of an individual relative to the total population (Individual within the total population).

The unbiased estimators of Weir and Cockerham's $F$-statistics ($F_{IT}$, $F_{ST}$ and $F_{IS}$) (Weir and Cockerham, 1984) for each locus were calculated by jackknifing across 11 Ethiopian goat populations using the computer program FSTAT version 2.9.3 (Goudet, 2001). Significance levels were determined after 10,000 randomizations of alleles within populations ($F_{IS}$), of alleles in total populations ($F_{IT}$), and after 15,000 randomizations of genotypes among populations assumed no random mating within populations ($F_{ST}$). Results of $F$-statistics and their significance levels are given in **Table 23**.

The $F_{ST}$ estimator was significantly different from zero ($P < 0.01$ or $P < 0.001$) for each of 15 loci (except for one MAF35, $P = 0.4496$) and for multi-locus, which means that the genetic differentiation has occurred among populations although the gene differentiation among populations was low ($F_{ST}$ = 1.6%). The $F_{IS}$ estimator was not significantly different from zero for single locus, but a significant positive value for six loci (BM1818 and MAF209, $P<0.05$; ILSTS05, $P < 0.01$; and ILSTS44, INRA63 and MAF65, $P < 0.001$), which means that an inbreeding effect within population was detected at these six loci. The $F_{IS}$ estimator showed also non-significant negative values for three loci (ILSTS11, ILSTS87 and MAF35). $F_{IS}$ was significantly different from zero for multi-locus ($P < 0.001$). The values of $F_{IT}$ estimators were all positive except negative for two (ILSTS87, $P = 0.72600$ and MAF35, $P = 0.82270$) and significantly different from zero at $P < 0.05$ or $P < 0.01$ or $P < 0.001$ levels for all loci, but five (BMS1494, ILSTS11, INRA05, INRA132, and OarFCB304, $P > 0.05$). $F_{IT}$ was significantly different from zero for multi-locus ($P < 0.001$). As inbreeding effect within populations ($F_{IS}$) was detected at BM1818, MAF209, ILSTS05, ILSTS44, INRA63, and MAF65, gene differentiation among populations has resulted in significant $F_{IT}$ estimators except for BM1818, MAF209, ILSTS05, ILSTS44, INRA63, and MAF65, which caused by both inbreeding within population plus gene differentiation among populations.

Table 23. **Results of *F*-statistics (Weir and Cockerham, 1984) for 15 microsatellite loci by jackknifing across the 11 Ethiopian goat populations.**

| Locus | $F$ ($F_{IT}$) | $\Theta$ ($F_{ST}$) | $f$ ($F_{IS}$) |
|---|---|---|---|
| BM1818 | 0.050 (0.024)$^{*}$ | 0.010 (0.006)$^{**}$ | 0.040 (0.024)$^{*}$ |
| BMS1494 | 0.053 (0.034)$^{ns}$ | 0.015 (0.009)$^{***}$ | 0.038 (0.036)$^{ns}$ |
| ILSTS05 | 0.135 (0.038)$^{**}$ | 0.021 (0.007)$^{***}$ | 0.116 (0.037)$^{**}$ |
| ILSTS11 | 0.028 (0.028)$^{ns}$ | 0.036 (0.013)$^{***}$ | -0.009 (0.019)$^{ns}$ |
| ILSTS44 | 0.150 (0.032)$^{***}$ | 0.014 (0.014)$^{**}$ | 0.138 (0.034)$^{***}$ |
| ILSTS87 | -0.019 (0.033)$^{ns}$ | 0.017 (0.006)$^{***}$ | -0.037 (0.037)$^{ns}$ |
| INRA05 | 0.039 (0.049)$^{ns}$ | 0.021 (0.009)$^{***}$ | 0.018 (0.044)$^{ns}$ |
| INRA63 | 0.109 (0.043)$^{***}$ | 0.004 (0.004)$^{***}$ | 0.105 (0.042)$^{***}$ |
| INRA132 | 0.046 (0.047)$^{ns}$ | 0.015 (0.007)$^{***}$ | 0.031 (0.046)$^{ns}$ |
| MAF35 | -0.043 (0.025)$^{ns}$ | 0.000 (0.004)$^{ns}$ | -0.043 (0.022)$^{ns}$ |
| MAF65 | 0.084 (0.018)$^{***}$ | 0.017 (0.006)$^{***}$ | 0.068 (0.019)$^{***}$ |
| MAF209 | 0.083 (0.028)$^{**}$ | 0.017 (0.007)$^{***}$ | 0.068 (0.030)$^{*}$ |
| OarAE129 | 0.041 (0.017)$^{*}$ | 0.009 (0.005)$^{***}$ | 0.033 (0.016)$^{ns}$ |
| OarFCB304 | 0.025 (0.017)$^{ns}$ | 0.009 (0.003)$^{***}$ | 0.016 (0.017)$^{ns}$ |
| SRCRSP03 | 0.064 (0.031)$^{*}$ | 0.030 (0.010)$^{***}$ | 0.035 (0.034)$^{ns}$ |
| Multi-locus | 0.061 (0.011)$^{***}$ | 0.016 (0.003)$^{***}$ | 0.045 (0.012)$^{***}$ |

$^{ns}$ Not significant (P>0.05), * P < 0.05, ** P < 0.01, *** P < 0.001. Significance levels were determined after 10,000 randomizations of alleles within populations ($F_{IS}$), of alleles in total populations ($F_{IT}$), and after 15,000 randomizations of genotypes among populations assumed no random mating within populations ($F_{ST}$). Standard errors are given in parenthesis.

## 4.8. The extent and distribution of the genetic diversity within the Ethiopian populations

### 4.8.1. Mean number of alleles and heterozygosities

A total of 133 alleles were observed in the 11 Ethiopian populations. Among the Ethiopian populations, the most polymorphic marker was OarFCB304 with 23 alleles, while the least polymorphic markers were INRA132 and MAF35 with 3 alleles. Allele frequencies across all loci ranged from 1.04 to 95.00 (**Appendix 21** to **Appendix 35**). Private alleles meaning alleles that were unique to a single or observed in only one population were observed in the following populations: Abergalle (one), Arsi-Bale (one), Afar (two), Central Highland (two), Gumez (two), Hararghe Highland (two), Keffa (three), Long-eared Somali (three), North-West Highland (four) and Short-eared Somali (one). However, the frequencies of these private alleles were less than 0.05 in all cases.

The MNA ranged from 5.13 in Central Highland to 6.47 in Long-ear Somali when all the individuals in the population were considered (**Table 24** to **Table** 26). To remove any sample bias, the MNA for a random sample of 20 animals was calculated for each population with 250 replicate re-sampling of 20 individuals with replacement. The Abergalle and Central Highland were the least diverse population both with MNA of 4.65, while the Long-eared Somali was again the most diverse population with MNA of 5.28. In general the mean numbers of alleles were moderate and were comparable among all the populations from the different geographic regions of Ethiopia in both cases.

The expected heterozygosity ($H_E$) and observed heterozygosity ($H_O$) values averaged over loci showed an overall pattern similar to that observed for MNA per locus (**Table 24** to **Table** 26). The highest value for $H_E$ around 61% was observed within the Arsi-Bale population, while the lowest value around 55% was for Gumez population. The highest vale for $H_O$ around 58% was again observed within the Arsi-Bale population, while the lowest value around 52% was observed for Afar. In all cases average observed heterozygosities were lower than that expected under HWE.

Table 24. **Expected heterozygosity (*H*E), observed heterozygosity (*H*O) and mean number of alleles (MNA) of all 11 Ethiopian populations.**

| Population | Sample size | MNA (All ind) | MNA (20 ind) | $H_E$ ± s.e. | $H_O$ ± s.e. |
|---|---|---|---|---|---|
| Abergalle | 46 | 5.33 | 4.65 | 0.57 ± 0.056 | 0.54 ± 0.053 |
| Arsi-Bale | 46 | 5.33 | 4.87 | 0.61 ± 0.043 | 0.58 ± 0.045 |
| Afar | 44 | 5.47 | 4.77 | 0.56 ± 0.051 | 0.52 ± 0.049 |
| Central Highland | 40 | 5.13 | 4.65 | 0.57 ± 0.056 | 0.53 ± 0.058 |
| Gumez | 43 | 5.53 | 4.71 | 0.55 ± 0.057 | 0.55 ± 0.054 |
| Hararghe Highland | 40 | 5.93 | 5.10 | 0.58 ± 0.052 | 0.55 ± 0.053 |
| Keffa | 38 | 5.60 | 5.05 | 0.58 ± 0.050 | 0.56 ± 0.047 |
| Long-eared Somali | 48 | 6.47 | 5.28 | 0.59 ± 0.044 | 0.56 ± 0.044 |
| North West Highland | 45 | 5.40 | 4.67 | 0.57 ± 0.053 | 0.54 ± 0.049 |
| Small-eared Somali | 43 | 5.60 | 4.90 | 0.58 ± 0.054 | 0.57 ± 0.055 |
| Woyto-Guji | 46 | 5.67 | 4.99 | 0.58 ± 0.052 | 0.54 ± 0.049 |
| Mean of all populations | | 5.59 | 4.88 | 0.58 ± 0.015 | 0.55 ± 0.015 |

Table 25. **The expected heterozygosity (*H*E) and observed heterozygosity (*H*O) per locus and population for Ethiopian goats.**

| Pop | BM1818 | | BMS1494 | | ILSTS05 | | ILSTS11 | | ILSTS44 | | ILSTS87 | | INRA05 | | INRA63 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ |
| ABG | 0.630 | 0.705 | 0.522 | 0.532 | 0.130 | 0.124 | 0.696 | 0.701 | 0.609 | 0.661 | 0.304 | 0.292 | 0.522 | 0.542 | 0.522 | 0.634 |
| ABA | 0.739 | 0.781 | 0.432 | 0.536 | 0.304 | 0.349 | 0.717 | 0.689 | 0.619 | 0.623 | 0.500 | 0.482 | 0.739 | 0.626 | 0.522 | 0.582 |
| AFA | 0.523 | 0.553 | 0.523 | 0.439 | 0.295 | 0.390 | 0.614 | 0.627 | 0.477 | 0.685 | 0.273 | 0.306 | 0.432 | 0.543 | 0.705 | 0.633 |
| CEH | 0.750 | 0.708 | 0.475 | 0.526 | 0.050 | 0.096 | 0.725 | 0.657 | 0.575 | 0.690 | 0.450 | 0.477 | 0.450 | 0.562 | 0.350 | 0.572 |
| GUM | 0.767 | 0.777 | 0.349 | 0.302 | 0.186 | 0.173 | 0.605 | 0.620 | 0.558 | 0.641 | 0.333 | 0.336 | 0.628 | 0.574 | 0.698 | 0.658 |
| HAH | 0.675 | 0.630 | 0.500 | 0.531 | 0.200 | 0.222 | 0.700 | 0.690 | 0.475 | 0.695 | 0.450 | 0.402 | 0.550 | 0.623 | 0.650 | 0.622 |
| KEF | 0.500 | 0.520 | 0.500 | 0.605 | 0.289 | 0.321 | 0.605 | 0.694 | 0.579 | 0.672 | 0.421 | 0.369 | 0.684 | 0.646 | 0.711 | 0.705 |
| LES | 0.771 | 0.742 | 0.438 | 0.473 | 0.271 | 0.339 | 0.667 | 0.684 | 0.583 | 0.653 | 0.563 | 0.449 | 0.625 | 0.590 | 0.542 | 0.687 |
| NWH | 0.667 | 0.733 | 0.568 | 0.523 | 0.156 | 0.145 | 0.622 | 0.629 | 0.644 | 0.680 | 0.378 | 0.404 | 0.622 | 0.596 | 0.533 | 0.655 |
| SES | 0.605 | 0.744 | 0.512 | 0.503 | 0.186 | 0.173 | 0.791 | 0.708 | 0.721 | 0.701 | 0.462 | 0.494 | 0.651 | 0.599 | 0.558 | 0.685 |
| WGJ | 0.652 | 0.678 | 0.500 | 0.562 | 0.239 | 0.267 | 0.717 | 0.698 | 0.500 | 0.649 | 0.283 | 0.262 | 0.435 | 0.569 | 0.543 | 0.614 |
| **Average** | **0.662** | **0.688** | **0.483** | **0.503** | **0.210** | **0.236** | **0.678** | **0.672** | **0.576** | **0.668** | **0.401** | **0.388** | **0.576** | **0.588** | **0.576** | **0.641** |
| **SD** | **0.093** | **0.087** | **0.059** | **0.079** | **0.080** | **0.102** | **0.061** | **0.033** | **0.074** | **0.025** | **0.095** | **0.082** | **0.106** | **0.034** | **0.108** | **0.042** |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Table 26. **The expected heterozygosity (*H*E) and observed heterozygosity (*H*O) per locus and population for Ethiopian goats.**

| Pop | INRA132 | | MAF35 | | MAF65 | | MAF209 | | OarAE129 | | OaeFCB304 | | SRCRSP03 | | Mean (SD) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ | $H_O$ | $H_E$ |
| ABG | 0.478 | 0.498 | 0.196 | 0.178 | 0.783 | 0.796 | 0.478 | 0.567 | 0.826 | 0.800 | 0.696 | 0.748 | 0.756 | 0.740 | 0.543 (0.053) | 0.568 (0.056) |
| ABA | 0.478 | 0.534 | 0.217 | 0.232 | 0.844 | 0.869 | 0.543 | 0.626 | 0.727 | 0.776 | 0.696 | 0.716 | 0.652 | 0.663 | 0.582 (0.045) | 0.606 (0.043) |
| AFA | 0.386 | 0.515 | 0.136 | 0.129 | 0.773 | 0.861 | 0.477 | 0.465 | 0.773 | 0.858 | 0.705 | 0.706 | 0.636 | 0.645 | 0.515 (0.049) | 0.557 (0.051) |
| CEH | 0.425 | 0.500 | 0.125 | 0.119 | 0.757 | 0.860 | 0.650 | 0.578 | 0.781 | 0.791 | 0.650 | 0.728 | 0.700 | 0.698 | 0.528 (0.058) | 0.571 (0.056) |
| GUM | 0.651 | 0.621 | 0.140 | 0.131 | 0.711 | 0.832 | 0.488 | 0.488 | 0.860 | 0.841 | 0.605 | 0.618 | 0.674 | 0.626 | 0.550 (0.054) | 0.549 (0.057) |
| HAH | 0.700 | 0.530 | 0.125 | 0.120 | 0.800 | 0.887 | 0.425 | 0.562 | 0.842 | 0.800 | 0.700 | 0.728 | 0.462 | 0.651 | 0.550 (0.053) | 0.579 (0.052) |
| KEF | 0.605 | 0.519 | 0.105 | 0.101 | 0.730 | 0.848 | 0.605 | 0.626 | 0.789 | 0.818 | 0.737 | 0.691 | 0.605 | 0.586 | 0.564 (0.047) | 0.581 (0.050) |
| LES | 0.458 | 0.524 | 0.250 | 0.252 | 0.872 | 0.856 | 0.500 | 0.561 | 0.766 | 0.843 | 0.563 | 0.597 | 0.563 | 0.631 | 0.562 (0.044) | 0.592 (0.044) |
| NWH | 0.467 | 0.497 | 0.200 | 0.182 | 0.795 | 0.841 | 0.467 | 0.533 | 0.800 | 0.825 | 0.689 | 0.691 | 0.556 | 0.677 | 0.544 (0.049) | 0.574 (0.053) |
| SES | 0.442 | 0.506 | 0.163 | 0.151 | 0.884 | 0.864 | 0.442 | 0.472 | 0.810 | 0.845 | 0.628 | 0.616 | 0.744 | 0.674 | 0.573 (0.055) | 0.582 (0.054) |
| WGJ | 0.543 | 0.538 | 0.200 | 0.182 | 0.761 | 0.861 | 0.630 | 0.628 | 0.783 | 0.834 | 0.783 | 0.731 | 0.587 | 0.594 | 0.544 (0.049) | 0.578 (0.052) |
| **Mean** | **0.512** | **0.526** | **0.169** | **0.162** | **0.792** | **0.852** | **0.519** | **0.555** | **0.796** | **0.821** | **0.677** | **0.688** | **0.630** | **0.653** | **0.551** | **0.576** |
| **SD** | **0.100** | **0.035** | **0.046** | **0.049** | **0.056** | **0.024** | **0.077** | **0.060** | **0.037** | **0.026** | **0.062** | **0.053** | **0.088** | **0.045** | **(0.015)$^{s.e.}$** | **(0.015)$^{s.e.}$** |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

### 4.9. Genetic relationship between the Ethiopian populations

Two measures of genetic distances, Nei's (1972) standard genetic distance **(**$D_S$) and $D_A$ genetic distance (Nei *et al.,* 1983) were used to estimate the genetic distances for each pair of populations.

#### 4.9.1. Nei's standard genetic distance ($D_S$)

The Nei's standard genetic distance ($D_S$) values for Ethiopian populations with only one reference (as out-group) population are shown in **Table 27**. The smallest $D_S$ for Ethiopian goat populations was -0.001 ± 0.004 between the Hararghe Highland and Short-ear Somali, followed by the value of 0.004 ± 0.004 between the Central Highland and North West Highland. The highest $D_S$ for Ethiopian population was 0.055 ± 0.018 observed between Abergalle and Afar populations. Smaller $D_S$ values were observed for populations within neighbouring regions. On the other hand the $D_S$ genetic distances observed between the populations in different geographic regions of the country were relatively higher.

#### 4.9.2. Nei's $D_A$ genetic distances

The genetic distances of the Ethiopian populations were determined using Nei's $D_A$ index of genetic distance with one reference (as out-group) population, as in the case for the Nei's $D_S$ genetic distance, the populations with smallest genetic distance were Hararghe Highland and Short-eared Somali with genetic distance of 0.024, followed by the value of 0.026 again between the Central Highland and North West Highland (**Table 28**). The highest $D_A$ among Ethiopian populations was between Keffa and Afar with genetic distance of 0.065. As in the case for the Nei's $D_S$ genetic distance, the $D_A$ genetic distances observed between populations from the neighbouring regions were small and those from relatively far apart regions of the country were relatively larger.

Table 27. **Nei's standard genetic distance (*D*S) matrix for 11 Ethiopian populations and 1 reference population from West Africa.**

| Pop | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | GUB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | 0.013 | 0.018 | 0.008 | 0.014 | 0.014 | 0.016 | 0.014 | 0.007 | 0.013 | 0.014 | 0.078 |
| ABA | 0.034 | - | 0.018 | 0.006 | 0.009 | 0.007 | 0.011 | 0.005 | 0.006 | 0.007 | 0.007 | 0.080 |
| AFA | 0.055 | 0.047 | - | 0.014 | 0.016 | 0.006 | 0.025 | 0.011 | 0.019 | 0.005 | 0.014 | 0.080 |
| CEH | 0.010 | 0.019 | 0.035 | - | 0.012 | 0.005 | 0.007 | 0.008 | 0.004 | 0.006 | 0.013 | 0.085 |
| GUM | 0.031 | 0.031 | 0.046 | 0.030 | - | 0.015 | 0.013 | 0.005 | 0.009 | 0.009 | 0.008 | 0.086 |
| HAH | 0.031 | 0.014 | 0.010 | 0.008 | 0.033 | - | 0.010 | 0.005 | 0.008 | 0.004 | 0.012 | 0.077 |
| KEF | 0.036 | 0.023 | 0.046 | 0.022 | 0.036 | 0.021 | - | 0.009 | 0.007 | 0.012 | 0.011 | 0.071 |
| LES | 0.037 | 0.012 | 0.026 | 0.026 | 0.019 | 0.011 | 0.030 | - | 0.006 | 0.003 | 0.007 | 0.081 |
| NWH | 0.014 | 0.011 | 0.035 | 0.004 | 0.022 | 0.012 | 0.012 | 0.016 | - | 0.007 | 0.008 | 0.085 |
| SES | 0.032 | 0.019 | 0.014 | 0.017 | 0.024 | -0.001 | 0.030 | 0.005 | 0.007 | - | 0.006 | 0.078 |
| WGJ | 0.034 | 0.012 | 0.033 | 0.036 | 0.021 | 0.020 | 0.022 | 0.011 | 0.019 | 0.015 | - | 0.063 |
| GUB | 0.220 | 0.211 | 0.201 | 0.222 | 0.202 | 0.200 | 0.211 | 0.210 | 0.224 | 0.198 | 0.163 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, GUB – Guinea Bissau.

Table 28. **Nei's *D*A genetic distance matrix for 11 Ethiopian populations and 1 reference population from West Africa.**

| Pop | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | GUB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | | | | | | | | | | | |
| ABA | 0.049 | - | | | | | | | | | | |
| AFA | 0.061 | 0.060 | - | | | | | | | | | |
| CEH | 0.034 | 0.039 | 0.051 | - | | | | | | | | |
| GUM | 0.054 | 0.049 | 0.056 | 0.055 | - | | | | | | | |
| HAH | 0.048 | 0.040 | 0.038 | 0.040 | 0.049 | - | | | | | | |
| KEF | 0.055 | 0.050 | 0.065 | 0.051 | 0.054 | 0.051 | - | | | | | |
| LES | 0.059 | 0.041 | 0.052 | 0.058 | 0.050 | 0.041 | 0.057 | - | | | | |
| NWH | 0.034 | 0.044 | 0.051 | 0.026 | 0.048 | 0.044 | 0.044 | 0.052 | - | | | |
| SES | 0.048 | 0.037 | 0.033 | 0.038 | 0.052 | 0.024 | 0.051 | 0.033 | 0.033 | - | | |
| WGJ | 0.051 | 0.030 | 0.043 | 0.050 | 0.044 | 0.041 | 0.046 | 0.035 | 0.041 | 0.033 | - | |
| GUB | 0.216 | 0.210 | 0.198 | 0.225 | 0.191 | 0.195 | 0.198 | 0.179 | 0.227 | 0.193 | 0.170 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, GUB – Guinea Bissau.

### 4.9.3. Phylogenetic relationship among Ethiopian goat populations

The phylogenetic relationship between the 11 Ethiopian goat populations was determined using $D_S$ and $D_A$ genetic distances and their UPGMA with out reference population and with one reference population as out-group is shown in the form of NJ trees in (**Figure 31** to **Figure** 38). The $D_S$ phylogenetic tree with one reference population (**Figure 31**) shows nine relatively clear groups. The first cluster being Abergalle and Central Highland with Abergalle far away from the Central Highland, the second cluster North-west Highland with very near branch to Central Highland, the third cluster Keffa, the fourth cluster Arsi-Bale, the fifth cluster Long-eared Somali, the sixth cluster Short-eared Somali with very close branch to the seventh cluster, which is Afar and Haraghe Highland, Afar being very far from the Haraghe Highland, the eighth cluster Gumez, and the ninth cluster Woyto-Guji with the bootstrap values ranging from 25% to76%. Unlike the $D_S$ NJ tree the $D_A$ phylogenetic tree with one reference population (**Figure 32**) showed similar clustering except for the clustering of Central Highland together with North-west Highland instead of Central Highland with Abergalle and Afar together with Short-eared Somali instead of Afar with Haraghe Highland with bootstrap values ranging from 15 to 88%. A similar topology is observed for the phylogenetic trees constructed from $D_S$ and $D_A$ genetic distances without any reference population (**Figure 33** and **Figure** 34) with bootstrap values ranging from 6% to 79% and 33% to 89% respectively with a little difference for the $D_A$. On the other hand all the UPGMA from $D_S$ distance for with and without out-group showed consistent clustering of Central Highland with North-West Highland and Hararghe Highland with Short-eared Somali with Long-eared Somali very close to them (**Figure 35** and **Figure** 36) with bootstrap values ranging from 7% to 69% for with one reference population, while with bootstrap values ranging from 7% to 52% for with out reference population. The UPGMA from $D_A$ distance for with and without out-group showed consistent clustering of Central Highland with North-West Highland and Hararghe Highland with Short-eared Somali (**Figure 37** and **Figure** 38) with bootstrap values ranging from 26% to 82% for with one reference population, while with bootstrap values ranging from 31% to 82% for with out reference population.

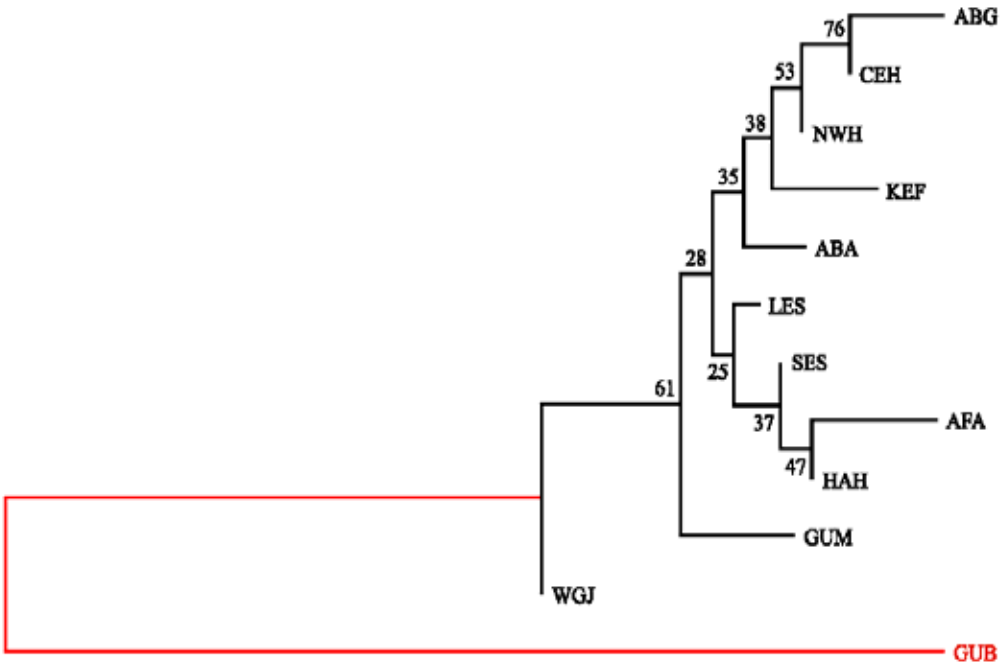


Figure 31. **Phylogenetic tree representing the genetic relationship between the 11 Ethiopian populations with GUB as an out-group population.** The NJ tree was constructed using the genetic distance $D_S$ calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, GUB – Guinea Bissau)

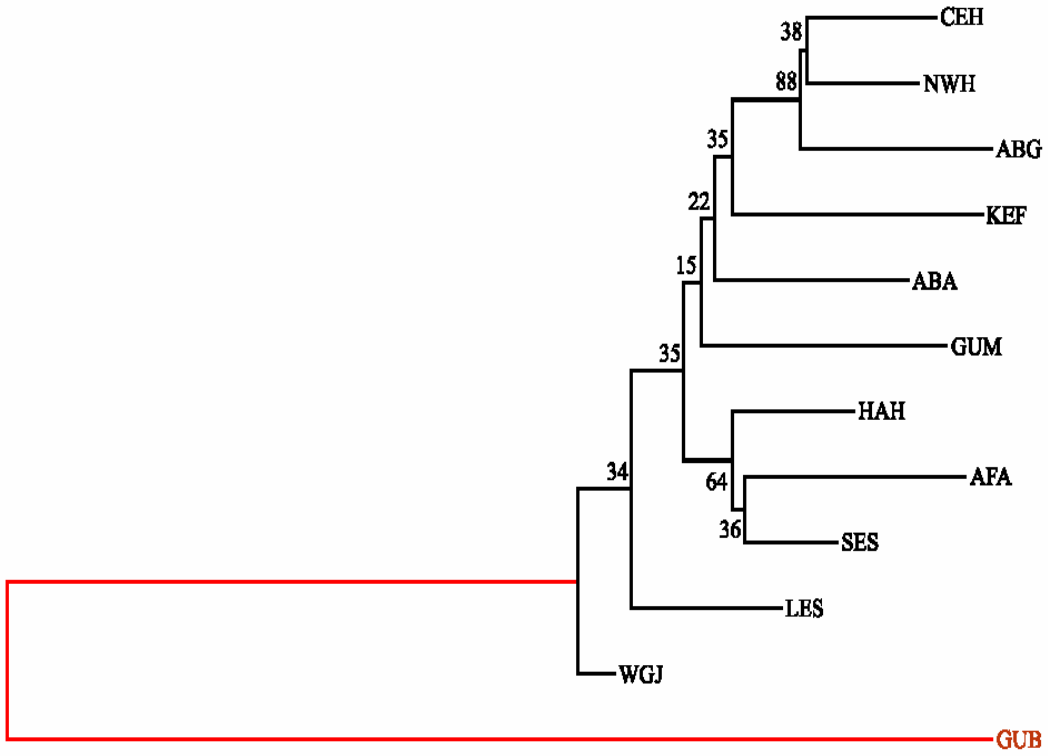


Figure 32. **Phylogenetic tree representing the genetic relationship between the 11 Ethiopian populations with GUB as an out-group population.** The NJ tree was constructed using the genetic distance $D_A$ calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, GUB – Guinea Bissau)

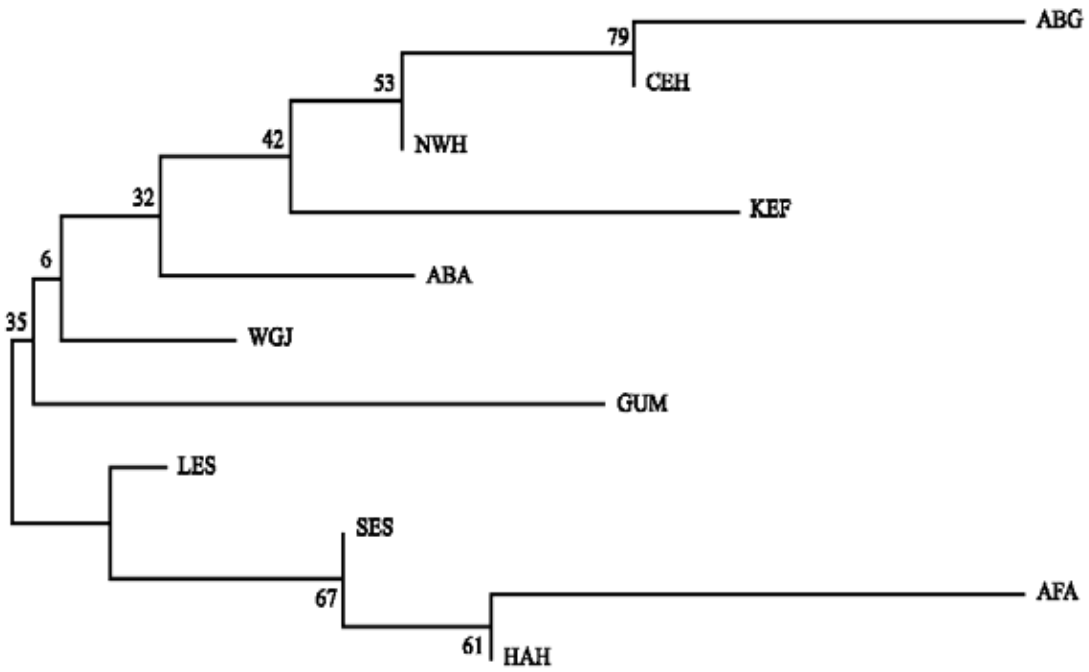


Figure 33. **Phylogenetic tree representing the genetic relationship between the 11 Ethiopian populations.** The NJ tree was constructed using the genetic distance $D_S$ calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

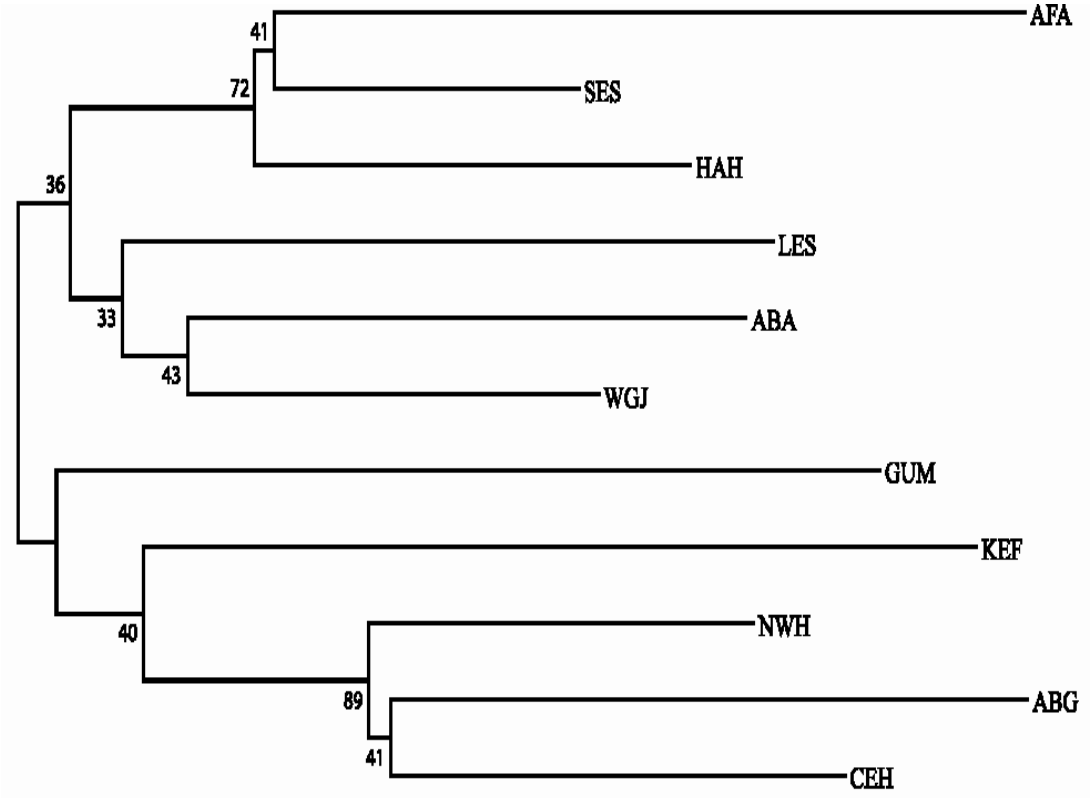


Figure 34. **Phylogenetic tree representing the genetic relationship between the 11 Ethiopian populations.** The NJ tree was constructed using the genetic distance DA calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

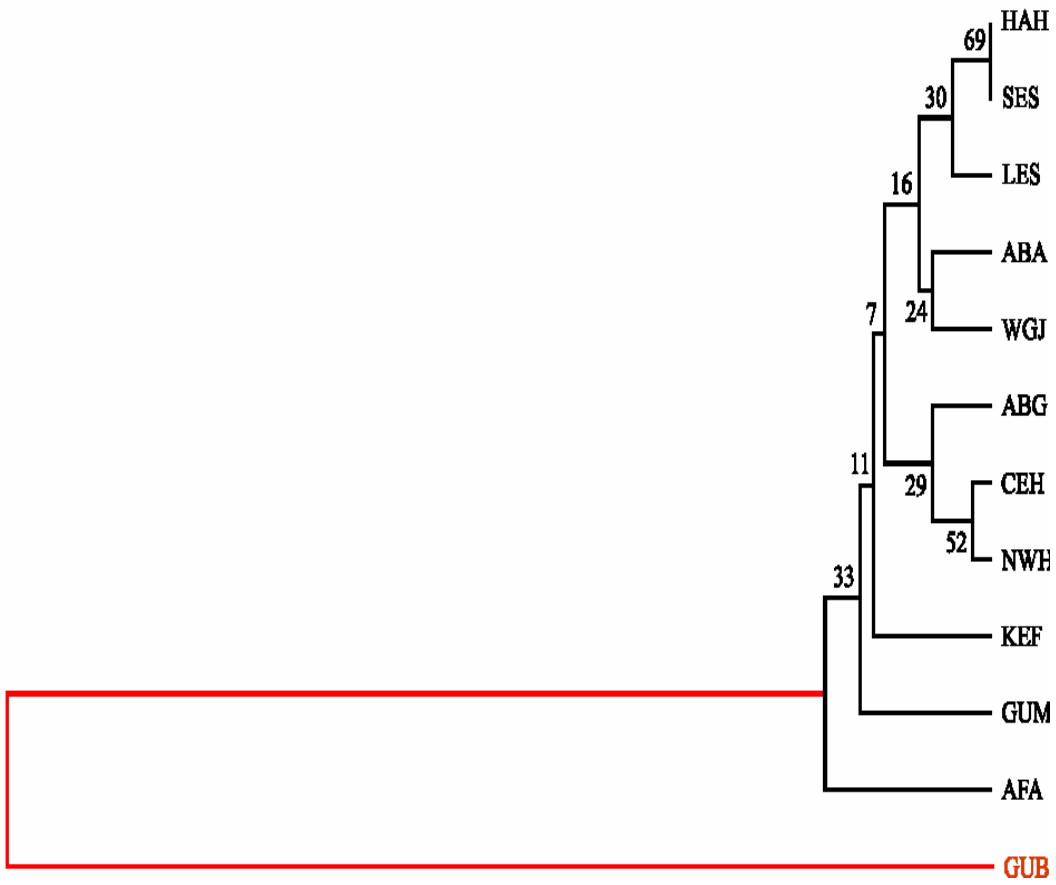


Figure 35. **UPGMA tree representing the genetic relationship between the 11 Ethiopian populations with GUB as an out-group population.** The NJ tree was constructed using the genetic distance DS calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji, GUB –Guinea Bissau)

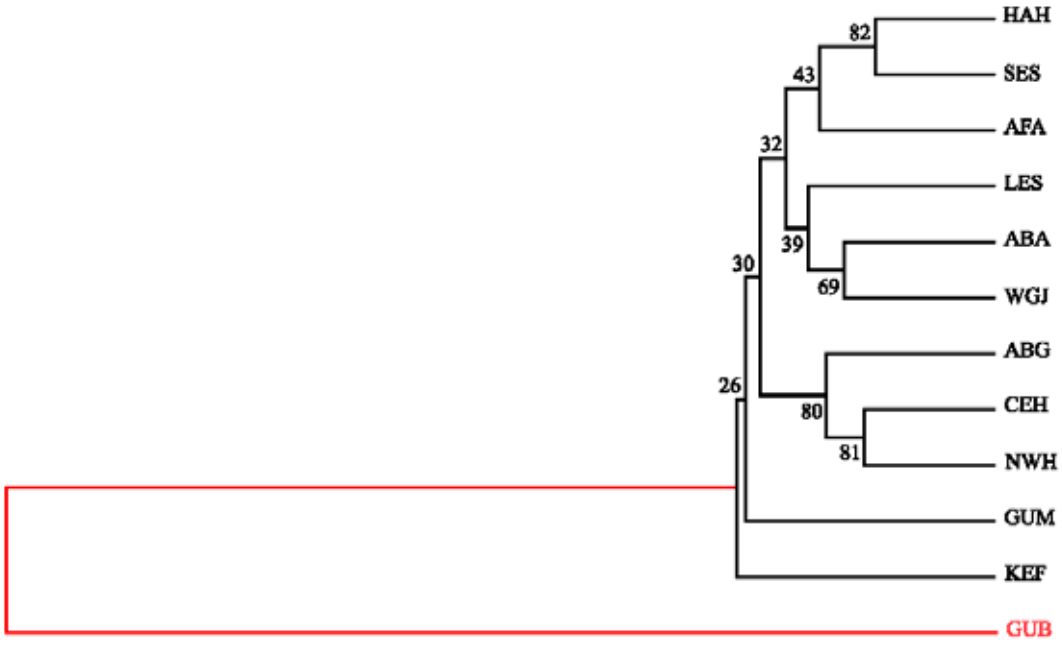


Figure 36. ***D*A-UPGMA tree representing the genetic relationship between the 11 Ethiopian populations with GUB as an out-group population.** The NJ tree was constructed using the genetic distance DA calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci

(ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji, GUB –Guinea Bissau)

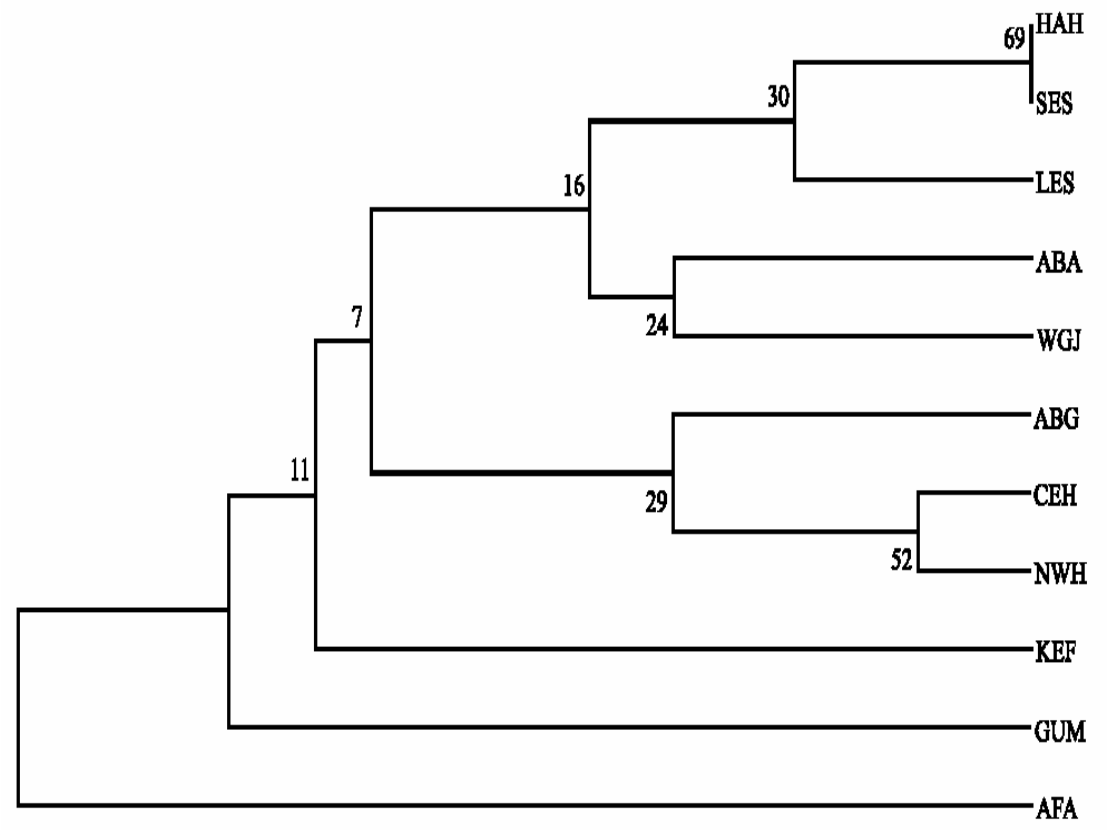


Figure 37. ***D*S-UPGMA tree representing the genetic relationship between the 11 Ethiopian populations.** The NJ tree was constructed using the genetic distance DS calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

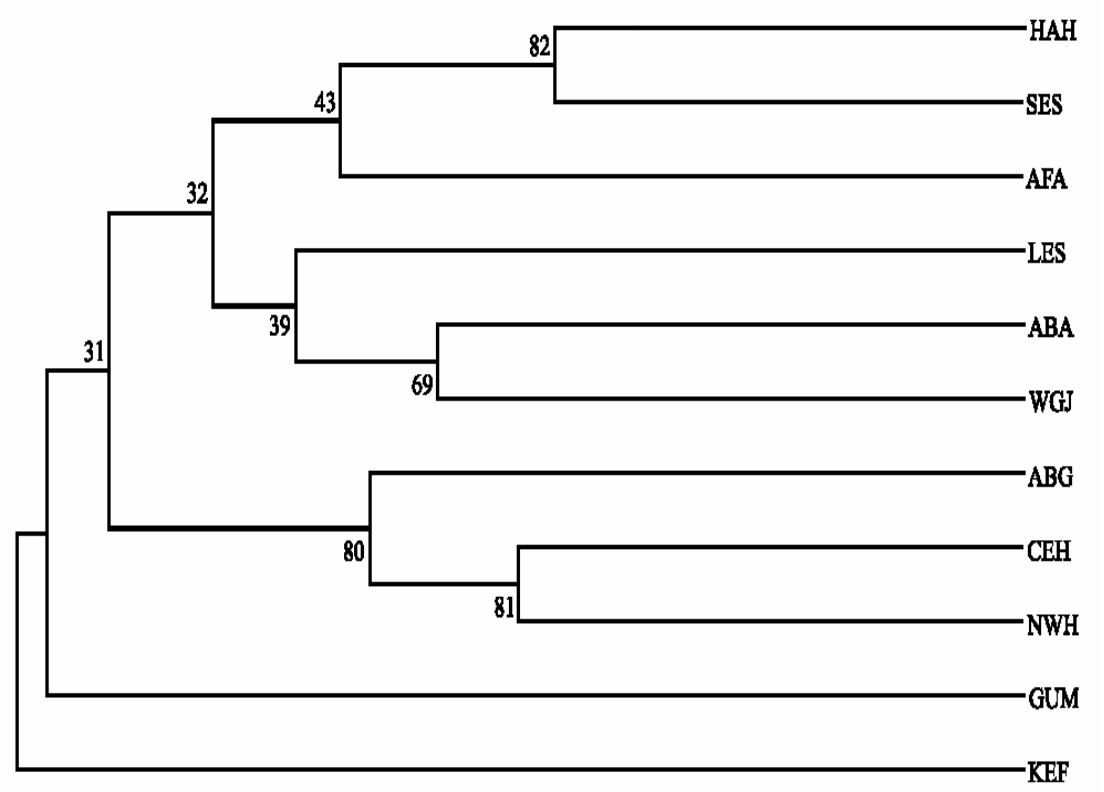


Figure 38. ***D*A-UPGMA tree representing the genetic relationship between the 11 Ethiopian populations.** The NJ tree was constructed using the genetic distance DA calculated from allele frequencies observed at 15 microsatellite loci. The numbers at the juncture of two branches are the percentage bootstrap values obtained after 1000 replications of re-sampled loci.

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

### 4.9.4. Principal component analysis (PCA)

**Table 29** shows the results of the PC analysis of the allele frequency data for the 11 Ethiopian goat populations using 15 microsatellite markers. The variation accounted for by PC1, PC2 and PC3 was 24.64%, 18.62% and 15.42%, respectively. The PC1 separated the Abergalle, Central Highland, Keffa, North-west Highland, Gumez and Arsi-Bale on one end from Woyto-Guji, Long-eared Somali, Hararghe Highland, Short-eared Somali, and

Afar populations on the other extreme. PC2 ranked Woyto-Guji, Gumez, Long-eared Somali, Arsi-Bale, and Keffa on one end and Short-eared Somali, North-west Highland, Hararghe Highland, Abergalle, Afar, and Central Highland on the other extreme. PC3 ranked Keffa, Arsi-Bale, Hararghe Highland, North-west Highland and Woyto-Guji on one end and Long-eared Somali, Short-eared Somali, Afar, Abergalle and Gumez, on the other extreme with Central Highland at the center.

**Figure 39** to **Figure** 41 is two-dimensional plots of PC1 with PC2, PC1 with PC3 and PC2 with PC3, respectively, showing the same thing from the Table 26 in figure form.

Table 29. **Ranked PC values of allele frequencies from 11 Ethiopian populations using 15 microsatellite markers.**

| % Variation accounted | 24.64 | | 18.62 | | 15.42 |
|---|---|---|---|---|---|
| Cumulative % | 24.64 | | 43.26 | | 58.68 |
| Breed | PC1 | Breed | PC2 | Breed | PC3 |
| ABG | -0.122 | WGJ | -0.1 | KEF | -0.106 |
| CEH | -0.065 | GUM | -0.085 | ABA | -0.071 |
| KEF | -0.055 | LES | -0.06 | HAH | -0.034 |
| NWH | -0.053 | ABA | -0.057 | NWH | -0.016 |
| GUM | -0.031 | KEF | -0.018 | WGJ | -0.011 |
| ABA | -0.027 | SES | 0.016 | CEH | 0 |
| WGJ | 0.021 | NWH | 0.021 | LES | 0.009 |
| LES | 0.051 | HAH | 0.055 | SES | 0.015 |
| HAH | 0.061 | ABG | 0.062 | AFA | 0.027 |
| SES | 0.066 | AFA | 0.069 | ABG | 0.065 |
| AFA | 0.154 | CEH | 0.097 | GUM | 0.123 |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji.

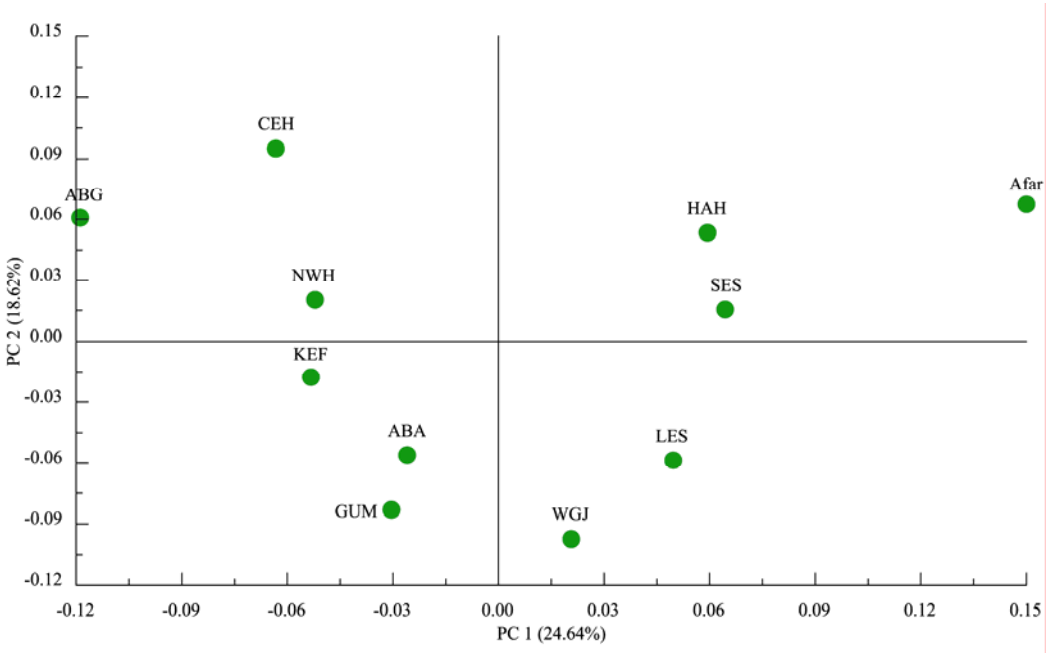


Figure 39. **Two-dimensional graph representing the relationship between PC1 and PC2 using allele frequency data from 15 microsatellite loci typed in 11 Ethiopian goat populations.**

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

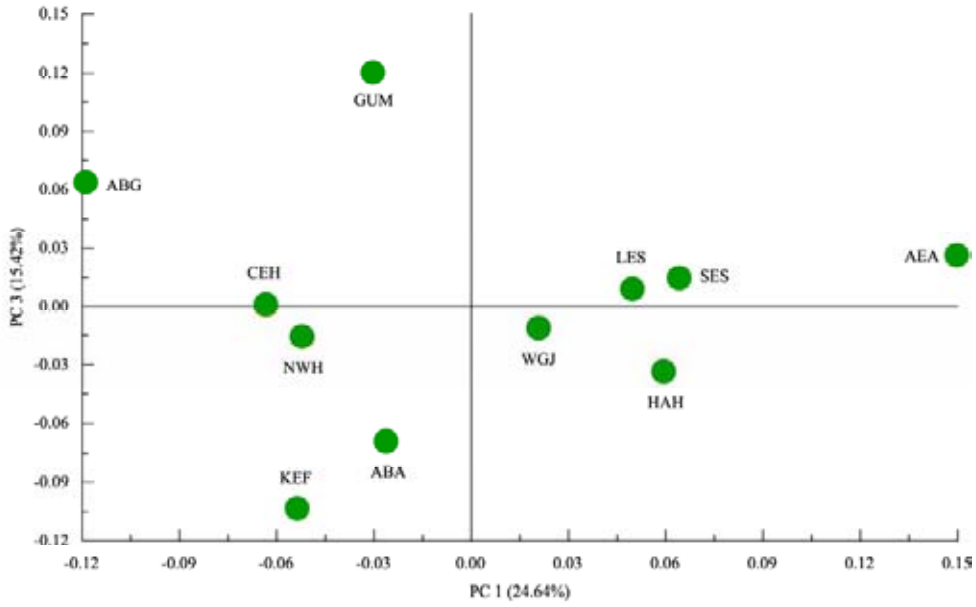


Figure 40. **Two-dimensional graph representing the relationship between PC1 and PC3 using allele frequency data from 15 microsatellite loci typed in 11 Ethiopian goat populations.**

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

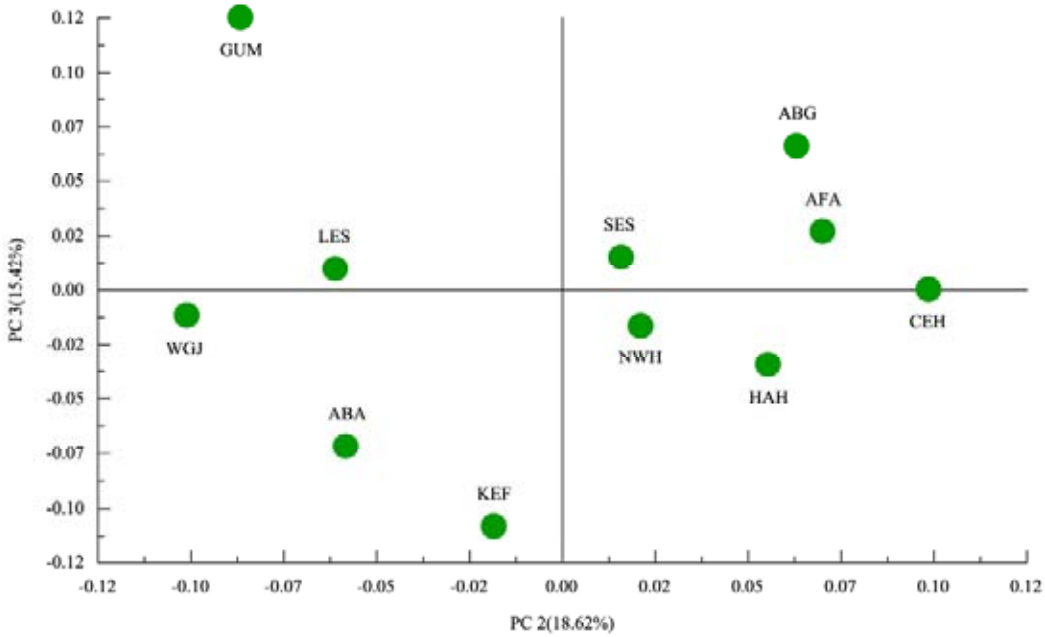


Figure 41. **Two-dimensional graph representing the relationship between PC2 and PC3 using allele frequency data from 15 microsatellite loci typed in 11 Ethiopian goat populations.**

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

### 4.9.5. Inference of population structure

Figure **42** and **Figure** 43 show the population structure, displayed with population Q-matrix and individual Q-matrix, respectively, inferred from allelic frequencies at 15 microsatellite loci. In this case there was no clear population structure inferred from Ethiopian populations by applying any K values.

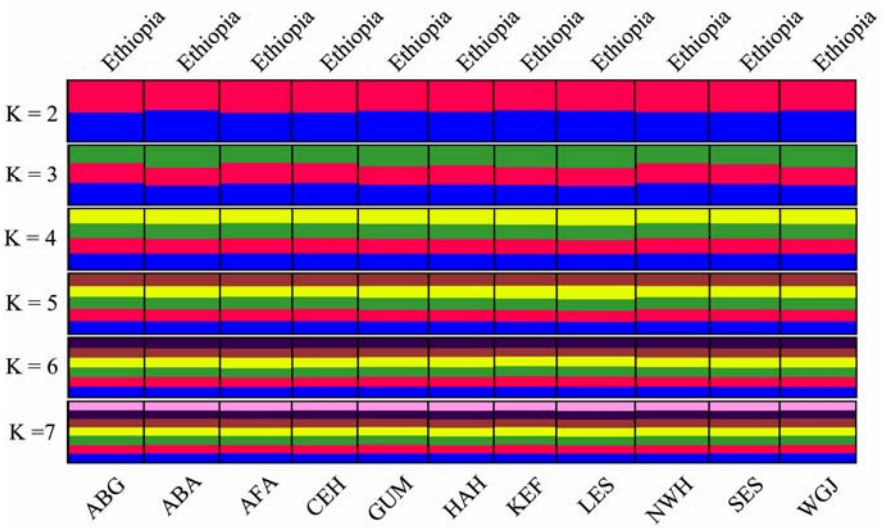


**Figure 42. Estimated population structure displayed with population Q-matrix.** Populations were partitioned in to K-coloured segments, which represent population's estimated average membership fractions in the K clusters. The black lines separate the different populations used for the study. Populations are labeled below the figure with their country affiliations above the figure. Five runs at each K produced nearly similar population membership coefficients

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

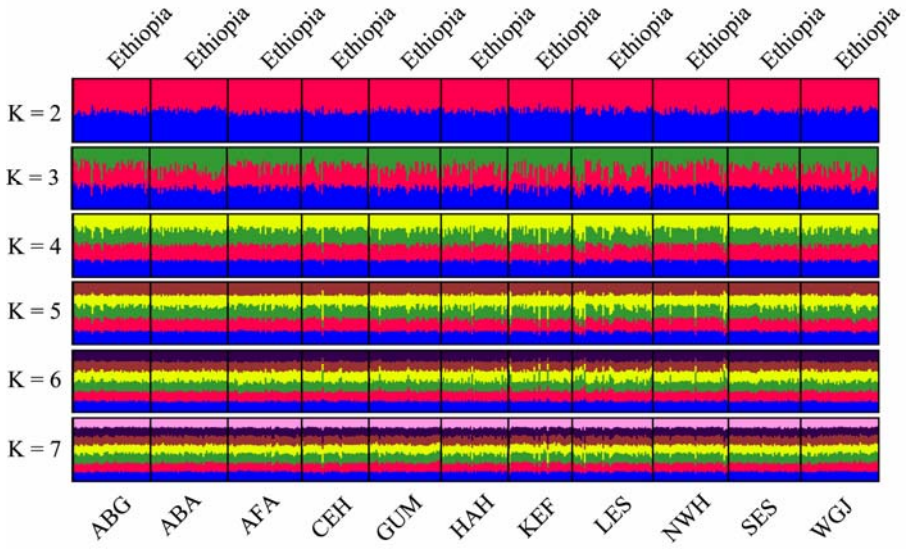


Figure 43. **Estimated population structure displayed with individual Q-matrix.** Each individual was represented by a thin vertical line, which is partitioned in to K coloured

segments, which represent individual's estimated average membership fractions in the K clusters. The black lines separate the individuals of different populations used for the study. Populations are labeled below the figure with their country affiliations above the figure. Five runs at each K produced nearly similar individual membership coefficients.

(ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji)

### 4.9.6. Analysis of gene flow among populations

Gene flow (also known as gene migration) is the transfer of genes from one population to another, resulted from movement of individuals or their gametes. Migration into or out of a population may be responsible for a marked change in gene pool frequencies (the number of individual members with a particular trait). Immigration may result in the addition of new genetic material to the established gene pool of a particular species or population, and conversely emigration results in the removal of genetic material. There are a number of factors that affect the rate of gene flow between different populations. One of the most significant factors is mobility, and animals tend to be more mobile than plants. Greater mobility of an individual tends to give it greater migratory potential.

Gene flow is usually expressed as migration rate *m*, defined as the proportion of alleles of migrant origin in a population per generation. Gene flow is particularly difficult to monitor directly, but usually is inferred from spatial distribution of genetic markers by statistical approaches. Most of the approaches are based on equilibrium expectations derived from theoretical models of population structure under neutrality theory, e.g. the 'island model' wherein a species is assumed to be subdivided into populations (demes or islands) of equal size *N*, all of which exchange alleles with equal probabilities (random migration between separate populations); or the 'stepping-stone' model wherein only adjacent demes (neighbouring populations) exchange alleles. Allelic frequencies in finite populations are influenced by random genetic drift also, which is a function of effective population size. The influence of drift and gene flow are difficult to tease apart and thus most statistical procedures applied to spatial genetic information only permit estimates of *N*m, which can be interpreted as the absolute number of individuals exchanged between populations per generation. Also *Nm* is of particular interest because under neutral theory,

the level of divergence among populations at equilibrium is a function of the numbers of migrants rather than the proportions of individuals exchanged. The most common approach to estimate *N*m is from *F-statistics* defined as the following formula:
$F_{ST}$ = 1/ (1 + 4*N*m) or *N*m = (1-$F_{ST}$)/4$F_{ST}$ (Wright, 1969).

The number of migrants per generation (*N*m), which is an indirect estimator of gene flow, was calculated for all pairs of populations using the above equation. The calculations were performed using the computer program GENETIX version 4.05 (available at http://www.univ-montp2.fr/~genetix/genetix/genetix.htm), and results are given in **Table 30**. For the 11 Ethiopian goat populations the highest *N*m was estimated between two geographically adjacent populations: Short-eared Somali & Haraghe Highland (999999), followed by 120.88 between North-West Highland and Central Highland populations, followed by 82.19 between Short-eared Somali & Long-eared Somali, while the lowest *N*m was estimated between Afar and Abergalle populations (6.06). The level of gene flow was relatively low between geographically distant or isolated populations (for example the *N*m was 6.06 between Abergalle and Afar, 7.3 between Arsi-Bale and Afar).

Table 30, **The number of migrants per generation, (Wright, 1969) among 11 goat populations estimated from Weir and Cockerham's (1984) $F_{ST}$ estimator using the equation: $F_{ST}$ = 1/ (1+4*N*m) or *N*m = (1-$F_{ST}$)/4$F_{ST}$.**

| | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | | | | | | | | | | | |
| ABA | 10.65 | | | | | | | | | | |
| AFA | 6.06 | 7.43 | | | | | | | | | |
| CEH | 35.79 | 19.2 | 9.63 | | | | | | | | |
| GUM | 10.27 | 10.53 | 7.04 | 10.89 | | | | | | | |
| HAH | 11.19 | 27.04 | 35.26 | 48.98 | 9.99 | | | | | | |
| KEF | 9.8 | 16.12 | 7.38 | 16.2 | 9.18 | 17.28 | | | | | |
| LES | 9.52 | 32.34 | 13.49 | 14 | 16.69 | 33.53 | 12.4 | | | | |
| NWH | 24.55 | 31.95 | 9.6 | 120.88 | 14.99 | 30.95 | 30.51 | 22.32 | | | |
| SES | 10.98 | 19.21 | 24.32 | 21.67 | 13.43 | 999999.00 | 11.83 | 82.19 | 53 | | |
| WGJ | 10.22 | 30.54 | 10.44 | 9.89 | 15.32 | 18.63 | 16.3 | 34.71 | 19.43 | 24.68 | |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji.

## 4.10. Genetic differentiation

### 4.10.1. Test on allele size information content

The mutation process at microsatellite loci typically occurs at high rates and with stepwise changes in allele sizes, features that may introduce bias when using classical measures of population differentiation based on allele identity (e.g., $F_{ST}$, Nei's $D_S$ genetic distance). Allele size-based measures of differentiation, assuming a stepwise mutation process [e.g., Slatkin's $R_{ST}$ (Slatkin, 1995), Goldstein *et al.*'s $\delta\mu^2$ (Goldstein *et al.*, 1995)] may better reflect differentiation at microsatellite loci, but they suffer high sampling variance (Hardy *et al.*, 2003). The relative efficiency of allele size- vs. allele identity-based statistics depends on the relative contributions of mutations vs. drift to population differentiation.

Hardy *et al.* (2003) developed a method based on a randomization procedure of allele sizes to determine whether stepwise-like mutations contributed to genetic differentiation. This test can be applied to any microsatellite data set designed to assess population differentiation and can be interpreted as testing whether $F_{ST} = R_{ST}$.

The test indicates whether allele sizes provide information on population differentiation given a data set, that is, whether shifts in allele sizes resulting from stepwise-like mutations contribute to population differentiation. Contribution of stepwise-like mutations to genetic differentiation requires (i) that the mutation process is at least partially SMM-like and (ii) that the mutation rate, $\mu$, is large enough relative to the effect of drift and migration (*e.g.* $\mu \geq m$, otherwise new mutations are quickly spread beyond their native population by migration). The principle of the test is based on obtaining a distribution of a statistic under the null hypothesis ($H_O$) that differences in allele sizes do not contribute to population differentiation. Therefore, a randomization procedure whereby the different allele sizes observed at a locus for a given data set is randomly permuted among allelic states. Throughout the randomization procedure, genotypes are defined in terms of allelic states and are not modified, but allele sizes are randomly reassigned among allelic states. After such a randomization, any two genes originally having the same allele size remain identical, although it can be for another allele size, whereas any two genes originally bearing different alleles of small size difference may bear alleles of large size

difference, or reciprocally. Hence, the allele identity information is kept intact but not the allele size information. Under the null hypothesis, the randomization procedure should not affect the expectation of a measure of differentiation such as $R_{ST}$. On the contrary, if allele sizes contribute to genetic differentiation, the $R_{ST}$ computed after allele size permutation (hereafter called $pR_{ST}$) would depend solely on allele identity/nonidentity and hence have a smaller expectation than the value computed before randomization. The test can thus be designed by comparing the observed $R_{ST}$ value (before randomization) to the distribution of $pR_{ST}$ values obtained for all possible configurations of allele size permutations. From this comparison, a probability that the null hypothesis holds can be estimated as the proportion of $pR_{ST}$ values larger than the observed $R_{ST}$ (one-tailed test). On a multi-locus $R_{ST}$ estimate, permuting allele sizes within each locus can carry out the test. It is noteworthy that the test makes no assumptions on the mutation model: A significant result ($R_{ST}$ significantly > $pR_{ST}$) suggests that mutations contributed to genetic differentiation, and that the mutation process follows at least partially a SMM and $F_{ST}$ is likely to provide a biased estimate of gene flow parameters. A non-significant result ($R_{ST}$ not significantly different from $pR_{ST}$) would suggest that allele size is not informative for population differentiation, because the mutation process is not stepwise-like and/or because mutations had not contributed to differentiation. In this case, allele identity-based statistics such as $F_{ST}$ *and* Nei's $D_S$ *genetic* distance should be preferential over allele size-based ones such as $R_{ST}$ and $\delta\mu^2$.

The global $R_{ST}$ statistics were computed and $pR_{ST}$ values were obtained by applying the allele size permutation tests for our dataset. Significance levels of $R_{ST}$ values were determined after 20,000 random permutations of the allele sizes. The analyses were performed with computer program SPAGeDi (Hardy and Vekemans, 2002), and the results are in **Table 31**. The $R_{ST}$ was not significantly ($P < 0.05$) different from $pR_{ST}$ values for either the $R_{ST}$ or for the multi-locus test. The tests indicate that mutations have not contributed to the genetic differentiations among Ethiopian goat populations, and genetic drift was responsible for the genetic differentiations among Ethiopian goat populations. Therefore, the allele identity based statistics such as $F_{ST}$ (Weir and Cockerham, 1984), Nei's standard genetic distance $D_S$ (Nei, 1972), $D_A$ genetic distance (Nei *et al.,* 1983) and $D_C$ (Cavalli-Sforza and Edwards, 1967) could be used to explain genetic differentiations and genetic relationships between goat populations.

Table 31. **Differentiation among Ethiopian goat populations, estimated by global *R*ST, mean *pR*ST values per locus and for a multilocus average.**

| Locus | No. of alleles | $R_{ST}$ | ${}_{P}R_{ST}$ (95% C.I.) |
|---|---|---|---|
| BM1818 | 12 | $0.009^{ns}$ | 2.56E-05 (-0.008, 0.012) |
| BMS1494 | 9 | $0.033^{ns}$ | -2.27E-05 (-0.008, 0.012) |
| ILSTS05 | 5 | $0.011^{ns}$ | -7.54E-05 (-0.008, 0.012) |
| ILSTS11 | 10 | $0.049^{ns}$ | 4.67E-05 (-0.008, 0.013) |
| ILSTS44 | 5 | $0.019^{ns}$ | -5.15E-05 (-0.008, 0.012) |
| ILSTS87 | 8 | $0.025^{ns}$ | 2.82E-05 (-0.008, 0.013) |
| INRA05 | 5 | $0.002^{ns}$ | 7.49E-06 (-0.008, 0.012) |
| INRA63 | 8 | $0.014^{ns}$ | -5.47E-05 (-0.009, 0.014) |
| INRA132 | 3 | $0.016^{ns}$ | 1.46E-05 (-0.008, 0.012) |
| MAF35 | 3 | $0.000^{ns}$ | 1.32E-05 (-0.007, 0.011) |
| MAF65 | 12 | $0.018^{ns}$ | -5.45E-05 (-0.009, 0.014) |
| MAF209 | 4 | $0.019^{ns}$ | -5.09E-05 (-0.009, 0.013) |
| OarAE129 | 20 | $0.030^{ns}$ | 1.47E-05 (-0.009, 0.013) |
| OarFCB304 | 23 | $0.001^{ns}$ | 2.97E-05 (-0.008, 0.013) |
| SRCRSP03 | 6 | $0.035^{ns}$ | 3.63E-06 (-0.008, 0.013) |
| All loci | 133 | $0.024^{ns}$ | 9.57E-06 (-0.005, 0.006) |

The 95% confidence interval given with $pR_{ST}$ is the 95% central $pR_{ST}$ values obtained after 20,000 random permutations of the allele sizes. *P*-values of allele size permutation tests on $R_{ST}$ are denoted as follows: ns, non-significant ($P > 0.05$).

### 4.10.2. Genetic differentiation between populations

**Table 32** shows two measures of population sub-division, the $G_{ST}$ (Nei, 1973) and $F_{ST}$ (Weir and Cockerham, 1984) values across populations. The values for the two measures of population differentiation were similar with the $G_{ST}$ values being slightly higher. The $G_{ST}$ values ranged from 0.010 for locus MAF35 to 0.044 for locus ILSTS11, while $F_{ST}$ ranged from 0.000 to 0.037 for the same loci respectively. The overall values were 0.026 for $G_{ST}$ and 0.016 for $F_{ST}$, showing that 97.4 to 98.4% of the total genetic differentiation was due to differences between individuals within populations.

**Table 33** and **Table** 34 show the $F_{ST}$ values and their level of significance for pairs of Ethiopian populations alone and with one out-group population respectively. The least genetically differentiated population pairs were the Short-eared Somali/Hararghe Highland, North-West Highland/Central Highland, and Short-eared Somali/Long-eared Somali with $F_{ST}$ value of -0.001, 0.002 and 0.003 respectively in both cases. The population pair with the highest genetic differentiation was Afar/Abergalle with $F_{ST}$ value of 0.040 in both cases. The genetic differentiation observed among the population pairs were highly significant ($P < 0.001$) for all except for two pairs ($P < 0.01$) and, eight pairs ($P < 0.05$) even where low $F_{ST}$ values were obtained, indicating that there was significant level of genetic differentiation between the Ethiopian goat populations.

The $F_{ST}$ values calculated for each population pair and each locus determine the ability of each locus to differentiate between population pairs. **Table 35** and **Appendix 36** show the efficiency of the 15 microsatellite loci to differentiate between the 55 population pair combinations. Out of the 15 loci, ILSTS11 was able to differentiate 72.73% of the population pair combinations (40 out of 55 population pair combinations) at $P < 0.05$ or lower, making it the most informative marker. The next informative marker was MAF65, which was able to differentiate 69.09% of the population pairs (38 out of 55 population pair combinations) at $P < 0.05$ or lower. On the other hand both BM1818 and ILSTS44 were able to differentiate only 20% of the population pairs (11 out of 55 population pair combinations each) at $P < 0.05$ or lower followed by MAF35 which was able to differentiate only 3.6% of the population pairs (2 out of 55 population pair combinations) at $P < 0.05$ or lower, which makes it the least informative marker.

Table 32. **Measure of population differentiation in Ethiopian goat populations**.

| Loci | $G_{ST}$ | $F_{ST}$ |
|---|---|---|
| BM1818 | 0.021 | 0.011 |
| BMS1494 | 0.025 | 0.015 |
| ILSTS05 | 0.031 | 0.021 |
| ILSTS11 | 0.044 | 0.037 |
| ILSTS44 | 0.024 | 0.014 |
| ILSTS87 | 0.025 | 0.017 |
| INRA05 | 0.029 | 0.021 |
| INRA63 | 0.015 | 0.004 |
| INRA132 | 0.025 | 0.015 |
| MAF35 | 0.010 | 0.000 |
| MAF65 | 0.027 | 0.017 |
| MAF209 | 0.026 | 0.017 |
| OarAE129 | 0.019 | 0.009 |
| OarFCB304 | 0.019 | 0.009 |
| SRCRSP03 | 0.038 | 0.030 |
| **Overall loci** | **0.026** | **0.016** |

Table 33. **Pair wise tests of population differentiation between the 11 Ethiopian goat populations. $F_{ST}$ values (below the diagonal) and their test of significance (above the diagonal) obtained after 55,000 permutations. * = $P$ < 0.05, ** = $P$ < 0.01, *** = $P$ < 0.001.**

| Pop | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | *** | *** | * | *** | *** | *** | *** | *** | *** | *** |
| ABA | 0.023 | - | *** | *** | *** | *** | *** | *** | *** | *** | * |
| AFA | 0.040 | 0.033 | - | *** | *** | * | *** | *** | *** | * | *** |
| CEH | 0.007 | 0.013 | 0.025 | - | *** | * | *** | *** | ns | ** | *** |
| GUM | 0.024 | 0.023 | 0.034 | 0.022 | - | *** | *** | *** | *** | *** | *** |
| HAH | 0.022 | 0.009 | 0.007 | 0.005 | 0.024 | - | *** | ** | *** | ns | *** |
| KEF | 0.025 | 0.015 | 0.033 | 0.015 | 0.027 | 0.014 | - | *** | *** | *** | *** |
| LES | 0.026 | 0.008 | 0.018 | 0.018 | 0.015 | 0.007 | 0.020 | - | *** | ns | * |
| NWH | 0.010 | 0.008 | 0.025 | 0.002 | 0.016 | 0.008 | 0.008 | 0.011 | - | * | *** |
| SES | 0.022 | 0.013 | 0.010 | 0.011 | 0.018 | -0.001 | 0.021 | 0.003 | 0.005 | - | * |
| WGJ | 0.024 | 0.008 | 0.023 | 0.025 | 0.016 | 0.013 | 0.015 | 0.007 | 0.013 | 0.010 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji.

Table 34. **Pair wise test of population differentiation between the 11 Ethiopian and 1 reference population. $F_{ST}$ values (below the diagonal) and their test of significance (above the diagonal) obtained after 66,000 permutations.**

| Pop | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | ABG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | *** | *** | * | *** | *** | *** | *** | ** | *** | *** | *** |
| ABA | 0.023 | - | *** | *** | *** | ** | *** | *** | *** | *** | * | *** |
| AFA | 0.040 | 0.033 | - | *** | *** | * | *** | *** | *** | * | *** | *** |
| CEH | 0.007 | 0.013 | 0.025 | - | *** | * | *** | *** | ns | ** | *** | *** |
| GUM | 0.024 | 0.023 | 0.034 | 0.022 | - | *** | *** | *** | *** | *** | *** | *** |
| HAH | 0.022 | 0.009 | 0.007 | 0.005 | 0.024 | - | *** | * | *** | ns | ** | *** |
| KEF | 0.025 | 0.015 | 0.033 | 0.015 | 0.027 | 0.014 | - | *** | *** | *** | *** | *** |
| LES | 0.026 | 0.008 | 0.018 | 0.018 | 0.015 | 0.007 | 0.020 | 0 | *** | ns | * | *** |
| NWH | 0.010 | 0.008 | 0.025 | 0.002 | 0.016 | 0.008 | 0.008 | 0.011 | - | * | *** | *** |
| SES | 0.022 | 0.013 | 0.010 | 0.011 | 0.018 | -0.001 | 0.021 | 0.003 | 0.005 | - | * | *** |
| WGJ | 0.024 | 0.008 | 0.023 | 0.025 | 0.016 | 0.013 | 0.015 | 0.007 | 0.013 | 0.010 | - | *** |
| GUB | 0.142 | 0.130 | 0.134 | 0.143 | 0.137 | 0.130 | 0.135 | 0.132 | 0.142 | 0.129 | 0.110 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji,, GUB – Guinea Bissau.

Table 35. **Summary of genotypic differentiation tests for all the 55 population combination pairs of Ethiopian populations for all the loci.**

| Locus | Percentage of population pairs that could be distinguished by the marker, $P < 0.05$ |
|---|---|
| BM1818 | 20.00 |
| BMS1494 | 50.91 |
| ILSTS05 | 36.36 |
| ILSTS11 | 72.73 |
| ILSTS44 | 20.00 |
| ILSTS87 | 50.91 |
| INRA05 | 40.00 |
| INRA63 | 29.09 |
| INRA132 | 34.55 |
| MAF35 | 3.60 |
| MAF65 | 69.09 |
| MAF209 | 34.55 |
| OarAE129 | 30.91 |
| OarFCB304 | 63.64 |
| SRCRSP03 | 60.00 |
| **Average for all population** | **41.21** |

### 4.10.3. Analysis of Isolation by Distance (IBD)

IBD is a fast and simple application for analyzing (genetic) isolation by (geographic) distance. The genetic similarity or differentiation among individuals or populations can be ascertained using a number of statistical techniques (re-viewed by Bohonak, 1999; Neigel, 1997; Roderick, 1996; Slatkin, 1985). When populations can be defined a priori, one option is to analyze genetic 'isolation by distance' (Wright, 1943) by plotting the genetic similarity (or distance) among population pairs as a function of the geographic distance between those pairs.

Qualitative and statistical analyses of isolation by distance can reveal much about population genetic structure. The primary use for plots of (genetic) isolation by (geographic) distance is to assess whether more distant population pairs are more different genetically. However, these plots can also be used to test the validity of simpler models of population structure (e.g. island or hierarchical island models). Isolation by distance analyses may help separate the effects of population history from ongoing gene flow, and test the explanatory power of alternative dispersal pathways and the influence of geographic features or specific life-history traits on population differentiation can also be tested (Bohonak, 2002).

Studies of isolation by distance typically seek to ascertain (i) whether there is a statistically significant relationship between genetic distance (similarity) and geographic distance, and (ii) the strength of this relationship. Significance is usually assessed by asking whether the pair-wise genetic distance matrix is correlated with the pair-wise geographic distance matrix using a Mantel test (Manly, 1994). For the genetic distance matrix *A* and the geographic distance matrix *B*, the test statistic is calculated as $Z = Zi,jAijBij$. IBD also reports an alternative statistic, *r*, which provides a standardized *Z* that ranges from -1 to 1 (Manly, 1994). Significance is assessed by comparing *Z* actual to a distribution of *Z* scores obtained by randomizing rows/columns of the *B* matrix and holding *A* constant. The IBD application provides one-tailed *P*-values for this distribution.

A logical way to quantify the strength of the isolation by distance relationship is to calculate the slope and intercept of genetic similarity or distance against geographic distance. Based on simulations, Hellberg (1994) suggested that reduced major axis (RMA)

regression is more appropriate for this purpose than standard ordinary least squares (OLS) regression.

The Ethiopian goat populations are found in a wide range of production systems all over the country (**Figure 19**) and the genetic differentiation may not have resulted from geographic isolation. The correlation between genetic distance $F_{ST}/(1-F_{ST})$ of Rousset (1997) and geographic distance (km) among the Ethiopian goat populations was assessed using Mantel test (Manly, 1986, 1994) with 100,000 randomizations. The test was implemented using a computer program IBD (Bohonak, 2002; available at http://www.bio.sdsu.edu/pub/andy/ibd.html). $F_{ST}$ distances were calculated with computer program FSTAT version 2.9.3 (Goudet, 2001) according to Weir and Cockerham (1984).

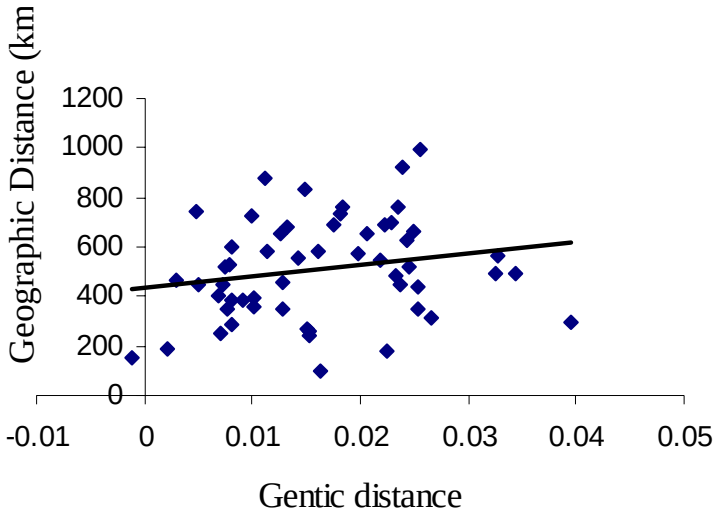


Figure 44. **Genetic differentiation among 11 Ethiopian goat populations.** Multi-locus estimates of pair-wise genetic distance (FST) is plotted against geographic distance (km). The regression is Y = 4723.4x + 431.05, Z = 475.97, r = 0.206, P < 0.1065. FST was calculated according to Weir and Cockerham (1984).

**Figure 44** illustrates the correlation between genetic distance $F_{ST}/(1–F_{ST})$ and geographic distance (km) among Ethiopian goat populations for different geographic regions. The geographic distance among Ethiopian goat populations ranges between 96.75 km to 993.31 km, and no significant correlation was detected between genetic distance and geographic distance at the entire geographic regions.

### 4.10.4. Analysis of molecular variance (AMOVA)

In order to understand the partitioning levels of the genetic diversity of the Ethiopian goats, an analysis of molecular variance (AMOVA) was conducted. The result of the AMOVA (**Table 36**) revealed that 98.36% of the total genetic diversity existed among the individuals within populations and only 1.64% of the total genetic diversity accounted for differences among populations.

Table 36. **Analysis of molecular variance (AMOVA) for 11 Ethiopian goat populations, using genotype data from 15 microsatellite loci.**

| Source of variation | Degree of freedom | Sum of squares | Variance components | Percentage variation |
|---|---|---|---|---|
| Among populations | 10 | 104.527 | 0.07109 | 1.64 |
| Within populations | 947 | 4037.97 | 4.26396 | 98.36 |

### 4.10.5. Individual specific analysis

Individuals were assigned to source populations using the Cornuet *et al.* (1999) GeneClass2 program, which allows different estimation procedures including the Bayesian (Rannala and Mountain, 1997), and frequency (Paetkau *et al*., 1995) as well as six distance methods; the Nei's $D_A$ (Nei *et al*., 1983), Cavalli-Sforza and the Edwards chord distance (1967), the Nei Minimum distance (Nei, 1973), the Nei $D_S$ (Nei, 1972), the Jin and Chakraborty's (1993) $D_{AS}$ and the $\delta\mu^2$ distance method of Goldstein *et al.* (1995).

The efficiency of the different methods to assign individuals to their source population was tested and the results are shown in **Table 37**. In all cases the percentage of individuals being correctly assigned to their source populations ranged from the least 13.57% with Goldstein *et al.* (1995) $\delta\mu^2$ method to the best (in this case) 35.07% with the Bayesian method.

The percentage of correct assignment of individuals to their source populations using a single locus was very low ranged from 9.19% for ILSTS05 to 15.87% for ILSTS87 (**Table 38**). However, when the five loci with the highest (in this case) individuals i.e. ILSTS87, MAF65, BMS1494, ILSTS11 and OarAE129, were used together the percentage

of individuals correctly assigned was increased to 28.39% (**Table 38**). Increasing the number of microsatellite loci to 15 resulted in correct assignment of only 35.07% of the individuals.

**Table 39** shows the results of the population assignment test obtained using frequency data from the 11 Ethiopian populations. Only 35.07% of the 479 individuals were assigned to the corrected source population. The population with the highest level of correct assignment was the Gumez with 55.81% of the individuals being correctly assigned, followed by the Keffa population with 47.37% of the individuals being correctly assigned to the source population. In general all the Ethiopian populations had low levels of correct assignment with the Hararghe Highland population having the lowest level of assignment (7.50%).

Table 37. **The accuracy of different methods to assign individuals to the source population**.

| Methods | Software | % Correct assignment |
|---|---|---|
| Bayesian | GeneClass | 35.07 |
| Maximum likelihood | WHICHRUN | 33.90 |
| Frequency | GeneClass | 33.19 |
| The Nei *et al*. (1983) $D_A$ | GeneClass | 32.36 |
| Cavalli-Sforza and the Edwards chord distance | GeneClass | 31.52 |
| The Nei standard distance ($D_{S)}$ | GeneClass | 27.56 |
| Nei Minimum distance | GeneClass | 26.72 |
| The Chakraborty and Jin $D_{AS}$ | GeneClass | 25.05 |
| The Goldstein *et al*. (1995) $(d\mu)^2$ | GeneClass | 13.57 |

Table 38. **Percent of correct individual-population assignment (only for Ethiopian populations) estimated for each microsatellite locus using the Bayesian method.**

| Locus | Percent of correct assignment |
|---|---|
| ILSTS87* | 15.87 |
| MAF65* | 15.66 |
| BMS1494* | 15.45 |
| ILSTS11* | 15.45 |
| OarAE129* | 15.24 |
| INRA63 | 14.82 |
| ILSTS44 | 14.61 |
| OarFCB304 | 14.61 |
| INRA132 | 13.78 |
| SRCRSP03 | 13.78 |
| MAF209 | 13.57 |
| BM1818 | 12.94 |
| INRA05 | 9.81 |
| MAF35 | 9.81 |
| ILSTS05 | 9.19 |
| *When these five loci taken together | 28.39 |

Table 39. **Frequency of correct assignment of individuals to the 11 Ethiopian populations using genotype information from 15 microsatellite loci using the Bayesian method.**

| Populations | No. | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | 46 | **45.65** | 10.87 | 0.00 | 6.52 | 0.00 | 2.17 | 4.35 | 4.35 | 15.22 | 2.17 | 8.70 |
| ABA | 46 | 6.52 | **36.96** | 2.17 | 4.35 | 0.00 | 4.35 | 6.52 | 10.87 | 2.17 | 15.22 | 10.87 |
| AFA | 44 | 4.55 | 0.00 | **43.18** | 2.27 | 6.82 | 18.18 | 2.27 | 2.27 | 2.27 | 11.36 | 6.82 |
| CEH | 40 | 5.00 | 12.50 | 2.50 | **32.50** | 5.00 | 10.00 | 10.00 | 0.00 | 15.00 | 7.50 | 0.00 |
| GUM | 43 | 2.33 | 2.33 | 0.00 | 4.65 | **55.81** | 4.65 | 6.98 | 2.33 | 9.30 | 0.00 | 11.63 |
| HAH | 40 | 2.50 | 12.50 | 20.00 | 2.50 | 7.50 | **7.50** | 7.50 | 7.50 | 5.00 | 20.00 | 7.50 |
| KEF | 38 | 2.63 | 0.00 | 2.63 | 2.63 | 10.53 | 2.63 | **47.37** | 5.26 | 10.53 | 7.89 | 7.89 |
| LES | 48 | 6.25 | 6.25 | 6.25 | 6.25 | 6.25 | 8.33 | 2.08 | **31.25** | 0.00 | 18.75 | 8.33 |
| NWH | 45 | 13.33 | 6.67 | 6.67 | 13.33 | 2.22 | 2.22 | 4.44 | 8.89 | **28.89** | 4.44 | 8.89 |
| SES | 43 | 2.33 | 4.65 | 11.63 | 4.65 | 4.65 | 11.63 | 4.65 | 11.63 | 11.63 | **23.26** | 9.30 |
| WGJ | 46 | 2.17 | 10.87 | 4.35 | 0.00 | 2.17 | 8.70 | 10.87 | 8.70 | 8.70 | 10.87 | **32.61** |

No. = number of animals; ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji.

### 4.10.6. Marginal loss of genetic diversity

Weitzman (1992, 1993) has suggested a method that uses genetic and non-genetic information to calculate a maximum-likelihood tree and the current diversity of a group of species and to Asses the expected change in diversity over a certain time horizon. This method allows evaluation of the change in the group's diversity when the degree of endangerment of one or several species is reduced (marginal diversity). Compared with other approaches (Crozier, 1992; Faith, 1992) mostly descriptive and illustrative, the Weitzman method provides a well-defined diversity metric, allowing quantitative assessment of the consequences of alternative conservation activities as an aid to decision-making in conservation. An application to livestock breeds was presented by Thaon d'Arnoldi *et al.* (1998), and the approach was used in studies on European cattle breeds (Bremond, 2001; Cañon *et al*., 2001), on African cattle breeds (Reist-Marti *et al.*, 2003) and European pig breeds (Laval *et al*., 2001).

Application of the Weitzman approach showed that removal of a random population from the total Ethiopian goat populations would result in a relative loss of genetic 'diversity' between 4 and 24%.

The marginal loss of genetic diversity after a removal of Afar, Abergalle, Gumez and Keffa populations from the total set were 24.32%, 19.22%, 16.59%, and 12.99%, respectively (**Table 40**). Removal of other populations had a lower impact being (Arsi-Bale; -6.23%; Woyto-Guji; -6.08%; Central Highland and North-West Highland combined; -5.33%; Hararghe Highland and Short-eared Somali combined; -4.73%; Long-eared Somali; -4.13% only).

Table 40. **Marginal loss of genetic diversity according to the Weitzman approach.**

| $F_{ST}$ V(S) = 0.133200 | | | |
|---|---|---|---|
| i | V(Si) | ΔV = V(Si) – V(S) | %V = ΔV/V(S) |
| AFA | 0.1008 | -0.0324 | -24.32 |
| ABG | 0.1076 | -0.0256 | -19.22 |
| GUM | 0.1111 | -0.0221 | -16.59 |
| KEF | 0.1159 | -0.0173 | -12.99 |
| ABA | 0.1249 | -0.0083 | -6.23 |
| WGJ | 0.1251 | -0.0081 | -6.08 |
| CEH&NWH | 0.1261 | -0.0071 | -5.33 |
| HAH&SES | 0.1269 | -0.0063 | -4.73 |
| LES | 0.1277 | -0.0055 | -4.13 |

i population.

V(S) Genetic diversity of the total set of populations.

V(Si) Genetic diversity of the total set, after removal of population i.

AFA- Afar, ABG- Abergalle, GUM- Gumez, KEF- Keffa, ABA- Arsi-Bale, WGJ- Woyto-Guji, CEH – Central Highland, NWH – North-West Highland, HAH – Hararghe Highland, SES – Short-eared Somali, LES- Long-eared Somali.

# CHAPTER FIVE

## 5.0. DISCUSSION

This is the first time that indigenous goat genetic resources of Ethiopia have been studied in details at molecular level using microsatellite loci. These results provide important information for the future management, breed genetic improvement and conservation of Ethiopian goat genetic resources.

The classification and characterization of livestock species/type is necessary before desired genetic materials can be preserved for future generations. Much genetic variability, the raw material of breed genetic improvement has been lost in developed countries. There is no doubt that preserving genetic diversity, for either immediate or future use, is important.

### 5.1. Within population genetic diversity

Genetic diversity is a fundamental component of biodiversity, forming the basis of species and ecosystem. It represents all of the genetically determined differences that occur between individuals of a species in the expression of a particular trait or a set of traits. Since species are composed of populations that exist somewhat independently each other, genetic diversity exists both within and among populations of species.

A total of 159 alleles were detected at 15 microsatellite loci in 776 goat individuals of 20 populations sampled from 20 locations in Ethiopia, Kenya, Botswana, Guinea Bissau, Egypt, Turkey, Italy, Saudi Arabia and Mongolia. The total number of alleles sampled in a population across all loci ranged between 80 in West African Dwarf (WAD) goat population from Guinea Bissau and 116 in Long-eared Somali goat population from Ethiopia. The mean number of alleles per locus (MNA) across all populations was 5.46 ± 0.158, varying between 4.67 ± 2.74 in Guinea Bissau WAD goat population and 6.47± 4.03 in Long-eared Somali goat population. The MNA for a uniformed sample size with 20 animals after random re-sampling procedure with 250 replacements over individuals was 4.88 ± 0.08, varying between 4.19 ± 2.26 in Guinea Bissau population and 5.74 ± 2.89 in Hair goat population. In general, the mean numbers of alleles were comparable among all the populations from different geographic regions. The sample size of each population ranges from 22 individuals in Baladie population from Egypt to 48 individuals in Long-eared Somali population from Ethiopia. The number of alleles at each locus in the Ethiopian populations ranged from

4 to 23, which was slightly higher than that of the Swiss goat breeds (3 to 19) (Saitbekova, 1999), Asian goat populations (5 to 18) (Barker, 2001) and Chinese goat populations (4 to 19 and 4 to 13) (Li *et al.*, 2002; Li and Valentini, 2004).

The mean observed heterozygosity ($H_O$) and expected heterozygosity ($H_E$) for Ethiopian goats was 0.551 ± 0.015 and 0.576 ± 0.015, respectively, across all loci and all population. The observed heterozygosity across all loci ranged between 0.515 ± 0.047 in Ethiopian Afar population to 0.625 ± 0.050 in Saudi Arabian Ardi population. The highest expected heterozygosity (gene diversity) across all loci was observed in Hair goat population from Turkey (0.642 ± 0.055) and the lowest was 0.523 ± 0.058 in WAD (Guinea Bissau) population. The differences of both values among the goat populations were not statistically significant ($P > 0.05$), and it also seemed that the values were not affected by sample size.

The mean number of alleles per locus (MNA) and the expected heterozygosity ($H_E$) are two fundamental parameters in the population genetics commonly used to quantify the genetic variability within populations. The level of genetic diversity ($H_E$ = 0.58 ± 0.015) obtained for the Ethiopian goat populations in the present study was similar to that previously reported in the Swiss goat breeds ($H_E$ = 0.59, Saitbekova *et al.*, 1999), and Asian goats ($H_E$ = 0.52, Barker *et al.*, 2001), but less than that reported for Chinese goats ($H_E$ = 0.716, Li *et al.*, 2002; and $H_E$ = 0.707, Li and Alessio, 2004).

The genetic diversity of the Ethiopian goat populations was not high, as indicated by the moderate MNA and expected heterozygosities observed for those populations. The average number of alleles per locus was 8.87 (ranged from 4 to 23) for the 11 Ethiopian goat populations at the 15 microsatellite loci used in the study. The results of the AMOVA showed that most of the diversity for the Ethiopian goat populations is found within populations, rather than between populations (**Table 36**).

Seven Ethiopian goat populations showed a significant inbreeding coefficient ($F_{IS}$) ($P < 0.05$ for Abergalle, Hararghe Highland, Long-eared Somali, North-West Highland and Woyto-Guji populations, and $P < 0.01$ for Afar and Central Highland populations), and none of the Ethiopian goat populations showed out-breeding effect. This inbreeding could be a consequence of the management systems (like tethering of the animals in the case of small flock size during crop season, herding of the flock by children and/or women, where the flock size is larger and some localized transhumance during the dry season), traditional/cultural practices (like preference for a specific coat colour at different cultural or religious ceremonies) and also ethnic/cultural barriers that

could affect the between flock breeding-males(bucks) exchange. Tethering minimises the number of different bucks that any particular female (doe) can mate with, leading to service by only males those preferred by the owner; while women and children herders usually avoid the mixing of different flocks or exchange of bucks from different areas even at the watering point. Localised transhumance and ethnic/cultural barriers also lead to minimal encounters between different flocks. In case each sampled populations/types have been subdivided into subpopulations, which contains fewer heterozygotes than predicted despite the fact that all subdivisions are in Hardy-Weinberg equilibrium, can also lead to inbreeding (Wahlund effect). All these systems and events tend to increase the rate of inbreeding.

### 5.2. Between population genetic diversity

Genetic differentiation between the goat populations studied was mainly resulted from the genetic drift other than mutations at allele sizes among populations. Wright's fixation indices or *F*-statistics (Wright, 1969) are commonly used for describing genetic differentiations among populations. The unbiased estimators of *F*-statistics ($F_{IT}$, $F_{ST}$ and $F_{IS}$) were calculated for each locus across all populations with method of Weir and Cockerham (1984), and their significance levels were tested by permutation procedures. $F_{IS}$ statistic, an estimator for describing local inbreeding effect within subpopulations, was significantly different from zero at six out of 15 loci (BM1818 and MAF209 at $P < 0.05$; ILSTS05 at $P < 0.01$; ILSTS44, INRA63 and MAF65 at $P < 0.001$) and across all loci at $P < 0.001$ level, indicating that the Ethiopian goat populations sampled for this study were generally at risk of inbreeding. The $F_{IT}$ estimator, describing the inbreeding as well as the subdivision in the total population, was significantly different from zero at eight out of 15 loci (BM1818, OarAE129 and SRCRSP03 at $P < 0.05$; ILSTS05 and MAF209 at $P < 0.01$; ILSTS44, INRA63 and MAF65 at $P < 0.001$) and across all loci at $P < 0.001$ level. As local inbreeding effect was detected only at six of 15 loci, significant $F_{IT}$ statistic at these loci was merely resulted from the subdivision among goat populations. $F_{ST}$ estimator is commonly called the coefficient of gene differentiation (analogue to $G_{ST}$ of Nei, 1973) and used to measure the degree of genetic subdivision among populations. $F_{ST}$ estimator was significantly different from zero at 14 out of 15 loci (BM1818 and ILSTS44 at $P < 0.01$; BMS1494, ILSTS05 ILSTS11, ILSTS87, INRA05, INRA63, INRA132, MAF65, MAF209 OarAE129, OarFCB304

and SRCRSP03 at $P < 0.001$) and across all loci at $P < 0.001$, which means that genotype frequencies among populations are not randomly distributed, meaning that all the Ethiopian goat populations have been genetically subdivided even though the mean $F_{ST}$ value across all loci was very low (0.016) compared to previously reported in Asian goats ($F_{ST}$ = 0.143, Barker *et al.,* 2001) and Chinese goats ($F_{ST}$ = 0.105, Li *et al.,* 2002; and $F_{ST}$ = 0.054, Li and Valentini, 2004). Similar value was obtained for the coefficient of gene differentiation using the method of Nei (1973) ($G_{ST}$ = 0.026). Both values of $F_{ST}$ and $G_{ST}$ indicate that the level of gene differentiation between Ethiopian goat populations was very low. Only 1.6% to 2.6% of total variation was contributed by genetic subdivision among populations while 97.4% to 98.4% of total variation accounts for genetic variations among individuals within populations. This figure went up to 6.4% ($F_{ST}$) and 7.4% ($G_{ST}$) when all the 20 goat populations are included in the analysis. These differentiation values are much lower compared with those generally recorded in goats and other domestic animal populations [e.g. 9.9% between Spanish dog breeds (Jordan *et al*., 1992), 10% between European cattle (MacHugh *et al.,* 1998), 17% between Swiss goat breeds (Saitbekova *et al.,* 1999), 8% between Spanish horses (Canon *et al.,* 2000), 10.5% between Chinese indigenous goat populations (Li *et al*., 2002), and as well as 7.7% between Chinese indigenous pig breeds (Yang *et al*., 2003)].

Although the magnitude of gene differentiation among populations was low, the pair-wise genetic differentiation between the majorities (52 out of 55 pair-wise combinations) of Ethiopian goat populations was significant at $P < 0.001$, 0.01 or 0.05 levels (**Table 33** and **Table** 34). There was no genetic differentiation ($P > 0.05$) between three pair-wise combinations of Ethiopian goat populations. These populations (Central Highland/North-West Highland, Hararghe Highland/Short-eared Somali, and Long-eared Somali/Short-eared Somali) are located geographically neighbouring to each other (**Figure 19**), therefore they seem to have experienced frequent gene flow between each other. The Ethiopian goats are a fairly agile species that has developed adaptations, e.g. long legs that enable them to walk long distances. This could have also facilitated this genetic exchange between the goats from different locations within the country. Also the large difference of $F_{ST}$ values observed between the goat populations from Ethiopia compared to those from other countries may well be a consequence of the complexity and diversity of the eco-geographic and sociological barriers, and recent human artificial selection, which may differ in the case of different developed countries.

The highest number of migrants per generation was estimated between Short-eared Somali and Hararghe Highland, where it shows that these two populations are not genetically different at all, followed by *N*m = 120.88 between North-West Highland and Central Highland populations, and followed by *N*m = 82.19 between Short-eared Somali and Long-eared Somali. The lowest number of migrants was estimated between Afar and Abergalle populations (*N*m = 6.06). As expected, gene flow was low between populations located in distant geographic regions. For instance, the *N*m was 7.04 between Gumez and Afar, 7.38 between Keffa and Afar and 9.52 between Long-eared Somali and Abergalle populations. In the case of Afar and Abergalle populations, even though they are not geographically distant from each other, there is a highland topology between the two respective lowland areas, which keeps them isolated.

A clear population structure was inferred in all goat populations considered in this study by using a multivariate analysis (PCA **Figure 26** and **Figure** 27). All 20 goat populations sampled in this study were largely classified into three major clusters. All the Ethiopian plus two Kenyan populations form one cluster; Italian, Mongolian, Saudi Arabian, Egyptian and Botswana form another cluster, and the Guinea Bissau alone form the third cluster. The genetic variation within and among these groups were further quantified by a hierarchical analysis of molecular variance (AMOVA), and the results showed that 89.99% to 91.15% of genetic variation accounted for variation within populations, 1.59% to 2.82% for variation among populations within groups, and 7.20% to 7.32% for variations among groups (**Table 16**). The genetic variation among populations within groups was the lowest when grouped according to their country of origin into nine populations (1.59%) and the highest when all the 20 goat populations were considered together as one group (6.38%). The higher genetic variation among groups was observed between groups according to their country of origin (nine groups 7.32%), and the lower was between group of three (all Ethiopian plus two Kenyan populations, Italian, Mongolian, Saudi Arabian, Egyptian and Botswana populations, and the Guinea Bissau populations 7.20%).

When the AMOVA is conducted only for the Ethiopian goat populations, the results showed that 98.36% of genetic variation accounted for variation within populations and only 1.64% for variation among populations.

### 5.3. Genetic relationship between goat populations

The phylogenetic relationship between the goat populations were determined using Nei's standard genetic distance ($D_S$, Nei, 1972) and Nei *et al.*,'s $D_A$ genetic distance (Nei, 1983) (**Table 10** and **Table** 11). Phylogenetic reconstruction was performed using NJ (Saitou and Nei, 1987) and UPGMA method (Sneath and Sokl, 1973) to summarize the genetic distances between goat populations. The UPGMA method with both Nei's standard genetic distance ($D_S$, **Figure 22**) and Nei *et al.*'s $D_A$ genetic distance (Nei, 1983) (**Figure 23**); and the $D_A$ NJ basically classified the goat populations into three clear phylogenetic clusters concordant with the population structure inferred from the admixture analysis (STRUCTURE, **Figure 29** and **Figure** 30). As expected that different populations from geographically distinct regions, all Ethiopian goat populations (together with the two neighboring Kenyan populations) and the other seven populations (from West Africa, South Africa, North Africa, Turkey, Saudi Arabia, Italy and Mongolia) clearly emerged as distinct lineages different from the West African (Guinea Bissau) standing completely alone (with relatively large genetic distance), suggesting different origins of ancestral populations. These population structures correspond likely to their geographic distributions. A very similar pattern of genetic relationship among the goat populations was observed by principal component analysis (PCA). In general, all the Ethiopian goat populations were more closely related to each other than to non-Ethiopian counterparts. The two exceptions are the Kenyan populations, which are in a cluster with the Ethiopian groups. The two populations were sampled from the northern Kenya close to the boarder between Ethiopia and Kenya, suggesting a common ancestry. All the reference populations (except the two Kenyan populations and the WAD goat population from Guinea Bissau) cluster closely together suggesting a different ancestry. The close clustering of the Ethiopian populations with the two Kenyans ones is in accordance with the work of Chenyambuga (2002). (Nei, Tajima, and Tateno, 1983)

When the Ethiopian goat populations are treated with one out-group (WAD, from Guinea Bissau) or alone for the above analysis, the UPGMA method with Nei's standard genetic distance ($D_S$,) basically classified the goat populations in to eight phylogenetic clusters (**Figure 35** and **Figure** 37), while the $D_A$ genetic distance for both with and without out-group population classified them differently (**Figure 36** and **Figure** 38). Both the NJ trees from $D_S$ and $D_A$ distances for both with and without out-

group population were not consistent with their clustering (**Figure 31** to **Figure 36**). As one origin of evolution was expected for Ethiopian goats, it seemed that the UPGMA performs better and consistent compared to the NJ tree methods. The UPGMA assumes an equal rate of evolution between lineages unlike the neighbor-joining method (Nei *et al.*, 1983).

Comparison of the genetic diversity of the Ethiopian goat populations with that of goats from West Africa, South Africa, Mongolian, Saudi Arabian, Egypt, Italy and Turkey indicated that the Ethiopian goat populations are less genetically diverse (**Table 10** and **Table** 11). This could be the result of being geographically very close to each other, much animal movement across the regions (zones & districts) and in some cases the results of strong human selection pressure for certain economic traits.

The Ethiopian goats, either as separated populations or when grouped as one, were found to have similar diversity with that of the non-Ethiopian populations in terms of MNA per locus and expected heterozygosity (**Table 9**).

There were no significant deviations from HW equilibrium expectations in all the populations studied, also the 15 loci analyzed may be assumed to be unlinked, as linkage disequilibrium did not yield any significant comparison for locus pairs across all populations. The slightly higher than expected ($P < 0.05$) deviation from linkage equilibrium observed in the Arsi-Bale, Guinea Bissau, Long-eared Somali, Italian Alpine Mongolia, Boran-Galla, Short-eared Somali, Afar, Small-East Africa and Woyto-Guji samples could indicate population subdivision, recent introduction of a non-random subset of genotypes or habitat-specific selection (King *et al*., 2001).

The assumptions of the sign test, and the consequences of violating the assumptions, have been discussed by Cornuet and Luikart (1996). It is worth reiterating here that, in test for bottlenecks, loci that are not in Hardy-Weinberg proportions should be used only with caution because they could bias the test results. For example, a locus could deviate from Hardy-Weinberg proportions by having an excess of heterozygote due to strong over-dominance selection. Such a locus might also have selection-induced heterozygosity excess and, therefore, should be used with caution in the sign test for heterozygosity excess. For the data sets studied here, excluding loci that were not in Hardy-Weinberg proportions (**Table 4** and**Table** 5) did not change the results of the sign test. Generally, the Wilcoxon test didn't detect any recent bottleneck effect in the studied Populations. But if there was any undetected bottleneck effect, it should be either not very severe or not very recent.

The present classification of Ethiopian goats is based on physical description and management system. Results of the present study using autosomal microsatellites suggest a close relationship of Ethiopian goat populations. It appears that all Ethiopian goat populations were seen to cluster more closely together, suggesting a common origin and that there is extensive gene flow between the North-West Highland/Central Highland, Short-eared Somali/Hararghe Highland and Long-eared Somali/Short-eared Somali populations. The Tswana from Botswana was seen to cluster with the Italian Alpine, suggesting that crossbreeding between the local Tswana population and the exotic Alpine from Italy has been done or undergoing for improved milk production purpose. The WAD population from Guinea Bissau was always clustered alone, suggesting a different origin from the Ethiopian populations or a differently developed population as trypanotolerant dwarf goats, which agrees with the suggestion of Wilson (1991) that the West African Dwarf (WAD) goats arose in the Fouta Djallon region of Guinea, as did the trypanotolerant N'Dama cattle. The other breeds (Italian Alpine, Baladie, Hair, Mongolia and Ardi) cluster together, suggesting most probably a common ancestry for them.

Given that no true wild goats in the strictest sense occur in Africa, even though Muzzolini (2000) asserts on the presence of goat progenitors in Africa which still needs to be demonstrated, it is generally agreed that the first goat came into Africa via Egypt (Epstein, 1971), this possibly through the Gulf of Suez (Close, 2002). It is also believed that prior to 5th millennium BC goats had reached Egypt (Mason, 1984) and from there they dispersed southwards along the River Nile and westward to what is today the Sahara region. The principal component analysis of our study (**Figure 26** and **Figure** 27) showed that the Ethiopian and Kenyan breeds cluster with the West African Guinea Bissau population, which might be an indication that the Eastern and Western Africa goats crossed the straits of Bab-el-Mandeb and entered the continent of Africa via the horn of Africa. Here, the Baladie population from Egypt is consistently clustered with the European and Asian populations, which agrees with Epstein (1971). Additional archaeological data, including more accurate dating of archaeological goat remains, is required to assess these theories.

The total variation accounted for by the first, second and third principal components, following the principal component analysis of the Ethiopian goat populations, was 58.68%. This figure increased to 60.23% when the reference populations were included in the analysis. This variation (58.68%) was slightly lower

compared to that observed among twelve Chinese indigenous goat populations (60%, Li *et al.*, 2002) and for sub-Saharan Africa camel populations (72%, Mburu *et al*., 2003), but higher compared to that observed for sub-Saharan Africa cattle breeds (54%, Hanotte *et al.*, 2002).

Our microsatellite based molecular genetic study doesn't support the distinction of eleven Ethiopian goat populations as indicated in the study published in 1996 using morphological traits (Farm-Africa, 1996). Instead, the results indicate eight separate genetic entities: the Arsi-Bale, Gumez, Keffa, Woyto-Guji, Abergalle, Afar, Highland goats (previously separated as Central and North West Highland) and Eastern and Southeastern goats (previously separated as Hararghe Highland, Short-eared Somali and Long-eared Somali goats). This shows that how the principal component analysis goes with the quantitative traits like body weight, height etc. but doesn't go with the qualitative traits like horn shape, ear length. This might be the reason that the results of molecular characterization are different from phenotypic characterization.

Weitzman analysis indicated that about 75% of the total genetic diversity of the Ethiopian goats is present in four breeds; Afar, Abergalle, Gumez and Keffa, with marginal loss of diversity of 24.32%, 19.22%, 16.59% and 12.99%, respectively, while removal of other populations had only a lower impact (Arsi-Bale; -6.23%; Woyto-Guji; -6.08%; Central Highland and North-West Highland combined; -5.33%; Hararghe Highland and Short-eared Somali combined; -4.73%; Long-eared Somali; -4.13%).

# CHAPTER SIX

## 6.0 SUMMARY AND CONCLUSIONS

### 6.1 Summary

Ethiopia has the largest livestock population among all the countries in Africa. Ethiopia with its great variation in agro-ecological zones represents a potential reservoir of sheep and goat diversity. They inhabit a wider range of environments extending from a tropical to cool temperate climates. According to FAO (1994), there are approximately 21.7 and 16.7 million sheep and goats, respectively. Sheep and goats are distributed in all agro-ecological zones of the country although the majority of the sheep population is concentrated in the highlands. The majority of the goat population is found in large flocks in the arid and semi-arid lowlands where pastoralists in the south, east, and west keep them for milk production, for slaughter and for sale. Goats in the highlands are widely distributed in the mixed crop-livestock production systems with small flock size as a means of cash earnings and meat. Although there are severe environmental constraints to increase goat productivity, there is considerable potential for improved goat production in the country, where goat milk, meat, and skin are valued commodities.

Ethiopia has been long back recognized as a source of genetic diversity in plants and animals. It appears that Ethiopia has served as a gateway of domestic animal genetic materials from Asia to Africa, and its diverse ecology leads to further diversification and development of different genotypes.

The identification, description, classification and naming of livestock types are the first step in the assessment of a country's livestock resources.

Work done in 1990s on physical description and management system has shown the presence of 14 goat types in Ethiopia and Eritrea (Workneh, 1992; Alemayehu, 1993; Nigatu, 1994), out of which 11 are in Ethiopia. This survey was the very first step in breed characterization of Ethiopian goats through differentiating a heterogeneous population into relatively homogeneous sub-populations, which may then be named as breeds, types or subtypes. However, it remains unclear how different at the genetic level are these 11 goat types. The availability of new molecular techniques for large-scale genetic characterization is providing unique opportunities to tackle the above

issues. These techniques are able to reveal genetic distances between populations allowing a better understanding of how these populations evolved.

In recent years, a range of innovations in molecular genetics has been developed for the study of genetic variation and evolution of populations using DNA genotyping information. The most utilized DNA marker for population genetics of livestock is microsatellite. Microsatellite markers, also called short tandem repeats (STRs) or simple sequence repeats (SSRs), are a relatively new class of genetic marker. Over a few years they have become a tool of choice to address population genetics and demographic questions (Sunnucks, 2000). They allow the study of genetic diversity and differentiation of closely related populations. Microsatellite, as genetic markers, have been applied successfully in the study of genetic variation of livestock including between and among European and African cattle breeds (Machugh *et al*., 1997; Okoma *et al*., 1998), between Swiss goat breeds (Saitbekova *et al.*, 1999), between Indian goat breeds (Ganai and Yadav, 2001), between Chinese indigenous goat populations (Li *et al.*, 2002, Li and Valentini 2004) and between Indian buffalo populations (Arora R *et al.*, 2004,). Unfortunately, for goats, throughout the world very limited work has been done to determine genetic variation among breeds/populations and in fact, information on microsatellite polymorphism of Ethiopian goats is completely lacking.

Accordingly the overall goal of this project is to obtain detailed information on the genetic diversity and differentiation of the indigenous goat populations of Ethiopia with the following specific objectives:

1. To quantify genetic diversity within and among Ethiopian goat populations.
2. To clarify the evolutionary genetic relationships among the Ethiopian goat populations.
3. To clarify the origin and relationships of Ethiopian goat populations through the study of non-Ethiopian breeds of references.
4. To compare the morphological classification of the Ethiopian goat populations with the one based on molecular genetic characterization.

For this purpose, 11 indigenous goat populations/types of Ethiopia and nine reference breeds (four from other African countries viz. two from Kenya, one from Guinea Bissau and one from Botswana), and five from non-African countries (one breed from Turkey, Egypt, Italy, Mongolia and Saudi Arabia each) were included in this study.

The sampled animals of Ethiopian goats were selected to represent the type/population as much as possible based on the available information of phenotypic characteristics and geographical distribution from previous survey by Farm-Africa (FARM – Africa, 1996). The finally identified animals for the study were sampled from distantly located villages and sampling was made 1-5 animals per flock. For all populations/types/breeds, approximately equal number of females and males were sampled.

DNA was extracted either from peripheral blood lymphocytes (PBL) or FTA® blood cards (Whatman® Bioscience). Blood was collected from jugular vein by puncher using 10 ml capacity vacutainer tubes containing EDTA (100 µl neat). Every time a new needle was used for each animal to avoid cross contamination between samples. The blood was mixed with EDTA by gently inverting the tube about three to five times. In the laboratory blood samples were centrifuged at 2000 rpm for 15 – 20 minutes the same day evening. The plasma was discarded and the buffy-coat containing the peripheral blood lymphocytes (PBL) was transferred with a wide-mouth Pasteur pipette into a 2 ml cryotube. Then 1 ml of 8 M urea was added into the tube, the content of the tube mixed by gentle inversions five times. The tube properly labeled and stored in upright position at room temperature, ready for DNA extraction.

In the case of sampling using the FTA cards, the blood in the tube was placed on ice and transferred onto FTA classic cards (Whatman® Bioscience) on the same day. After drying, these cards were transferred into bags containing silica gel for storage and future use.

The 15 microsatellite loci (BM1818, BMS1494, ILSTS05, ILSTS11, ILSTS44, ILSTS87, INRA05, INRA63, NRA132, MAF35, MAF65, MAF 209, OarAE129, OarFCB304, and SRCRSP03) were analyzed at ILRI, Nairobi, Kenya. All PCR amplifications were performed in a total volume of 10 µl (except for some FTA® blood disks, where a total volume of 20 µl was used) on GeneAMP®PCR system 9700 (Applied Biosystem) thermocycler. All amplification conditions included an initial denaturing step of 5 min at 95°C, followed by 35 cycles of 30 seconds at 95°C, 1 minute at an appropriate primer annealing temperature and 1 minute at 72°C. Duration of final extensions was 10 minutes at 72°C. The PCR products were analyzed by electrophoresis on a 4.25% denaturing polyacrylamide gel using a 377 ABI automatic DNA sequencer (ABI PRISM 377, Applied Biosystems). The GeneScan™ 350-

TAMRA served as an internal size standard in each lane. A control sample was included in addition to the GeneScan™ 350-TAMRA for further standardization of fragment sizes. Fragment lengths were identified using the GeneScan® software (version 3.1.2 Applied Biosystems), while allele sizes identifications were performed by GENOTYPER® software (version 2.0 Applied Biosystems) using the third order least squares method for size calling.

A number of analyses were performed using allelic variation obtained at 15 microsatellite loci. Allele frequencies at each locus were calculated for each population. Microsatellite variability such as the observed and expected number of alleles at each locus for each population, mean number of alleles per locus, observed heterozygosity and expected heterozygosity were calculated to quantify the genetic variation within the population. Statistics such as fixation indices or *F*-statistics were performed to describe the population structure and genetic differentiation among populations. Phylogenetic analysis and multivariate analysis such as principal component analysis were performed to infer the evolutionary relationship among populations.

Results indicated that 39 out of 300 i.e.13% of the comparisons ($P < 0.05$;) deviated significantly from HWE, a figure higher than the 5% that would be expected from type I error alone. Application of sequential Bonferroni adjustment for multiple testing reduced this number of significant locus-population combinations to only three. Out of the 15 loci, there was no locus that deviated from HWE for all the populations and also no population was observed to deviate from HWE for all the loci.

The linkage disequilibrium tests within each of the 20 populations were performed for the 105 pair-wise combinations of the 15 loci. Assuming a 5% type I error rate ($P < 0.05$), about 5 to 6 pairs of loci are expected to be in linkage disequilibrium. On average 6% of the locus pairs were in linkage disequilibrium within the populations.

The global exact test for linkage disequilibrium across all 20 populations (105 pair-wise comparisons) yielded significant deviations in nine comparisons at $P < 0.05$ . Only one out of these nine tests remained significant following application of sequential Bonferroni corrections.

The recent bottleneck effect was inferred for each population using a Wilcoxon's signed rank test (Cornuet and Luikart, 1996). Generally, there was no recent bottleneck effect detected in any population for both Wilcoxon's signed rank test and quantitative graphical method. Interestingly, ten populations out of the total 20

populations showed a significant deficiency of heterozygosity. This suggests that these populations are not at mutation-drift equilibrium but instead have experienced a recent expansion in population size or perhaps a recent entry of rare alleles from genetically distinct immigrants.

A total of 159 alleles were detected at 15 microsatellite loci in 776 goat individuals of 20 populations sampled from 20 locations in Ethiopia, Kenya, Botswana, Guinea Bissau, Egypt, Turkey, Italy, Saudi Arabia and Mongolia.

The expected MNA across all populations was 4.60 ± 0.08 and 7.42 ± 0.16 for SMM and IAM, respectively, ranging between 4.05 ± 1.59 in Guinea Bissau and 5.77 ± 2.70 in Hair for SMM, and between 6.39 ± 2.34 in Ardi and 9.29 ± 4.85 in Hair populations for IAM. The results indicate that the stepwise mutation model (SMM) is more appropriate for this particular dataset and should be preferred in the model-based algorithms such as the inference of the recent population bottleneck.

The total number of alleles sampled in a population across all loci ranged between 80 in West African Dwarf (WAD) goat population from Guinea Bissau and 116 in Long-eared Somali goat population from Ethiopia. The mean number of alleles per locus (MNA) across all populations was 5.46 ± 0.09, varying between 4.67 ± 2.74 in Guinea Bissau WAD goat population and 6.47± 4.03 in Long-eared Somali goat population. The MNA for a uniformed sample size with 20 animals after random re-sampling procedure with 250 replacements over individuals was 4.88 ± 0.08 varying between 4.19 ± 2.26 in Guinea Bissau population and 5.74 ± 2.89 in Hair goat population.

The mean observed heterozygosity ($H_O$) and expected heterozygosity ($H_E$) for Ethiopian goats was 0.551 ± 0.015 and 0.576 ± 0.015, respectively, across all loci and all population. The observed heterozygosity across all loci ranged between 0.515 ± 0.047 in Ethiopian Afar population to 0.625 ± 0.050 in Saudi Arabian Ardi population. The highest expected heterozygosity (gene diversity) across all loci was observed in Hair goat population from Turkey (0.642 ± 0.055) and the lowest was 0.523 ± 0.058 in WAD (Guinea Bissau) population. The differences of both values among the goat populations were not statistically significant ($P > 0.05$), and it also seemed that the values were not affected by sample size.

The level of genetic diversity ($H_E$ = 0.58 ± 0.015) obtained for the Ethiopian goat populations in the present study was similar to that previously reported in Swiss goat breeds ($H_E$ = 0.59, Saitbekova *et al.,* 1999), and Asian goats ($H_E$ = 0.52, Barker *et*

*al*., 2001), but less than that reported for Chinese goats ($H_E$ = 0.716, Li *et al*., 2002; and $H_E$ = 0.707, Li and Alessio, 2004).

Seven out of 11 populations showed a significant inbreeding coefficient ($F_{IS}$) ($P < 0.05$). This inbreeding could be a consequence of the management systems, traditional/cultural practices and also ethnic/cultural barriers that could affect the between flock breeding-males (backs) exchange. In case each sampled populations/types have been subdivided into subpopulations, which contains fewer heterozygotes than predicted despite the fact that all subdivisions are in Hardy-Weinberg equilibrium, can also lead to inbreeding (Wahlund effect). All these systems and events tend to increase the rate of inbreeding.

$F_{IS}$ statistic, an estimator for describing local inbreeding effect within subpopulations, was significantly different from zero at six out of 15 loci and across all loci at $P < 0.05$ level, indicating that the Ethiopian goat populations sampled for this study were generally at risk of inbreeding. $F_{ST}$ estimator was significantly different from zero at 14 out of 15 loci at $P < 0.01$, which means that genotype frequencies among populations are not randomly distributed, meaning that all the Ethiopian goat populations have been genetically subdivided even though the mean $F_{ST}$ value across all loci was very low (0.016) compared to previously reported in Asian goats ($F_{ST}$ = 0.143, Barker *et al.*, 2001) and Chinese goats ($F_{ST}$ = 0.105, Li *et al.*, 2002 and $F_{ST}$ = 0.054, Li and Valentini, 2004). Similar value was obtained for the coefficient of gene differentiation using the method of Nei (1973) ($G_{ST}$ = 0.026). Only 1.6% to 2.6% of total variation was contributed by genetic subdivision among populations while 97.4% to 98.4% of total variation accounts for genetic variations among individuals within populations. This figure went up to 6.4% ($F_{ST}$) and 7.4% ($G_{ST}$) when all the 20 goat populations are included in the analysis. These differentiation values are much lower compared with those generally recorded in goats and other domestic animal populations [e.g. 9.9% between Spanish dog breeds (Jordan *et al*., 1992), 10% between European cattle (MacHugh *et al.,* 1998), 17% between Swiss goat breeds (Saitbekova *et al.*, 1999), 8% between Spanish horses (Canon *et al.*, 2000), 10.5% between Chinese indigenous goat populations (Li *et al*., 2002), and as well as 7.7% between Chinese indigenous pig breeds (Yang *et al*., 2003)].

The highest number of migrants per generation was estimated between Short-eared Somali and Hararghe Highland, where it shows that these two populations are not genetically different at all, followed by ($N$m = 120.88) between North-West Highland

and Central Highland populations, and followed by ($Nm$ = 82.19) between Short-eared Somali and Long-eared Somali.

A clear population structure was inferred in all goat populations considered in this study by using a multivariate analysis. All 20 goat populations sampled in this study were largely classified into three major clusters. All the Ethiopian plus two Kenyan populations form one cluster; Italian, Mongolian, Saudi Arabian, Egyptian and Botswana form another cluster, and the Guinea Bissau alone form the third cluster.

The UPGMA method with both Nei's standard genetic distance and Nei *et al.*'s $D_A$ genetic distance (Nei, 1983); and the $D_A$ NJ basically classified the goat populations into three clear phylogenetic clusters concordant with the population structure inferred from the Admixture analysis.

As expected for different populations from geographically distinct regions, all Ethiopian goat populations (together with the two neighboring Kenyan populations) and the other 7 populations (from West Africa, South Africa, North Africa, Turkey, Saudi Arabia, Italy and Mongolia) clearly emerged as distinct lineages different from the West African (Guinea Bissau) standing completely alone as an out-group (with relatively large genetic distance), suggesting different origins of ancestral populations. In general, all the Ethiopian goat populations were more closely related to each other than to non-Ethiopian counterparts. All the reference populations (except the two Kenyan populations and the WAD goat population from Guinea Bissau) cluster closely together suggesting a different ancestry. The close clustering of the Ethiopian populations with the two Kenyans ones is in accordance with the work of Chenyambuga (2002). (Nei, Tajima, and Tateno, 1983)

When the Ethiopian goat populations are treated with one out-group (WAD, from Guinea Bissau) or alone for the above analysis, the UPGMA method with Nei's standard genetic distance basically classified the goat populations in to eight phylogenetic clusters, while the $D_A$ genetic distance for both with and without out-group population classified them differently. As one origin of evolution was expected for Ethiopian goats, it seemed that the UPGMA performs better and consistent compared to the NJ trees methods.

Comparison of the genetic diversity of the Ethiopian goat populations with that of goats from West Africa, South Africa, Mongolian, Saudi Arabian, Egypt, Italy and Turkey indicated that the Ethiopian goat populations are less genetically diverse. This could be the result of being geographically very close to each other, much animal

movement across the regions (zones & districts) and in some cases the results of strong human selection pressure for certain economic traits.

The Ethiopian goats, either as separated populations or when grouped as one, were found to have similar diversity with that of the non-Ethiopian populations in terms of MNA per locus and expected heterozygosity.

The slightly higher than expected ($P < 0.05$) deviation from linkage equilibrium observed in some of the samples could indicate population subdivision, recent introduction of a non-random subset of genotypes or habitat-specific selection (King *et al*., 2001

Results of the present study using autosomal microsatellites suggest a close relationship of Ethiopian goat populations. It appears that all Ethiopian goat populations were seen to cluster more closely together, suggesting a common origin and that there is extensive gene flow especially between the North-West Highland/Central Highland, Short-eared Somali/Hararghe Highland and Long-eared Somali/Short-eared Somali populations.

The WAD population from Guinea Bissau was always clustered alone as an out-group, suggesting a different origin from the Ethiopian populations or a differently developed population as trypanotolerant dwarf goats, which agrees with the suggestion of Wilson (1991) that the West African Dwarf (WAD) goats arose in the Fouta Djallon region of Guinea, as did trypanotolerant N'Dama cattle. The other breeds (Italian Alpine, Baladie, Hair, Mongolia and Ardi) cluster together, suggesting most probably common ancestry for them.

The principal component analysis of this study showed that the Ethiopian and Kenyan breeds cluster with the West African Guinea Bissau population, which might be an indication that the Eastern and Western Africa goats crossed the straits of Bab-el-Mandeb and entered the continent of Africa via the horn of Africa. Here, the Baladie population from Egypt is consistently clustered with the European and Asian populations, which agrees with Epstein (1971). Additional archaeological data, including more accurate dating of archaeological goat remains, is required to assess these theories.

The total variation accounted for by the first, second and third principal components, following the principal component analysis of the Ethiopian goat populations, was 58.68%. This figure increased to 60.23% when the reference populations were included in the analysis. This variation (58.68%) was slightly lower

compared to that observed among twelve Chinese indigenous goat populations (60%, Li *et al.*, 2002) and for sub-Saharan Africa camel populations (72%, Mburu *et al*., 2003), but higher compared to that observed for sub-Saharan Africa cattle breeds (54%, Hanotte *et al.*, 2002).

Weitzman analysis indicated that about 75% of the total genetic diversity of the Ethiopian goats is present in four breeds; Afar, Abergalle, Gumez and Keffa, with marginal loss of diversity of 24.32%, 19.22%, 16.59% and 12.99% respectively, while removal of other populations had only a lower impact.

Results presented in this study show that phenotypic characteristics subject to environmental influences could lead to a heterogeneity that is not reflected at the genetic level. Accordingly our microsatellite based molecular genetic study does not support the distinction of eleven Ethiopian goat populations as indicated in the study published in 1996 using morphological traits (Farm-Africa, 1996). Instead, following the analysis of 15 microsatellite loci, the result indicate eight separate genetic entities, the Arsi-Bale, Gumez, Keffa, Woyto-Guji, Abergalle, Afar, Highland Goats (previously separated as Central and North West Highland) and the goats from the previously known Hararghe, Southeastern Bale and Southern Sidamo provinces (Hararghe Highland, Short-eared Somali and Long-eared Somali goats). It appears likely that all the Ethiopian goat populations have a common origin and that there is extensive gene flow between them.

The present results show that the variation among Ethiopian goat populations is reasonable and therefore, within breed selection should be conducted to improve these eight breeds and lines with different characters should be established to conserve the genetic diversity in the future.

The results from this study have also shown that more than 60% of the goat populations showed significant inbreeding effect, which leads to loss of genetic diversity. Therefore, measures like education (how to exchange the breeding males between flocks or introduction of new breeding-males and removing of old ones within a certain period of time) to the farming communities, pastoralists or goat owners are highly important.

The results from this study have shown that the goat genetic resource of Ethiopia represents a substantial genetic resource and as Weitzman method provides a well-defined diversity metric, allowing quantitative assessment of the consequences of alternative conservation activities as an aid to decision-making in conservation, here it

is also recommended to put more emphasis on the four (Afar, Abergalle, Gumez and Keffa) breeds in which more than 70% of the total genetic diversity of the Ethiopian goats is present; to maximizing diversity conserved and guide breeding improvement activities for meat and milk production.

Other beneficial uses of present day populations will emerge as genome maps are completed and scientists learn to identify and manipulate genes involved in mechanisms of growth, reproduction and disease resistance tolerance. Indigenous animal genetic resources that have evolved for 9,000-12,000 years are an integral component of our heritage and should be nurtured and preserved for future generations.

### 6.2 Conclusions

Phenotypic characterization of individual animals of particular breeds has been a routine process. However, the results presented in this study show that phenotypic characteristics subject to environmental influences could lead to a heterogeneity that is not reflected at the genetic level. Accordingly our microsatellite based molecular genetic study does not support the distinction of eleven Ethiopian goat populations. Instead, following the analysis of 15 microsatellite loci, the result indicate eight separate genetic entities: the Arsi-Bale, Gumez, Keffa, Woyto-Guji, Abergalle, Afar, Highland Goats (previously separated as Central and North West Highland) and the goats from the previously known Hararghe, Southeastern Bale and Southern Sidamo provinces (Hararghe Highland, Short-eared Somali and Long-eared Somali goats). This study revealed that there is no genetic differentiation between Hararghe Highland and Short-eared Somali, and insignificant differentiation between Central Highland and Northwest Highland; as well as between Long-eared Somali and Short-eared Somali populations. It appears likely that all the Ethiopian goat populations have a common origin and that there is extensive gene flow between the North-West Highland/Central Highland, Short-eared Somali/Hararghe Highland and Long-eared Somali/Short-eared Somali populations.

Furthermore, it can also be concluded from this study that microsatellite marker is a powerful genetic tool in estimation of genetic variation within and among closely related populations. The present results show that the variation among Ethiopian goat populations is reasonable and therefore within breed selection should be conducted to

improve these eight breeds and lines with different characters should be established to conserve the genetic diversity in the future

### 6.3. Recommendation

The results from this study have shown that more than 60% of the goat populations showed significant inbreeding effect, which leads to loss of genetic diversity. Therefore, measures like education (how to exchange the breeding males between flocks or introduction of new breeding-males and removing of old ones within a certain period of time) to the farming communities, pastoralists or goat owners are highly vital. The results from this study have shown that the goat genetic resources of Ethiopia represent a substantial genetic resource and as Weitzman method provides a well-defined diversity metric, allowing quantitative assessment of the consequences of alternative conservation activities as an aid to decision-making in conservation, here it is also recommended to put more emphasis on the four (Afar, Abergalle, Gumez and Keffa) breeds in which more than 70% of the total genetic diversity of the Ethiopian goats is present; to maximizing diversity conserved and guide breeding improvement activities for meat and milk production.

Other beneficial uses of present day populations will emerge as genome maps are completed and scientists learn to identify and manipulate genes involved in mechanisms of growth, reproduction and disease resistance tolerance. Indigenous animal genetic resources that have evolved for 9,000-12,000 years are an integral component of our heritage and should be nurtured and preserved for future generations.

# REFERENCES


Aboagye, G. (1992). Characterization and evaluation Sanga and West African shorthorn cattle on the Accra plains. In: Africa Animal Genetic Resources: Their characterization, conservation and utilization. Proceedings of the research planning workshop held at ILCA, Addis Ababa, Ethiopia 19-21 February, 1992. ILCA (International Livestock Center for Arfica) Addis Ababa, Ethiopia. pp 73-75.

Aboagye, G. S., Tawah, C. L., and Rege, J. E. O. (1994). Shorthorn cattle of West and Central Africa. III. Physical, adaptive and special genetic characteristics. World Animal Review 78 (1): 22-32.

Akey, J. M., Zhang, K., Xiong, M., and Jin, L. (2003). The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium. *Mol Biol Evol* **20**(2)**,** 232-42.

Alados, C. L. (1985b). Distribution and status of the Spanish ibex (Capra pyrenaica Schinz). In Lovari (1985), pp. 204-11.

Alemayehu Reda (1993). Characterization (Phenotypic) of indigenous goats and goat husbandry practices in east and south-eastern Ethiopia. M.Sc thesis, Alemaya University of Agriculture, Ethiopia. 135pp.

Alexandrino, F. W., Knudson, K. L., and Leary, R. F. (1983). Adaptive significance of the differences in the tissue-specific expression of phosphoglucomutase gene in rainbow trout. In: Proceedings of theNational Academy of Science of the USA 800: 1397-1400.

Allen, P. J., Amos, W., Pomeroy, P. P., and Twiss, S. D. (1995). Microsatellite variation in grey seals (Halichoerus grypus) shows evidence of genetic differentiation between two British breeding colonies. *Mol Ecol* **4**(6)**,** 653-62.

Allendorf, F. W., and Leary, R. F. (1986). Heterozygosity and fitness in natural populations of animals. In: Conservation Biology: The science of scarcity and diversity. (Edited by M.E. Soule) Sinauer, Sunderland, Massachusetts. pp 57-76.

Armour, J. A. L., and Jeffreys, A. J. (1992). Biology and applications of human minisatellite loci. Current Opinion in Genetics and Development 2: 850-856.

Arora, R., Lakhchaura, B. D., Prasad, R. B., M.S, T., and Vijh, R. K. (2004). Genetic diversity analysis of two buffalo populations of northern India using microsatellite markers. *Anim. Breed. Genet.* **121,** 111-118.

Arranz, J. J., Bayon, Y., and San Primitivo, F. (1998). Genetic relationships among Spanish sheep using microsatellites. *Anim Genet* **29**(6)**,** 435-40.

Avis, J. C. (1994). Molecular markers, Natural history and evolution. Chapman & Hall. New York, London. pp 511.

Ayalew, W. (1992). Preliminary survey of indigenous goat types and goat husbandery practices in Southern Ethiopia. MSc thesis, Alemaya University of Agriculture, Alemaya, Ethiopia. 153pp.

Baker, A. J., and Moeed, A. (1987). Rapid genetic differentiation and founder effect in colonizing populations of common Mynas (Acridotheres tristis). Evolution 41:525-538.

Balloux, F., and Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Mol Ecol* **11**(2)**,** 155-65.

Banks, M. A., and Eichert, W. (2000). WHICHRUN (version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. *J Hered* **91**(1)**,** 87-9.

Barbancho, M., Llanes, D., Morera, L., Garzon, R., and Rodero, A. (1984). Genetic markers in the blood of Spanish goat breeds. *Anim Blood Groups Biochem Genet* **15**(3)**,** 207-12.

Barendse, W., Armitage, S. M., Kossarek, L. M., Shalom, A., Kirkpatrick, B. W., Ryan, A. M., Clayton, D., Li, L., Neibergs, H. L., Zhang, N., and et al. (1994). A genetic linkage map of the bovine genome. *Nat Genet* **6**(3)**,** 227-35.

Barendse, W., Vaiman, D., Kemp, S. J., Sugimoto, Y., Armitage, S. M., Williams, J. L., Sun, H. S., Eggen, A., Agaba, M., Aleyasin, S. A., Band, M., Bishop, M. D., Buitkamp, J., Byrne, K., Collins, F., Cooper, L., Coppettiers, W., Denys, B., Drinkwater, R. D., Easterday, K., Elduque, C., Ennis, S., Erhardt, G., Li, L., and et al. (1997). A medium-density genetic linkage map of the bovine genome. *Mamm Genome* **8**(1)**,** 21-8.

Barker, J. S. F. (1994). A global protocol for determining genetic distances among domestic livestock breeds. In: Proceedings of the 5th world congress of genetic applied to livestock production. 7-12 August, 1994 Guelph, Canada. Vol 21: 501-508."

Barker, J. S. F. (2002). Relevance of animal genetic resources and differences to the plant sector. In: Groeneveld E. and Glodek P. (Eds), Animal Breeding and Animal Genetic Resources. Federal Agricultural Research Centre (FAL), Mariensee and Institute of Animal Genetics, Gottingen, Braunschweig, Germany, 15-21.

Barker, J. S. F., Tan, S. G., Moore, S. S., Mukherjee, T. K., Matheson, J.-L., and Selvaraj, O. S. (2001). Genetic variation within and relationships among populations of Asian goats (Capra hircus). *Journal of Animal Breeding and Genetics* **118**(4)**,** 213-233.

Beja-Pereira, A., Alexandrino, P., Bessa, I., Carretero, Y., Dunner, S., Ferrand, N., Jordana, J., Laloe, D., Moazami-Goudarzi, K., Sanchez, A., and Canon, J. (2003). Genetic characterization of southwestern European bovine breeds: a historical and biogeographical reassessment with a set of 16 microsatellites. *J Hered* **94**(3)**,** 243-50.

Bermond, J., editor. (2001). Molekulargenetische Differenzierung verschiedener Rotviehpopulationen. Bundesministerium fur Verbraucherschutz, Ernahrung und Landwirtschaft, Munster-Hiltrup, Germany.

Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sunden, S. L., Hawkins, G. A., Toldo, S. S., Fries, R., Grosz, M. D., Yoo, J., and et al. (1994). A genetic linkage map for cattle. *Genetics* **136**(2)**,** 619-39.

Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Q Rev Biol* **74**(1)**,** 21-45.

Bohonak, A. J. (2002). IBD (Isolation by Distance): a program for analyses of isolation by distance. *J Hered* **93**(2)**,** 153-4.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* **32**(3)**,** 314-31.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., and Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**(6470)**,** 455-7.

Bradley, D. G., MacHugh, D. E., Cunningham, P., and Loftus, R. T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A* **93**(10)**,** 5131-5.

Brookfield, J. F. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* **5**(3)**,** 453-5.

Bruford, M. W., and Wayne, R. K. (1993). Microsatellites and their application to population genetic studies. *Curr Opin Genet Dev* **3**(6)**,** 939-43.

Buchanan, F. C., Adams, L. J., Littlejohn, R. P., Maddox, J. F., and Crawford, A. M. (1994). Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics* **22**(2)**,** 397-403.

Burke, T. (1989). DNA fingerprinting and other methods for the study of mating success. *Trends Ecol. Evol.*, **4:,** 139-144.

Canon, J., Alexandrino, P., Bessa, I., Carleos, C., Carretero, Y., Dunner, S., Ferran, N., Garcia, D., Jordana, J., Laloe, D., Pereira, A., Sanchez, A., and Moazami-Goudarzi, K. (2001). Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genet Sel Evol* **33**(3)**,** 311-32.

Canon, J., Checa, M. L., Carleos, C., Vega-Pla, J. L., Vallejo, M., and Dunner, S. (2000). The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Anim Genet* **31**(1)**,** 39-48.

Carson, H. L. (1983). The genetics of the founder effect. In: Genetics and Conservation: A reference for managing wild animal and plant populations. (Edited by C.M. Schonewald-Cox; S.M. Chambers; B.MacBryde and L.Thomas). Benjamin/Cummings, menlopark, CA. pp 189-200.

Cavalli Sforza, L. L., Menozzi, P., and A, P. (1994). "The history and geography of human genes," Princeton University Press, Princeton, New Jersey.

Cavalli-Sforza, L. L., and Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* **19**(3)**,** Suppl 19:233+.

Cavalli-Sforza, L. L., and Edwards, W. L. (1964). Analysis of human evolution. *Proceedings 11th International Congress on Genetics***,** 923-933.

Chakraborty, R., and Zhong, Y. (1994). Statistical power of an exact test of Hardy-Weinberg proportions of genotypic data at a multiallelic locus. *Hum Hered* **44**(1)**,** 1-9.

Charlesworth, D., Lyons E.E., Litchfield L.B., (1994). Inbreeding depression in two highly inbreeding populations of *Leaven-worthia. Proc R Soc Lond B Biol Sci* **258**: 209-214.

Charlesworth, D., and Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. Annual Review of Ecology and Systematics 18: 237-268.

Chenyambuga, S. (2002). Genetic characterisation of indigenous goat populations of sub-Saharan Africa using microsatelite DNA markers. PhD thesis, Department of Animal Science and Production, Sokoine University of Agriculture, Sokoine, Tanzania.

Chung, E. R., Kim, W. T., Kim, Y. S., and Kin, D. K. (1998). *Korean J. Dairy Sci.,* **21:,** 1-8.

Ciampolini, R., Moazami-Goudarzi, K., Vaiman, D., Dillmann, C., Mazzanti, E., Foulley, J. L., Leveziel, H., and Cianci, D. (1995). Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. *J Anim Sci* **73**(11)**,** 3259-68.

Close, A. E. (2002). Sinai, Sahara, Sahel: The Introduction of Domestic Caprines to Africa, In:Jennerstasse 8(Eds.) Tides of the desert, Africa Praehistorica 14, Heinrich Barth Institut, Koln, pp. 459-469.

Cluctton-Brock, J. (1981). Domestic Animals from Early Times (Heinemann and Br. Mus. of Nat. Hist., London).

Corbet, G. B. (1978). The mammals of the Palaearctic Region: a taxonomic review. British Museum. (National History), London, 314 pp.

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**(4)**,** 2001-14.

Cornuet, J. M., Piry, S., Luikart, G., Estoup, A., and Solignac, M. (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153(4), 1989-2000.

Crawford, A. M., Dodds, K. G., Ede, A. J., Pierson, C. A., Montgomery, G. W., Garmonsway, H. G., Beattie, A. E., Davies, K., Maddox, J. F., Kappes, S. W., and et al. (1995). An autosomal genetic linkage map of the sheep genome. *Genetics* **140**(2)**,** 703-24.

Crawford, A. M., Godds, K. G., and Pierson (1995). An autosomal genetic linkage of the sheep genome. *Genetics* **140,** 703-724.

Crozier R. H. (1992). Genetic diversity and the agony of choice. *Biological Conservation* **61,** 11-15.

Cunningham, E. P., Loftus, R. T., MacHugh, D. E., and Bradley, D. G. (1994). Molecular evolution of Africa, European and Asian cattle. In: Proceedings of the 5th world congress of genetic applied to livestock production. 7-12 August, 1994 Guelph, Canada. Vol 21: 86-89."".

Delgado, C., Rosegrant, M., Steinfeld, H., Ehui, S., and Courbois, C. (1999). "Livestock to 2020 – The Next Food Revolution". Food, Agriculture and the Environment Discussion Paper 28. IFPRI/FAO/ILRI.

Devendra, C., and Burns, M. (1983). Goat production in the tropics. Commonwealth Agricultural Beauro. Farnham house, Farnham royal, Slough SL2 3BN UK. pp**,** 183.

Dieringer, D., and Schlötterer, C. (2003). microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3 (1), 167-169**.

Donnelly, M. J., and Townson, H. (2000). Evidence for extensive genetic differentiation among populations of the malaria vector Anopheles arabiensis in Eastern Africa. *Insect Mol Biol* **9**(4)**,** 357-67.

Dowling, T. E., Moritz, C., Palmer, J. D., and Rieseberg, L. H. (1996). Nucleic Acids III: Analysis of fragments and restriction sites. In: Molecular systematics 2nd edition (Edited by D.M. Hills; C. moritz, and B.K. Mable). Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts, USA. pp 249-320.

Edwards, A., Civitello, A., Hammond, H. A., and Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* **49**(4)**,** 746-56.

Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T., and Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**(2)**,** 241-53.

Ellegren, H. (1993). Genome analysis with microsatellite markers. PhD dissertation. Swidish University of Agricultural Science.

Ellerman, J. R., and Morrison-Scott., T. C. S. (1966). Checklist of Palaearctic and Indian mammals. British Mus. (Nat. Hist.), London, 810 pp.

Epstein, H. (1971). The origion of the domestic animals of Africa. Vol. II. African publishing corporation. New York, London, Munich. pp 719.

Estoup, A., Garnery, L., Solignac, M., and Cornuet, J. M. (1995). Microsatellite variation in honey bee (Apis mellifera L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* **140**(2)**,** 679-95.

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**(2)**,** 479-91.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61,** 1-10.

Falconer, D. S., and Mackey, T. F. C. (1996). ""Introduction to quantitative Genetics," Longman Group Ltd, Harlow."

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**(4)**,** 1567-87.

Fan, B., Wang, Z. G., Li, Y. J., Zhao, X. L., Liu, B., Zhao, S. H., Yu, M., Li, M. H., Chen, S. L., Xiong, T. A., and Li, K. (2002). Genetic variation analysis within and among Chinese indigenous swine populations using microsatellite markers. *Anim Genet* **33**(6)**,** 422-7.

FAO (1994). Production Yearbook. Food and Agriculture Organization. Rome, Italy.

FAO (1995). World watch list for domestic animal diversity. 2nd edition. Food and Agricultural organization of the united nations, Rome, Italy. pp.769.

FAO (1999). The global strategy of the management of Farm Animal Genetic Resources. Executive Brief., FAO, Rome.

FAO (2000). Conserving and developing farm animal diversity. In State of the world animal genetic resources. *Http://www.fao.org/News/2000/001201-e.htm*.

Farm-Africa (1996). Goat types of Ethiopia and Eritrea. Physical discription and management systems. Published jointly by FARM-Africa, London, UK, and ILRI (International Livestock Research Institute), Nairobi, Kenya. 76pp.

Farris, J. S. (1970). Methods for computing Wagner trees. *Systematic Zoology* **34,** 21-34.

Farris, J. S. (1971). The hypothesis of non-specificity and taxonomic congruence. *Annual Review of Ecology and Systematics* **2,** 227-302.

Farris, J. S. (1981). Distance data in phylogenetic analysis. In: Advances in Cladistics. Proceedings of thr first meeting of the willi Henning Society (V.A. Funk and D.R. Brooks edition). New York Botanical garden, New York. pp 3-23.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**(6)**,** 368-76.

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

Fitch, W. M. (1971). Toward defining the course of evolution. Minimum change for a specific tree topology. *Systematic Zoology* **20,** 406-416.

Forbes, S. H., Hogg, J. T., Buchanan, F. C., Crawford, A. M., and Allendorf, F. W. (1995). Microsatellite evbolution in congeneric mammals: domestic and bighorn sheep. *Mol. Biol. & Evol.,* **12:,** 1106-1113.

Frankham, R. (1994). Conservation of genetic diversity for animal improvement. In: Proceeding of the 5th World Congress of Genetic Applied toLivestock Production. 7-12 August, 1994 Guelph, Canada. Vol 21: 385-392."

Franklin, I. R. (1980). Evolutionary change in small populations. In: Conservation Biology: An evolutionary-Ecological perspective. (Edited by M.E. Soule and B.A. Wilcox). Sinauer Associates, Sunderland, MA. pp 135-149.

Frantz, A. C., Pope, L. C., Carpenter, P. J., Roper, T. J., Wilson, G. J., Delahay, R. J., and Burke, T. (2003). Reliable microsatellite genotyping of the Eurasian badger (Meles meles) using faecal DNA. *Mol Ecol* **12**(6)**,** 1649-61.

Gall, C. (1981). Goat Production. Academic Press Inc., New York, USA. 619 pp.

Ganai, N. A., and Yaday, B. R. (2001). Genetic variation within and among three Indian breeds of goat using heterologous microsatellite markers. *Anim Biotechnol.* **(2),** 121-36.

Georges, M., Dietz, A. B., Mishra, A., Nielsen, D., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X., Leipold, H., Womack, J. E., and et al. (1993). Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. *Proc Natl Acad Sci U S A* **90**(3)**,** 1058-62.

Gerlach, G., and Musolf, K. F. (2000). Fragmentation of landscape as a cause for genetic subdivision in bank volves. *Conservation biology* **14,** 1066-1074.

Goldstein, D. B., A. Ruiz Linares, L.L. Cavalli-Sforza, and Feldman., M. W. (1995b). Genetic absolute dating based on microsatellites and the origin of modem humans. *Proc. Natl. Acad Sci. USA 92:6723-6727.*

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L., and Feldman, M. W. (1995a). An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**(1)**,** 463-71.

Gottelli, D., Sillero-Zubiri, C., Applebaum, G. D., Roy, M. S., Girman, D. J., Garcia-Moreno, J., Ostrander, E. A., and Wayne, R. K. (1994). Molecular genetics of the most endangered canid: the Ethiopian wolf Canis simensis. *Mol Ecol* **3**(4)**,** 301-12.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at http://www.unil.ch/izea/software/fstat.htm.

Grodzicker, T., Williams, J. L., Sharp, P., and Sambrook, J. (1974). Cold Spring Harbor Symp. Quant. Bio. 39: 439-446. .

Guo, S. W., and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**(2)**,** 361-72.

Gwakisa, P. S., Kemp, S. J., and Teale, A. J. (1994). Characterization of Zebu cattle breeds in Tanzania using random amplified polymorphic DNA markers. *Anim Genet* **25**(2)**,** 89-94.

Haberfeld, A., Cahaner, A., Yoffe, O., Plotsky, Y., and Hillel, J. (1991). DNA fingerprints of farm animals generated by microsatellite and minisatellite DNA probes. *Anim Genet* **22**(3)**,** 299-305.

Haldane, J. B. S. (1954). An exact test for randomness of mating. *Genetics* **52,** 631-635.

Hamada, H., Seidman, M., Howard, B. H., and Gorman, C. M. (1984). Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. *Mol Cell Biol* **4**(12)**,** 2622-30.

Hammond, K. (1993). Why Conserve animal genetic resources? Giversity 9(3): 30-35.

Hammond, K. (1994). Conservation of domestic animal diversity: Global overview. In: proceeding of the 5th world congress of genetic applied to livestock production. 7-12 August, 1994 Guelph, Canada. Vol 21: 423-430.

Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. (2002). African pastoralism: genetic imprints of origins and migrations. *Science* **296**(5566)**,** 336-9.

Hanotte, O., Burke, T., Armour, J. A., and Jeffreys, A. J. (1991). Hypervariable minisatellite DNA sequences in the Indian peafowl Pavo cristatus. *Genomics* **9**(4)**,** 587-97.

Hanslik, S., Harr, B., Brem, G., and Schlotterer, C. (2000). Microsatellite analysis reveals substantial genetic differentiation between contemporary New World and Old World Holstein Friesian populations. *Anim Genet* **31**(1)**,** 31-8.

Hardy, O. J. (2003). Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Mol Ecol* **12**(6)**,** 1577-88.

Hardy, O. J., Charbonnel, N., Freville, H., and Heuertz, M. (2003). Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163**(4)**,** 1467-82.

Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2,** 618-620.

Harris, D. R. (1961). The distribution and ancestory of the domestic goat. *The journal of Linnean Society of London* **173,** 79-91.

Hartl, D. L., and Clark, A. G. (1989). "Darwinian evolution in mendelian populations. Principles of population genetics; Sinauer Associates Inc, Saunderland, Massachusetts.

Hartl, D. L., and Clark, A. G. (1997). Principles of population genetics, 3rd ed. Sinauer Associates. Saunderland, Massachusetts. pp 542.

Hasegawa, M., Kishino, H., and Saitou, N. (1991). On the maximum likelihood method in molecular phylogenetics. *J Mol Evol* **32**(5)**,** 443-5.

Hedrick, P. W., Brussard, P. F., Allendorf, F. W., Beardmore, J., and Orzack, S. H. (1986). Protein variation, fitness, and captive propagation. *Zoo Biology* **5,** 91-99.

Hellberg, M. E. (1994). Relationships between inferred levels of gene flow and geographic distance in a philopatric coral, Balanophyllia elegans. Evolution 48: 1829-1854.

Hills, D., Schlaepfer, J., Comincini, S., MacLean, I., Dolf, G., Ferretti, L., Olsaker, I., and Williams, J. L. (2003). Sequence variation in the bovine and ovine PRNP genes. *Anim Genet* **34**(3)**,** 183-90.

Hodges, J. (1990). The global organization of animal genetic resources. In: Proceeding of the 4th World Congress on Genetics Applied to Livestock Production. Vol. XIV: 466-472.

Hole, F. (1996). The Context of Caprine Domestication in the Zagaros region. In Harris,D. R. (Ed). The Origions and Spread of Agriculture and Pastoralism in EuroAsia. Smithsonian Inst., Washington,DC. pp 263-281.**,** 390-412.

Hunter, M. L. J. (1996). Fundamentals of conservation biology. Blackwell science, Inc., Cambridge, Massachusetts, U.S.A. pp 482.

Iglesias, G. M., Soria, L. A., Goto, R. M., Jar, A. M., Miquel, M. C., Lopez, O. J., and Miller, M. M. (2003). Genotypic variability at the major histocompatibility complex (B and Rfp-Y) in Camperos broiler chickens. *Anim Genet* **34**(2)**,** 88-95.

Jin, L. & Chakraborty, R. 1993. Estimation of Genetic Distance and Coefficient of Gene Diversity from Single-Probe Multilocus DNA Fingerprinting Data. *Mol. Biol. Evol.* **11**: 120-127.

Jeffreys, A. J., and Morton, D. B. (1987). DNA fingerprints of dogs and cats. *Anim Genet* **18**(1)**,** 1-15.

Jeffreys, A. J., Royle, N. J., Wilson, V., and Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332**(6161)**,** 278-81.

Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L., and Armour, J. A. (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* **6**(2)**,** 136-45.

Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**(6006)**,** 67-73.

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J., and Ponder, B. A. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Am J Hum Genet* **39**(1)**,** 11-24.

Jianlin, H., Mburu, D., Ochieng, J., Kaufmann, B., Rege, J. E., and Hanotte, O. (2000). Application of New World Camelidae microsatellite primers for amplification of polymorphic loci in Old World camelids. *Anim Genet* **31**(6)**,** 404-6.

Joshi, M. B., Rout, P. K., Mandal, A. K., Tyler-Smith, C., Singh, L., and Thangaraj, K. (2004). Phylogeography and origin of Indian domestic goats. *Mol Biol Evol* **21**(3)**,** 454-62.

Kalinowski, S. T. (2002). Evolutionary and statistical properties of three genetic distances. *Mol Ecol* **11**(8)**,** 1263-73.

Jordana J., Piedrafita J., Sanchez A., and Puig P. (1992). Comparative F statistics analysis of the genetic structure of ten Spanish dog breeds. *Journal of Heredity* **83**: 367-74.

Kanga, S. (2003). Development and Application of genetic markers linked to a QTL controlling resistance to Trypanosomosis on bovine chromosome 7. Ph.D. Thesis. Department of Biochemistry, Kenyatta University.

Kantanen, J., Vilkki, J., Elo, K., and Maki-Tanila, A. (1995). Random amplified polymorphic DNA in cattle and sheep: application for detecting genetic variation. *Anim Genet* **26**(5)**,** 315-20.

Kappes, S. M., Keele, J. W., Stone, R. T., McGraw, R. A., Sonstegard, T. S., Smith, T. P., Lopez-Corrales, N. L., and Beattie, C. W. (1997). A second-generation linkage map of the bovine genome. *Genome Res* **7**(3)**,** 235-49.

Karp, A., Issac, P. G., and Ingram, D. S. (1998). Molecular tools for screening biodiversity. Plants and Animals. Chapman and Hall. London., Weinheim, New York, Tokyo, Melbourne, Madras. pp 498.

Kashi, Y., Lipkin, E., Darvasi, A., Nave, A., Gruenbaum, Y., Beckmann, J. S., and Soller, M. (1990a). Parentage identification in the bovine using "deoxyribonucleic acid fingerprints". *J Dairy Sci* **73**(11)**,** 3306-11.

Kashi, Y., Tikochinsky, Y., Genislav, E., Iraqi, F., Nave, A., Beckmann, J. S., Gruenbaum, Y., and Soller, M. (1990b). Large restriction fragments containing poly-TG are highly polymorphic in a variety of vertebrates. *Nucleic Acids Res* **18**(5)**,** 1129-32.

Kemp, S. J. (1992). The potential contribution of biotechnology in breed characterization. In: African animal genetic Resources:Proceedings of the research planning workshop held at ILCA, Addis Ababa, Ethiopia 19-21 February, 1992. ILCA (International Livestock Center for Arfica) Addis Ababa, Ethiopia. pp 23-28."

Kikkawa, Y., Yonekawa, H., Suzuki, H., and Amano, T. (1997). Analysis of genetic diversity of domestic water boffaloes and anoas based on variations in the metochondrial gene for cytochrome b. *Animal Genetics* **28,** 195-201.

Kimura, M., and Crow, J. F. (1964). The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics* **49,** 725-38.

King, T. L., Kalinowski, S. T., Schill, W. B., Spidle, A. P., and Lubinski, B. A. (2001). Population structure of Atlantic salmon (Salmo salar L.): a range-wide perspective from microsatellite DNA variation. *Mol Ecol* **10**(4)**,** 807-21.

Kuhnlein, U., Zadworny, D., Dawe, Y., Fairfull, R. W., and Gavora, J. S. (1989). DNA fingerprinting: A tool for determining genetic distances between strains of poultry. *Theoretical and Applied Genetics* **77,** 669-672.

Lacy, R. C. (1987). Loss of genetic variation from managed populations: Interacting effects of drift, mutation, immigration, selection and population subdivision. Conservation Biology 1: 143-158.

Lande, R. (1988). Genetics and demography in biological conservation. *Science* **241**(4872)**,** 1455-60.

Lande, R., and Barrowclaugh, G. F. (1987). Effective population size, genetic variation, and their use in population management. In: Viable populations for conservation (Edited by M.E. Soule). University press. pp 87-123.

Laval, G., Iannuccelli, N., Legault, C., Milan, D., Groenen, M. A., Giuffra, E., Andersson, L., Nissen, P. H., Jorgensen, C. B., Beeckmann, P., Geldermann, H., Foulley, J. L., Chevalet, C., and Ollivier, L. (2000). Genetic diversity of eleven European pig breeds. *Genet Sel Evol* **32**(2)**,** 187-203.

Leak, S., Faye, D., and Geerts, S. (2002). Genetic resistance to vectors and vector borne diseases in indigenous West African livestock. In: Integrated Control of Pathogenic Trypanosomes and their Vectors (ICPTV) newsletter, 6: 36-38.

Leberg, P. L. (1992). Effects of a population bottleneck on genetic variation as measured by alloyzme electrophoresis. evolution 46: 477-494.

Li, M. H., Zhao, S. H., Bian, C., Wang, H. S., Wei, H., Liu, B., Yu, M., Fan, B., Chen, S. L., Zhu, M. J., Li, S. J., Xiong, T. A., and Li, K. (2002). Genetic relationships among twelve Chinese indigenous goat populations based on microsatellite analysis. *Genet Sel Evol* **34**(6)**,** 729-44.

Li Xiang-Long, and Valentini, A. (2004). Genetic diversity of Chinese goat breeds based on microsatellite markers. *J. Anim. Breed. Genet.* **121,** 350-355.

Litt, M., and Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* **44**(3)**,** 397-401.

Loftus, R. T., Ertugrul, O., Harba, A. H., El-Barody, M. A., MacHugh, D. E., Park, S. D., and Bradley, D. G. (1999). A microsatellite survey of cattle from a centre of origin: the Near East. *Mol Ecol* **8**(12)**,** 2015-22.

Loftus, R. T., MacHugh, D. E., Ngere, L. O., Balain, D. S., Badi, A. M., Bradley, D. G., and Cunningham, E. P. (1994). Mitochondrial genetic variation in European, African and Indian cattle populations. *Anim Genet* **25**(4)**,** 265-71.

Lomker, R., and Simon, D. L. (1994). Costs of and inbreeding in conservation strategies for endangered breeds of cattle. In: Proceedings of the 5th World Congress on Genetics Applied to Livestock production, Vol. XIV: 434-442.

Louis, E. J., and Dempster, E. R. (1987). An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* **43**(4)**,** 805-11.

Luikart, G. (1997). Usefulness of molecular markers for detecting population bottlenecks and monitoring genetic change (PhD dissertation). Missoula, Montana: University o Montana.

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**(3)**,** 238-47.

Luikart, G., and Cornuet, J. M. (1998). Emperical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation biology* **12**(1)**,** 228-237.

Luikart, G., Gielly, L., Excoffier, L., Vigne, J. D., Bouvet, J., and Taberlet, P. (2001). Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc Natl Acad Sci U S A* **98**(10)**,** 5927-32.

Luo, J., Liu, Z. G., Yang, G. S., and Zhen, X. M. (2001). Evaluation of Genetic Relativeness and Diversity in Five Goat Breeds Using Randomly Amplified Polymorphic DNA (RAPD) Analysis. http://www.luresext.edu/goats/library/luo2001-1.html.

Lynch, M. (1988). Estimation of relatedness by DNA fingerprinting. *Mol Biol Evol* **5**(5)**,** 584-99.

MacHugh, D. E., Loftus, R. T., Bradley, D. G., Sharp, P. M., and Cunningham, P. (1994). Microsatellite DNA variation within and among European cattle breeds. *Proc R Soc Lond B Biol Sci* **256**(1345)**,** 25-31.

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**(5)**,** 333-40.

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**(3)**,** 1071-86.

Manly B.F.J. (1986). Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Res. Pop. Ecol.* **28,** 201-218.

Manly B.F.J. (1994). Multivariate statistical methods: a primer, 2nd ed. New York: Chapman & Hall.

Mannen, H., Nagata, Y., and Tsuji, S. (2001). Mitochondrial DNA reveal that domestic goat (Capra hircus) are genetically affected by two subspecies of bezoar (Capra aegagurus). *Biochem Genet* **39**(5-6)**,** 145-54.

Marshall, F. (2000). The origion and spread of domestic animals in East Africa. In: The origions and development of African livestock: archaeology, genetics, linguistics and ethinography (Edited by Blench, R.M. and MacDonald, K.C), UCL press. London, New York. pp 191-221.

Martin-Burriel, I., Garcia-Muro, E., and Zaragoza, P. (1999). Genetic diversity analysis of six Spanish native cattle breeds using microsatellites. *Anim Genet* **30**(3)**,** 177-82.

Mason, I. L. (1984). Goat. In: Mason I.L. (Ed), Evolution of domestic animals. Longman, Essex, England. pp 85-99.

Mburu, D. N., Ochieng, J. W., Kuria, S. G., Jianlin, H., Kaufmann, B., Rege, J. E., and Hanotte, O. (2003). Genetic diversity and relationships of indigenous Kenyan camel (Camelus dromedarius) populations: implications for their classification. *Anim Genet* **34**(1)**,** 26-32.

Meester, J., and Setzer., H. W. (1977). The mammals of Africa: an identification manual. Smithson. Inst. Press, Washington, D.C.

Meghen, C., MacHugh, D. E., and Bradley, D. G. (1994). Genetic characterization and West Africa cattle. World Animal Review 78(1): 59-66.

Miller, C. R., and Waits, L. P. (2003). The history of effective population size and genetic diversity in the Yellowstone grizzly (Ursus arctos): implications for conservation. *Proc Natl Acad Sci U S A* **100**(7)**,** 4334-9.

Mason, I.L. 1981. Breeds: In: Goat production. Ed: by C Gall. Chapter 3, Academic press.

Mills, S. L., and Smouse, P. E. (1994b). Demographic consequences of inbreeding in reminant populations. *American Naturalist* **144,** 412-431.

Minelli, A. (1993). Biological systematics. The state of the art. Capman & Hall. London, Glasgow, New York, Tokyo, Melbourne, Madras. pp 387.

Moazami-Goudarzi, K., and Laloe, D. (2002). Is a multivariate consensus representation of genetic relationships among populations always meaningful? *Genetics* **162**(1)**,** 473-84.

Moazami-Goudarzi, K., Laloe, D., Furet, J. P., and Grosclaude, F. (1997). Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites. *Anim Genet* **28**(5)**,** 338-45.

Mommens, G., Peelman, L. J., Vanzeveren, A., G. D. 'leteren, and Wissocq, N. (1999). Microsatellite variation between an African and five European taurine breeds results in a geographical phylogenetic tree with a bison outgroup. *J. Ani. Breed. Genet.* **116:5,** 325-330.

Moore, H. D. M. (1985). Storage of gametes and embryos. In: Advances in Animal consevation. The proceedings of a symposium held at the Zoological Society of London on 31 May - 1 June, 1984. No.54: 137-148.

Moore, S. S., Sargeant, L. L., King, T. J., Mattick, J. S., Georges, M., and Hetzel, D. J. (1991). The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**(3)**,** 654-60.

Mopper, S., Mitton, J. B., Whitham, T. J., Cobb, N. S., and Christensen, K. M. (1991). Genetic differentiation and heterozygosity in pinyon pine associated with resistance to herbivory and environmental stress. Evolution 45: 989-999.

Morin, P. A., Moore, J. J., Chakraborty, R., Jin, L., Goodall, J., and Woodruff, D. S. (1994). Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**(5176)**,** 1193-201.

Muzzolini, A. (2000). Livestock in Saharan art. In Blench, R.M.and Macdonald, K.C. (Eds.). The origion and development of African livestock, Archaeology, genetics, linguistics and ethnography. University College London press, New York. pp. 87-110.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and et al. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**(4796)**,** 1616-22.

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

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A* **70**(12)**,** 3321-3.

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

Nei, M. (1984). *Molecular Evolutionary Genetics. Colombia University press/New York.*

Nei, M. (1987). Molecular Evolutionary Genetics. Columbia University press, New York. pp 506.

Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annu Rev Genet* **30,** 371-403.

Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* **19**(2)**,** 153-70.

Nei, M., and Takezaki, N. (1994). Estimation of genetic distances and phylogenetic trees fromDNA analysis. *In: proceeding of the 5th world congress of genetic applied to livestock production. 7-12 August, 1994 Guelph, Canada. Vol 21: 405-412.*

Neigel, J. E. (1997). A comparison of alternative strategies for estimating gene flow from genetic markers. *Annu. Rev. Ecol. Syst.* **28:,** 105-128.

Nelson, K., and Soule, M. E. (1987). Genetic conservation of exploited fishes. In: Population genetics and fishery management (Edited by Nils Ryman and Fred Utter). Washington press. Seattle and London. pp 345-368.

Nguyen, T. C. (1990). Genetic systems of red cell blood groups in goats. *Anim Genet* **21**(3)**,** 233-45.

Nigatu Alemayehu (1994). Characterization of indigenous goat types of Eritrea , Northern and Western Ethiopia . MSc thesis, Alemaya University of Agriculture, Alemaya, Ethiopia. 136pp.

Notter, D. R., Mariante, A. d. S., and Sheng, Z. (1994). Modern approach to active conservation of domestic animal diversity. In: Proceeding of the 5th World Congress of Genetic Applied toLivestock Production. 7-12 August, 1994 Guelph, Canada. Vol 21: 509-516."

Okoma, M. A., Rege, J. E. O., Teale, A., and Hanotte, O. (1998). Genetic characterization of indigenous East African cattle breeds using microsatellite DNA markers. In:. *Proceeding of the 6th world congress on Genetics Applied to livestock production, Armidale, Australia, January 11-16, 1998.*

Oldenbroek, J. K. E. (1999). "Genebanks and the conservation of farm animal genetic resources" DLO, The Netherlands.

Ohta T and Kimura M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Cambr*. **22:** 201-204.

Ota, T. (1993). "Dispan: Genetic Distance and Phylogenetiv Analysis ". Pennsylvania State University, Pennsylvania.

Paetkau, D., Calvert, W., Stirling, I., and Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* **4**(3)**,** 347-54.

Paetkau, D., and Strobeck, C. (1995b). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol Ecol* **4**(4)**,** 519-20.

Page, R. D. M., and Holmes, E. C. (1998). Molecular evolution. A phylogenetic approach. Blackwell science. University press, Cambridge. pp 346.

Palsson, S. (2000). Microsatellite variation in Daphnia pulex from both sides of the Baltic Sea. *Mol Ecol* **9**(8)**,** 1075-88.

Pandey, A. K. (1999). Studies on DNA polymorphism in buffaloes using microsatellite markers. M.Sc. Thesis, NDRI, Deemed University, Karnal.

Park, S. D. E. (2001). Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection [ Ph.D. thesis ] University of Dublin.

Pemberton, J. M., Albon, S. D., Guinness, F. E., Clutton-Brock, T. H., and Berry, R. J. (1988). Genetic variation and juvenile survival in red deer. Evolution 42: 921-934.

Pepin, L., and Nguyen, T. C. (1994). Blood groups and protein polymorphisms in five goat breeds (Capra hircus). *Anim Genet* **25**(5)**,** 333-6.

Piry S, Alapetite A, Cornuet, J.-M., Paetkau D, Baudouin, L., and Estoup, A. (2004). GeneClass2: a software for genetic assignment and first generation migrants detection. *Journal of Heredity (accepted).*

Porter, V. (1996). Goats of the world (Farming press, Ipswich, U.K.).

Powell, W., Morgante, M., Dndre, C., Hanafey, M., Vogel, J., Tingey, S. V., and Rafalski, J. A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (Microsatellite) markers for germplasm analysis. *Molecular Breeding* **2:,** 225-238.

Pringle, H. (1998). NEOLITHIC AGRICULTURE:Reading the Signs of Ancient Animal Domestication. *Science* **282**(5393)**,** 1448.

Pringle, H. (1998b). NEOLITHIC AGRICULTURE:Reading the Signs of Ancient Animal Domestication. *Science* **282**(5393)**,** 1448-.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**(2)**,** 945-59.

Queller, D. C., Strassmann, J. E., and Hughes, C. R. (1993). Microsatellite and Kinship. *Trends in Ecology and Evolution* **8,** 285-288.

Ralls, K., Ballou, J. D., and Templeton, A. (1988). Estimation of lethal equivalents and the costs of inbreeding in mammals. Conservation Biology 2: 185-193.

Rannala, B., and Mountain, J. L. (1997). Detection immigration by using multilocus genotypes. Proceeding of the National Accademy of Sciences USA 94: 9197-9221.

Raymond, M., and Riousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact test and ecumenicsm. *J.Heredity,* **86,** 248-249.

Raymond, M., and Rousset, F. (1995). An exact test for population differentiation. *Evolution* **49,** 1280-1283.

Rege, J. E. O. (1992). Indigenous African small ruminants: a case for characterization and improvement. In: proceedings of the second Biennial conference of the African small ruminant research Network, 7 - december, 1992, AICC, Arusha, Tanzania. pp 205-211.

Rosenberg, N. A. (2004), Distruct: a program for the graphical display of structure results. http://www.cmb.usc.edu/~noahr/distruct.html).

Rege, J. E. O., Muigai, A. W. T., Chenyambuga, S., Kierstein, G., and Hanotte, O. (2002). Assesment of genetic diversity in African small ruminants: present status and future prospects. In: 53rd Annual meeting of the European association of animal production, Cairo, Egypt, September 1-4, 2002.

Reist-Marti, S. B., Simianer, H., Gibson, J., Hanotte, O., and O., R. J. E. (2003). Weitzman's Approach and Conservation of Breed Diversity: an Application to African Cattle Breeds. *Conservation Biology* **17:5p,** 1299-1311.

Rice, W. R. (1989). Analysing tables of ststistical tests. *Evolution* **43:,** 223-225.

Richards, R. I., and Sutherland, G. R. (1994). Simple repeat DNA is not replicated simply. *Nat Genet* **6**(2)**,** 114-6.

Ritz, L. R., Glowatzki-Mullis, M. L., MacHugh, D. E., and Gaillard, C. (2000). Phylogenetic analysis of the tribe Bovini using microsatellites. *Anim Genet* **31**(3)**,** 178-85.

Roderick GK (1997). Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Ann Rev Entomol* **41,** 325-352.

Rogers, J. S. (1972). Measures of genetic similarity and genetic distances. Studies in Genetics, Univ. Texas Publ., 7213, 145–153.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**(4)**,** 1219-28.

Rousset, F., and Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics* **140**(4)**,** 1413-9.

Roy, M. S., Geffen, E., Smith, D., Ostrander, E. A., and Wayne, R. K. (1994). Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Mol Biol Evol* **11**(4)**,** 553-70.

Rzhetsky, A., and Nei, M. (1992). Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. *J Mol Evol* **35**(4)**,** 367-75.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn GT, Mullis, K. B., and H.A., E. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science.* **239,** 487-91.

Saitbekova, N., Gaillard, C., Obexer-Ruff, G., and Dolf, G. (1999). Genetic diversity in Swiss goat breeds based on microsatellite analysis. *Anim Genet* **30**(1)**,** 36-41.

Saitou, N. (1988). Property and efficiency of the maximum likelihood method for molecular phylogeny. *J Mol Evol* **27**(3)**,** 261-73.

Saitou, N., and Imanishi, T. (1989). Relative efficiencies of the Fitch-Margoliash, Maximum-Parsimony, Maximum-Likelihood, Minimum-Evolution and Neighbour-Joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biology and Evolution* **6,** 514-525.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4)**,** 406-25.

Schlotterer, C., and Pemberton, J. (1994). In molecular Ecology and Evolution: approaches and Applications (eds. B. Schierwater, R., Streit, G.P., Wagner and R. Desalle), pp.203-14. Bir Khauser Verlag Basel, Switzerland.

Schlotterer, C., and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* **20**(2)**,** 211-5.

Schwartz, M. K., Mills, L. S., Ortega, Y., Ruggiero, L. F., and Allendorf, F. W. (2003). Landscape location affects genetic variation of Canada lynx (Lynx canadensis). *Mol Ecol* **12**(7)**,** 1807-16.

Seehausen, O., Koetsier, E., Schneider, M. V., Chapman, L. J., Chapman, C. A., Knight, M. E., Turner, G. F., van Alphen, J. J., and Bills, R. (2003). Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. *Proc R Soc Lond B Biol Sci* **270**(1511)**,** 129-37.

Selander, R. K. (1983). Evolutionary consequences of inbreeding. In: Genetics and Conservation: A reference for managing wild animal and plant populations. (Edited by C.M. Schonewald-Cox; S.M. Chambers; B.MacBryde and L.Thomas). Benjamin/Cummings, menlopark, CA. pp 201-215.."

Serikawa, T., Montagutelli, X., Simon-Chazottes, D., and Guenet, J. L. (1992). Polymorphisms revealed by PCR with single, short-sized, arbitrary primers are reliable markers for mouse and rat gene mapping. *Mamm Genome* **3**(2)**,** 65-72.

Shackleton, D. M., ed. (1997). in Wild Sheep and Goats and their Relatives: Status Survey and Conservation Action Plan for Caprinae (IUCN, Gland, Swizerland).

Shriver, M. D., Jin, L., Boerwinkle, E., Deka, R., Ferrell, R. E., and Chakraborty, R. (1995). A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* **12**(5)**,** 914-20.

Simm, G. (1998). Genetic improvement of cattle and sheep. Farming press. London. pp 433.

Skalova, H., Pechackova, S., Suzuki, J., Herben, T., Hara, T., Hadincova, V., and Krahulec, F. (1997). Within population genetic differentiation in traits affecting clonal growth: Festuca rubra in a mountain grassland. *J Evolution Biol* **10**(3)**,** 383-406.

Slatkin, M. (1985). Gene flow in natural populations. Annu Rev Ecol Syst 16: 393-430.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**(1)**,** 457-62.

Smith, A. B. (2000). "The origion and spread of domestic animals in Southern Africa. In: The origions and development of African livestock: archaeology, genetics, linguistics and ethinography (Edited by Blench, R.M. and MacDonald, K.C), UCL press. London, New York. pp 222-231."

Smith, C. (1984). Estimated costs of genetic conservation in farm livestock. In: Animal Genetic Resources Conservation by Management, Data Banks and Training. FAO Animal Production and Health Paper 44/1: 21–30.

Sneath, P. H. A., and Sokal, R. R. (1973). In: Numerical Taxonomy (pp; 230-234), W.H. Freeman and Copany, San Francisco, California, USA.

Soller, M., and Beckmann, J. S. (1982). In: Proc. 2nd world Cong. Genet. Appl livestock prod. Madrid, 6: 396-404.

Soule, M. E. (1980). Thresholds for survival: maintaining fitness and evolutionary potential. In: Conservation biology (Edited by M.E. Soule and B.A. Wilcox). Sinauer, Sunderland, Massachusetts, USA. pp 151-168.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**(3)**,** 503-17.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E., and Moyzis, R. K. (1991). Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics* **10**(3)**,** 807-15.

Sultana, S., Mannen, H., and Tsuji, S. (2003). Mitochondrial DNA diversity of Pakistani goats. *Animal Genetics* **34**(6)**,** 417-421.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution* **15**(5)**,** 199-203.

Swofford, D. L., and Oslen, G. J. (1990). Phylogeny reconstruction. In: Molecular systematics (Edited by D.M. Hills and C. Mortiz). Sinauer, Sunderland, Massachusetts. pp 411-501.

Swofford, D. L., Oslen, G. J., Waddell, P. J., and Hills, D. M. (1996). Phylogenetic inference. In: Molecular systematics 2nd ed (Edited by D.M. Hills and C. Mortiz). Sinauer, Sunderland, Massachusetts. pp 407-514.

Taberlet, P., Camarra, J. J., Griffin, S., Uhres, E., Hanotte, O., Waits, L. P., Dubois-Paganon, C., Burke, T., and Bouvet, J. (1997). Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Mol Ecol* **6**(9)**,** 869-76.

Takada, T., Kikkawa, Y., Yonekawa, H., Kawakami, S., and Amano, T. (1997). Bezoar (Capra aegagrus) is a matriarchal candidate for ancestor of domestic goat (Capra hircus): evidence from the mitochondrial DNA diversity. *Biochem Genet* **35**(9-10)**,** 315-26.

Takezaki, N., and Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**(1)**,** 389-99.

Tanksley, S. D., Young, N. D., Peterson, A. H., and Bonierbale, M. W. (1989). RFLP mapping in plant breeding: new tools for an old science. *Biotechnology* **7,** 257-264.

Tateno, Y., Nei, M., and Tajima, F. (1982). Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. *J Mol Evol* **18**(6)**,** 387-404.

Tauz, D. (1989). Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acid Research* **17,** 6463-6471.

Taylor, A. C., Sherwin, W. B., and Wayne, R. K. (1994). Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat Lasiorhinus krefftii. *Mol Ecol* **3**(4)**,** 277-90.

Teale, A. J., Wambugu, J., Gwakisa, P. S., Stranzinger, G., Bradley, D., and Kemp, S. J. (1995). A polymorphism in randomly amplified DNA that differentiates the Y chromosomes of Bos indicus and Bos taurus. *Anim Genet* **26**(4)**,** 243-8.

Templeton, A. R. (1986). Coadaptation and outbreeding depression. In: Conservation Biology. The Science of scarcity and diversity (Edited by M.E. Soule). Sinauer, Sunderland, Massachusetts. pp 105-116.

Thaon d'Arnoldi, C., J. L. Foulley, and Ollivier., L. (1998). An overview of the Weitzman approach to diversity. *Genetics Selection Evolution* **30,** 149-161.

Tiedemann, R., Hardy, O., Vekemans, X., and Milinkovitch, M. C. (2000). Higher impact of female than male migration on population structure in large mammals. *Mol Ecol* **9**(8)**,** 1159-63.

Uerpmann, H.-P. (1996). Animal Domestication-accident or intention? In: Harris, D.R. (Ed), The origions and spread of agriculture and pastoralism in Euroasia. Smithsonian press, Washington D.C pp 227-237.

Vaiman, D., Imam-Ghali, M., Moazami-Goudarzi, K., Guerin, G., Grohs, C., Leveziel, H., and Saidi-Mehtar, N. (1994a). Conservation of a syntenic group of microsatellite loci between cattle and sheep. *Mamm Genome* **5**(5)**,** 310-4.

Vaiman, D., Mercier, D., Moazami-Goudarzi, K., Eggen, A., Ciampolini, R., Lepingle, A., Velmala, R., Kaukinen, J., Varvio, S. L., Martin, P., and et al. (1994b). A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. *Mamm Genome* **5**(5)**,** 288-97.

Vaiman, D., Schibler, L., Bourgeois, F., Oustry, A., Amigues, Y., and Cribiu, E. P. (1996). A genetic linkage map of the male goat genome. *Genetics* **144**(1)**,** 279-305.

Valdes, A. M., Slatkin, M., and Freimer, N. B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**(3)**,** 737-49.

Van Oosterhout, C., Hutchinson, B., Wills, B., and Shipley, P. (2003). Micro-Checker: Microsatellite Data Checking Software (Last updated 4 August 2003). Available from http://www.microchecker.hull.ac.uk.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**(21)**,** 4407-14.

Wayne, R. K., Gilbert, D., and Lehman, N. (1991). Conservation genetics of endangered Isle Royale gray wolf. *Conservation Biology:* **5,** 41-51.

Weber, J. L. (1990). Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics* **7**(4)**,** 524-30.

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**(3)**,** 388-96.

Weir, B. S. (1996). Genetic Data Analysis II: Methods for Discrete Population Genetic Data.

Weir, B. S., and Basten, C. J. (1990). Sampling strategies for distances between DNA sequences. *Biometrics* **46**(3)**,** 551-82.

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38,** 1358-1370.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. (1992). A second-generation linkage map of the human genome. *Nature* **359**(6398)**,** 794-801.

Weitzman M.L. (1998). The Noah's ark problem, Econometrica 66, 1279-1298.

Weitzman, M. L. (1992). On diversity, Quarterly Journal of Economics 107, 363-405.

Weitzman, M. L. (1993). What to preserve? An application of diversity theory to crane conservation, Quarterly Journal of Economics 108, 157-183.

Welsh, J., and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* **18**(24)**,** 7213-8.

Williams, J. G., Hanafey, M. K., Rafalski, J. A., and Tingey, S. V. (1993). Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol* **218,** 704-40.

Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* **18**(22)**,** 6531-5.

Williams, J. G. K., Rafalski, J. A., and Tingey, S. V. (1993). In Wu R (ed), Methods in Enzymology 218:. *Methods Enzymol.* **86:,** 704-740.

Wilson, G. A., and Strobeck, C. (1999). Genetic variation within and relatedness among wood and plains bison populations. *Genome* **42**(3)**,** 483-96.

Wilson, T. R. (1991). FAO animal health and production paper no. 88, pp. 231.

Wollny, C. B. A. (2003). The need to conserve farm animal genetic resources in Africa: should policy makers be concerned? *Ecological Economics* 45: 341-351

Workneh Ayalew (1992). Preliminary survey of indigenous goat types and goat husbandery practices in Southern Ethiopia. MSc thesis, Alemaya University of Agriculture, Alemaya, Ethiopia. 153pp.

Wright, S. (1943). ISOLATION BY DISTANCE. *Genetics* **28**(2)**,** 114-138.

Wright, S. (1969). Evolution and the genetics of populations. University of Chicago press, Chicago.

Yadav, B. R., S.De, and Mitra, A. (1998). Bovine microsatellite primers reveal DNA polymorphism in river buffalo (*Bubalus Bubalis*). In: Indo-Brazilian Workshop on Agricultural Biotechnology held at NDRI, Karnal, April 27-29, 1998, p-26.

Yadav, B. R., Thakur, M. K., Singhal, A., and Singh R.P (2001). Genetic variation on three breeds of water buffalo analyzed by RAPD primers. In: 5th ADNAT convention: A two days symposium on perspectives in genome analysis held at CCMB, Hyderabad, Feb.23-24. Abst. P. 40.

Yang, L. Zhao, S. H. Li, K. Peng, Z. Z. and Montgomery, G. W. (1999). Determination of genetic relationships among five indigenous Chinese goat breeds with six microsatellite markers. *Anim. Genet* **30**(6): 452-5.

Yang, S.L., Wang, Z.G., Liu, B., Zhang, G.X., Zhao, S.H., Yu, M., Fan, B., Li, M.H., Xiong, T.A., and Li, K. (2003). Genetic variation and relationships of eighteen Chinese indigenous pig breeds. *Genet Sel. Evol*. **35**(6): 657-71.

Zeder, M. A. and B. Hesse (2000). "The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science* **287**(5461): 2254-7.

# APPENDICES

Appendix 1. **Summary of linkage disequilibrium between all markers across all 20 populations.**

| | Linkage between markers | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | BM1818 & BMS1494 | | | | | | | | | | * | | | * | | | | | | | |
| 2 | BM1818 & ILSTS05 | | | | | | | | | | | | | | * | | | | | | |
| 3 | BM1818 & ILSTS11 | | | * | | ** | * | | | | | * | | | | | | | | | |
| 4 | BM1818 & ILSTS44 | | | | | ** | | | | | | | | | * | | | | | | |
| 5 | BM1818 & ILSTS87 | | | | | | | | | | | | | | | | | | | | |
| 6 | BM1818 & INRA05 | | | * | | | | | | | | | | | | | | | | | |
| 7 | BM1818 & INRA132 | | | | | | | ** | | | | ** | | | | | | | | * | * |
| 8 | BM1818 & INRA63 | | | | | | | | | | | | | * | | | | | | | * |
| 9 | BM1818 & MAF209 | | | ** | | | | | * | | * | | | | | | | | | | |
| 10 | BM1818 & MAF35 | | | | | | | | | | | | | | | | | | | | |
| 11 | BM1818 & MAF65 | | | | | | | | | | | | | | | | | | | | |
| 12 | BM1818 & OarAE129 | | ** | | | | | | | | | * | | | | | | | | | |
| 13 | BM1818 & OarFCB304 | | | | | | | | | | | * | * | | | | | | | * | |
| 14 | BM1818 & SRCRSP03 | | | | | | | | | | | | * | | | | | | | | |
| 15 | BMS1494 & ILSTS05 | | | | | | | | | | | | | | | | * | | | | * |
| 16 | BMS1494 & ILSTS11 | | | | ** | | | | | | ** | | | | | | | | | | |
| 17 | BMS1494 & ILSTS44 | | | | | | | | | | | | | | | | * | | | | |
| 18 | BMS1494 & ILSTS87 | | | | | | | | | ** | | | | | | | | | | | |
| 19 | BMS1494 & INRA05 | | * | | | | | | * | | | | | | | | | | | | |
| 20 | BMS1494 & INRA132 | | | | | | | | | | | | | | | | | | | | |
| 21 | BMS1494 & INRA63 | | | | | | | | | | | | | | | | | | | | |
| 22 | BMS1494 & MAF209 | | * | | * | | | | | | | | | | | | | | | | |
| 23 | BMS1494 & MAF35 | | | | | | | | * | | | | | | | | * | | | | |
| 24 | BMS1494 & MAF65 | | | | | * | | | * | | | | * | | | | | | * | | |
| 25 | BMS1494 & OarAE129 | | | | | * | | | * | | | | | | | | | | | | |
| 26 | BMS1494 & OarFCB304 | | * | | | | | | | | | | | * | * | | | | | | |
| 27 | BMS1494 & SRCRSP03 | | | | | | | | | | | | | | | | | | | | |
| 28 | ILSTS05 & ILSTS11 | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | |

| | Linkage between markers | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 29 | ILSTS05 & ILSTS44 | | | | | | | | | | | | | | | | | | | | |
| 30 | ILSTS05 & ILSTS87 | | | | | | | | | | | | | | | | | | | | ** |
| 31 | ILSTS05 & INRA05 | | | | | | | | | | | | | | * | | * | | | | |
| 32 | ILSTS05 & INRA132 | | | | | | | | | | | | | | | | | | | | |
| 33 | ILSTS05 & INRA63 | | * | | | | | | | | | | | | | | | | | | * |
| 34 | ILSTS05 & MAF209 | | | | | | | | | | * | | | | | | | | | | |
| 35 | ILSTS05 & MAF35 | | | | * | | | | | | | | | | | | | | | | * |
| 36 | ILSTS05 & MAF65 | | | | | | | | | | | | | | | | | | | | |
| 37 | ILSTS05 & OarAE129 | | * | | | | | | | | | | | | | * | | | | | |
| 38 | ILSTS05 & OarFCB304 | | | | | | | | | | | | | | | | | | | | ** |
| 39 | ILSTS05 & SRCRSP03 | | | | | | | | | | | | | | ** | | * | * | | | |
| 40 | ILSTS11 & ILSTS44 | | | | | | | | | | | | | | | | | | | | |
| 41 | ILSTS11 & ILSTS87 | | | | | | | | | | | | | | * | | | | | | |
| 42 | ILSTS11 & INRA05 | | | | | | | | | | | | | | | | | | | | |
| 43 | ILSTS11 & INRA132 | | | | | | | | | | | | | | | | | | | | |
| 44 | ILSTS11 & INRA63 | * | | | * | | | | | | | | | | | | | | | | |
| 45 | ILSTS11 & MAF209 | | | | | | | | | | * | | | | | | | | | | |
| 46 | ILSTS11 & MAF35 | | | | | | | | | | | | | | | | | | | | |
| 47 | ILSTS11 & MAF65 | | | | | | | | | | | | | | | | | | | | |
| 48 | ILSTS11 & OarAE129 | | | | | | | | | | * | | | | | | | | | | |
| 49 | ILSTS11 & OarFCB304 | | * | | | | | ** | | | | | | | | | | | | | |
| 50 | ILSTS11 & SRCRSP03 | | | | | | | | | | * | | | | | | | | | | |
| 51 | ILSTS44 & ILSTS87 | | * | | | | | | | | | | * | | | | | | | | |
| 52 | ILSTS44 & INRA05 | | | | | | | | | | | | | | | | | | | | |
| 53 | ILSTS44 & INRA132 | | | | | | | | ** | | | | | | | | * | | | * | * |
| 54 | ILSTS44 & INRA63 | | | | | | | | | | | | | | ** | | | | | | |
| 55 | ILSTS44 & MAF209 | | | | | | | | | | | | | | * | | | | | | |
| 56 | ILSTS44 & MAF35 | | | | | | | | | | | * | | | | | | | | | |
| 57 | ILSTS44 & MAF65 | | | | | | | | | | | * | | | | | | | | | |
| 58 | ILSTS44 & OarAE129 | | | | | | | * | | * | | | | | | | | | | | |
| 59 | ILSTS44 & OarFCB304 | | | | * | | * | | | | | | | | | | | | | | |

| | Linkage between markers | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 60 | ILSTS44 & SRCRSP03 | | | | | | | | | | | | | | | | | | | | |
| 61 | ILSTS87 & INRA05 | | | | * | | * | | | | | | | | | | | | | | |
| 62 | ILSTS87 & INRA132 | | | | | | | | | | | | | * | * | | | | | | |
| 63 | ILSTS87 & INRA63 | | | | | * | | | | | * | | | | | | | | | | |
| 64 | ILSTS87 & MAF209 | | | | | | | | | | | | | | | | | | | | |
| 65 | ILSTS87 & MAF35 | | | | | | | | | | | | | | | | | | | | |
| 66 | ILSTS87 & MAF65 | | | | | | | | | | | | | | | | | | | | |
| 67 | ILSTS87 & OarAE129 | | | | | | | | * | | | | | | | | | | * | | * |
| 68 | ILSTS87 & OarFCB304 | | | | | | | | | | | | | | | | | | | | |
| 69 | ILSTS87 & SRCRSP03 | | | | | | | | | | | | ** | | | | | | | | |
| 70 | INRA05 & INRA132 | | | | | | | | | | | | | | | | | | | | |
| 71 | INRA05 & INRA63 | | | | | | | | | | | | | ** | | | | | | | |
| 72 | INRA05 & MAF209 | | | | | | | | | | | | | * | | | | | * | | |
| 73 | INRA05 & MAF35 | | * | * | | | | | | | | | | | | | | | | | |
| 74 | INRA05 & MAF65 | | | | | | | | | | | | | | | | | | | | |
| 75 | INRA05 & OarAE129 | | | * | | | | | | | | | | | | | | | | ** | |
| 76 | INRA05 & OarFCB304 | | * | | | | | | | | | | | | | | | | | | |
| 77 | INRA05 & SRCRSP03 | | | | | | | | | | | | | | | | | | | | |
| 78 | INRA132 & MAF209 | | | | | | | | | | | | | | | | | | | | |
| 79 | INRA132 & MAF35 | | | | | | | | | | | | * | | | | | | | | |
| 80 | INRA132 & MAF65 | | | * | | | | | | | | | | | ** | | * | | | | |
| 81 | INRA132 & OarAE129 | | | | | | | | | | | | | | | | * | | | | |
| 82 | INRA132 & OarFCB304 | | | | | | | | | * | | | | | | | | | | | |
| 83 | INRA132 & SRCRSP03 | | | | | | | | | | | | | | | | | * | | | |
| 84 | INRA63 & INRA132 | * | | | | | | | | | | | | | | | | | | | |
| 85 | INRA63 & MAF209 | | | | | | | | | | | | | | * | | | | | | |
| 86 | INRA63 & MAF35 | | | | | | | | | | | | | | | | ** | | | | |
| 87 | INRA63 & MAF65 | | | * | | | | * | | | | | | | | | | | | | |
| 88 | INRA63 & OarAE129 | | * | | | | | | | | | | | | | | | | | | |
| 89 | INRA63 & OarFCB304 | | | | | | | | | | | | | | | | | | | | |
| 90 | INRA63 & SRCRSP03 | | | | | | | | | | | | | | | | | | | | |

| | Linkage between markers | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ | BGA | SEA | GUB | TSW | ALP | ARD | BAL | HAI | MON |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 91 | MAF209 & OarAE129 | | | | | | | | | | | | | | | | | | | | |
| 92 | MAF209 & OarFCB304 | * | | | | | | | ** | | | | * | * | | | | | | | |
| 93 | MAF209 & SRCRSP03 | | | | | | | | | | | | | | | | | | | | |
| 94 | MAF35 & MAF209 | | | | | | | * | | | | | | | | | | | | | |
| 95 | MAF35 & MAF65 | | | | | | * | | | | | | * | | | | | | | | |
| 96 | MAF35 & OarAE129 | | | | | | | | | | | | | | | | | * | | | |
| 97 | MAF35 & OarFCB304 | ** | | | | | | | | * | | | | | | | | | | | |
| 98 | MAF35 & SRCRSP03 | | | | | | | | | | | * | | | | | | | | | |
| 99 | MAF65 & MAF209 | | | | | | | | | | | | | | | | | | | | |
| 100 | MAF65 & OarAE129 | | | | | | | | | | | | | | | | | | | | |
| 101 | MAF65 & OarFCB304 | | | | | | | | | | | | | | | | | | | | |
| 102 | MAF65 & SRCRSP03 | | | | | | | | | | | | | | | | | | | * | |
| 103 | OarAE12 & OarFCB304 | | | | | | | | * | | | | | | | | | * | | | |
| 104 | OarAE12 & SRCRSP03 | | | | | | | | | | | | | | | | | | | | |
| 105 | OarFCB3 & SRCRSP03 | | | | | | | | | | | | | | | | | | | | |

* = $P < 0.05$; ** = $P < 0.01$.

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 2. **Matrix of all pair-wise combinations of loci showing the linkage disequilibrium (for 20 populations). *P* values condensed across all populations.** For simplicity, the matrix only shows levels of significance.

| | BM 1818 | BM 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BM1818 | | | | | | | | | | | | | | | |
| BMS1494 | | | | | | | | | | | | | | | |
| ILSTS05 | | | | | | | | | | | | | | | |
| ILSTS11 | * | * | | | | | | | | | | | | | |
| ILSTS44 | * | | | | | | | | | | | | | | |
| ILSTS87 | | | | | | | | | | | | | | | |
| INRA05 | | | | | | | | | | | | | | | |
| INRA63 | | | | | | | | | | | | | | | |
| INRA132 | | | | | | * | | | | | | | | | |
| MAF35 | | | | | | | | | | | | | | | |
| MAF65 | | | | | | | | | | | | | | | |
| MAF209 | * | | | | | | | | | | | | | | |
| OarAE129 | * | * | | | | | | | | | | | | | |
| OarFCB304 | | | | | | | | | | | | ** | | | |
| SRCRSP03 | | | ** | | | | | | | | | | | | |

* = $P < 0.05$; ** = $P < 0.01$.

Appendix 3. Allele frequency for locus BM1818 for all the 20 populations.

| Allele | 244 | 246 | 250 | 252 | 254 | 256 | 258 | 260 | 262 | 264 | 266 | 268 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | 1.09 | - | - | 7.61 | - | 7.61 | 51.09 | 7.61 | 13.04 | 6.52 | 5.43 | - |
| ABA | 3.26 | - | - | 5.43 | - | 5.43 | 40.22 | 7.61 | 20.65 | 6.52 | 4.35 | 6.52 |
| AFA | - | 1.14 | - | 1.14 | 1.14 | 3.41 | 65.91 | 9.09 | 3.41 | 6.82 | 5.68 | 2.27 |
| CEH | 5.00 | - | - | 5.00 | 1.25 | 2.50 | 51.25 | 6.25 | 10.00 | 11.25 | 7.50 | - |
| GUM | - | - | - | 8.14 | - | 8.14 | 41.86 | 13.95 | 10.47 | 8.14 | 8.14 | 1.16 |
| HAH | - | - | - | 2.50 | 2.50 | 7.50 | 58.75 | 11.25 | 10.00 | 5.00 | 2.50 | - |
| KEF | - | - | - | 6.58 | - | 5.26 | 68.42 | 7.89 | 5.26 | 3.95 | 2.63 | - |
| LES | 3.13 | 1.04 | 1.04 | 6.25 | 5.21 | 9.38 | 47.92 | 6.25 | 11.46 | 3.13 | 4.17 | 1.04 |
| NWH | 1.11 | - | - | 13.33 | 2.22 | 1.11 | 45.56 | 5.56 | 20.00 | 6.67 | 4.44 | - |
| SES | 1.16 | - | - | 6.98 | 4.65 | 4.65 | 46.51 | 10.47 | 15.12 | 6.98 | 2.33 | 1.16 |
| WGJ | 1.09 | - | - | 3.26 | 2.17 | 2.17 | 54.35 | 5.43 | 13.04 | 7.61 | 4.35 | 6.52 |
| BGA | - | - | 1.43 | 2.86 | 8.57 | 18.57 | 41.43 | 12.86 | 7.14 | 4.29 | 1.43 | 1.43 |
| SEA | - | - | - | 5.13 | 12.82 | 6.41 | 52.56 | 7.69 | 8.97 | - | 6.41 | - |
| GUB | 6.52 | - | - | 2.17 | 6.52 | 2.17 | 36.96 | 8.70 | - | 9.78 | 22.83 | 4.35 |
| TSW | - | - | - | 2.08 | 6.25 | 27.08 | 45.83 | 4.17 | 6.25 | 4.17 | 4.17 | - |
| ALP | - | - | - | - | - | 35.48 | 30.65 | 6.45 | 9.68 | 11.29 | 6.45 | - |
| ARD | - | - | - | - | - | 12.00 | 52.00 | 18.00 | 12.00 | 4.00 | 2.00 | - |
| BAL | - | - | 11.36 | 4.55 | 13.64 | 4.55 | 34.09 | 18.18 | 6.82 | 4.55 | 2.27 | - |
| HAI | - | - | 9.68 | 9.68 | 14.52 | 8.06 | 25.81 | 19.35 | 9.68 | 1.61 | 1.61 | - |
| MON | - | - | - | 10.81 | 21.62 | 14.86 | - | 16.22 | 16.22 | 2.70 | 14.86 | 2.70 |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 4. Allele frequency for locus BMS1494 for all the 20 populations.

| Allele | 244 | 246 | 256 | 264 | 270 | 276 | 280 | 282 | 284 | 288 |
|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | - | 9.78 | - | - | 65.22 | 18.48 | - | 6.52 | - |
| ABA | - | - | 4.55 | - | - | 57.95 | 36.36 | - | 1.14 | - |
| AFA | - | - | - | 1.14 | - | 71.59 | 22.73 | - | 4.55 | - |
| CEH | 1.25 | 1.25 | 11.25 | - | - | 65.00 | 21.25 | - | - | - |
| GUM | - | - | 2.33 | - | - | 82.56 | 13.95 | - | 1.16 | - |
| HAH | - | - | 3.75 | - | - | 60.00 | 33.75 | - | 2.50 | - |
| KEF | - | - | 5.26 | - | - | 57.89 | 21.05 | - | 14.47 | 1.32 |
| LES | - | - | 2.08 | - | - | 68.75 | 23.96 | - | 4.17 | 1.04 |
| NWH | - | - | 6.82 | - | - | 62.50 | 29.55 | 1.14 | - | - |
| SES | - | - | - | - | - | 61.63 | 34.88 | - | 3.49 | - |
| WGJ | - | - | 3.26 | - | - | 58.70 | 30.43 | - | 7.61 | - |
| BGA | - | - | - | - | - | 65.71 | 32.86 | - | 1.43 | - |
| SEA | - | - | 2.56 | - | - | 70.51 | 12.82 | - | 14.10 | - |
| GUB | - | - | 15.56 | - | - | 11.11 | 7.78 | - | 65.56 | - |
| TSW | - | - | 17.24 | 1.72 | - | 31.03 | 48.28 | - | 1.72 | - |
| ALP | - | - | 3.33 | - | - | 48.33 | 46.67 | - | 1.67 | - |
| ARD | - | - | - | - | 4.00 | 60.00 | 36.00 | - | - | - |
| BAL | - | - | 4.55 | - | - | 52.27 | 43.18 | - | - | - |
| HAI | - | - | 12.90 | - | - | 37.10 | 40.32 | - | 9.68 | - |
| MON | - | - | 3.85 | - | - | 65.38 | 28.21 | - | 2.56 | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 5. Allele frequency for locus ILSTS05 for all the 20 populations.

| Allele | 178 | 180 | 182 | 184 | 186 | 188 | 190 |
|---|---|---|---|---|---|---|---|
| ABG | - | 1.09 | 93.48 | 5.43 | - | - | - |
| ABA | 1.09 | - | 78.26 | 20.65 | - | - | - |
| AFA | 3.41 | 1.14 | 76.14 | 18.18 | 1.14 | - | - |
| CEH | - | - | 95.00 | 5.00 | - | - | - |
| GUM | 1.16 | - | 90.70 | 8.14 | - | - | - |
| HAH | - | - | 87.50 | 12.50 | - | - | - |
| KEF | - | - | 80.26 | 19.74 | - | - | - |
| LES | 7.29 | - | 80.21 | 12.50 | - | - | - |
| NWH | - | - | 92.22 | 7.78 | - | - | - |
| SES | - | - | 90.70 | 8.14 | 1.16 | - | - |
| WGJ | 2.17 | - | 84.78 | 13.04 | - | - | - |
| BGA | - | - | 94.29 | 5.71 | - | - | - |
| SEA | - | - | 91.03 | 8.97 | - | - | - |
| GUB | 14.13 | - | 85.87 | - | - | - | - |
| TSW | 34.48 | - | 37.93 | 27.59 | - | - | - |
| ALP | 19.35 | - | 58.06 | 22.58 | - | - | - |
| ARD | 18.00 | - | 74.00 | 6.00 | - | - | 2.00 |
| BAL | 6.82 | - | 75.00 | 13.64 | - | - | 4.55 |
| HAI | 17.74 | 1.61 | 64.52 | 14.52 | - | 1.61 | - |
| MON | 16.67 | - | 61.54 | 14.10 | - | 6.41 | 1.28 |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 6. Allele frequency for locus ILSTS11 for all the 20 populations.

| Allele | 261 | 265 | 267 | 269 | 271 | 273 | 275 | 277 | 279 | 281 | 283 | 285 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | - | - | 42.39 | 13.04 | 3.26 | 2.17 | 3.26 | 32.61 | 3.26 | - | - |
| ABA | - | - | - | 46.74 | 4.35 | - | - | 10.87 | 28.26 | 7.61 | 2.17 | - |
| AFA | - | - | - | 15.91 | 6.82 | - | 4.55 | 15.91 | 56.82 | - | - | - |
| CEH | - | - | - | 45.00 | 12.50 | - | 1.25 | 2.50 | 36.25 | 2.50 | - | - |
| GUM | - | - | - | 37.21 | 3.49 | - | 1.16 | 9.30 | 48.84 | - | - | - |
| HAH | - | - | - | 32.50 | 2.50 | - | 1.25 | 17.50 | 42.50 | 3.75 | - | - |
| KEF | - | 1.32 | 1.32 | 50.00 | 15.79 | - | - | 6.58 | 17.11 | 7.89 | - | - |
| LES | - | - | 2.08 | 39.58 | - | - | 1.04 | 21.88 | 34.38 | 1.04 | - | - |
| NWH | - | - | - | 53.33 | 6.67 | - | 5.56 | 4.44 | 28.89 | 1.11 | - | - |
| SES | - | - | - | 29.07 | 5.81 | - | 2.33 | 16.28 | 43.02 | 1.16 | 2.33 | - |
| WGJ | - | - | - | 32.61 | 3.26 | - | - | 27.17 | 35.87 | 1.09 | - | - |
| BGA | - | - | - | 21.43 | 2.86 | - | - | 27.14 | 48.57 | - | - | - |
| SEA | - | - | - | 42.11 | 1.32 | - | 2.63 | 14.47 | 35.53 | 3.95 | - | - |
| GUB | - | - | 8.70 | - | - | - | - | 5.43 | 57.61 | 28.26 | - | - |
| TSW | - | - | - | 17.86 | 5.36 | - | 28.57 | 10.71 | 26.79 | 5.36 | 5.36 | - |
| ALP | - | - | - | 25.81 | 4.84 | - | - | 19.35 | 38.71 | 11.29 | - | - |
| ARD | - | - | 34.00 | 2.00 | 4.00 | - | 6.00 | 6.00 | 48.00 | - | - | - |
| BAL | - | - | - | 27.27 | 4.55 | - | 4.55 | 27.27 | 31.82 | 4.55 | - | - |
| HAI | - | - | 4.84 | 19.35 | 8.06 | 3.23 | 4.84 | 25.81 | 32.26 | - | - | 1.61 |
| MON | 2.56 | - | - | 26.92 | 2.56 | 7.69 | - | 34.62 | 25.64 | - | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 7. Allele frequency for locus ILSTS44 for all the 20 populations.

| Allele | 151 | 153 | 155 | 157 | 159 | 161 | 169 | 173 |
|---|---|---|---|---|---|---|---|---|
| ABG | - | - | 21.74 | 31.52 | 44.57 | 2.17 | - | - |
| ABA | - | - | 34.52 | 50.00 | 11.90 | 3.57 | - | - |
| AFA | - | - | 39.77 | 35.23 | 19.32 | 5.68 | - | - |
| CEH | - | - | 33.75 | 33.75 | 30.00 | 2.50 | - | - |
| GUM | - | 1.16 | 39.53 | 44.19 | 11.63 | 3.49 | - | - |
| HAH | - | - | 38.75 | 35.00 | 18.75 | 7.50 | - | - |
| KEF | - | - | 36.84 | 42.11 | 13.16 | 7.89 | - | - |
| LES | - | 2.08 | 46.88 | 33.33 | 14.58 | 3.13 | - | - |
| NWH | - | - | 37.78 | 36.67 | 22.22 | 3.33 | - | - |
| SES | - | - | 40.70 | 30.23 | 20.93 | 8.14 | - | - |
| WGJ | - | - | 38.04 | 43.48 | 15.22 | 3.26 | - | - |
| BGA | - | - | 41.43 | 40.00 | 11.43 | 7.14 | - | - |
| SEA | 1.28 | - | 29.49 | 43.59 | 16.67 | 8.97 | - | - |
| GUB | - | - | 6.52 | 85.87 | - | 7.61 | - | - |
| TSW | - | 3.45 | 77.59 | 10.34 | - | 1.72 | 6.90 | - |
| ALP | - | 9.68 | 80.65 | 4.84 | 1.61 | - | 3.23 | - |
| ARD | - | - | 70.00 | 28.00 | - | - | 2.00 | - |
| BAL | - | - | 65.91 | 29.55 | 2.27 | - | 2.27 | - |
| HAI | - | - | 72.58 | 11.29 | 9.68 | - | 6.45 | - |
| MON | - | - | 83.33 | 7.69 | 1.28 | - | 1.28 | 6.41 |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 8. Allele frequency for locus ILSTS87 for all the 20 populations.

| Allele | 141 | 143 | 145 | 147 | 149 | 151 | 153 | 155 | 157 | 159 | 161 | 163 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | - | 9.78 | 83.70 | 1.09 | 1.09 | - | - | - | - | 3.26 | 1.09 |
| ABA | - | - | 9.78 | 70.65 | 6.52 | - | - | 1.09 | - | 9.78 | 2.17 | - |
| AFA | - | - | 4.55 | 82.95 | 2.27 | 7.95 | - | - | - | - | 2.27 | - |
| CEH | - | - | 7.50 | 71.25 | 7.50 | - | - | - | - | 8.75 | 5.00 | - |
| GUM | - | - | 8.33 | 80.95 | 7.14 | 1.19 | - | 1.19 | - | 1.19 | - | - |
| HAH | - | - | 13.75 | 76.25 | 2.50 | 3.75 | - | 2.50 | - | 1.25 | - | - |
| KEF | - | - | 3.95 | 78.95 | 3.95 | - | - | - | - | 5.26 | 7.89 | - |
| LES | - | - | 22.92 | 70.83 | 3.13 | - | - | 1.04 | - | - | 2.08 | - |
| NWH | - | - | 3.33 | 76.67 | 7.78 | 3.33 | - | - | - | 6.67 | 1.11 | 1.11 |
| SES | - | - | 16.67 | 69.23 | 6.41 | 2.56 | - | 2.56 | - | 1.28 | 1.28 | - |
| WGJ | - | - | 1.09 | 85.87 | 3.26 | 1.09 | - | - | - | 3.26 | 3.26 | 2.17 |
| BGA | - | - | 14.29 | 80.00 | 1.43 | - | - | 1.43 | - | 2.86 | - | - |
| SEA | - | - | - | 85.90 | 6.41 | 1.28 | 1.28 | - | - | - | 5.13 | - |
| GUB | - | - | 5.43 | 82.61 | 4.35 | - | 7.61 | - | - | - | - | - |
| TSW | - | - | 9.26 | 70.37 | 1.85 | 14.81 | 1.85 | 1.85 | - | - | - | - |
| ALP | - | - | 4.84 | 8.06 | 1.61 | 20.97 | 27.42 | 37.10 | - | - | - | - |
| ARD | - | 4.00 | 2.00 | 46.00 | - | 8.00 | - | 38.00 | 2.00 | - | - | - |
| BAL | - | 13.64 | 11.36 | 47.73 | 11.36 | 2.27 | 2.27 | 11.36 | - | - | - | - |
| HAI | 4.84 | 11.29 | 9.68 | 32.26 | 12.90 | 3.23 | 9.68 | 6.45 | 3.23 | 6.45 | - | - |
| MON | 3.95 | 10.53 | 18.42 | 9.21 | 1.32 | 7.89 | 3.95 | 44.74 | - | - | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 9. Allele frequency for locus INRA05 for all the 20 populations.

| Allele | 131 | 139 | 141 | 143 | 145 | 149 |
|---|---|---|---|---|---|---|
| ABG | - | 60.87 | 29.35 | 1.09 | 8.70 | - |
| ABA | - | 41.30 | 44.57 | 9.78 | 4.35 | - |
| AFA | 1.14 | 42.05 | 53.41 | 3.41 | - | - |
| CEH | - | 43.75 | 50.00 | 6.25 | - | - |
| GUM | - | 58.14 | 27.91 | 12.79 | 1.16 | - |
| HAH | - | 35.00 | 50.00 | 10.00 | 5.00 | - |
| KEF | - | 46.05 | 36.84 | 7.89 | 9.21 | - |
| LES | - | 46.88 | 43.75 | 6.25 | 3.13 | - |
| NWH | - | 51.11 | 37.78 | 6.67 | 4.44 | - |
| SES | - | 51.16 | 37.21 | 3.49 | 8.14 | - |
| WGJ | - | 60.87 | 22.83 | 10.87 | 5.43 | - |
| BGA | - | 61.43 | 32.86 | 4.29 | 1.43 | - |
| SEA | - | 75.64 | 14.10 | 8.97 | 1.28 | - |
| GUB | - | 50.00 | - | 44.57 | 5.43 | - |
| TSW | - | 58.93 | 8.93 | 14.29 | 17.86 | - |
| ALP | - | 67.74 | 11.29 | 17.74 | 3.23 | - |
| ARD | - | 14.00 | 48.00 | 36.00 | 2.00 | - |
| BAL | - | 54.55 | 31.82 | 6.82 | 4.55 | 2.27 |
| HAI | - | 46.77 | 35.48 | 12.90 | 4.84 | - |
| MON | - | 71.79 | 26.92 | 1.28 | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 10. Allele frequency for locus INRA63 for all the 20 populations.

| Allele | 157 | 161 | 163 | 165 | 167 | 169 | 171 | 173 | 175 |
|---|---|---|---|---|---|---|---|---|---|
| ABG | - | - | 6.52 | 8.70 | 54.35 | 25.00 | 5.43 | - | - |
| ABA | - | - | 3.26 | 5.43 | 57.61 | 29.35 | 4.35 | - | - |
| AFA | - | - | 2.27 | 14.77 | 47.73 | 35.23 | - | - | - |
| CEH | - | - | 7.50 | 7.50 | 62.50 | 17.50 | 5.00 | - | - |
| GUM | 9.30 | - | - | 15.12 | 52.33 | 20.93 | 2.33 | - | - |
| HAH | - | - | 3.75 | 8.75 | 51.25 | 33.75 | 1.25 | 1.25 | - |
| KEF | 6.58 | 1.32 | 3.95 | 11.84 | 44.74 | 28.95 | 1.32 | 1.32 | - |
| LES | - | - | 10.42 | 8.33 | 50.00 | 20.83 | 9.38 | 1.04 | - |
| NWH | - | - | 4.44 | 11.11 | 46.67 | 34.44 | 3.33 | - | - |
| SES | - | - | 8.14 | 8.14 | 43.02 | 34.88 | 5.81 | - | - |
| WGJ | - | - | 5.43 | 8.70 | 54.35 | 29.35 | 2.17 | - | - |
| BGA | - | - | 2.86 | 4.29 | 52.86 | 38.57 | 1.43 | - | - |
| SEA | - | - | 15.38 | 11.54 | 39.74 | 28.21 | 5.13 | - | - |
| GUB | - | - | 4.35 | 17.39 | 64.13 | 14.13 | - | - | - |
| TSW | - | - | 10.71 | 28.57 | 12.50 | 41.07 | 7.14 | - | - |
| ALP | - | 8.06 | - | 27.42 | 30.65 | 14.52 | 17.74 | 1.61 | - |
| ARD | - | - | 2.00 | 36.00 | 48.00 | 12.00 | - | - | 2.00 |
| BAL | - | - | 11.36 | 27.27 | 38.64 | 20.45 | - | 2.27 | - |
| HAI | - | - | 11.29 | 29.03 | 25.81 | 24.19 | 9.68 | - | - |
| MON | - | 1.28 | 6.41 | 20.51 | 47.44 | 16.67 | 7.69 | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 11. Allele frequency for locus INRA132 for all the 20 populations.

| Allele | 143 | 145 | 147 | 151 | 153 | 157 |
|---|---|---|---|---|---|---|
| ABG | 60.87 | 36.96 | 2.17 | - | - | - |
| ABA | 52.17 | 44.57 | 3.26 | - | - | - |
| AFA | 40.91 | 56.82 | 2.27 | - | - | - |
| CEH | 63.75 | 31.25 | 5.00 | - | - | - |
| GUM | 46.51 | 38.37 | 15.12 | - | - | - |
| HAH | 50.00 | 47.50 | 2.50 | - | - | - |
| KEF | 59.21 | 36.84 | 3.95 | - | - | - |
| LES | 45.83 | 52.08 | 2.08 | - | - | - |
| NWH | 56.67 | 43.33 | - | - | - | - |
| SES | 50.00 | 50.00 | - | - | - | - |
| WGJ | 41.30 | 54.35 | 4.35 | - | - | - |
| BGA | 40.00 | 58.57 | 1.43 | - | - | - |
| SEA | 30.77 | 65.38 | 3.85 | - | - | - |
| GUB | 45.65 | 38.04 | 16.30 | - | - | - |
| TSW | 61.11 | 24.07 | 9.26 | 1.85 | - | 3.70 |
| ALP | 53.23 | 46.77 | - | - | - | - |
| ARD | 66.00 | 20.00 | 10.00 | - | 4.00 | - |
| BAL | 68.18 | 22.73 | 4.55 | - | - | 4.55 |
| HAI | 67.74 | 32.26 | - | - | - | - |
| MON | 64.71 | 30.88 | 2.94 | - | - | 1.47 |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 12. Allele frequency for locus MAF35 for all the 20 populations.

| Allele | 101 | 105 | 107 | 109 | 115 |
|---|---|---|---|---|---|
| ABG | - | 90.22 | 9.78 | - | - |
| ABA | - | 86.96 | 11.96 | 1.09 | - |
| AFA | - | 93.18 | 6.82 | - | - |
| CEH | - | 93.75 | 6.25 | - | - |
| GUM | - | 93.02 | 6.98 | - | - |
| HAH | - | 93.75 | 5.00 | 1.25 | - |
| KEF | - | 94.74 | 5.26 | - | - |
| LES | - | 85.42 | 14.58 | - | - |
| NWH | - | 90.00 | 10.00 | - | - |
| SES | - | 91.86 | 8.14 | - | - |
| WGJ | - | 90.00 | 10.00 | - | - |
| BGA | - | 88.57 | 11.43 | - | - |
| SEA | - | 93.59 | 6.41 | - | - |
| GUB | - | 98.91 | - | - | 1.09 |
| TSW | - | 89.29 | 8.93 | 1.79 | - |
| ALP | 1.67 | 95.00 | 1.67 | 1.67 | - |
| ARD | - | 88.00 | 12.00 | - | - |
| BAL | - | 90.91 | 9.09 | - | - |
| HAI | - | 91.94 | 8.06 | - | - |
| MON | - | 92.31 | 7.69 | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 13. Allele frequency for locus MAF65 for all the 20 populations.

| Allele | 114 | 116 | 118 | 120 | 122 | 124 | 126 | 128 | 130 | 132 | 134 | 136 | 138 | 140 | 144 | 148 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | 1.09 | 19.57 | 3.26 | 26.09 | - | 27.17 | - | - | 2.17 | 17.39 | 3.26 | - | - | - | - |
| ABA | - | 7.78 | 18.89 | 5.56 | 12.22 | 4.44 | 21.11 | - | 14.44 | 5.56 | 10.00 | - | - | - | - | - |
| AFA | - | 18.18 | 4.55 | 19.32 | 17.05 | 3.41 | 12.50 | - | 2.27 | 15.91 | 6.82 | - | - | - | - | - |
| CEH | - | 13.51 | 9.46 | 6.76 | 21.62 | - | 18.92 | - | 4.05 | 16.22 | 9.46 | - | - | - | - | - |
| GUM | - | 1.32 | 11.84 | 2.63 | 27.63 | 1.32 | 22.37 | - | 1.32 | 17.11 | 5.26 | 6.58 | 2.63 | - | - | - |
| HAH | - | 5.00 | 13.75 | 5.00 | 18.75 | 7.50 | 11.25 | - | 10.00 | 16.25 | 10.00 | 1.25 | 1.25 | - | - | - |
| KEF | - | 1.35 | 4.05 | 18.92 | 14.86 | - | 17.57 | - | 5.41 | 24.32 | 9.46 | 4.05 | - | - | - | - |
| LES | - | 5.32 | 3.19 | 5.32 | 26.60 | 4.26 | 19.15 | - | 11.70 | 11.70 | 9.57 | 3.19 | - | - | - | - |
| NWH | - | 3.41 | 7.95 | 15.91 | 25.00 | - | 17.05 | 1.14 | - | 18.18 | 10.23 | 1.14 | - | - | - | - |
| SES | - | 8.14 | 8.14 | 5.81 | 15.12 | 2.33 | 12.79 | - | 10.47 | 25.58 | 11.63 | - | - | - | - | - |
| WGJ | - | 3.26 | 7.61 | 4.35 | 15.22 | 5.43 | 23.91 | - | 7.61 | 19.57 | 10.87 | - | 2.17 | - | - | - |
| BGA | - | 4.41 | 4.41 | 1.47 | 19.12 | 1.47 | 14.71 | - | 10.29 | 29.41 | 13.24 | 1.47 | - | - | - | - |
| SEA | - | 9.21 | 2.63 | 6.58 | 26.32 | - | 11.84 | 2.63 | 9.21 | 19.74 | 10.53 | - | 1.32 | - | - | - |
| GUB | - | - | - | - | 27.17 | - | 3.26 | - | 1.09 | 32.61 | 29.35 | 4.35 | 1.09 | 1.09 | - | - |
| TSW | - | - | 14.81 | 5.56 | 3.70 | 1.85 | 44.44 | - | - | 14.81 | 7.41 | 7.41 | - | - | - | - |
| ALP | - | - | 8.62 | 1.72 | 41.38 | - | - | - | - | 22.41 | 24.14 | 1.72 | - | - | - | - |
| ARD | - | 8.33 | 2.08 | - | 35.42 | 12.50 | 8.33 | - | 8.33 | - | 18.75 | - | - | - | 4.17 | 2.08 |
| BAL | - | 4.55 | 13.64 | 9.09 | 20.45 | 4.55 | 31.82 | - | - | 6.82 | 9.09 | - | - | - | - | - |
| HAI | - | - | 11.29 | 4.84 | 14.52 | 3.23 | 25.81 | 4.84 | 9.68 | 3.23 | 17.74 | 1.61 | - | 3.23 | - | - |
| MON | 2.86 | - | 24.29 | 42.86 | 10.00 | - | 7.14 | 1.43 | 4.29 | 7.14 | - | - | - | - | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 14. Allele frequency for locus MAF209 for all the 20 populations.

| Allele | 102 | 104 | 106 | 108 |
|---|---|---|---|---|
| ABG | 30.43 | 57.61 | 11.96 | - |
| ABA | 22.83 | 51.09 | 26.09 | - |
| AFA | 14.77 | 70.45 | 14.77 | - |
| CEH | 27.50 | 57.50 | 15.00 | - |
| GUM | 30.23 | 65.12 | 4.65 | - |
| HAH | 26.25 | 60.00 | 12.50 | 1.25 |
| KEF | 38.16 | 47.37 | 10.53 | 3.95 |
| LES | 16.67 | 60.42 | 22.92 | - |
| NWH | 23.33 | 63.33 | 13.33 | - |
| SES | 17.44 | 69.77 | 12.79 | - |
| WGJ | 20.65 | 50.00 | 29.35 | - |
| BGA | 8.57 | 70.00 | 21.43 | - |
| SEA | 14.10 | 37.18 | 48.72 | - |
| GUB | 2.17 | 63.04 | 34.78 | - |
| TSW | 13.79 | 65.52 | 20.69 | - |
| ALP | 9.68 | 88.71 | 1.61 | - |
| ARD | 20.83 | 62.50 | 16.67 | - |
| BAL | 34.09 | 65.91 | - | - |
| HAI | 17.74 | 77.42 | 4.84 | - |
| MON | 20.51 | 79.49 | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 15. Allele frequency for locus OarAE129 for all the 20 populations.

| Allele | 139 | 141 | 143 | 145 | 147 | 149 | 151 | 153 | 155 | 157 | 159 | 161 | 163 | 165 | 167 | 169 | 171 | 173 | 175 | 177 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | 27.17 | 14.13 | - | - | - | - | - | - | 14.13 | 29.35 | 8.70 | - | - | - | - | 3.26 | 1.09 | - | 2.17 | - |
| ABA | 35.23 | 20.45 | - | - | - | - | - | - | 21.59 | 12.50 | 5.68 | - | - | 1.14 | - | 3.41 | - | - | - | - |
| AFA | 12.50 | 9.09 | - | - | - | 1.14 | - | 3.41 | 18.18 | 25.00 | 10.23 | - | - | 3.41 | 1.14 | 13.64 | 2.27 | - | - | - |
| CEH | 35.94 | 15.63 | - | - | - | - | - | 1.56 | 7.81 | 20.31 | 14.06 | - | - | - | - | 1.56 | - | 1.56 | 1.56 | - |
| GUM | 30.23 | 13.95 | - | - | - | 2.33 | - | - | 9.30 | 13.95 | 5.81 | - | - | 15.12 | 2.33 | 4.65 | - | 1.16 | 1.16 | - |
| HAH | 34.21 | 13.16 | - | - | - | - | - | - | 13.16 | 22.37 | 7.89 | - | - | 2.63 | - | 2.63 | 2.63 | 1.32 | - | - |
| KEF | 28.95 | 22.37 | - | - | 2.63 | - | - | 1.32 | 3.95 | 22.37 | 5.26 | - | 2.63 | 1.32 | 2.63 | 3.95 | - | - | 2.63 | - |
| LES | 22.34 | 13.83 | - | 1.06 | - | - | 2.13 | 2.13 | 20.21 | 22.34 | 3.19 | - | 1.06 | 2.13 | 2.13 | 5.32 | - | - | 1.06 | 1.06 |
| NWH | 30.00 | 21.11 | 1.11 | 5.56 | - | - | - | 1.11 | 10.00 | 16.67 | 8.89 | 1.11 | - | - | 1.11 | 1.11 | - | - | 2.22 | - |
| SES | 25.00 | 17.86 | - | - | - | 1.19 | - | 1.19 | 16.67 | 15.48 | 10.71 | - | - | - | 1.19 | 8.33 | - | - | 2.38 | - |
| WGJ | 27.17 | 17.39 | - | - | - | - | - | 3.26 | 17.39 | 18.48 | 2.17 | - | - | 2.17 | 4.35 | 3.26 | 1.09 | - | 3.26 | - |
| BGA | 17.24 | 22.41 | - | - | - | - | - | 1.72 | 13.79 | 25.86 | 5.17 | - | - | 3.45 | 3.45 | 5.17 | - | 1.72 | - | - |
| SEA | 22.97 | 5.41 | - | - | - | - | - | 2.70 | 31.08 | 27.03 | 5.41 | - | - | 5.41 | - | - | - | - | - | - |
| GUB | - | - | - | - | - | - | - | - | 40.91 | 26.14 | 5.68 | - | - | 1.14 | 18.18 | 6.82 | - | 1.14 | - | - |
| TSW | 62.07 | 3.45 | - | - | - | - | - | 1.72 | 12.07 | 18.97 | - | - | - | - | 1.72 | - | - | - | - | - |
| ALP | 46.77 | - | - | - | - | - | - | 11.29 | 17.74 | 6.45 | - | - | 1.61 | 1.61 | 14.52 | - | - | - | - | - |
| ARD | 4.00 | 2.00 | - | - | - | - | - | - | 48.00 | - | 6.00 | - | - | 4.00 | 14.00 | 14.00 | 8.00 | - | - | - |
| BAL | 2.27 | 2.27 | - | - | - | 9.09 | - | - | 34.09 | 27.27 | - | - | - | 4.55 | 18.18 | 2.27 | - | - | - | - |
| HAI | 12.90 | 1.61 | - | - | - | 3.23 | - | 6.45 | 30.65 | 17.74 | 1.61 | - | - | - | 24.19 | 1.61 | - | - | - | - |
| MON | 19.74 | 7.89 | - | - | - | 14.47 | - | - | 22.37 | 9.21 | 3.95 | 15.79 | - | - | 5.26 | 1.32 | - | - | - | - |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 16. Allele frequency for locus OarFCB304 for all the 20 populations.

| Allele | 120 | 124 | 128 | 132 | 134 | 136 | 138 | 140 | 142 | 144 | 146 | 148 | 150 | 154 | 156 | 158 | 160 | 162 | 164 | 166 | 168 | 170 | 174 | 178 | 180 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | 2.17 | - | - | - | 2.17 | - | 45.65 | 9.78 | 5.43 | 14.13 | - | - | 13.04 | - | 2.17 | - | 2.17 | 2.17 | - | 1.09 | - | - | - | - |
| ABA | 1.09 | - | - | - | - | 3.26 | - | 50.00 | 14.13 | 3.26 | 7.61 | - | 3.26 | 6.52 | - | - | 2.17 | 8.70 | - | - | - | - | - | - | - |
| AFA | - | 1.14 | - | - | - | 3.41 | - | 50.00 | 14.77 | 3.41 | 10.23 | - | - | 12.50 | - | 3.41 | - | - | - | 1.14 | - | - | - | - | - |
| CEH | - | 3.75 | - | - | - | 5.00 | - | 47.50 | 8.75 | 1.25 | 12.50 | - | 2.50 | 16.25 | - | 2.50 | - | - | - | - | - | - | - | - | - |
| GUM | - | - | - | - | 1.16 | 3.49 | 1.16 | 59.30 | 13.95 | 4.65 | 1.16 | - | - | 11.63 | - | - | - | 3.49 | - | - | - | - | - | - | - |
| HAH | - | 1.25 | 1.25 | - | - | 3.75 | - | 48.75 | 5.00 | 1.25 | 13.75 | - | - | 6.25 | 1.25 | 12.50 | 1.25 | - | - | - | - | 2.50 | - | - | 1.25 |
| KEF | - | - | - | - | - | - | - | 50.00 | 22.37 | 3.95 | 2.63 | - | 2.63 | 9.21 | - | 7.89 | 1.32 | - | - | - | - | - | - | - | - |
| LES | - | - | - | - | - | 4.17 | - | 62.50 | 9.38 | 1.04 | - | - | - | 6.25 | 1.04 | 3.13 | 2.08 | 2.08 | 1.04 | 1.04 | 4.17 | 1.04 | 1.04 | - | - |
| NWH | - | 1.11 | - | - | - | 2.22 | - | 50.00 | 11.11 | 8.89 | 3.33 | - | - | 21.11 | - | 2.22 | - | - | - | - | - | - | - | - | - |
| SES | - | - | - | 1.16 | - | 4.65 | - | 60.47 | 9.30 | 1.16 | 5.81 | - | - | 8.14 | - | 5.81 | 2.33 | - | - | 1.16 | - | - | - | - | - |
| WGJ | - | - | - | - | - | 4.35 | - | 47.83 | 16.30 | 5.43 | 8.70 | - | - | 8.70 | - | - | 2.17 | - | 3.26 | 1.09 | - | - | 2.17 | - | - |
| BGA | - | - | - | - | - | - | - | 62.86 | 4.29 | 4.29 | 5.71 | - | - | 14.29 | - | - | 2.86 | 1.43 | - | 4.29 | - | - | - | - | - |
| SEA | - | - | - | - | - | 2.56 |  | 51.28 | 15.38 | 5.13 | - | - | - | 8.97 | - | 5.13 | 7.69 | - | 3.85 | - | - | - | - | - | - |
| GUB | - | - | - | 6.52 | - | - | 8.70 | 26.09 | 3.26 | 3.26 | - | - | - | - | 3.26 | 10.87 | 30.43 | 4.35 | 1.09 | 2.17 | - | - | - | - | - |
| TSW | - | - | - | 3.45 | - | 32.76 | 1.72 | 36.21 | 18.97 | - | - | - | - | - | - | - | 3.45 | - | - | - | - | 3.45 | - | - | - |
| ALP | - | - | - | - | - | 12.90 | 11.29 | 50.00 | 12.90 | - | - | - | - | 9.68 | 3.23 | - | - | - | - | - | - | - | - | - | - |
| ARD | - | - | 2.00 | 4.00 | 4.00 | 2.00 | 2.00 | 66.00 | 8.00 | 2.00 | 2.00 | - | - | 4.00 | 2.00 | - | - | - | - | - | 2.00 | - | - | - | - |
| BAL | - | - | - | 4.55 | - | 22.73 | 6.82 | 27.27 | 6.82 | 4.55 | - | - | - | 11.36 | - | - | - | 2.27 | 4.55 | 2.27 | 2.27 | - | - | 4.55 | - |
| HAI | - | - | 3.23 | 8.06 | - | 20.97 | 1.61 | 35.48 | 8.06 | - | - | 1.61 | - | 4.84 | - | - | - | - | 4.84 | 8.06 | 3.23 | - | - | - | - |
| MON | - | - | 6.41 | 1.28 | - | 7.69 | 14.10 | 37.18 | 1.28 | - | - | - | 1.28 | 1.28 | - | - | - | 10.26 | 12.82 | 1.28 | 1.28 | - | - | - | 3.85 |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 17. Allele frequency for locus SRCRSP03 for all the 20 populations.

| Allele | 111 | 115 | 117 | 119 | 121 | 123 | 125 |
|---|---|---|---|---|---|---|---|
| ABG | 27.78 | - | - | 13.33 | 30.00 | 28.89 | - |
| ABA | 38.04 | - | - | 4.35 | 41.30 | 16.30 | - |
| AFA | 19.32 | - | - | 17.05 | 53.41 | 10.23 | - |
| CEH | 18.75 | - | - | 18.75 | 46.25 | 16.25 | - |
| GUM | 52.33 | - | - | 10.47 | 30.23 | 6.98 | - |
| HAH | 19.23 | - | 2.56 | 15.38 | 53.85 | 7.69 | 1.28 |
| KEF | 36.84 | - | - | 1.32 | 52.63 | 9.21 | - |
| LES | 40.63 | - | - | 5.21 | 44.79 | 8.33 | 1.04 |
| NWH | 27.78 | - | - | 8.89 | 46.67 | 16.67 | - |
| SES | 26.74 | - | 1.16 | 17.44 | 47.67 | 6.98 | - |
| WGJ | 45.65 | - | - | 3.26 | 44.57 | 6.52 | - |
| BGA | 24.29 | 2.86 | - | 8.57 | 51.43 | 12.86 | - |
| SEA | 20.51 | - | - | 8.97 | 55.13 | 15.38 | - |
| GUB | 31.52 | - | - | 15.22 | 53.26 | - | - |
| TSW | 5.17 | - | 3.45 | 10.34 | 81.03 | - | - |
| ALP | 19.35 | - | - | 6.45 | 70.97 | 3.23 | - |
| ARD | 14.00 | - | - | 16.00 | 60.00 | 10.00 | - |
| BAL | 22.73 | - | - | 13.64 | 59.09 | 4.55 | - |
| HAI | 9.68 | - | - | 16.13 | 62.90 | 6.45 | 4.84 |
| MON | 1.28 | - | - | 24.36 | 64.10 | 1.28 | 8.97 |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-Ear Somali, NWH – North-West Highland, SES – Small-Ear Somali, WGJ – Woyto-Guji, BGA– Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW – Tswana, ALP – Italian Alpine, ARD- Ardi, BAL – Baladie, HAI- Hair, MON– Mongolia.

Appendix 18.: Test for genetic differentiation between pairs of 20 populations for each locus.

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABA & ABG | ns | * | ** | ** | *** | ** | ** | ns | ns | ns | *** | ns | ns | * | * |
| AFA&ABG | ns | ** | ** | *** | ** | ns | *** | ns | * | ns | *** | ns | ** | ns | ** |
| AFA &ABA | ** | * | ns | *** | ns | *** | ns | * | ns | ns | *** | * | *** | * | ** |
| CEH &ABG | ns | ns | ns | ns | ns | ** | *** | ns | ns | ns | *** | ns | ns | ns | * |
| CEH &ABA | ns | | ** | * | * | ns | ns | ns | ns | ns | * | ns | ns | ** | ** |
| CEH &AFA | ns | *** | ** | *** | ns | *** | ns | * | * | ns | ns | ns | ** | ns | ns |
| GUM &ABG | ns | * | ns | ** | *** | ns | ** | ** | ** | ns | * | ns | *** | * | *** |
| GUM &ABA | ns | ** | ns | ** | ns | ns | ns | *** | * | ns | *** | *** | ** | ns | * |
| GUM & AFA | * | ns | ns | * | ns | ns | ** | ** | ** | ns | *** | ** | *** | * | *** |
| GUM & CEH | ns | * | ns | * | ns | * | * | ** | * | ns | * | ns | * | ** | *** |
| HAH & ABG | ns | ns | ns | ** | ** | ns | ** | ns | ns | ns | *** | ns | ns | * | *** |
| HAH & ABA | * | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | *** | ** |
| HAH & AFA | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ** | ns | * | ns | ns |
| HAH & CEH | ns | * | ns | ** | ns | ** | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| HAH & GUM | ns | * | ns | ns | ns | ns | ** | ** | ** | ns | ns | ns | ns | *** | *** |
| KEF & ABG | ns | ns | ** | ns | *** | * | ns | ns | ns | ns | *** | ns | ns | ** | *** |
| KEF & ABA | ** | ** | ns | * | ns | ns | | * | ns | ns | *** | ** | * | ** | ns |
| KEF & AFA | ns | * | ns | *** | ns | ** | ** | ns | ns | ns | * | *** | *** | ns | *** |
| KEF & CEH | ns | ** | * | ns | ns | ns | * | * | ns | ns | * | ns | ns | ** | *** |
| KEF & GUM | ns | ** | * | *** | ns | * | * | ns | * | ns | * | * | * | ** | ** |
| KEF & HAH | ns | * | | *** | ns | ** | ns | ns | ns | ns | ns | ns | ns | ** | ** |
| LES & ABG | ns | ns | ** | *** | *** | ns | * | ns | ns | ns | *** | * | ns | ** | *** |
| LES & ABA | ns | ns | * | ** | ns | ** | ns | ns | ns | ns | * | ns | ns | ** | ns |
| LES & AFA | * | ns | ns | *** | ns | *** | ns | *** | ns | ns | ** | ns | ns | * | ** |
| LES & CEH | ns | * | ** | *** | ns | ns | ns | ns | ns | ns | ns | ns | ns | *** | ** |
| LES & GUM | ns | ns | ns | * | ns | * | ns | *** | ** | ns | * | ** | * | ns | ns |
| LES & HAH | ns | ns | * | ns | ns | *** | ns | ns | ns | * | ns | ns | ns | ** | * |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LES & KEF | ns | ns | * | *** | ns | *** | ns | * | ns | ns | * | ** | ns | * | ns |
| NWH & ABG | ns | * | ns | ns | * | ** | ns | ns | ns | ns | *** | ns | ns | ns | ns |
| NWH & ABA | ns | | * | ** | ns | ns | ns | ns | ns | ns | *** | ns | ns | *** | ns |
| NWH & AFA | *** | ** | * | *** | ns | ns | * | ns | * | ns | * | ns | *** | ns | ns |
| NWH & CEH | ns | ns | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns |
| NWH & GUM | * | ** | ns | * | ns | ns | ns | ** | *** | ns | ns | ns | ** | ns | ** |
| NWH & HAH | * | ns | ns | ** | ns | * | ns | ns | ns | ns | ** | ns | ns | *** | ns |
| NWH & KEF | * | *** | * | * | ns | ns | ns | ns | ns | ns | ns | * | ns | * | ns |
| NWH & LES | ns | * | ** | *** | ns | *** | ns | ns | ns | ns | *** | ns | ns | ** | ns |
| SES & ABG | ns | *** | ns | ** | ** | ns | ns | ns | ns | ns | *** | ns | ns | * | *** |
| SES & ABA | ns | ns | * | * | * | ns | ns | ns | ns | ns | * | * | ns | ** | ** |
| SES & AFA | ns | ns | ns | ns | ns | * | * | * | ns | ns | * | ns | ns | ns | ns |
| SES & CEH | ns | *** | ns | ** | ns | * | * | ns | * | ns | ns | ns | ns | ns | ns |
| SES & GUM | ns | ** | ns | ns | ns | ns | * | *** | *** | ns | ** | ns | ** | * | ** |
| SES & HAH | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| SES & KEF | ns | ** | * | *** | ns | ** | ns | ns | * | ns | * | ** | ns | ns | ** |
| SES & LES | ns | ns | * | * | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | * |
| SES & NWH | ns | ** | ns | ** | ns | * | ns | ns | ns | ns | ** | ns | ns | * | ns |
| WGJ & ABG | ns | ns | ns | *** | *** | ns | * | ns | ns | ns | *** | * | ns | ns | *** |
| WGJ & ABA | ns | ns | ns | ** | ns | * | * | ns | ns | ns | * | ns | ns | * | ns |
| WGJ & AFA | ns | ns | ns | ** | ns | ns | *** | ns | ns | ns | ** | * | * | ns | *** |
| WGJ & CEH | ns | ** | ns | *** | ns | * | ** | ns | * | ns | ns | ns | ns | * | *** |
| WGJ & GUM | * | ** | ns | * | ns | ns | ns | ** | * | ns | ns | *** | * | * | ns |
| WGJ & HAH | ns | ns | ns | ns | ns | ** | ** | ns | ns | ns | ns | * | ns | *** | ** |
| WGJ & KEF | * | ns | ns | *** | ns | ns | ns | ns | ns | ns | * | ** | ns | ** | ns |
| WGJ & LES | ns | ns | ns | ns | ns | *** | * | ns | ns | ns | ns | ns | ns | ** | ns |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WGJ & NWH | ns | * | ns | *** | ns | ns | ns | ns | * | ns | ** | * | ns | * | * |
| WGJ & SES | ns | ns | ns | ns | ns | ** | ns | ns | ns | ns | ns | * | ns | ns | ** |
| BGA & ABG | ** | ** | ns | *** | *** | ns | ns | ns | * | ns | *** | ** | ns | * | * |
| BGA & ABA | *** | ns | * | *** | ns | ns | ns | ns | ns | ns | ** | * | ns | ** | ns |
| BGA & AFA | ** | ns | * | ns | ns | ** | ns | ns | ns | ns | *** | ns | ns | * | ns |
| BGA & CEH | *** | ** | ns | *** | ns | * | ns | ns | ** | ns | ns | ** | ns | ** | ns |
| BGA & GUM | * | * | ns | * | ns | ns | ns | *** | *** | ns | ns | *** | ns | * | ** |
| BGA & HAH | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ns | * | ns | *** | ns |
| BGA & KEF | ** | ** | * | *** | ns | * | ns | ns | * | ns | ns | *** | ns | *** | ns |
| BGA & LES | ns | ns | * | * | ns | ns | ns | * | ns | ns | ns | ns | ns | * | ns |
| BGA & NWH | *** | * | ns | *** | ns | * | ns | ns | * | ns | ** | * | ns | * | ns |
| BGA & SES | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| BGA & WGJ | ** | ns | ns | ns | ns | ** | ns | ns | ns | ns | ns | * | ns | * | * |
| SEA &ABG | ** | ns | ns | ** | ** | ** | *** | ns | *** | ns | *** | *** | ** | ** | * |
| SEA & ABA | *** | *** | ns | ns | ns | *** | *** | * | * | ns | *** | * | ** | *** | ns |
| SEA & AFA | ** | * | * | *** | ns | * | *** | ** | ns | ns | * | *** | ** | ** | ns |
| SEA & CEH | *** | *** | ns | ** | ns | ** | *** | ns | *** | ns | ns | *** | *** | *** | ns |
| SEA & GUM | *** | * | ns | ns | ns | ** |  | *** | ** | ns | ** | *** | *** | * | *** |
| SEA & HAH | ns | ** | ns | ns | ns | *** | *** | ns | ns | ns | ns | *** | * | *** | ns |
| SEA & KEF | ** | ns | ns | ** | ns | * | *** | * | ** | ns | * | *** | *** | ns | * |
| SEA & LES | ns | ns | * | ns | ns | *** | *** | ns | ns | ns | ns | ** | ns | ns | * |
| SEA & NWH | *** | *** | ns | ns | ns | * | ** | ns | ** | ns | * | *** | *** | ** | ns |
| SEA & SES | ns | *** | ns | ns | ns | *** | *** | ns | * | ns | ns | *** | *** | ns | ns |
| SEA & WGJ | ** | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | * | * | * | ** |
| SEA & BGA | * | ** | ns | ** | ns | *** | * | * | ns | ns | ns | *** | * | ** | ns |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GUB & ABG | *** | *** | *** | *** | *** | ** | *** | * | ** | ** | *** | *** | *** | *** | *** |
| GUB & ABA | *** | *** | *** | *** | *** | *** | *** | ** | ** | *** | *** | *** | *** | *** | *** |
| GUB & AFA | *** | *** | *** | *** | *** | *** | *** | * | ** | * | *** | *** | *** | *** | ** |
| GUB & CEH | ** | *** | *** | *** | *** | *** | *** | ns | * | * | *** | *** | *** | *** | *** |
| GUB & GUM | *** | *** | *** | *** | *** | * | *** | *** | ns | * | *** | *** | *** | *** | *** |
| GUB & HAH | *** | *** | *** | *** | *** | ** | *** | * | ** | * | *** | *** | *** | *** | * |
| GUB & KEF | *** | *** | *** | *** | *** | *** | *** | ** | * | * | *** | *** | *** | *** | *** |
| GUB & LES | *** | *** | *** | *** | *** | *** | *** | *** | *** | *** | *** | ** | *** | *** | ** |
| GUB & NWH | *** | *** | *** | *** | *** | *** | *** | ** | *** | ** | *** | *** | *** | *** | *** |
| GUB & SES | *** | *** | *** | *** | *** | *** | *** | *** | *** | ** | *** | *** | *** | *** | * |
| GUB & WGJ | *** | *** | *** | *** | *** | *** | *** | * | ** | *** | *** | *** | *** | *** | *** |
| GUB & BGA | *** | *** | *** | *** | *** | ** | *** | *** | *** | *** | *** | ns | *** | *** | *** |
| GUB & SEA | *** | *** | *** | *** | *** | ** | *** | *** | *** | * | *** | *** | *** | *** | *** |
| TSW & ABG | * | *** | *** | *** | *** | ** | *** | *** | * | ns | *** | ns | *** | *** | *** |
| TSW & ABA | *** | ** | *** | *** | *** | *** | *** | *** | * | ns | *** | ns | *** | *** | *** |
| TSW & AFA | ** | *** | *** | *** | *** | ns | *** | *** | ** | ns | *** | ns | *** | *** | *** |
| TSW & CEH | ** | *** | *** | *** | *** | *** | *** | *** | ns | ns | *** | ns | ** | *** | *** |
| TSW & GUM | ** | *** | *** | *** | *** | * | *** | *** | * | ns | ns | ** | *** | *** | *** |
| TSW & HAH | ns | ** | *** | *** | *** | ns | *** | *** | ** | ns | *** | ns | ** | *** | * |
| TSW & KEF | ** | *** | *** | *** | *** | *** | ** | *** | ns | ns | ** | ** | *** | *** | *** |
| TSW & LES | ns | *** | *** | *** | *** | *** | *** | *** | ** | ns | *** | ns | ** | *** | *** |
| TSW & NWH | *** | ** | *** | *** | *** | ** | *** | ** | *** | ns | *** | ns | *** | *** | *** |
| TSW & SES | ns | *** | *** | *** | *** | ns | *** | ** | *** | ns | *** | ns | *** | *** | *** |
| TSW & WGJ | ** | *** | *** | *** | *** | *** | * | *** | ** | ns | ** | ns | ** | *** | *** |
| TSW & BGA | ns | *** | *** | *** | *** | ** | *** | *** | *** | ns | *** | ns | *** | *** | *** |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TSW & SEA | * | *** | *** | *** | *** | *** | ** | ** | *** | | *** | ** | *** | *** | *** |
| TSW & GUB | *** | *** | *** | *** | *** | *** | *** | *** | * | ** | *** | * | *** | *** | *** |
| ALP & ABG | *** | ** | *** | *** | *** | *** | *** | *** | ns | * | *** | *** | *** | *** | *** |
| ALP & ABA | *** | ns | *** | ns | *** | *** | *** | *** | ns | * | *** | *** | *** | *** | ** |
| ALP & AFA | *** | ** | ** | ** | *** | *** | *** | *** | ns | ns | *** | ** | *** | *** | |
| ALP & CEH | *** | ** | *** | *** | *** | *** | *** | *** | ns | ns | *** | *** | *** | *** | ** |
| ALP & GUM | *** | *** | *** | ** | *** | *** | ns | *** | *** | ns | *** | ** | *** | ** | *** |
| ALP & HAH | *** | | *** | ns | *** | *** | *** | *** | ns | ns | *** | ** | *** | *** | ns |
| ALP & KEF | *** | ** | *** | ** | *** | *** | ** | *** | ns | ns | *** | *** | *** | *** | * |
| ALP & LES | *** | * | * | ** | *** | *** | *** | *** | ns | ** | *** | *** | *** | ** | * |
| ALP & NWH | *** | ns | *** | *** | *** | *** | ** | *** | ns | * | *** | ** | *** | *** | * |
| ALP & SES | *** | ns | *** | Ns | *** | *** | *** | *** | ns | ns | *** | * | *** | *** | * |
| ALP & WGJ | *** | ns | *** | ns | *** | *** | ns | *** | ns | * | *** | *** | *** | *** | * |
| ALP & BGA | ** | ns | *** | * | *** | *** | ** | *** | ns | * | *** | ** | *** | *** | ns |
| ALP & SEA | *** | *** | *** | ns | *** | *** | ns | *** | * | | *** | *** | *** | *** | ns |
| ALP & GUB | *** | *** | *** | *** | *** | *** | *** | *** | *** | ns | * | *** | *** | *** | * |
| ALP & TSW | ns | ns | ns | *** | ns | *** | ns | *** | ** | ns | *** | ** | ** | *** | * |
| ARD & ABG | ns | *** | *** | *** | *** | *** | *** | *** | * | ns | *** | ns | *** | ** | ** |
| ARD & ABA | * | ns | *** | *** | *** | *** | *** | *** | ** | ns | *** | ns | *** | ** | ** |
| ARD & AFA | ns | * | ** | *** | *** | *** | *** | ** | *** | ns | *** | ns | *** | * | ns |
| ARD & CEH | ** | ** | *** | *** | *** | *** | *** | ** | ns | ns | *** | ns | *** | ** | ns |
| ARD & GUM | ns | *** | *** | *** | *** | *** | *** | *** | * | ns | *** | ns | *** | ns | *** |
| ARD & HAH | ns | ns | *** | *** | *** | *** | ** | *** | ** | ns | *** | ns | *** | ** | ns |
| ARD & KEF | ns | *** | *** | *** | *** | *** | *** | ** | ns | ns | *** | ns | *** | ** | ** |
| ARD & LES | ns | * | ns | *** | *** | *** | *** | *** | *** | ns | *** | ns | *** | ns | ** |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ARD & NWH | *** | * | *** | *** | *** | *** | *** | *** | *** | ns | *** | ns | *** | ** | ns |
| ARD & SES | ns | ns | *** | *** | *** | *** | *** | *** | *** | ns | *** | ns | *** | ns | ns |
| ARD & WGJ | * | ** | ** | *** | *** | *** | *** | *** | *** | ns | *** | ns | *** | ** | *** |
| ARD & BGA | ns | ns | *** | *** | *** | *** | *** | *** | *** | ns | *** | ns | *** | * | ns |
| ARD & SEA | ** | *** | *** | *** | *** | *** | *** | *** | *** | ns | *** | *** | *** | ** | ns |
| ARD & GUB | *** | *** | * | *** | *** | *** | *** | ns | * | * | *** | *** | *** | *** | ** |
| ARD & TSW | ns | *** | *** | *** | ns | *** | *** | *** | ns | ns | *** | ns | *** | *** | ** |
| ARD & ALP | ** | ns | * | *** | *** | *** | *** | *** | *** | ns | *** | ** | *** | * | ns |
| BAL & ABG | *** | ** | ** | ** | *** | *** | ns | * | ns | ns | ns | * | *** | *** | ** |
| BAL & ABA | *** | ns | ns | * | ** | *** | ns | ** | * | ns | ns | *** | *** | *** | * |
| BAL & AFA | *** | ** | ns | * | ** | *** | * | * | ** | ns | * | *** | *** | *** | ns |
| BAL & CEH | *** | ns | ** | ** | *** | *** | ns | * | ns | ns | ns | ** | *** | *** | ns |
| BAL & GUM | *** | *** | * | * | * | *** | ns | *** | ** | ns | ns | ns | *** | *** | ** |
| BAL & HAH | ** | ns | * | ns | ** | *** | ns | * | ** | ns | * | * | *** | *** | ns |
| BAL & KEF | *** | ** | * | ** | ** | *** | ns | ns | ns | ns | * | * | *** | *** | * |
| BAL & LES | ns | ns | ns | ns | * | *** | ns | * | ** | ns | * | *** | *** | *** | ns |
| BAL & NWH | *** | ns | ** | ** | *** | *** | ns | * | ** | ns | ns | ** | *** | *** | ns |
| BAL & SES | ns | ns | ** | ns | *** | ** | ns | * | *** | ns | * | ** | *** | *** | ns |
| BAL & WGJ | *** | ns | ns | ns | ** | *** | ns | * | ** | ns | ns | *** | *** | *** | * |
| BAL & BGA | ns | ns | ** | ns | * | *** | ns | ** | *** | ns | ** | *** | *** | *** | ns |
| BAL & SEA | * | *** | ** | ns | *** | *** | ns | ns | *** | ns | ** | *** | *** | *** | ns |
| BAL & GUB | *** | *** | *** | *** | *** | *** | *** | ns | ** | ** | *** | *** | * | *** | ns |
| BAL & TSW | ** | ns | *** | * | ns | ** | ** | * | ns | ns | ns | *** | *** | ** | ** |
| BAL & ALP | *** | ns | * | ns | ** | *** | * | *** | ** | ns | *** | ** | *** | ** | ns |
| BAL & ARD | ** | ns | ns | *** | ns | ** | *** | ns | ns | ns | *** | ** | *** | ** | ns |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HAI & ABG | *** | ** | *** | *** | *** | *** | * | ** | ns | ns | ** | * | *** | *** | *** |
| HAI & ABA | *** | ** | *** | *** | *** | *** | ns | *** | ns | ns | * | *** | *** | *** | *** |
| HAI & AFA | *** | *** | * | * | *** | *** | * | *** | ** | ns | *** | ns | *** | *** | ns |
| HAI & CEH | *** | *** | *** | *** | *** | *** | ns | *** | ns | ns | *** | * | *** | *** | * |
| HAI & GUM | *** | *** | *** | *** | *** | *** | ns | *** | *** | ns | ** | ns | *** | *** | *** |
| HAI & HAH | *** | * | *** | * | *** | *** | ns | *** | * | ns | * | ns | *** | *** | ns |
| HAI & KEF | *** | * | *** | *** | *** | *** | ns | ** | ns | ns | *** | ** | *** | *** | *** |
| HAI & LES | ** | *** | | ** | *** | *** | ns | ** | * | ns | ** | ** | *** | *** | *** |
| HAI & NWH | *** | *** | *** | *** | *** | *** | ns | ** | ns | ns | *** | ns | *** | *** | ** |
| HAI & SES | * | *** | *** | * | *** | *** | ns | * | ns | ns | *** | ns | *** | *** | * |
| HAI & WGJ | *** | * | ** | ** | *** | *** | ns | *** | ** | ns | * | *** | *** | *** | *** |
| HAI & BGA | ns | *** | *** | * | *** | *** | ns | *** | ** | ns | ** | * | *** | *** | * |
| HAI & SEA | ** | *** | *** | ** | *** | *** | ** | ns | *** | ns | *** | *** | *** | *** | * |
| HAI & GUB | *** | *** | *** | *** | *** | *** | *** | *** | *** | ** | *** | *** | *** | *** | *** |
| HAI & TSW | ** | ns | ** | *** | ns | *** | ** | ns | * | ns | ** | * | *** | ** | * |
| HAI & ALP | *** | * | ns | ** | * | *** | * | ** | ns | ns | *** | ns | ** | *** | ns |
| HAI & ARD | *** | *** | ns | *** | * | *** | *** | ** | * | ns | *** | ns | *** | ** | ns |
| HAI & BAL | ns | * | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| MON & ABG | *** | ns | *** | *** | *** | *** | * | ns | ns | ns | *** | *** | *** | *** | *** |
| MON & ABA | *** | ns | *** | *** | *** | *** | *** | * | ns | ns | *** | *** | *** | *** | *** |
| MON & AFA | *** | ns | ** | *** | *** | *** | ** | ** | * | ns | *** | *** | *** | *** | *** |
| MON & CEH | *** | ns | *** | *** | *** | *** | ** | ns | ns | ns | *** | *** | *** | *** | *** |
| MON & GUM | *** | ns | *** | *** | *** | *** | ** | ** | * | ns | *** | * | *** | *** | *** |
| MON & HAH | *** | ns | *** | ** | *** | *** | *** | * | ns | ns | *** | ** | *** | *** | *** |
| MON & KEF | *** | ns | *** | *** | *** | *** | *** | * | ns | ns | *** | *** | *** | *** | *** |
| MON & LES | *** | ns | ** | *** | *** | *** | ** | ns | ns | ns | *** | *** | *** | *** | *** |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MON & NWH | *** | ns | *** | *** | *** | *** | ** | ns | ns | ns | *** | *** | *** | *** | *** |
| MON & SES | *** | ns | *** | *** | *** | *** | ** | ns | * | ns | *** | ** | *** | *** | *** |
| MON & WGJ | *** | ns | *** | * | *** | *** | ** | * | * | ns | *** | *** | *** | *** | *** |
| MON & BGA | *** | ns | *** | ** | *** | *** | | ** | * | ns | *** | *** | *** | *** | *** |
| MON & SEA | *** | * | *** | *** | *** | *** | * | ns | ** | ns | *** | *** | *** | *** | *** |
| MON & GUB | *** | *** | *** | *** | *** | *** | *** | * | ** | ** | *** | *** | *** | *** | *** |
| MON & TSW | *** | *** | ** | *** | ns | *** | *** | ** | ns | ns | *** | *** | *** | *** | ** |
| MON & ALP | *** | ns | ns | *** | * | *** | *** | * | ns | ns | *** | ns | *** | *** | *** |
| MON & ARD | *** | ns | ns | *** | * | *** | *** | * | ns | ns | *** | *** | *** | *** | *** |
| MON & BAL | *** | ns | ns | * | * | *** | * | ns | ns | ns | *** | ns | *** | ** | *** |
| MON & HAI | *** | ** | ns | * | * | *** | *** | ns | ns | ns | *** | ns | *** | *** | ns |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ– Woyto-Guji, BGA- Boran Galla, SEA – Small East African, GUB – Guinea Bissau, TSW- Tswana, ALP- Italian Alpine, ARD- Ardi, BAL- Baladie, HAI- Hair, MON- Mongolia.

**=P<0.05; **=P<0.01; ***=P<0.001*

Appendix 19. **Summary of linkage disequilibrium between all markers across all Ethiopian populations.**

| | Linkage between markers | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | BM1818 & BMS1494 | | | | | | | | | | * | |
| 2 | BM1818 & ILSTS05 | | | | | | | | | | | |
| 3 | BM1818 & ILSTS11 | | | * | | ** | * | | | | | * |
| 4 | BM1818 & ILSTS44 | | | | | ** | | | | | | |
| 5 | BM1818 & ILSTS87 | | | | | | | | | | | |
| 6 | BM1818 & INRA05 | | | * | | | | | | | | |
| 7 | BM1818 & INRA132 | | | | | | | ** | | | | ** |
| 8 | BM1818 & INRA63 | | | | | | | | | | | |
| 9 | BM1818 & MAF209 | | | ** | | | | | * | | * | |
| 10 | BM1818 & MAF35 | | | | | | | | | | | |
| 11 | BM1818 & MAF65 | | | | | | | | | | | |
| 12 | BM1818 & OarAE129 | | ** | | | | | | | | | ** |
| 13 | BM1818 & OarFCB304 | | | | | | | | | | | * |
| 14 | BM1818 & SRCRSP03 | | | | | | | | | | | |
| 15 | BMS1494 & ILSTS05 | | | | | | | | | | | |
| 16 | BMS1494 & ILSTS11 | | | | ** | | | | | | ** | |
| 17 | BMS1494 & ILSTS44 | | | | | | | | | | | |
| 18 | BMS1494 & ILSTS87 | | | | | | | | | ** | | |
| 19 | BMS1494 & INRA05 | | * | | | | | | * | | | |
| 20 | BMS1494 & INRA132 | | | | | | | | | | | |
| 21 | BMS1494 & INRA63 | | | | | | | | | | | |
| 22 | BMS1494 & MAF209 | | * | | * | | | | | | | |
| 23 | BMS1494 & MAF35 | | | | | | | | * | | | |
| 24 | BMS1494 & MAF65 | | | | | * | | | * | | | |
| 25 | BMS1494 & OarAE129 | | | | | * | | | * | | | |
| 26 | BMS1494 & OarFCB304 | | * | | | | | | | | | |
| 27 | BMS1494 & SRCRSP03 | | | | | | | | | | | |
| 28 | ILSTS05 & ILSTS11 | | | | | | | | | | | |
| 29 | ILSTS05 & ILSTS44 | | | | | | | | | | | |
| 30 | ILSTS05 & ILSTS87 | | | | | | | | | | | |
| 31 | ILSTS05 & INRA05 | | | | | | | | | | | |
| 32 | ILSTS05 & INRA132 | | | | | | | | | | | |
| 33 | ILSTS05 & INRA63 | | * | | | | | | | | | |
| 34 | ILSTS05 & MAF209 | | | | | | | | | | * | |
| 35 | ILSTS05 & MAF35 | | | | * | | | | | | | |
| 36 | ILSTS05 & MAF65 | | | | | | | | | | | |
| 37 | ILSTS05 & OarAE129 | | * | | | | | | | | | |
| 38 | ILSTS05 & OarFCB304 | | | | | | | | | | | |
| 39 | ILSTS05 & SRCRSP03 | | | | | | | | | | | |

| | Linkage between markers | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | ILSTS11 & ILSTS44 | | | | | | | | | | | |
| 41 | ILSTS11 & ILSTS87 | | | | | | | | | | | |
| 42 | ILSTS11 & INRA05 | | | | | | | | | | | |
| 43 | ILSTS11 & INRA132 | | | | | | | | | | | |
| 44 | ILSTS11 & INRA63 | * | | | * | | | | | | | |
| 45 | ILSTS11 & MAF209 | | | | | | | | | | * | |
| 46 | ILSTS11 & MAF35 | | | | | | | | | | | |
| 47 | ILSTS11 & MAF65 | | | | | | | | | | | |
| 48 | ILSTS11 & OarAE129 | | | | | | | | | | * | |
| 49 | ILSTS11 & OarFCB304 | | * | | | | | ** | | | | |
| 50 | ILSTS11 & SRCRSP03 | | | | | | | | | | * | |
| 51 | ILSTS44 & ILSTS87 | | * | | | | | | | | | |
| 52 | ILSTS44 & INRA05 | | | | | | | | | | | |
| 53 | ILSTS44 & INRA132 | | | | | | | | ** | | | |
| 54 | ILSTS44 & INRA63 | | | | | | | | | | | |
| 55 | ILSTS44 & MAF209 | | | | | | | | | | | |
| 56 | ILSTS44 & MAF35 | | | | | | | | | | | * |
| 57 | ILSTS44 & MAF65 | | | | | | | | | | | * |
| 58 | ILSTS44 & OarAE129 | | | | | | | * | | * | | |
| 59 | ILSTS44 & OarFCB304 | | | | * | | * | | | | | |
| 60 | ILSTS44 & SRCRSP03 | | | | | | | | | | | |
| 61 | ILSTS87 & INRA05 | | | | | | * | | | | | |
| 62 | ILSTS87 & INRA132 | | | | * | | | | | | | |
| 63 | ILSTS87 & INRA63 | | | | | * | | | | | * | |
| 64 | ILSTS87 & MAF209 | | | | | | | | | | | |
| 65 | ILSTS87 & MAF35 | | | | | | | | | | | |
| 66 | ILSTS87 & MAF65 | | | | | | | | | | | |
| 67 | ILSTS87 & OarAE129 | | | | | | | | * | | | |
| 68 | ILSTS87 & OarFCB304 | | | | | | | | | | | |
| 69 | ILSTS87 & SRCRSP03 | | | | | | | | | | | |
| 70 | INRA05 & INRA132 | | | | | | | | | | | |
| 71 | INRA05 & INRA63 | | | | | | | | | | | |
| 72 | INRA05 & MAF209 | | | | | | | | | | | |
| 73 | INRA05 & MAF35 | | * | * | | | | | | | | |
| 74 | INRA05 & MAF65 | | | | | | | | | | | |
| 75 | INRA05 & OarAE129 | | | * | | | | | | | | |
| 76 | INRA05 & OarFCB304 | | * | | | | | | | | | |
| 77 | INRA05 & SRCRSP03 | | | | | | | | | | | |
| 78 | INRA132 & MAF209 | | | | | | | | | | | |
| 79 | INRA132 & MAF35 | | | | | | | | | | | |
| 80 | INRA132 & MAF65 | | | * | | | | | | | | |
| 81 | INRA132 & OarAE129 | | | | | | | | | | | |
| 82 | INRA132 & OarFCB304 | | | | | | | | | * | | |

| | Linkage between markers | ABG | ABA | AFA | CEH | GUM | HAH | KEF | LES | NWH | SES | WGJ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 83 | INRA132 & SRCRSP03 | | | | | | | | | | | |
| 84 | INRA63 & INRA132 | * | | | | | | | | | | |
| 85 | INRA63 & MAF209 | | | | | | | | | | | |
| 86 | INRA63 & MAF35 | | | | | | | | | | | |
| 87 | INRA63 & MAF65 | | | * | | | | * | | | | |
| 88 | INRA63 & OarAE129 | | * | | | | | | | | | |
| 89 | INRA63 & OarFCB304 | | | | | | | | | | | |
| 90 | INRA63 & SRCRSP03 | | | | | | | | | | | |
| 91 | MAF209 & OarAE129 | | | | | | | | | | | |
| 92 | MAF209 & OarFCB304 | * | | | | | | | ** | | | |
| 93 | MAF209 & SRCRSP03 | | | | | | | | | | | |
| 94 | MAF35 & MAF209 | | | | | | | * | | | | |
| 95 | MAF35 & MAF65 | | | | | | * | | | | | |
| 96 | MAF35 & OarAE129 | | | | | | | | | | | |
| 97 | MAF35 & OarFCB304 | ** | | | | | | | | * | | |
| 98 | MAF35 & SRCRSP03 | | | | | | | | | | | * |
| 99 | MAF65 & MAF209 | | | | | | | | | | | |
| 100 | MAF65 & OarAE129 | | | | | | | | | | | |
| 101 | MAF65 & OarFCB304 | | | | | | | | | | | |
| 102 | MAF65 & SRCRSP03 | | | | | | | | | | | |
| 103 | OarAE12 & OarFCB304 | | | | | | | | * | | | |
| 104 | OarAE12 & SRCRSP03 | | | | | | | | | | | |
| 105 | OarFCB3 & SRCRSP03 | | | | | | | | | | | |

ABG – Abergalle, ABA – Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM – Gumez, HAH – Hararghe Highland, KEF – Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

**=P<0.05; **=P<0.01*

Appendix 20. **Matrix of all pair-wise combinations of loci showing the linkage disequilibrium (for 11 Ethiopian populations).** *P* values condensed across all populations. For simplicity, the matrix shows only levels of significance.

| | BM 1818 | BM 1494 | ILST S05 | ILST S11 | ILST S44 | ILST S87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCPSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BM1818 | | | | | | | | | | | | | | | |
| BMS1494 | | | | | | | | | | | | | | | |
| ILSTS05 | | | | | | | | | | | | | | | |
| ILSTS11 | ** | | | | | | | | | | | | | | |
| ILSTS44 | * | | | | | | | | | | | | | | |
| ILSTS87 | | | | | | | | | | | | | | | |
| INRA05 | | | | | | | | | | | | | | | |
| INRA63 | | | | * | | * | | | | | | | | | |
| INRA132 | | | | | | | | | | | | | | | |
| MAF35 | | | | | | | | | | | | | | | |
| MAF65 | | | | | | | | | | | | | | | |
| MAF209 | ** | | | | | | | | | | | | | | |
| OarAE129 | ** | * | | | | | | | | | | | | | |
| OFCB304 | | * | | | | | | | | | | * | | | |
| SRCRS03 | | | | | | | | | | | | | | | |

- = $P < 0.05$; ** = $P < 0.01$.

Appendix 21. Allele frequency for locus BM1818 for all the 11 Ethiopian populations.

| Allele | 244 | 246 | 250 | 252 | 254 | 256 | 258 | 260 | 262 | 264 | 266 | 268 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | 1.09 | - | - | 7.61 | - | 7.61 | 51.09 | 7.61 | 13.04 | 6.52 | 5.43 | - |
| ABA | 3.26 | - | - | 5.43 | - | 5.43 | 40.22 | 7.61 | 20.65 | 6.52 | 4.35 | 6.52 |
| AFA | - | 1.14 | - | 1.14 | 1.14 | 3.41 | 65.91 | 9.09 | 3.41 | 6.82 | 5.68 | 2.27 |
| CEH | 5.00 | - | - | 5.00 | 1.25 | 2.50 | 51.25 | 6.25 | 10.00 | 11.25 | 7.50 | - |
| GUM | - | - | - | 8.14 | - | 8.14 | 41.86 | 13.95 | 10.47 | 8.14 | 8.14 | 1.16 |
| HAH | - | - | - | 2.50 | 2.50 | 7.50 | 58.75 | 11.25 | 10.00 | 5.00 | 2.50 | - |
| KEF | - | - | - | 6.58 | - | 5.26 | 68.42 | 7.89 | 5.26 | 3.95 | 2.63 | - |
| LES | 3.13 | 1.04 | 1.04 | 6.25 | 5.21 | 9.38 | 47.92 | 6.25 | 11.46 | 3.13 | 4.17 | 1.04 |
| NWH | 1.11 | - | - | 13.33 | 2.22 | 1.11 | 45.56 | 5.56 | 20.00 | 6.67 | 4.44 | - |
| SES | 1.16 | - | - | 6.98 | 4.65 | 4.65 | 46.51 | 10.47 | 15.12 | 6.98 | 2.33 | 1.16 |
| WGJ | 1.09 | - | - | 3.26 | 2.17 | 2.17 | 54.35 | 5.43 | 13.04 | 7.61 | 4.35 | 6.52 |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 22. Allele frequency for locus BMS1494 for all the 11 Ethiopian populations.

| Allele | 244 | 246 | 256 | 264 | 276 | 280 | 282 | 284 | 288 |
|---|---|---|---|---|---|---|---|---|---|
| ABG | - | - | 9.78 | - | 65.22 | 18.48 | - | 6.52 | - |
| ABA | - | - | 4.55 | - | 57.95 | 36.36 | - | 1.14 | - |
| AFA | - | - | - | 1.14 | 71.59 | 22.73 | - | 4.55 | - |
| CEH | 1.25 | 1.25 | 11.25 | - | 65.00 | 21.25 | - | - | - |
| GUM | - | - | 2.33 | - | 82.56 | 13.95 | - | 1.16 | - |
| HAH | - | - | 3.75 | - | 60.00 | 33.75 | - | 2.50 | - |
| KEF | - | - | 5.26 | - | 57.89 | 21.05 | - | 14.47 | 1.32 |
| LES | - | - | 2.08 | - | 68.75 | 23.96 | - | 4.17 | 1.04 |
| NWH | - | - | 6.82 | - | 62.50 | 29.55 | 1.14 | - | - |
| SES | - | - | - | - | 61.63 | 34.88 | - | 3.49 | - |
| WGJ | - | - | 3.26 | - | 58.70 | 30.43 | - | 7.61 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 23. Allele frequency for locus ILSTS05 for all the 11 Ethiopian populations.

| Allele | 178 | 180 | 182 | 184 | 186 |
|---|---|---|---|---|---|
| ABG | - | 1.09 | 93.48 | 5.43 | - |
| ABA | 1.09 | - | 78.26 | 20.65 | - |
| AFA | 3.41 | 1.14 | 76.14 | 18.18 | 1.14 |
| CEH | - | - | 95.00 | 5.00 | - |
| GUM | 1.16 | - | 90.70 | 8.14 | - |
| HAH | - | - | 87.50 | 12.50 | - |
| KEF | - | - | 80.26 | 19.74 | - |
| LES | 7.29 | - | 80.21 | 12.50 | - |
| NWH | - | - | 92.22 | 7.78 | - |
| SES | - | - | 90.70 | 8.14 | 1.16 |
| WGJ | 2.17 | - | 84.78 | 13.04 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 24. Allele frequency for locus ILSTS11 for all the 11 Ethiopian populations.

| Allele | 265 | 267 | 269 | 271 | 273 | 275 | 277 | 279 | 281 | 283 |
|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | - | 42.39 | 13.04 | 3.26 | 2.17 | 3.26 | 32.61 | 3.26 | - |
| ABA | - | - | 46.74 | 4.35 | - | - | 10.87 | 28.26 | 7.61 | 2.17 |
| AFA | - | - | 15.91 | 6.82 | - | 4.55 | 15.91 | 56.82 | - | - |
| CEH | - | - | 45.00 | 12.50 | - | 1.25 | 2.50 | 36.25 | 2.50 | - |
| GUM | - | - | 37.21 | 3.49 | - | 1.16 | 9.30 | 48.84 | - | - |
| HAH | - | - | 32.50 | 2.50 | - | 1.25 | 17.50 | 42.50 | 3.75 | - |
| KEF | 1.32 | 1.32 | 50.00 | 15.79 | - | - | 6.58 | 17.11 | 7.89 | - |
| LES | - | 2.08 | 39.58 | - | - | 1.04 | 21.88 | 34.38 | 1.04 | - |
| NWH | - | - | 53.33 | 6.67 | - | 5.56 | 4.44 | 28.89 | 1.11 | - |
| SES | - | - | 29.07 | 5.81 | - | 2.33 | 16.28 | 43.02 | 1.16 | 2.33 |
| WGJ | - | - | 32.61 | 3.26 | - | - | 27.17 | 35.87 | 1.09 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 25. Allele frequency for locus ILSTS44 for all the 11 Ethiopian populations.

| Allele | 153 | 155 | 157 | 159 | 161 |
|---|---|---|---|---|---|
| ABG | - | 21.74 | 31.52 | 44.57 | 2.17 |
| ABA | - | 34.52 | 50.00 | 11.90 | 3.57 |
| AFA | - | 39.77 | 35.23 | 19.32 | 5.68 |
| CEH | - | 33.75 | 33.75 | 30.00 | 2.50 |
| GUM | 1.16 | 39.53 | 44.19 | 11.63 | 3.49 |
| HAH | - | 38.75 | 35.00 | 18.75 | 7.50 |
| KEF | - | 36.84 | 42.11 | 13.16 | 7.89 |
| LES | 2.08 | 46.88 | 33.33 | 14.58 | 3.13 |
| NWH | - | 37.78 | 36.67 | 22.22 | 3.33 |
| SES | - | 40.70 | 30.23 | 20.93 | 8.14 |
| WGJ | - | 38.04 | 43.48 | 15.22 | 3.26 |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 26. Allele frequency for locus ILSTS87 for all the 11 Ethiopian populations.

| Allele | 145 | 147 | 149 | 151 | 155 | 159 | 161 | 163 |
|---|---|---|---|---|---|---|---|---|
| ABG | 9.78 | 83.70 | 1.09 | 1.09 | - | - | 3.26 | 1.09 |
| ABA | 9.78 | 70.65 | 6.52 | - | 1.09 | 9.78 | 2.17 | - |
| AFA | 4.55 | 82.95 | 2.27 | 7.95 | - | - | 2.27 | - |
| CEH | 7.50 | 71.25 | 7.50 | - | - | 8.75 | 5.00 | - |
| GUM | 8.33 | 80.95 | 7.14 | 1.19 | 1.19 | 1.19 | - | - |
| HAH | 13.75 | 76.25 | 2.50 | 3.75 | 2.50 | 1.25 | - | - |
| KEF | 3.95 | 78.95 | 3.95 | - | - | 5.26 | 7.89 | - |
| LES | 22.92 | 70.83 | 3.13 | - | 1.04 | - | 2.08 | - |
| NWH | 3.33 | 76.67 | 7.78 | 3.33 | - | 6.67 | 1.11 | 1.11 |
| SES | 16.67 | 69.23 | 6.41 | 2.56 | 2.56 | 1.28 | 1.28 | - |
| WGJ | 1.09 | 85.87 | 3.26 | 1.09 | - | 3.26 | 3.26 | 2.17 |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 27. Allele frequency for locus INRA05 for all the 11 Ethiopian populations.

| Allele | 131 | 139 | 141 | 143 | 145 |
|---|---|---|---|---|---|
| ABG | - | 60.87 | 29.35 | 1.09 | 8.70 |
| ABA | - | 41.30 | 44.57 | 9.78 | 4.35 |
| AFA | 1.14 | 42.05 | 53.41 | 3.41 | - |
| CEH | - | 43.75 | 50.00 | 6.25 | - |
| GUM | - | 58.14 | 27.91 | 12.79 | 1.16 |
| HAH | - | 35.00 | 50.00 | 10.00 | 5.00 |
| KEF | - | 46.05 | 36.84 | 7.89 | 9.21 |
| LES | - | 46.88 | 43.75 | 6.25 | 3.13 |
| NWH | - | 51.11 | 37.78 | 6.67 | 4.44 |
| SES | - | 51.16 | 37.21 | 3.49 | 8.14 |
| WGJ | - | 60.87 | 22.83 | 10.87 | 5.43 |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 28. Allele frequency for locus INRA63 for all the 11 Ethiopian populations.

| Allele | 157 | 161 | 163 | 165 | 167 | 169 | 171 | 173 |
|---|---|---|---|---|---|---|---|---|
| ABG | - | - | 6.52 | 8.70 | 54.35 | 25.00 | 5.43 | - |
| ABA | - | - | 3.26 | 5.43 | 57.61 | 29.35 | 4.35 | - |
| AFA | - | - | 2.27 | 14.77 | 47.73 | 35.23 | - | - |
| CEH | - | - | 7.50 | 7.50 | 62.50 | 17.50 | 5.00 | - |
| GUM | 9.30 | - | - | 15.12 | 52.33 | 20.93 | 2.33 | - |
| HAH | - | - | 3.75 | 8.75 | 51.25 | 33.75 | 1.25 | 1.25 |
| KEF | 6.58 | 1.32 | 3.95 | 11.84 | 44.74 | 28.95 | 1.32 | 1.32 |
| LES | - | - | 10.42 | 8.33 | 50.00 | 20.83 | 9.38 | 1.04 |
| NWH | - | - | 4.44 | 11.11 | 46.67 | 34.44 | 3.33 | - |
| SES | - | - | 8.14 | 8.14 | 43.02 | 34.88 | 5.81 | - |
| WGJ | - | - | 5.43 | 8.70 | 54.35 | 29.35 | 2.17 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 29. Allele frequency for locus INRA132 for all the 11 Ethiopian populations.

| Allele | 143 | 145 | 147 |
|---|---|---|---|
| ABG | 60.87 | 36.96 | 2.17 |
| ABA | 52.17 | 44.57 | 3.26 |
| AFA | 40.91 | 56.82 | 2.27 |
| CEH | 63.75 | 31.25 | 5.00 |
| GUM | 46.51 | 38.37 | 15.12 |
| HAH | 50.00 | 47.50 | 2.50 |
| KEF | 59.21 | 36.84 | 3.95 |
| LES | 45.83 | 52.08 | 2.08 |
| NWH | 56.67 | 43.33 | - |
| SES | 50.00 | 50.00 | - |
| WGJ | 41.30 | 54.35 | 4.35 |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 30. Allele frequency for locus MAF35 for all the 11 Ethiopian populations.

| Allele | 105 | 107 | 109 |
|---|---|---|---|
| ABG | 90.22 | 9.78 | - |
| ABA | 86.96 | 11.96 | 1.09 |
| AFA | 93.18 | 6.82 | - |
| CEH | 93.75 | 6.25 | - |
| GUM | 93.02 | 6.98 | - |
| HAH | 93.75 | 5.00 | 1.25 |
| KEF | 94.74 | 5.26 | - |
| LES | 85.42 | 14.58 | - |
| NWH | 90.00 | 10.00 | - |
| SES | 91.86 | 8.14 | - |
| WGJ | 90.00 | 10.00 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 31. Allele frequency for locus MAF65 for all the 11 Ethiopian populations.

| Allele | 116 | 118 | 120 | 122 | 124 | 126 | 128 | 130 | 132 | 134 | 136 | 138 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | 1.09 | 19.57 | 3.26 | 26.09 | - | 27.17 | - | - | 2.17 | 17.39 | 3.26 | - |
| ABA | 7.78 | 18.89 | 5.56 | 12.22 | 4.44 | 21.11 | - | 14.44 | 5.56 | 10.00 | - | - |
| AFA | 18.18 | 4.55 | 19.32 | 17.05 | 3.41 | 12.50 | - | 2.27 | 15.91 | 6.82 | - | - |
| CEH | 13.51 | 9.46 | 6.76 | 21.62 | - | 18.92 | - | 4.05 | 16.22 | 9.46 | - | - |
| GUM | 1.32 | 11.84 | 2.63 | 27.63 | 1.32 | 22.37 | - | 1.32 | 17.11 | 5.26 | 6.58 | 2.63 |
| HAH | 5.00 | 13.75 | 5.00 | 18.75 | 7.50 | 11.25 | - | 10.00 | 16.25 | 10.00 | 1.25 | 1.25 |
| KEF | 1.35 | 4.05 | 18.92 | 14.86 | - | 17.57 | - | 5.41 | 24.32 | 9.46 | 4.05 | - |
| LES | 5.32 | 3.19 | 5.32 | 26.60 | 4.26 | 19.15 | - | 11.70 | 11.70 | 9.57 | 3.19 | - |
| NWH | 3.41 | 7.95 | 15.91 | 25.00 | - | 17.05 | 1.14 | - | 18.18 | 10.23 | 1.14 | - |
| SES | 8.14 | 8.14 | 5.81 | 15.12 | 2.33 | 12.79 | - | 10.47 | 25.58 | 11.63 | - | - |
| WGJ | 3.26 | 7.61 | 4.35 | 15.22 | 5.43 | 23.91 | - | 7.61 | 19.57 | 10.87 | - | 2.17 |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 32. Allele frequency for locus MAF209 for all the 11 Ethiopian populations.

| Allele | 102 | 104 | 106 | 108 |
|---|---|---|---|---|
| ABG | 30.43 | 57.61 | 11.96 | - |
| ABA | 22.83 | 51.09 | 26.09 | - |
| AFA | 14.77 | 70.45 | 14.77 | - |
| CEH | 27.50 | 57.50 | 15.00 | - |
| GUM | 30.23 | 65.12 | 4.65 | - |
| HAH | 26.25 | 60.00 | 12.50 | 1.25 |
| KEF | 38.16 | 47.37 | 10.53 | 3.95 |
| LES | 16.67 | 60.42 | 22.92 | - |
| NWH | 23.33 | 63.33 | 13.33 | - |
| SES | 17.44 | 69.77 | 12.79 | - |
| WGJ | 20.65 | 50.00 | 29.35 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 33. Allele frequency for locus OarAE129 for all the 11 Ethiopian populations.

| Allele | 139 | 141 | 143 | 145 | 147 | 149 | 151 | 153 | 155 | 157 | 159 | 161 | 163 | 165 | 167 | 169 | 171 | 173 | 175 | 177 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | 27.17 | 14.13 | - | - | - | - | - | - | 14.13 | 29.35 | 8.70 | - | - | - | - | 3.26 | 1.09 | - | 2.17 | - |
| ABA | 35.23 | 20.45 | - | - | - | - | - | - | 21.59 | 12.50 | 5.68 | - | - | 1.14 | - | 3.41 | - | - | - | - |
| AFA | 12.50 | 9.09 | - | - | - | 1.14 | - | 3.41 | 18.18 | 25.00 | 10.23 | - | - | 3.41 | 1.14 | 13.64 | 2.27 | - | - | - |
| CEH | 35.94 | 15.63 | - | - | - | - | - | 1.56 | 7.81 | 20.31 | 14.06 | - | - | - | - | 1.56 | - | 1.56 | 1.56 | - |
| GUM | 30.23 | 13.95 | - | - | - | 2.33 | - | - | 9.30 | 13.95 | 5.81 | - | - | 15.12 | 2.33 | 4.65 | - | 1.16 | 1.16 | - |
| HAH | 34.21 | 13.16 | - | - | - | - | - | - | 13.16 | 22.37 | 7.89 | - | - | 2.63 | - | 2.63 | 2.63 | 1.32 | - | - |
| KEF | 28.95 | 22.37 | - | - | 2.63 | - | - | 1.32 | 3.95 | 22.37 | 5.26 | - | 2.63 | 1.32 | 2.63 | 3.95 | - | - | 2.63 | - |
| LES | 22.34 | 13.83 | - | 1.06 | - | - | 2.13 | 2.13 | 20.21 | 22.34 | 3.19 | - | 1.06 | 2.13 | 2.13 | 5.32 | - | - | 1.06 | 1.06 |
| NWH | 30.00 | 21.11 | 1.11 | 5.56 | - | - | - | 1.11 | 10.00 | 16.67 | 8.89 | 1.11 | - | - | 1.11 | 1.11 | - | - | 2.22 | - |
| SES | 25.00 | 17.86 | - | - | - | 1.19 | - | 1.19 | 16.67 | 15.48 | 10.71 | - | - | - | 1.19 | 8.33 | - | - | 2.38 | - |
| WGJ | 27.17 | 17.39 | - | - | - | - | - | 3.26 | 17.39 | 18.48 | 2.17 | - | - | 2.17 | 4.35 | 3.26 | 1.09 | - | 3.26 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 34. Allele frequency for locus OarFCB304 for all the 11 Ethiopian populations.

| Allele | 120 | 124 | 128 | 132 | 134 | 136 | 138 | 140 | 142 | 144 | 146 | 150 | 154 | 156 | 158 | 160 | 162 | 164 | 166 | 168 | 170 | 174 | 180 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABG | - | 2.17 | - | - | - | 2.17 | - | 45.65 | 9.78 | 5.43 | 14.13 | - | 13.04 | - | 2.17 | - | 2.17 | 2.17 | - | 1.09 | - | - | - |
| ABA | 1.09 | - | - | - | - | 3.26 | - | 50.00 | 14.13 | 3.26 | 7.61 | 3.26 | 6.52 | - | - | 2.17 | 8.70 | - | - | - | - | - | - |
| AFA | - | 1.14 | - | - | - | 3.41 | - | 50.00 | 14.77 | 3.41 | 10.23 | - | 12.50 | - | 3.41 | - | - | - | 1.14 | - | - | - | - |
| CEH | - | 3.75 | - | - | - | 5.00 | - | 47.50 | 8.75 | 1.25 | 12.50 | 2.50 | 16.25 | - | 2.50 | - | - | - | - | - | - | - | - |
| GUM | - | - | - | - | 1.16 | 3.49 | 1.16 | 59.30 | 13.95 | 4.65 | 1.16 | - | 11.63 | - | - | - | 3.49 | - | - | - | - | - | - |
| HAH | - | 1.25 | 1.25 | - | - | 3.75 | - | 48.75 | 5.00 | 1.25 | 13.75 | - | 6.25 | 1.25 | 12.50 | 1.25 | - | - | - | - | 2.50 | - | 1.25 |
| KEF | - | - | - | - | - | - | - | 50.00 | 22.37 | 3.95 | 2.63 | 2.63 | 9.21 | - | 7.89 | 1.32 | - | - | - | - | - | - | - |
| LES | - | - | - | - | - | 4.17 | - | 62.50 | 9.38 | 1.04 | - | - | 6.25 | 1.04 | 3.13 | 2.08 | 2.08 | 1.04 | 1.04 | 4.17 | 1.04 | 1.04 | - |
| NWH | - | 1.11 | - | - | - | 2.22 | - | 50.00 | 11.11 | 8.89 | 3.33 | - | 21.11 | - | 2.22 | - | - | - | - | - | - | - | - |
| SES | - | - | - | 1.16 | - | 4.65 | - | 60.47 | 9.30 | 1.16 | 5.81 | - | 8.14 | - | 5.81 | 2.33 | - | - | 1.16 | - | - | - | - |
| WGJ | - | - | - | - | - | 4.35 | - | 47.83 | 16.30 | 5.43 | 8.70 | - | 8.70 | - | - | 2.17 | - | 3.26 | 1.09 | - | - | 2.17 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Small-eared Somali, WGJ – Woyto-Guji.

Appendix 35. **Allele frequency for locus SRCRSP03 for all the 11 Ethiopian populations.**

| Allele | 111 | 117 | 119 | 121 | 123 | 125 |
|---|---|---|---|---|---|---|
| ABG | 27.78 | - | 13.33 | 30.00 | 28.89 | - |
| ABA | 38.04 | - | 4.35 | 41.30 | 16.30 | - |
| AFA | 19.32 | - | 17.05 | 53.41 | 10.23 | - |
| CEH | 18.75 | - | 18.75 | 46.25 | 16.25 | - |
| GUM | 52.33 | - | 10.47 | 30.23 | 6.98 | - |
| HAH | 19.23 | 2.56 | 15.38 | 53.85 | 7.69 | 1.28 |
| KEF | 36.84 | - | 1.32 | 52.63 | 9.21 | - |
| LES | 40.63 | - | 5.21 | 44.79 | 8.33 | 1.04 |
| NWH | 27.78 | - | 8.89 | 46.67 | 16.67 | - |
| SES | 26.74 | 1.16 | 17.44 | 47.67 | 6.98 | - |
| WGJ | 45.65 | - | 3.26 | 44.57 | 6.52 | - |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji.

Appendix 36. **Test for genetic differentiation between pairs of Ethiopia goat populations for each locus.**

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABA & ABG | ns | * | ** | ** | *** | ** | ** | ns | ns | ns | *** | ns | ns | * | * |
| AFA &ABG | ns | ** | ** | *** | ** | ns | *** | ns | * | ns | *** | ns | ** | ns | ** |
| AFA &ABA | ** | * | ns | *** | ns | *** | ns | * | ns | ns | *** | * | *** | * | ** |
| CEH &ABG | ns | ns | ns | ns | ns | ** | *** | ns | ns | ns | *** | ns | ns | ns | * |
| CEH &ABA | ns | ns | ** | * | * | ns | ns | ns | ns | ns | * | ns | ns | * | ** |
| CEH &AFA | ns | *** | ** | *** | ns | *** | ns | * | * | ns | ns | ns | *** | ns | ns |
| GUM &ABG | ns | * | ns | ** | *** | ns | ** | ** | ** | ns | * | ns | *** | * | *** |
| GUM &ABA | ns | ** | ns | ** | ns | ns | ns | *** | * | ns | *** | *** | ** | ns | * |
| GUM & AFA | * | ns | ns | * | ns | ns | ** | ** | ** | ns | *** | ** | *** | * | *** |
| GUM & CEH | ns | * | ns | * | ns | * | * | ** | * | ns | * | ns | * | ** | *** |
| HAH & ABG | ns | ns | ns | ** | ** | ns | ** | ns | ns | ns | *** | ns | ns | * | *** |
| HAH & ABA | * | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ns | *** | ** |
| HAH & AFA | ns | ns | ns | * | ns | ns | ns | ns | ns | ns | ** | ns | * | ns | ns |
| HAH & CEH | ns | * | ns | ** | ns | ** | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| HAH & GUM | ns | * | ns | ns | ns | ns | ** | ** | ** | ns | ns | ns | ns | *** | *** |
| KEF & ABG | ns | ns | * | ns | *** | * | ns | ns | ns | ns | *** | ns | ns | ** | *** |
| KEF & ABA | ** | ** | ns | ns | ns | ns | ns | * | ns | ns | *** | ** | * | ** | ns |
| KEF & AFA | ns | * | ns | *** | ns | ** | ** | ns | ns | ns | ** | *** | *** | | *** |
| KEF & CEH | ns | ** | * | ns | ns | ns | * | * | ns | ns | * | | ns | ** | *** |
| KEF & GUM | ns | ** | * | *** | ns | * | * | ns | * | ns | * | * | ns | ** | ** |
| KEF & HAH | ns | * | ns | *** | ns | ** | ns | ns | ns | ns | ** | | * | ** | ** |
| LES & ABG | ns | ns | ** | *** | *** | ns | * | ns | ns | ns | *** | * | ns | ** | *** |
| LES & ABA | ns | ns | * | ** | ns | ** | ns | ns | ns | ns | ** | ns | ns | ** | ns |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFC B304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LES & AFA | * | ns | ns | *** | ns | *** | ns | *** | ns | ns | ** | ns | ns | * | ** |
| LES & CEH | ns | * | ** | *** | ns | *** | ns | ns | ns | ns | ns | ns | ns | *** | ** |
| LES & GUM | ns | ns | ns | * | ns | * | ns | *** | ** | ns | * | ** | * | ns | ns |
| LES & HAH | ns | ns | * | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ** | * |
| LES & KEF | ns | ns | * | *** | ns | *** | ns | ** | ns | * | * | ** | ns | * | ns |
| NWH & ABG | ns | * | ns | ns | * | ** | ns | ns | ns | ns | *** | ns | ns | ns | ns |
| NWH & ABA | ns | ns | * | * | ns | ns | ns | ns | ns | ns | *** | ns | ns | *** | ns |
| NWH & AFA | *** | ** | * | *** | ns | ns | * | ns | * | ns | * | ns | *** | ns | ns |
| NWH & CEH | ns | ns | ns | ns | ns | ns | ns | ns | * | ns | ns | ns | ns | ns | ns |
| NWH & GUM | * | ** | ns | * | ns | ns | ns | ** | *** | ns | ns | ns | ** | ns | ** |
| NWH & HAH | * | ns | ns | ** | ns | * | ns | ns | ns | ns | ** | ns | ns | *** | ns |
| NWH & KEF | * | ** | * | * | ns | ns | ns | ns | ns | ns | ns | * | ns | * | ns |
| NWH & LES | ns | * | ** | *** | ns | *** | ns | ns | ns | ns | ** | ns | ns | ** | ns |
| SES & ABG | ns | ** | ns | ** | ** | ns | ns | ns | ns | ns | *** | ns | ns | * | *** |
| SES & ABA | ns | ns | * | * | * | ns | ns | ns | ns | ns | * | * | ns | ** | ** |
| SES & AFA | ns | ns | * | ns | ns | * | * | * | ns | ns | * | ns | ns | ns | ns |
| SES & CEH | ns | *** | ns | ** | ns | * | * | ns | * | ns | ns | ns | ns | ns | ns |
| SES & GUM | ns | ** | ns | ns | ns | ns | * | *** | *** | ns | ** | ns | ** | * | ** |
| SES & HAH | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns |
| SES & KEF | ns | ** | * | *** | ns | ** | ns | ns | * | ns | * | ** | ns | ns | ** |
| SES & LES | ns | ns | * | * | ns | ns | ns | ns | ns | ns | ns | ns | ns | ns | * |
| SES & NWH | ns | ** | ns | ** | ns | * | ns | ns | ns | ns | ** | ns | ns | * | ns |
| WGJ & ABG | ns | ns | ns | *** | *** | ns | * | ns | * | ns | *** | * | ns | ns | *** |
| WGJ & ABA | ns | ns | ns | ** | ns | * | * | ns | ns | ns | * | ns | ns | * | ns |
| WGJ & AFA | ns | ns | ns | ** | ns | ns | *** | ns | ns | ns | ** | * | * | ns | *** |

| Population pairs | BM 1818 | BMS 1494 | ILSTS 05 | ILSTS 11 | ILSTS 44 | ILSTS 87 | INRA 05 | INRA 63 | INRA 132 | MAF 35 | MAF 65 | MAF 209 | OarAE 129 | OarFCB 304 | SRCRSP 03 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WGJ & CEH | ns | ** | ns | *** | ns | * | ** | ns | * | ns | ns | ns | ns | * | *** |
| WGJ & GUM | * | ** | ns | * | ns | ns | ns | ** | * | ns | ns | *** | * | * | ns |
| WGJ & HAH | ns | ns | ns | ns | ns | ** | ** | ns | ns | ns | ns | * | ns | *** | ** |
| WGJ & KEF | * | ns | ns | *** | ns | ns | ns | ns | ns | ns | * | ** | ns | ** | ns |
| WGJ & LES | ns | ns | ns | ns | ns | *** | * | ns | ns | ns | ns | ns | ns | ** | ns |
| WGJ & NWH | ns | * | ns | *** | ns | ns | ns | ns | * | ns | ** | * | ns | * | * |
| WGJ & SES | ns | ns | ns | ns | ns | ** | ns | ns | ns | ns | ns | * | ns | ns | ** |

ABG – Abergalle, ABA– Arsi-Bale, AFA- Afar, CEH – Central Highland, GUM- Gumez, HAH – Hararghe Highland, KEF- Keffa, LES – Long-eared Somali, NWH – North-West Highland, SES – Short-eared Somali, WGJ – Woyto-Guji.
**=P<0.05; **=P<0.01; ***=P<0.001*


Kalinowski, S. T. (2002). Evolutionary and statistical properties of three genetic distances. *Mol Ecol* **11**(8)**,** 1263-73.

Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* **19**(2)**,** 153-70.

Takezaki, N., and Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**(1)**,** 389-99.

Weir, B. S., and Basten, C. J. (1990). Sampling strategies for distances between DNA sequences. *Biometrics* **46**(3)**,** 551-82.