Forensic Identification of six of Tanzanian Populations Using the Extended Haplotype Markers

A thesis submitted in the partial fulfilment of the requirements for the degree of Magister of Scientiae in the Department of Biotechnology, University of the Western Cape

Hadija Saidi Mwema

May 2011

Supervisors:

Dr. Maria Eugenia D'Amato (PhD)

Prof. Sean Davison (PhD)

Key words

Population

DNA

Y chromosome

Tanzania

STR

Loci

Extended haplotype

Genotyping

Database

Polymerase Chain Reaction (PCR)

Electrophoresis

Allele frequency

Discrimination capacity

Polymorphism

Forensic

Genetic Diversity

Abstract

Forensic Identification of Six Tanzanian Populations Using the Extended Haplotype Markers

Hadija Saidi Mwema

MSc Thesis, Department of Biotechnology, University of the Western Cape

13/05/2011

The aim of the present study was to evaluate the power of discrimination and genetic (diversity) parameters in the Y chromosome extended haploytpe markers in populations of Tanzania for forensic and populations studies. Eleven Y chromosome extended haplotype markers were selected for this study, these includes Minimal haplotypes markers i.e. DYS19, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS389I/II and two additional markers DYS438 and DYS439. Six populations of Tanzania were investigated under this study. These populations were selected based on the language family categories; Niger Congo (Kuria and Sukuma), Nilo Saharan (Luo and Maasai) and Afro Asiatic (Iraqw and Alagwa).

Buccal swabs were collected from unrelated males from Mwanza province (Sukuma), Mara (Kuria and Luo), Arusha (Maasai and Iraqw) and Dodoma province (Alagwa). Samples were typed using ABI 377 Genetic Analyser (Applied Biosystem) followed by analysis using softwares Gelprocessor, GeneScan 3.0.0 (Applied Biosystems) and Genotyper 3.7 (Applied Biosystems). The data obtained were analysed by GenePop 4.0, Arlequin 3.11 and Genetix v.4.05.2 software packages. Analyses such as AMOVA, Fst population pairwise comparison, Factorial component Analysis were used to obtain Allele frequency, haplotype frequency, gene diversities among various loci and levels of gene flow between populations.

For the overall individuals, the highest Gene Diversity value was 0.8251 (DYS385) and the lowest was 0.2723 (DYS392). The overall Haplotype Diversity was 0.9984 and Discrimination capacity resulted 84.27%. A total of 225 distinct haplotypes were

identified in 267 individuals, 28 were shared, the most frequent haplotype was present in 5 individuals. The levels of genetic diversity for the haplotypes per group as revealed by haplotype diversities confirmed that the most diverse group being Sukuma, Kuria, Iraqw, Maasai, Luo and Alagwa being the least diverse. The Discrimination capacity of these set of markers showed the highest value in Sukuma population (100%) subsequently followed by Iraqw, Luo, Maasai, Kuria and Alagwa (78.38%) being the lowest. Analysis of Molecular Variance showed a significant differentiation among populations, 93.96% of variance was found within population and 6.04% among population. Population pairwise results between all population pairs (except Sukuma and kuria and Alagwa and Luo) showed significant results ($P < 0.05$).

Genetic heterogeneity that was found among Tanzanian populations could not be attributed to language barriers but was largely being contributed by a limited level of gene flow between these populations due to different ethnical, social, cultural and historical backgrounds between them. All Y chromosome extended haplotype loci used in this study (except DYS392 and DYS391 which showed the lowest level of polymorphism) were found to be likely useful for forensic application in Tanzania. Furthermore the extended haplotype markers used in this study may be useful in the establishment of the National DNA database following the enactment of the Human DNA Legislation in Tanzania (http://www.parliament.go.tz).

Declaration

I declare that The Forensic identification of Six Tanzanian Populations using the extended Haplotype Markers is my own work and it has not been submitted for any degree or examination from any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full Name: Hadija Saidi Mwema

Signed:.....

Date: Thirteenth day of May, year two thousand and eleven

Acknowledgements

I am grateful to males who donated their DNA samples for this study from High schools in Mwanza, Mara, Arusha and Dodoma provinces of Tanzania.

This thesis would not have been finalized without supervision, professional support and guidance. I would like to acknowledge Professor Sean Davidson and Dr. Maria Eugenia D'amato for their sincere supervision throughout this work.

I wish to express my gratitude to my family; Mom, Dad and my husband for taking care of the kids while I was absent.

Special thanks to the Late Chief Government Chemist of Tanzania, Dr. Ernest N. M. Mashimba, who released me for this work, and passed away during my studies, May God rest his soul in eternal peace.

Amongst friends and colleagues, I would like to acknowledge colleagues at the UWC, UWC-FDL laboratory and colleagues at GCLA in Tanzania for their encouragement, cooperation and support throughout these studies.

I am profoundly indebted to Belgium Technical Cooperation (BTC) for granting the scholarship which enabled the accomplishment of this study.

This thesis is an outcome of several months of hard work, hardship, and sacrifices. The achievement to this goal was upon guidance, care and protection of the Almighty GOD. May the glory and honour be given to HIM.

To:

My Son Saidi.

He endured my absence for this work while he was just two years of age.

List of Abbreviation

STR	**Short Tandem Repeat**
Y-STR	**Y- chromosome Short Tandem Repeat**
DNA	**Deoxyribonucleic Acid**
GCLA	**Government Chemist Laboratory Agency of Tanz**
PCR	**Polymerase Chain Reaction**
AMPS	**Ammonium Persulfate**
dNTPs	**Deoxy Nucleotide Tri Phosphates**
EDTA	**Ethylene Diamine Tetra Acetic Di-Sodium Salt**
UWC	**University of the Western Cape**
FDL	**Forensic DNA Laboratory**
DC	**Discrimination Capacity**
HD	**Haplotype Diversity**
D	**Gene Diversity**

bp	**base pairs**
PCR	**Polymerase Chain Reaction**
TEMED	**N, N, N', N' Tetramethyl-EthyleneDiamine**
BSA	**Bovine Serum Albumin**
MgCl2	**Magnesium Chloride**
YHRD	**Y-Chromosome Haplotype Reference Database**
SWGDAM	**Scientific Working Group on DNA Analysis Methods**
NRY	**Non-recombining Y-chromosome**
SNP	**Single Nucleotide Polymorphism**
mtDNA	**Mitochondrial DNA**
PAR	**Pseudo-Autosomal Region**

List of Tables

Haplotypes shared by more than one Kuria males (n = 57)

Table 4.10

Haplotypes shared by more than one Luo males (n = 34)

Table 4.11

Haplotypes shared by more than Maasai males (n = 51)

Table 4.12

Haplotypes shared by more than one Iraqw males (n = 67)

Table 4.13

Haplotypes shared by more than one Alagwa males (n = 37)

Table 4.14

Haplotypes shared among males in the six population groups of Tanzania

Table 4.15

List of Y- STR extended haplotypes observed in population of Tanzania (n=267)

Table 4.16

Y- STR haplotype pairwise Fst values below diagonal and corresponding P values above diagonal (P values < 0.05 are bolded) for six populations of Tanzania (Locus DYS385a,b) was excluded from the analysis.

List of Figures

Luo (Mara province), Maasai and Iraqw (Arusha province) and Alagwa (Dodoma province).

Figure 4.1

Graphical representation of a typical electropherogram of the Extended Haplotype where each locus is labelled.

Figure 4.2

Allele frequency of DYS389I

Figure 4.3

Allele frequency of DYS389II

Figure 4.4 Allele

Allele frequency of DYS438

Figure 4.5

Allele frequency of DYS391

Figure 4.6

Allele frequency of DYS 439

Figure 4.7

Allele frequency of DYS393

Figure 4.8

Allele frequency of DYS390

Figure 4.9

Allele frequency of DYS392

Figure 4.10

Allele frequency of DYS19

Figure 4.11

Graphical presentation of Gene diversity values across populations studied.

Figure 4.12.

A Graphical presentation of gene diversity values across loci in the overall population of Tanzania.

Figure 4.13.

Worldwide distribution of the most common 'extended haplotype' among Sukuma males from Mwanza province (Red dots on the map indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.14.

Worldwide distribution of the most common 'extended haplotype' among Kuria males from Mara province (Red dots on the map indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.15.

Worldwide distribution of the 2nd most common 'extended haplotype'among Kuria males from Mara province (blue dots on the map

indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.16.

Worldwide distribution of the most common 'extended haplotype' among Luo males from Mara province (blue dots on the map indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.17.

Worldwide distribution of the 2nd most common extended haplotype among Luo males from Mara province (Red dots on the map indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.18.

Worldwide distribution of the most common 'extended haplotype' among Maasai males from Arusha province (blue dots on the map indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.19.

Worldwide distribution of the 2nd most common 'extended haplotype'among Maasai males from Arusha province (blue dots on the

map indicate populations for which extended haplotype data is available, but in which no match could be found).

Figure 4.20.

Worldwide distribution of the most common 'extended haplotype' among Iraqw males from Arusha province (blue dots on the map indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.21.

Worldwide distribution of the 2nd most common 'extended haplotype' among Iraqw males from Arusha province (blue dots on the map indicate populations for which extended haplotype data is available, but in which no match could be found).

Figure 4.22.

Worldwide distribution of the most common 'extended haplotype' among Alagwa males from Dodoma province (Red dots on the map indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found).

Figure 4.23.

Worldwide distribution of the 2nd most common 'extended haplotype'among Alagwa males from Dodoma province (Red dots on the map indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no match could be found).

Figure 4.24.

Factorial Component Analysis of the extended haplotypes data for the different samples of Tanzanian population (locus DYS385 was excluded in the analysis).

Table of contents

Title: Forensic Identification of six Tanzanian Populations using the Extended Haplotype Markers

Chapter 1

Introduction

Research in Y-chromosome Short Tandem Repeats polymorphisms and applications has been increasingly studied in the world (Kayser *et al*., 1997; Ayoub *et al*., 2000; Kayser *et al*., 2004; D'Amato *et al*., 2011). The invaluable application of Y chromosomal STR systems is due to their potential in detecting and discriminating male DNA. Human Y-chromosomal short tandem repeat polymorphisms (Y-STRs) or microsatellites are useful in resolving and relating male lineages in forensics especially in sexual assault cases where there is a large proportion of mixed male/female stains (Prinz *et al*., 1996; Redd *et al*., 2002), genealogical (Kayser *et al*., 2007), evolutionary (Jobling and Tyler-Smith, 2003), population (e.g. D'Amato *et al*., 2008) and anthropological studies (Knijiff *et al*., 2000). Also it was indicated that the chromosome Y STRs are useful in comparing closely related populations (Knijff *et al*., 1997).

According to Butler (2005) population genetics is the study of inherited variation and its modulation in time and space, it quantifies the variation observed within a population group or among different population groups in terms of alleles and genotypes. He further elaborated that there is great genetic variation that exists within species at the individual

nucleotide level and it is approximated that 10 million nucleotides can differ between individuals in humans. Studies in population genetics are important while constructing databases the samples mostly taken from a range of ethnical and geographical background, this provides an important information of unrelated lineages on a given population for forensic application (Roewer *et al*., 2001).

Africa is considered to be the origin of all modern humans approximately 100,000 years ago (Cann et al., 1987). Various types of diversities such as genetic, linguistic, ethnical and cultural diversities are found within African continent. Studies (Kayser et al., 2001, Tshkoff and Willium, 2002; Reed and Tishkoff, 2006, Omran et al., 2008) have indicated that African populations contains substantial amount of human genetic diversity, both within and between populations. Other studies (Newman, 1995; Heine and Nurse, 2000; Blench, 2006) have indicated that Africa contains not only genetic but also cultural and linguistic diversities. It is therefore important to study the levels of genetic diversities that are found in African populations for various applications.

The population of Tanzania has been estimated to contain about 128 ethno-linguistic groups (http://www.ethnologue.com), these groups fall into four language families categories; Niger Congo, Afro Asiatic, Nilo Saharan and Khoesan (Ambrose, 1982; Newman, 1995; Heine and Nurse, 2000; Blench, 2006). Previous studies in human genetics using different Y chromosome markers have been conducted in some of Tanzanian population groups on tracing human migration pattern (Henn *et al*.,2008) and testing associations between genetic and linguistic or geographic differeferentiation (Knight *et al*., 2003; Wood *et al*., 2005; Tishkoff *et al*., 2007; Tishkoff *et al*., 2009). More genetic studies are required in different population groups of Tanzania, because of extensive levels of ethnically diverse populations. In Tanzania Forensic DNA analysis is the responsibility of the Government Chemist Laboratory Agency (GCLA) (http://www.gcla.go.tz) in collaboration with Tanzanian Police Force (http://www.policeforce.go.tz). Currently Y chromosome STR markers are not widely used in forensic casework in Tanzania.

The aim of the present study was to evaluate the power of discrimination and genetic (diversity) parameters in the Y chromosome extended haploytpe markers in populations

of Tanzania for forensic and populations studies. Eleven Y chromosome extended haplotype markers were selected for this study, these includes Minimal haplotypes markers i.e. DYS19, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS389I/II and two additional markers DYS438 and DYS439. Six populations of Tanzania were investigated under this study. These populations were selected based on the language family categories; Niger Congo (Kuria and Sukuma), Nilo Saharan (Luo and Maasai) and Afro Asiatic (Iraqw and Alagwa). Analyses such as AMOVA, Fst population pairwise comparison, Factorial component Analysis by using Genepop, Arlequine and Genetix software packages were used to obtain Allele frequency, haplotype frequency, gene diversities among various loci and levels of gene flow between populations. The information content of these parameters were used to study population diversities in Tanzania from historical perspectives and discrimination potential of the markers for forensic application in Tanzania.

Chapter 2

Literature Review

2.1 General overview of DNA

As suggested by Butler (2005) the information in the Deoxyribonucleic Acid (DNA) are used for cell self replication and construction of enzyme required by organism to function and reproduce. DNA is a genetic material that is located within the nucleus of the cell (except red blood cell which lack nuclei). DNA is divided into dense packets of Chromosomes and protection proteins called histones. Within the body every cell have the same chromosome makeup because of precision of chromosome distribution during the cell division process, hence the cell from various tissues such as blood, hair, skin, Saliva, and semen have the same DNA content with the same genetic information (Crow *et al*., 1996).

The structure of a DNA is composed of two strands with nucleotide units that made up of nucleobase, deoxyribose sugar and phosphate. Deoxyribose sugar and phosphate makes the backbone and the bases attached to it. There are four types of nucleobases: Adenine

(A), Thymine (T), Cytosine (C) and Guanine (G). Hydrogen bonds are formed between the bases in two stands; Adenine always pairs with Thymine, Guanine always with Cytosine (Watson and Click, 1953). The various combinations of these bases results in the diverse biological differences among human beings and all living creatures.

Normal human genome consists of 23 pairs of chromosomes per cell, one member of each pair comes from the mother and the other from the father. Chromosomes 1-22 are usually not involved in sex determination and are termed autosomes while chromosomes X and Y are the sex chromosomes. Females are designated XX because they contain two copies of the chromosome X while males are designated XY and contain one copy of chromosome X and Y. Somatic cells contain two sets of each chromosome, and are termed diploid. Conversely, gametes, reproductive cells have only a single set of chromosomes and are designated as haploid (Crow *et al*., 1996).

Approximately 99.7% of human DNA is invariable between individuals. The remaining 0.3% (approximately one million nucleotides) is variable from one individual to the other. These differences provide the opportunity for using DNA sequence information for such endeavors as human identification purposes (Butler, 2001).

2.2 DNA fingerprinting methods

There are two important categories of tandem repeat have been used widely in genetics these are minisatellites (also referred to as variable number tandem repeats, VNTRs) and microsatellites (also referred to as Short Tandem Repeats, STRs). The general structure of VNTRs and STRs is the same; variation that exists between different alleles is caused by a difference in the number of repeat units. It is for this reason the tandem repeat polymorphisms are known as length polymorphisms (Goodwin *et al*., 2007).

Minisatellites (VNTR) are made up of repeated sequences that can vary in unit length from 6 to 100 bases. The variation in number of repeat creates allele that range in size from 500bp to over 30kb. These units can be repeated two to several hundred times at each minisatellite. Thousands of different minisatellites are scattered throughout the genome,

but they are often clustered near the telomere, or end of the chromosome (Saad, 2005). VNTR loci are particularly convenient as markers for human identification because they have very large number of different alleles and very high mutation rates leading to changes in length (Crow *et al*., 1996). VNTR loci are determined by a technique called VNTR profiling which involves specific enzymes that cut DNA into small fragments including the VNTR to be analyzed. The fragmented DNA is then placed in a small well of a semisolid gel. Each of a different DNA samples is placed in a different well. Additional wells receive various known DNA samples to serve as controls and fragments size indicators. The gel is then placed in an electric field and the DNA migrates away from the wells. The smaller the fragments, the more rapidly it moves (Crow *et al*., 1996)

Microsatellites (STR) are made up of a unit that can vary in length from 1 to 7 bases and repeats range from 50 to 300 pb. This unit is repeated 5 to 100 times at each microsatellite locus. Thousands of different microsatellites are randomly scattered throughout the genome, but not in a specific area (Saad, 2005). Microsatellites are determined by PCR based DNA typing. DNA amplification by PCR provides increased sensitivity, thus allowing small amounts of DNA to be analyzed, even from archival and partially degraded samples (Saad, 2005).

2.3 Short Tandem Repeats (STRs)

STR have become the most useful DNA markers because: they are robust, they can be applied to the analysis of a wide range of biological material; the results generated in different laboratories are easily compared; they are highly discriminatory, especially when analysing a large number of loci simulataneously (multiplexing); they are very sensitive, requiring only a few cells for a successful analysis (easily amplified by polymerase chain reaction); and there is a large number of STRs throughout the genome that do not appear to be under any selective pressure (Goodwin *et al*., 2007).

STR loci are spread throughout the genome including the 22 autosomal chromosomes and in the X and Y sex chromosomes, and usually occur in different sizes and their sequences are designated by the length of the core repeat unit and the number of contiguous repeat units. For example dinucleotide repeats have two nucleotides while trinucleotides have

three nucleotides in the repeat unit. Tetranucleotide repeats have become the most popular STR markers for human identification because of their reduced amount of stutter percentage (Butler, 2005).The number of repeats in STR markers can be highly variable among individuals, which make them effective for human identification purposes (Butler, 2005).

STRs are often divided into several categories based on the repeat pattern. Simple repeats contain units of identical length and sequence, compound repeats comprise two or more adjacent simple repeats, and complex repeats may contain several repeat blocks of variable unit length as well as variable intervening sequences. Complex hypervariable repeats also exist with numerous non-consensus alleles that differ in both size and sequence and therefore challenging to genotype reproducibly (Butler, 2005).

Hundreds of STRs have been identified in the last decade, analysed and selected for forensic casework, specifically in human identification (Moretti *et al*., 2001). Autosomal STRs, although useful for human identity testing, have limitations that decrease their effectiveness for use in sexual assault cases. In sexual assault cases, biological material available for analysis is often a mixture of material from the male assailant and the female victim. The availability of Y chromosome specific STRs has in indeed helped to overcome this problem.

2.4 Y chromosome

Quintana-Murci and Fellous (2001) elaborated the structure of Y chromosome as being the smallest in human genome (approximately 60 Mb in length) representing around 2%–3% of a haploid genome, and is composed of two pseudoautosomal regions (PAR 1 and 2) which are located in the short (Yp) and long(Yq) arms. In these regions the Y chromosome pairs and exchanges genetic material with the X chromosome during male meiosis in the same manner as autosomal genes. PAR1 and PAR2 represent the 5% of the entire chromosome.

In each male there is only one Y chromosome. The male determining gene (SRY gene) resides in the male specific (MSY) or nonrecombining (NRY) region near the distal end of

the short arm. There is growing evidence that the MSY region of chromosomes may have some ability to recombine within itself (Buckleton *et al*., 2005).

About (95%) of the length of the Y chromosome is made by the Non-Recombining Y (NRY) region, in this region there is no X- Y crossing during meiosis stage of cell division. The Non-Recombining region of the Y chromosome includes the euchromatic and heterochromatic portion. The heterochromatic portion is considered genetically inert while the euchromatic portion has numerous highly repeated sequences but also contains some genes responsible for important biological functions (Quintana-Murci and Fellous, 2001).

The sequences in the NRY region of the Y chromosome is inherited in a patrilineal manner and do not recombine during meiosis they are passed down from one generation to another without changing, except when there is occasional mutation (Knijff, 2000). This non-independent segregation of genetic markers on the Y chromosome, which is in sharp contradistinction to the independently segregating behavior of commonly used autosomal STR markers, results in reduced genetic variability (Hall and Ballantyne, 2003). In Y chromosome markers the genetic information is referred to as a haplotype rather than genotype because there is only a single allele per individual.

Y chromosome contains the large number of polymorphisms including variable number and short tandem repeats (Jobling *et al*., 1998). It is responsible for important biological roles such as sex determination and male fertility. It has also become a powerful tool in studying human populations and evolutionary pathways. Since the nonrecombining portion of the Y retains a record of the mutational events that have occurred along male lineages throughout evolution, therefore the study of the different mutations this molecule has accumulated along its evolution may be highly informative in deducing the histories of human populations (Quintana-Murci and Fellous, 2001).

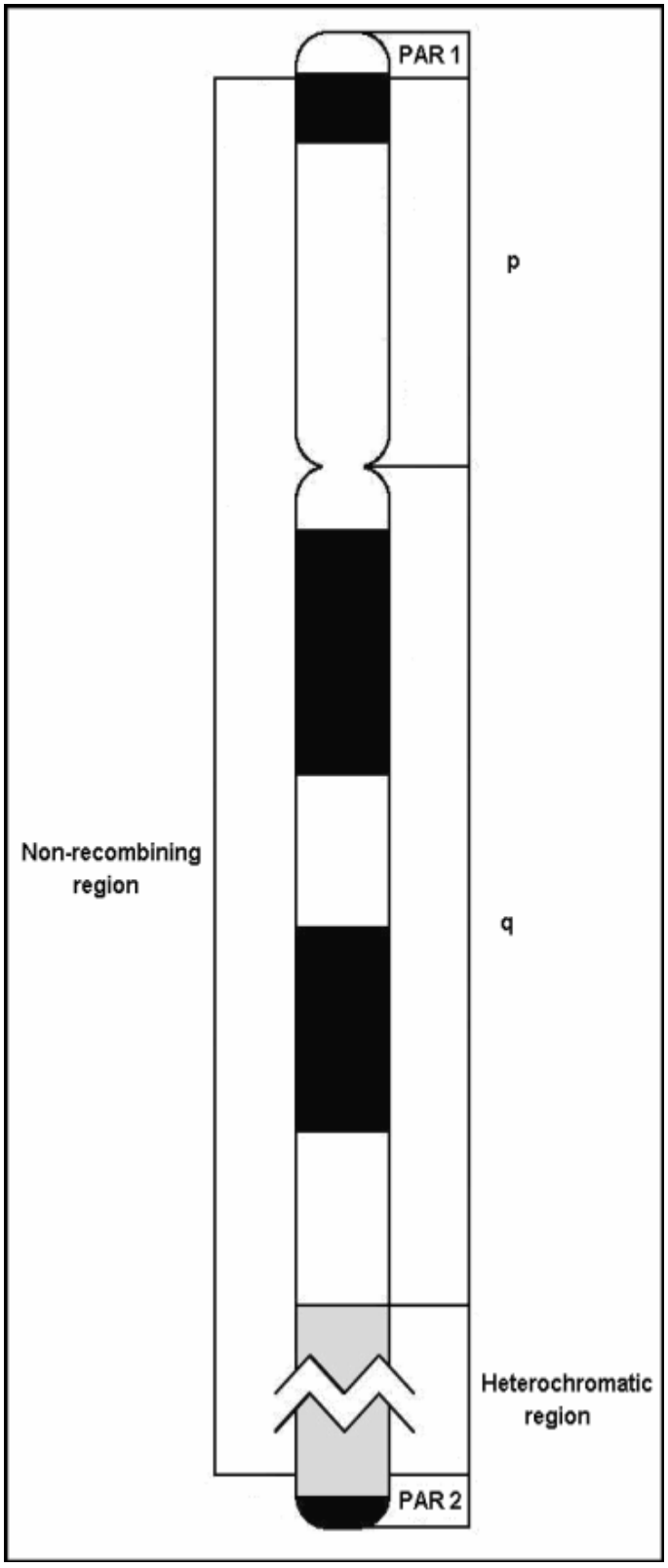

Figure 2.1. Simplified Y-chromosome structure representing the 3 regions: pseudo-autosomal (PAR1 and PAR2) in the short (p) and long (q) arms, heterochromatic and the non-recombining (Jobling and Tyler-Smith, 2000).

2.4.1 Y chromosome STR

The number of Y chromosome short tandem repeat (Y-STR) loci available for use in human identity testing has increased considerably since the turn of the century. In the 1990s only few Y STR markers were characterized and available for use. The first STR locus to be identified on the Y chromosome was DYS19 (Roewer *et al*., 1992).

A series of highly polymorphic Y-specific microsatellites have been identified and tested on different population samples. These markers show high levels of heterogeneity within and between populations and thus very useful for population genetic, evolutionary and forensic applications (Knijff *et al*., 1997).

A core set was of Y-STR loci was selected for human identity testing in 1997 that continue to serve as 'minimal haplotype' loci (Kayser *et al*., 1997; Pascali *et al*.,1998). The minimal haplotype is defined by the single copy Y-STR loci DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, and highly polymorphic multicopy loci DYS385 a/b (Schneider *et al*., 1999).

In 2004 The Scientific Working Group on DNA Analysis Methods (SWGDAM) voted to adopt the 11 Y-STR loci for forensic casework analysis. The decision was based on availability to the scientific community and the large amount of published performance and database information for most to these loci. The committee encourages further study of additional loci as to their suitability for forensic use. The first nine loci comprise the minimal haplotype complement of markers, plus two other additional markers DYS 438 and DYS 439 (Ayub *et al*., 2000; Ballantyne *et al*., 2004).

The limitation of Y-STRs compared with autosomal STRs is a reduced power of discrimination due to a lack of recombination throughout most of the Y-chromosome (Mullero *et al*., 2006). But this has an advantage because male relatives share for several generations an identical Y-STR profile, the Y chromosome analysis has a potential to make inferences on the population of origin of a given DNA profile (Roewer *et al*., 2009).

Y-STRs have various applications in forensic DNA analysis. The advantages of Y-STR analysis over autosomal STRs includeds: a) male profile can be obtained in the presence of large amounts of femaleDNA; b) differential extraction of sperm and non-sperm fraction is not necessary; c) analysis of azoospermic semen samples from vasectomized males is feasible; e) the number of male contributors often can be determined in multiple rape cases because of the haploid nature of the Y-STRs; 5) rapid exclusion of suspects can occur; f) interpretation is simplified due to single allele per locus profile; g) in deficient paternities and h) multigeneration male lineage studies can be performed (Shewale *et al*., 2004).

Due to the duplicated, palindromic regions of the Y chromosome, some Y-STR loci occur more than once and when amplified with a locus specific set of primers produce more than one PCR product. This fact can lead to confusion in terms of counting the number of loci present in a haplotype. A single set of primers can produce two amplicons, which may be thought of as 'two loci' for a Y chromosome haplotype. For example the Y-STR locus DYS385 is present in two regions along the long arm of the Y chromosome. These duplicated regions are located about 40,000 bp apart and can generate two different alleles when amplified with a single set of primers. The two alleles are typically labelled 'a' and 'b' with 'a' designation going to the smaller sized allele. It also possible to have both 'a' and 'b' alleles be the same size in which case only a single peak would appear in an electropherogram. Due to the presence of two alleles, this duplicated locus is usually referred to as DYS 385a/b. Two PCR products can also be generated at the locus DYS389I using a single set of primers resulting in DYS389II which is a subset of DYS389I (Butler, 2005).

In some cases duplications or even triplications of a Y-STR locus have been reported, particularly for DYS 19. It is important to keep this fact in mind so that two peaks at the DYS 19 locus are not automatically interpreted as coming from a mixture of two males. Both of these issues, mutations and duplications of loci impact analysis and therefore confusing mixture interpretation, suggest that analysis of additional Y-STR loci can be helpful in these situations (Butler, 2005).

2.4.2 Y-STR Database

The importance of DNA databases are for comparison purposes to understand how frequent and how rare a crime scene DNA profile may be in a particular population and are often generated by individual forensic laboratories to assess variation in common local populations. The primary goal of generating a population database is to find all common alleles and sample these alleles multiple times in order to reliably estimate the frequency of alleles present in the population under consideration (Roewer *et al*., 2000).

The largest and most widely used Y-STR database was created by Lutz Roewer and colleagues at Humbult University in Berlin, Germany and has been available online since 2000 (Roewer *et al*., 2001). The information in this database comes from 89 collaborating institutions located in 36 different countries.

Since the Y-STR haplotype database (YHRD) project was initiated, four major objectives have been pursued: first was to establish a standardized, highly informative and stain-sensitive haplotyping method for mapped, sequenced and `multiplexable' Y-chromosomal STRs; second was to introduce a means of minimum quality control for forensic laboratories using Y-STRs; third was to assess the extent of population stratification among males of European origin, but recently also to assess the extent of population stratification among males in various population of the world; and finally, to obtain reliable estimates of Y-STR haplotype frequencies for use in forensic practice (Roewer *et al*., 2000).

As of December 30th, 2010 the database contains 91,601 haplotypes with information from various markers (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635, and YGATAH4) as well as Y-SNPs from 710 populations sampled worldwide and can accessed in the websites http://www.ystr.org or http://www.yhrd.org.

The website database has become an important tool for use in comparing Y-STRs from different populations across the world. The website standardizes Y-STR nomenclature by

ensuring the successful completion of a quality assurance exercise. All Y-STR population data accepted is done so with the understanding that a population is defined as a group of more than 50 individuals living in the same area. When using Y-STRs it is advisable to understand the data obtained with a look towards the origin of the population (Willuweit and Roewer, 2007).

Other Y-STR databases are also accessible in the following websites www.ystr.org.usa, www.ybase.org and www. Appliedbiosystems.com/yfilerdatabase.

2.5 Polymerase Chain Reaction (PCR)

In the mid of 1990s, the technology changed to encompass the use of PCR (Mullis et al., 1986) in the analysis of STR loci. The PCR reaction is like a molecular photocopier, the specific region of DNA is replicated to yield many copies of particular sequence like in the enzymatic processes of DNA replication.

The reagents for PCR are DNA template, Oligonucleotide primers, deoxynucleotide triphosphates (dNTP), Magnesium chloride, buffer, and thermostable DNA polymerase such as *Taq* DNA polymerase. Bovine serum albmin (BSA) may also be added which binds soluble inhibitory factors that copurifies with DNA and also stabilize the *Taq* DNA polymerase. All ingredients are mixed in a reaction tube that is inserted into a thermal cycler which enables cyclical change in temperature (Mullis *et al*., 1986).

The amplification involves two oligonucleotide primers that flank the DNA segment to be amplified and repeated cycles of heat denaturation of the DNA, annealing of the primers to the complementary sequences and extension of the annealed primers with DNA polymerase (Saiki *et al*., 1988).

In the denaturation stage the reaction is heated to 94°C, the double stranded DNA molecule to melt forming two single stranded molecules. DNA melts at this temperature because hydrogen bonds that hold the two strands of the DNA molecule together are relatively weak. Between 50°C and 65°C the oligonucleotide primers anneal to the template DNA. These primers hybridize to opposite strands of the target sequence and

are oriented so as DNA synthesis by the polymerase proceeds across the region between the primers, effectively doubling the amount of that DNA segment in each circle this occurs at 72°C (Saiki *et al*., 1988).

PCR has been used in the examination of nucleotide sequence variations and chromosomal rearrangements, high efficiency cloning of genomic sequences, direct sequencing of mitochondrial and genomic DNAs and detection of viral pathogens (Saiki *et al*., 1988). Degraded DNA samples can also be amplified by PCR so is a useful tool especially in the analysis of forensic material, which may contain degraded DNA due to age of the sample, environmental exposure or chemical treatment (Budowle *et al*., 2000).

In polymerase chain reaction more than one region of DNA can be amplified simultaneously by adding more than one set of primer to the reaction mixture, this is commonly known as multiplexing. For multiplex reaction to work properly, the primer pairs need to be compatible i.e. the primer annealing temperatures should be similar and excessive regions of complementarity should be avoided to prevent the formation of primer-dimmers that will cause the primers to bind to one another instead of the template DNA. Multiplex PCR application is likely to require some degree of optimization in either the reagent components or thermal cycling conditions. Extensive optimization is normally required to obtain a good balance between amplicons of various loci being amplified (Butler, 2001).

2.6 General history of population of Tanzania

European exporation to the interior of Tanganyika (mainland Tanzania) started by the mid of nineteenth century. From 1880 to 1918 Tanganyika was colonised by Germany and from 1919 it became under British rule untill its independence in 1961 (Illife, 1979).

Two events in the nineteenth century had significant impact on population developments in Tanzania. First the Ngoni (Bantu) peoples arrived from southern Africa who found their way north along both sides of Lake Malawi between 1840 and 1845. The streams of migrants converged to the northeast of the Lake Malawi and formed two conguest states in the 1860s. The second event that affected population growth involved the expansion of

trade in the coast, with the ivory as the most valued commodity in the beginning. After 1860 the slaves become an important item of commerce to meet the labour needs of newly established plantations on Mauritius and Reunion Islands and to satisfy the ongoing demand for domestic in the Arabian Peninsula. The main caravan routes crossed central Tanzania (Newman, 1995).

The history of the late 19th-century disasters throughout East Africa has been told many times (Kjekshus, 1977; Koponen, 1988; Giblin, 1992). Among the causes that lead immigrants from one place to search for home into another place to settle and call their own, were diseases and competition for resources between neighbouring populations. This was contributed by the widespread of social breakdown in 19th Century, where el Nino drought patterns and the introduction of new diseases, like smallpox and syphilis, from the coastal caravan trade affected most of arrears in East Africa. The end result of these disasters was ecological collapse as people moved away from their home areas and tsetse fly bush recolonized the land, leading to an epidemic of sleeping sickness leading into further loss of habitable land. As a result, conflict between communities increased as they fought for their own survival both physically and socially (Shetler, 2007).

In the interior of eastern Africa, at various places Khoisan, Cushitic, Nilotic and Bantu peoples (an example of all four can be found within Tanzania) met one another to produce the most complex ethnolinguistic region in the continent (Sutton, 1974; Newman, 1995).

The diverse range of environments in Tanzania has made it possible for distinct groups practicing different economies to live side by side without rigorous competition. Thus it is possible for pastoralists to live on open savanna while agriculturalists are farming the adjacent woodlands and bush. Hunters and gathers can avoid competion with herders and farmers by exploiting ecological zones such as montane forests ad arid bush, and by developing exchange relationship with their neighbours (Ambrose, 1982).

2.6.1. Cushites (Afro Asiatic)

The Cushites originated from southern Ethiopian highlands, their movement into East Africa is estimated to have occurred between 3000 and 5000 years ago. The southern cushites in East Africa were agriculturalists. The ancestral southern Cushitic communities were pastoralist, the evidence of ancient Southern Cushitic vocabulary suggests that they herded goats, sheeps and cattles; in addition they were also cultivators (Ehret, 1974).

In Tanzania the modern speakers of this branch of Cushitic are found in the northern part of the country and are represented by isolated groups from the Mbulu highlands to the usambara hills. The populations that are Afro Asiatic include Asa, Iraqw, Gorowa, Alagwa, and Burunge (Ambrose, 1982).

2.6.1.1. Iraqw

The Iraqw are found in the Arusha and Manyara provinces of north-central Tanzania, with settlement in the Mbulu Highlands near Rift valley wall and south of Ngorongoro Crater. They are also called Erokh, Iraku, Kiiraqw, Mbulu, Mbulunge (http:www.ethnologue.com). The Iraqw language belongs to the South Cushitic branch of the Afro Asiatic family. The Iraqw are agriculturalists: maize, beans, red and white sorghum, sweet potatoes, millet; animal husbandry: cattle, sheep. Iraqw follows Christian, Muslim and Traditional religion.

Both the German and British colonial administration in Tanganyika (present day mainland Tanzania) described the Iraqw as particularly capable and knowledgeable farmers who practised soil conservation and cultivated their land intensively. The history of Iraqw intensive farming in the Mbulu Highlands has previously been associated with a growing population pressure in their homeland area, which in turn was thought to have acted as a primary motivation for the Iraqw to start using land more intensively for crop cultivation. According to this historiography, the Iraqw where confined to a relatively closed society in Mbulu Highlands because of the fear of hostilities and risk of cattle raids and attacks by other ethnic groups, mainly from their neithbouring pastoral Maasai and Datooga populations, who where in control of large parts of the Mbulu Highlands during the nineteenth century (Baumann, 1894). The view that Iraqw society had been isolated and confined drawn support from an argument by Fosbrooke that the Iraqw changed their

style of housing, from round houses with high conical thatched roofs to low roofed houses, as a defensive measure against Maasai raids during the early nineteenth century (Fosbrooke, 1954).

However during the middle of the twentieth century, the Iraqw settlement and political and economic influence expanded rapidly in the Mbulu Highlands and beyond, while the pastoral peoples (Datoga and Maasai) who had formerly dominated most of these areas were marginalised. This expansion process has in turn been interpreted as a consequence of the Iraqw no longer being forced to live in their densely populated nineteenth century homeland area.

2.6.1.2 Alagwa

The Alagwa is an ethnic group based at Kondoa District in Dodoma Province, central Tanzania. They are also called Chasi or Wasi. They predominantly occupy 15 villages in a rocky hill region that ranges in elevation between 1200 and 2200 meters above sea level (Chalcraft, 2005). Linguistically they are classified as Afro-Asiatic, Cushitic with alternate names Alagwaisi, Alagwase, Alawa, Asi, Chasi, Kialagwa and Wasi. About ninety percent of Alagwa are Muslim and follow the Muslim beliefs system. (http://www.ethnologue.com). They practise Agriculture as well as animal husbandry. Their principal crops are maize and millet. They work cooperatively to cultivate large fields, raise livestock, build houses and roads, and dig wells.

The Alagwa are a very reserved people; slow to receive outsiders. Some of their reserved behavior is evident in the way they build and care for their homes. Their houses are built of fired brick with distinctive grass roofs that peak high above the entry. The homes are often surrounded by high and sturdy fences made of grass, reeds, or sticks. Alagwa claim that they are owners of the rain, and believe that they can make rain. The process of rainmaking involves sacrificial domestic animals. In Alagwa culture rainmaking is not only a religious ceremony but also a magical one (Ten Raa, 1969; Chalcraft, 2005).

2.6.2 Nilotes (Nilo Saharan)

The expansion of the western Nilotes took place over 400 – 500 years and did not terminate until near the end of the nineteenth century (Newman, 1995). They moved from Upper Egypt to south through Uganda and the western highlands of Kenya, eventually as far as Lake Eyasi area in northern Tanzania. The Nilotes appears to have been few in east of the Kenya Rift Valley (Ambrose, 1982). In Tanzania the modern speakers of Nilo Saharan language includes Aramanik, Maasai, Datooga, Luo and Dorobo populations (http://www.ethnologue.com).

2.6.2.1 Luo

The Luo are found in Mara province of Tanzania. Alternate names of Luo are Dholuo, Kavirondo, and Kidjaluo (http://www.ethnologue.com). They follow Christian, Muslim as well as Traditional religion. Luo began to migrate from Southern Sudan looking for new land, fresh pastures, or better fishing grounds reaching as far as Tanzania coasts of Lake Victoria (Ehret, 1974).

Throughout the mid- and late twentieth century, they have lived in densely settled land of separated farmsteads, communities are linked by marriage and other kin ties. Although the Luo farm the land in order to produce an adequate food supply, they are first and foremost cattle herders. In East Africa Luo are also found in Kenya and Uganda (Ehret, 1974).

The Luo are distinctive in that they adapted to many different ecologies and economies as they spread throught the entire region. Unlike Maasai, the Luo did not define their ethnicity by their subsitence pattern as herders, hunters or farmers. In every place they went, Luo speakers interacted intimately with people speaking other languages and practicing other cultures thereby adapting new lifestyles and identity. In each place they had an inordinate influence on the people with whom they interacted, as the arrears they passed through Uganda and western Kenya now speak primarily Luo Language (Feierman, 1995).

2.6.2.2 Maasai

The Maasai are an indeginous African ethnic group of semi-nomadic people located in Arusha province, northern Tanzania. They speak Maa which is classified as a member of the Nilo Saharan, Eastern sudanic, Nilotic. Maasai follows Christian and Traditional religion (http://www.ethnologue.com). Majority of Maasai are pastoralist and have resisted adopting a more sedentary lifestyle, but some of them have adopted farming way of life (Berntsen, 1976).

The significant residential groupings are the locality, the village, and the polygynous homestead, or joint family. The life-style of the Maasai is oriented toward their herds of cattle, although sheep and goats play an important part in their diet, especially during the dry season, when milk is scarce. In East Africa Maasai also present in Kenya.

The Maasai developed nomadic herding, but some of them in Arusha province eventually adopted farming way of life. Migrations to find pastures and water give them opportuinities to access needed resources which were located mostly to their south, routes and cycles of herd movement tend to change, producing a patial drift that over time can markedly alter population distribution (Newman, 1995).

The Maasai maintain wide ranging links throughout their society based on age-sets that united all the men of a given age in a single comprehensive social institution. Economic responsibilities were divided functionally according to age and gender, with boys responsible for herding the cattle, young men for protecting them, women for milking household labour, and reproducing the family and older men for managing the family and its collective herds. Relations throughout pastoral society and with other society's beyond were also mediated by age, as elders sought among their age-mates for distant cattle partners to disperse their herds, and marriage partners to ensure their social survival (Spear and Waller, 1993). Pastoralists Maasai sought to maintain as wide a circle of exchange and marriage partners as possible by incorporating their agricultural and hunting and gathering neighbours within their age sets, while farmers married pastoral women and adopted pastoral men into their lineage. Each set of social relations was thus historical specific to a single means of production in order to mediate relations among them within a common mode of production (Spear and Waller, 1993). Most farmers with

whom Maasai interacted were Bantu or Kalenjin – Kikuyu, Chaga, Sonjo, Nandi and others. Women conducted frequent trade between pastoral and agricultural homesteads, bilingualism and intermarriage were common (Spear and Waller, 1993).

2.6.3 Bantu (Niger Congo)

Bantu speaking peoples first reached the intelacustrine area from their homeland West Africa 3,000 years ago. The first settlements congretated at Lake Victoria shore. Paleobotanical findings showing a rapid decline in forest cover suggesting that initial agriculturalists could have reached the lakes as early as 5,000 years ago and not later than 3,000 years ago. The Bantu cultivated yam sorghum and millet, also depended on fishing along with cattle keeping and thus gained the means for extending their settlements into the Eastern Africa (Newman, 1995).

Many of the Bantu speaking societies drew a major portion of their human ancestry from earlier African peoples of quite different languages and cultural backgrounds from southern Cushites and Southern Nilotes in Northern Tanzania; from central Sudanic and other Nilo Saharan speaking groups in the Great lake region (Ehret et al., 2001). In Tanzania the populations that their language have been categorized in Niger Congo language includes Ngoni, Sambaa, Nyamwezi, Sukuma, Chagga, Kuria, and Makonde populations (http://www.ethnologue.com).

2.6.3.1 Sukuma

Sukuma is one of the best known in Bantu language group that falls into Niger Congo language phylum (Haine and Nurse, 2000). The Sukuma are found in Mwanza province to the south of Lake Victoria (figure 2.2a), considered as a Bantu language with several mutually intelligible dialects. They are Pastoralists (cattle), agriculturalists (sorghum, millet, maize, rice, sweet potatoes, cassava, peanuts, beans, chick-peas, gourds, sunflowers, cotton, tobacco) also practice fishing (http://www.ethnologue.com).

Historically Sukuma lived in villages for protection from cattle raids in pre-colonial times and during the compulsory villagization period which started in 1976. Traditionally they were divided into number of chiefdoms whose rulers carried out religious rituals to ensure

the welfare of the people who otherwise acted on the advice of their elders. The Sukuma follows Christian, Muslim and Traditional regions (Tanner, 1999).

2.6.3.2 Kuria

Kuria are found in Mara province, Tarime and Musoma rural districts. They are also found in Kenya and they are classified as Niger-congo, Bantoid. Alternative names for this tribe are Igikuria, Ikikuria, Kikuria, Kurya, Kurye, Tende (www.ethnologue.com). Majority of Kurias are Christian and others follow traditional religion. In East Africa the Kuria are also found in Kenya (www.ethnologue.com). The Kuria people are mainly agriculturalists and pastoralists, Tanzanian Kurians more towards pastoralism. Food crops include maize, sorghum, cassava, bananas, finger millet, Irish potatoes, sweet potatoes and onions. Coffee and maize are the primary cash crops. Most of highlands families own only one or two head of cattle often dairy breeds that can be maintained in small cowshed and thus require no land for grazing many other highlanders have merely relocated their herds to the lowlands arrears, where pasture is still plentiful, often placing them in the care of relatives, or acquiring a wife in the lowlands and setting up a separate homestead there to enable her to look after the cattle, or putting the cattle in the lowland area under the Kuria system of stock associateship (Fleisher, 2002).

For Kuria pre – colonial times, inter-clan warfare and raiding were frequent occurrences with mutual hostile clans fighting among one another and with individual clans sometime even combining with outsiders' example Maasai aginst fellow Kuria or rival clan (Ruel, 1991). Among the Kuria, market oriented cattle raiding and inter-clan warfare are tightly linked phenomena, resulting in demographic, and environmental stress in their society (Keen, 1996).

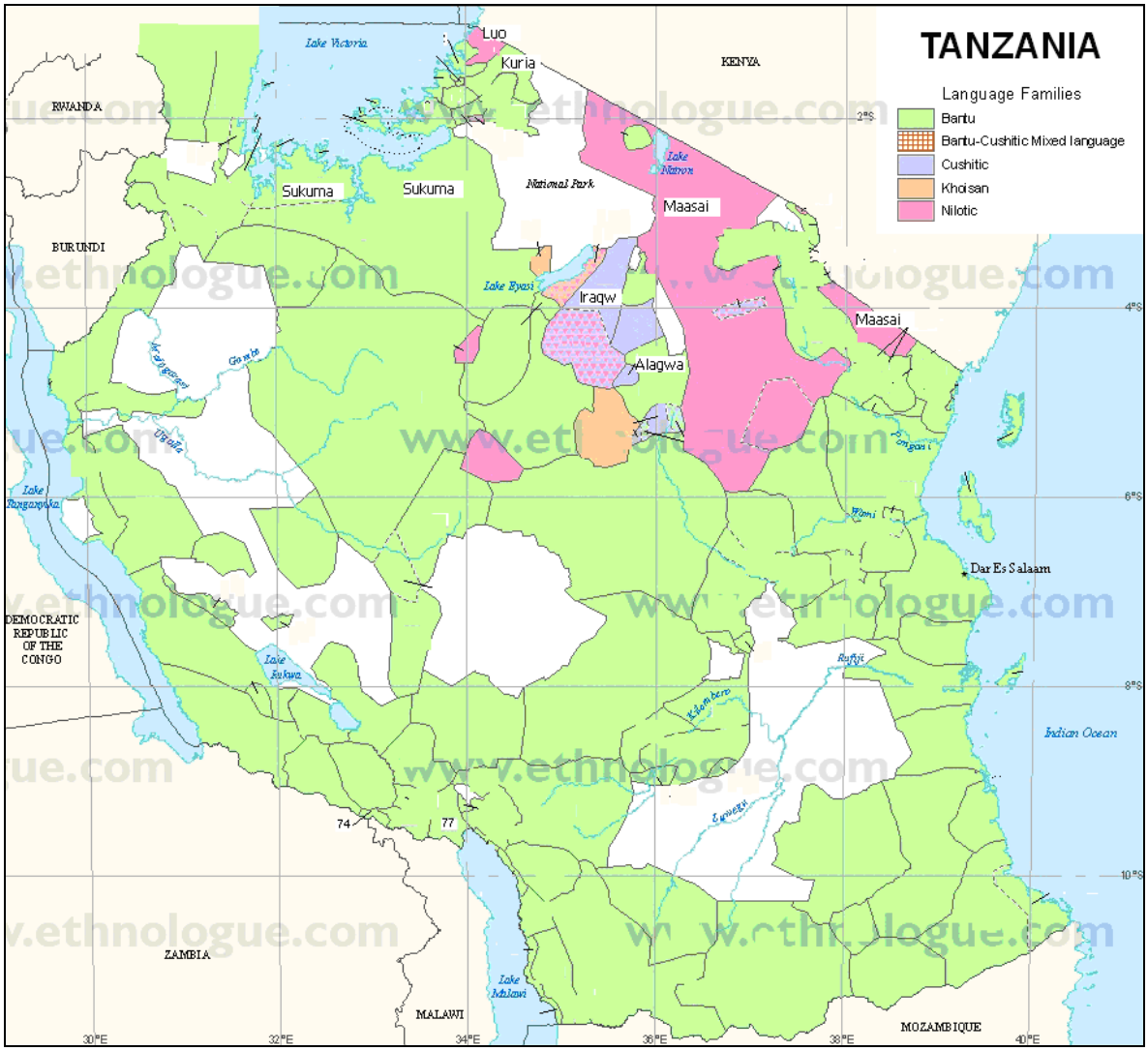

Figure 2.2a ; a map of Tanzania indicating four linguistic families Khoisan, Afro Asiatic, Nilo Saharan and Niger Kordofanian that are found in Tanzania and the six populations that are being studied in the present work (www.ethnologue.com).

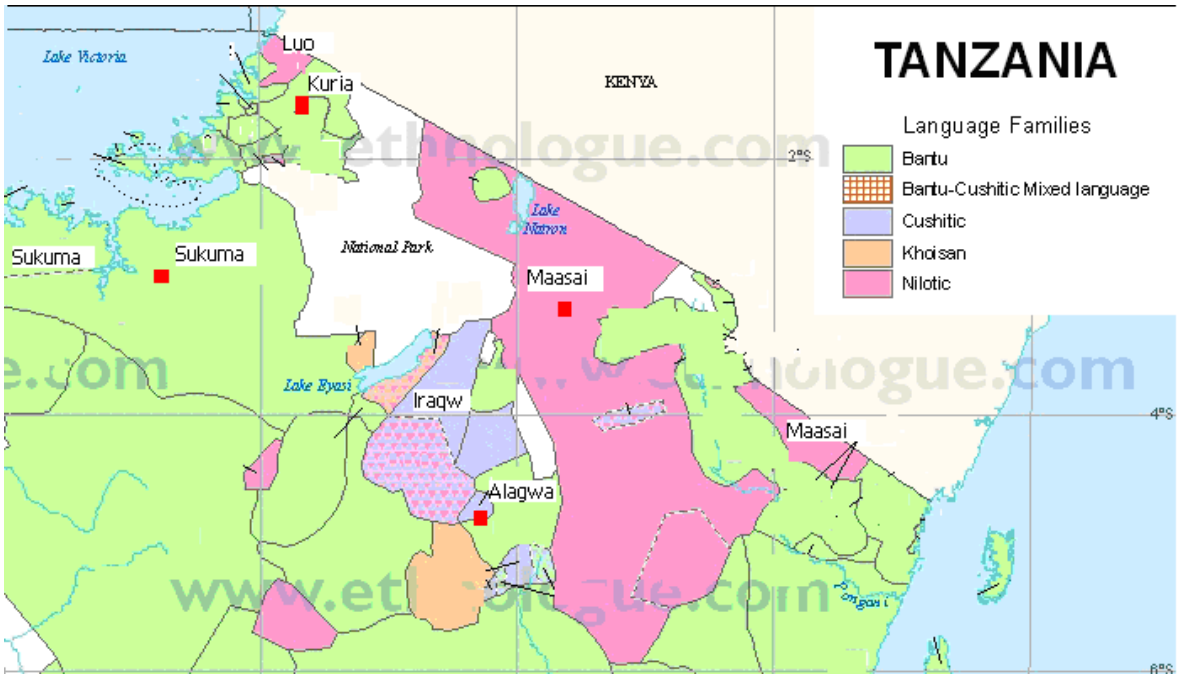

Figure 2.2 b ; a cross section of a map of Tanzania indicating six populations investigated in this study. Nilo Saharan (Luo and Maasai), Niger Congo (Kuria and Sukuma) and Afro Asiatic (Alagwa and Iraqw) (www.ethnologue.com). The red square indicates the regions where the samples were collected. Sukuma (Mwanza province), Kuria and Luo (Mara province), Maasai and Iraqw (Arusha province) and Alagwa (Dodoma province).

Chapter 3

Material and Methods

3.1 Sample collection

Buccal swabs were collected from unrelated males from Sukuma (21) in Mwanza province, Kuria (57) and Luo (34) residing in Mara province, Maasai (51) and Iraqw (67) residing in Arusha province and Alagwa (37) residing in Dodoma provinces of Tanzania (figure 2.2b). A total of 270 samples were collected with informed consent. Ethical approval was obtained from National Institute of Medical Research (NIMR) of the United Republic of Tanzania as well as the University of the Western Cape. Samples collected were then placed in folded perforated cardboard in an envelope and placed in the freezer set at -20°C.

3.2 Extraction and Quantitation of DNA

DNA was extracted from swabs using salting out method adopted and modified from (Medrano *et al*., 1990). The protocol essentially involves recovery of nuclei after disruption of cell membranes using detergents. Nuclei were then disrupted and protein material was salted out. DNA was recovered by ethanol precipitation (Medrano *et al*., 1990) (See Appendix). The DNA concentration and purity was measured using Nanodrop ND1000 spectrophotometer (See Appendix). DNA purity was obtained by the ratio of A_{260}/A_{280} where if the ratio falls in the range of 1.7 and 2.0, then the DNA is said to be of high quality.

3.3 Amplification.

3.3.1 Table A; Primer sequences (forward and reverse) for each of the Y-STR loci, fluorescent dye labels and final concentration of all primers used in multiplex reaction.

s/n	Y-STR Locus	Dye	Primer sequence (5'-3')	[Primer]	Length	Molecular Weight
1	DYS19	NED	F 5'ACT ACT GAG TTT CTG TTA TAG TGT TTT T 3'	1 μM	28	9292

			R 5' GTC AAT CTC TGC ACC TGG AAA T 3'	1 µM	22	6694
2	DYS385a/b	VIC	F 5' AGC ATG GGT GAC AGA GCT A 3' R 5' GCC AAT TAC ATA GTC CTC CTT TC 3'	0.2 µM 0.2 µM	19 23	6605 6910
3	DYS 389I/II	6-FAM	F 5' CCA ACT CTC ATC TGT ATT ATC TAT G 3' R 5' GTT ATC CCT GAG TAG TAG AAG AAT G 3'	0.4 µM 0.4 µM	25 25	8066 7745
4	DYS390	PET	F 5' CAA TGT GTA TAC TCA GAA ACA AGG 3' R 5' CAC ATA TAT TTT ACA CAT TTT TGG G 3'	0.8 µM 0.8 µM	24 25	8293 7621
5	DYS391	6-FAM	F 5' TTC AAT CAT ACA CCC ATA TCT GTC 3' R 5' GAT AGA GGG ATA GGT AGG CAG GC 3'	0.2 µM 0.2 µM	24 23	7731 7243
6	DYS392	NED	F 5' TAG AGG CAG TCA TCG CAG TG 3' R 5' GAC CTA CCA ATC CCA TTC CTT 3'	0.6 µM 0.6 µM	20 21	6893 6261
7	DYS 393	VIC	F 5' GTG GTC TTC TAC TTG TGT CAA TAC 3' R 5' GAA CTC AAG TCC AAA AAA TGA GG 3'	0.25 µM 0.25 µM	24 23	8028 7099
8	DYS438	6-FAM	F 5' CCA AAA TTA GTG GGG AAT AGT TG 3' R 5' GAT CAC CCA GGG TCT GGA GTT 3'	0.25 µM 0.25 µM	23 21	7676 6462
9	DYS439	6-FAM	F 5' TCG AGT TGT TAT GGT TTT AGG TCT 3' R 5' GTG GCT TGG AAT TCT TTT ACC C 3'	0.18 µM 0.18 µM	24 22	7935 6707

3.3.2. Table B; General information of the extended haplotype loci that were used (Butler et al., 2002). It has to be noted that primer sequence for the locus DYS390 was re-designed by Dr. Maria Eugenia D'Amato from the Forensic DNA Laboratory at the University of the Western Cape (UWC), following its failure in amplification in multiplex reaction.

s/n	Y-STR Loci	Genbank Accession	Reference Allele	Repeat structure	Allele range	Size range (bp)
1	DYS19	AC017019	15	$[TAGA]_3TAGG[TAGA]_{12}$	10-19	233-269
2	DYS385a/b	AC022486	11	GAAA	7-23	242-306
3	DYS389I	AC004617	12	$[TCTG]_3[TCTA]_9$	9-17	143-175
4	DYS389II	AC004617	29	$[TCTG]_5[TCTA]_{12}[TCTG]_3[TCTA]_9$	26-34	263-295
5	DYS 390	AC011289	24	$[TCTG]_8[TCTA]_{11}[TCTG]_1[TCTA]_4$	17-28	170-215
6	DYS391	AC011302	11	TCTA	7-14	93-121
7	DYS392	AC011745	13	TAT	6-16	290-320
8	DYS393	AC006152	12	AGAT	9-16	109-133
9	DYS438	AC002531	10	TTTTC	6-13	300-335
10	DYS439	AC002992	20	$[GATA]_2N_4[GATA]_3N_{14}[GATA]_1N_3[GATA]_1N_7[GATA]_{13}$	16-21	210-230

3.3.3 Multiplex Amplification

All loci were amplified in single multiplex reaction in a 10µl reaction volume. The reaction mixture comprising of 0.5µL superTherm Gold (5units/µL) (MEDOX *(PTY) Ltd),* 5ng genomic DNA, 1.5mM Magnesium Chloride buffer, 2mM of each dNTP (*Roche*), sterile water, 4mg/ml BSA, and 0.5% glycerol. Amplification were performed in a GeneAmp 2720 Thermal cycler (*Applied Biosystems*), under the following cyclic conditions: 95°C for 10 minutes; 32 cycles of 94°C for 30 seconds; primer annealing at 58°C for 1 minute and extension at 72°C for 1minute, extension of 1cycle at 68°C for 75 minute then to 12°C until the samples were removed from the thermal cycler and stored into fridge at 4°C.

3.4 Detection of amplified products

Analysis of amplified products was performed by mixing 1µL of amplified product with 1µL of loading mix consisted of de–ionized formamide; Dextran Blue dye (*Applied Biosystems*) and LIZ500 size standard (*Applied Biosystems*) in a ratio of 5:2:1 (µl). The mixture were then denatured using GeneAmp 2720 Thermal cycler PCR System set at 95 °C for 5 minutes. Immediately after denaturation, the samples were placed in ice for snap – cooling.

Thereafter the denatured products were electrophoretically separated on the ABI 377 Automatic Sequencer (*Applied Biosystems*). The data was collected with the ABI 377 collection software (*Applied Biosystems*), analyzed using GeneScan 3.0.0 software (*Applied Biosystems*). Allele designations were automatically assigned using Genotyper 3.7 software (*Applied Biosystems*).

3.5 Analysis

Summary statistics parameters such as allele frequencies and Gene diversity values for each locus per population group were calculated using the software GenePop version 4.0.7 (Rousset, 2007). Haplotype frequencies were obtained with the software Arlequin 3.11 (Excoffier, 2006). Discrimination capacity (DC) was calculated as the number of distinct haplotypes divided by the number of genotyped individuals, and was calculated using the formula below in Microsoft Excel spread sheet:

Discrimination Capacity = No distinct profiles/ Number of profiled individuals

Haplotype diversity HD was calculated in Microsoft Excel spread sheet using the following formula by Nei (1987): $HD = 1- \sum (Pi)^2 (n/n-1)$, where Pi = haplotype frequency and n = number of individuals.

3.5.1 Haplotype analysis and comparison of haplotypes with those in YHRD

Haplotypes that were obtained from each of the six populations were compared to those in YHRD. This was done in an attempt to establish possible links between these populations to other populations through their historical background, also to check the contribution of Tanzanian haplotype the YHRD. YHRD had been useful in assessment of

male population stratification among world wide populations by using Y STR haplotype frequency distributions. The database is also useful in generation of reliable frequency estimates for Y STR haplotypes to be used in quantitative assessment of matches in forensic casework.

3.5.2 AMOVA analysis

3.5.2.1 Non hierachial AMOVA analysis

Genetic variation among populations based on their ethnicity was estimated by non hierachial Analysis of Molecular Variance of six different populations based on their ethnicity. The method involves testing the significance of the variance components among populations (Va) and within populations (Vb) and ϕ-statistics (Fst), under the null hypothesis, samples are considered as drawn from a global population, with variation due to random sampling in the construction of populations. The hypothesis that is being tested was to check whether or not there is a significant genetic differentiation among these populations based on their ethinicity.

3.5.2.2 Hierachial AMOVA analysis

Genetic variation among the different populations based on their language groups was estimated by the partition of variance components (AMOVA) applying conventional Fst statistics using haplotype frequencies as implemented in Arlequin 3.11 (Excoffier, 2006). The populations were partitioned into three language groups each comprising of two populations; Afro Asiatic (Iraqw and Alagwa), Nilo Saharan (Luo and Maasai) and Niger Congo (Sukuma and Kuria). The hypothesis that is being tested is to check whether or not there is a significant genetic differentiation among the populations based on their language family categories (within populations, among populations within groups and among groups). The method involves testing the significance of the variance components and ϕ-statistics. Under the null hypothesis, samples are considered as drawn from a global population, with variation due to random sampling in the construction of populations. Each individual is allocated to the randomly chosen population, while holding sample sizes constant at the realized value. The variance components are estimated from each large number (normally 1000) of permuted matrices, this procedure is used to obtain

the null distribution and significance of Φst (FST) and Vc^2 (Excoffer *et al.,* 1992). To obtain a null distribution Φsc (FSC) and Vb^2 the method of permutation assumes the groups are real but the populations within them are not, this permutes individuals within groups without regard to population. And in the final analysis which is used to obtain the null distribution Φct (FCT) and Va^2 the method assumes the populations are real and the groupings are artificial while permuting whole populations across groups (Excoffier *et al*., 1992).

3.5.3 Factorial Component analysis

The degree of genetic relationship among populations was further investigated with FCA (Factorial Component Analysis) as implemented in Genetix v.4.05.2 (Belkhir K, 1999). In this method factorial coordinates are being used to establish the extent of genetic differentiation on subsamples (populations) using discrete variables (loci) which could have different states (alleles). Each population (subsamples) is situated along a factorial axis or on a factor plane relative to other population in function of allele distribution at the analysed loci (Guinand, 1995). The distance between populations along a factorial axis gives an idea of the genetic differentiation between populations. The FCA method has an added advantage as it allows independent estimates of the mean Fst i.e. eigenvalues, because each factorial axis is uncorrelated to the other, the distribution of the eigenvalues is equal to the distribution of different mean Fst estimates of the data set. This point is important because it allows the relationship linking the gene flow estimate to Fst (Guinand, 1995).

Chapter 4

Results and Discussion

4.1 DNA Extraction and Purification

Concentrations and purity readings were taken for all the DNA samples (See Appendix), it was observed that concentrations varied from 5.3 ng/µl to 203.8 ng/µl. DNA samples had purity readings ranging from 1.69 to 2.14 and maintained an average purity reading of 1.80.

4.2 Multiplex Amplification

Of the 270 DNA samples, multiplex amplification provided full haplotypes for 267 samples.

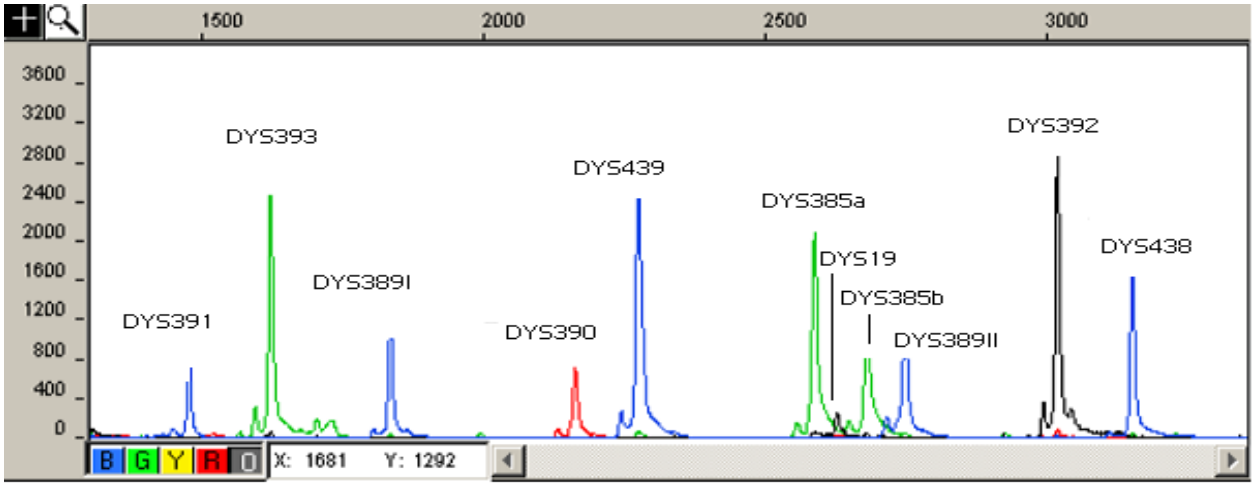

Figure 4.1: Graphical representation of a typical electropherogram of the Extended Haplotype where each locus is labelled.

4.3 Analysis of Allele and Haplotype Frequencies

Table 4.1 present allele and haplotype frequencies for extended haplotype loci among the Sukuma males from Mwanza province. Table 4.2 present allele and haplotype frequencies for the extended loci among the Kuria males from Mara province. Table 4.3 present allele and haplotype frequencies for the extended haplotype loci among Luo males from Mara province. Table 4.4 present allele and haplotype frequencies for the extended haplotype loci among Maasai Males from Arusha province. Table 4.5 present allele and haplotype frequencies for the extended haplotype locus among Iraqw males from Arusha province.

Table 4.6 present allele and haplotype frequencies for the extended haplotype loci among the Alagwa males from Dodoma province.

4.3.1 Allele and haplotype frequencies for Sukuma males

The highest *D* value for DYS385 among the Sukuma males was 0.8690 (Table 4.1). This was the highest gene diversity observed for all locus among the six sub-populations evaluated. The highest *D* value for a single-copy locus was 0.6381 (DYS393). The lowest *D* value was 0.0952 (DYS392). This was expected as almost 95% of the samples shared allele 11 for this locus, similar results were observed in South African Xhosa population where 99% of sampled population shared allele 11 in DYS392 locus (Leat *et al*., 2004).

Table 4.1. Allele frequencies for Extended haplotype loci among Sukuma males (n = 21)

Allele	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	H	DYS 385a, b
7										11,13	0.0952
8	0.048									15,15	0.0476
9	0.048									15,16	0.0952
10			0.571	0.095	0.048					16,16	0.0476
11			0.381	0.857	0.19				0.952	15,17	0.1429
12	0.143		0.048	0.048	0.619				0.048	16,17	0.1429
13	0.571				0.143	0.476		0.048		15,18	0.0952
14	0.19					0.381		0.095		16,18	0.0476
15						0.143		0.667		17,19	0.1429
16								0.143		14,20	0.0952
17								0.048		14,21	0.0476
18											
19											
20											
21							0.667				
22							0.048				
23											
24							0.143				
25							0.095				
26							0.048				
27											
28		0.143									
29		0.095									
30		0.429									
31		0.333									
36											
Gene Diversity	0.6429	0.7095	0.5524	0.2667	0.5857	0.6381	0.5476	0.5476	0.0952		0.869

H = Haplotypes

4.3.2 Allele and haplotype frequencies for Kuria males

The *D* value for DYS385 among the Kuria males was 0.7912 (Table 4.2). The highest *D* value for a single-copy locus was 0.7444 (DYS19). The lowest *D* value was 0.2538 (DYS392), which is to be expected as 86% of the samples shared allele 11 for this locus, similar results were observed in South African Xhosa population where 99% of sampled population shared allele 11 in DYS392 locus (Leat *et al*., 2004).

Table 4.2. Allele frequencies for Extended haplotype loci among Kuria males (n = 57)

Allele	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	H	DYS 385a, b
7										10,11	0.0351
8									0.018	12,13	0.0175
9			0.018						0.018	14,15	0.0175
10			0.719	0.07						15,15	0.0175
11			0.246	0.789	0.421			0.07	0.86	16,16	0.0351
12	0.105		0.018	0.14	0.439				0.105	14,17	0.0175
13	0.702				0.14	0.509		0.035		15,17	0.0351
14	0.175					0.211		0.14		16,17	0.2807
15	0.018					0.263		0.439		17,17	0.0526
16						0.018		0.175		15,18	0.0175
17								0.14		16,18	0.1754
18							0.053			17,18	0.1228
19										18,18	0.0351
20										15,19	0.0175
21							0.667			16,19	0.0526
22							0.053			14,20	0.0175
23										15,20	0.0351
24							0.07			16,20	0.0175
25							0.105				
26							0.053				
27											
28		0.07									
29		0.07									
30		0.404									
31		0.333									
32		0.105									
33		0.018									
Gene Diversity	0.4737	0.7174	0.4292	0.3584	0.6216	0.6385	0.5407	0.7444	0.2538		0.7912

H = Haplotypes

4.3.3 Allele and haplotype frequencies for Luo males

The *D* value for DYS385 among the Luo males was 0.8266 (Table 4.3). The highest *D* value for a single-copy-locus was 0.6916 (DYS19). The lowest *D* value was 0.1159 (DYS392), with 94% of the samples shared allele 11 for this locus, as expected the least variable loci have allele frequency distributions with a high frequency for one allele and few additional alleles. Similar results with highest level of polymorphism in DYS385 locus and lowest with DYS392 was also observed in Mozambique population (Alves *et al*., 2003)

Table 4.3. Allele frequencies for Extended haplotype loci among Luo males (n = 37)

Allele	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	H	DYS 385a, b
7										11,11	0.0882
8			0.029						0.029	11,12	0.0588
9			0.088							12,12	0.0294
10			0.882	0.265						11,13	0.0294
11				0.471	0.353				0.941	12,13	0.0294
12	0.206			0.265	0.559			0.029	0.029	12,16	0.0294
13	0.647				0.059	0.676				16,16	0.0882
14	0.147				0.029	0.294		0.147		14,17	0.0294
15						0.029		0.5		15,17	0.0588
16								0.118		16,17	0.2059
17								0.206		17,17	0.1471
18										14,18	0.0294
19										15,18	0.0588
20							0.029			16,18	0.0294
21							0.618			17,18	0.0294
22							0.088			14,19	0.0294
23							0.029			16,20	0.0294
24							0.176				
25							0.029				
26							0.029				
27		0.029									
28		0.059									
29											
30		0.559									
31		0.265									
32		0.088									
Gene Diversity	0.533	0.6239	0.2193	0.6578	0.5758	0.4688	0.5936	0.6916	0.1159		0.8266

H = Haplotypes

4.3.4 Allele and haplotype frequencies for Maasai males

The highest *D* value among the Maasai males was 0.8094 (DYS385) (Table 4.4). The highest *D* value for a single-copy-locus was 0.7176 (DYS19). The lowest *D* value was 0.1145 (DYS392). Gene diversity values for all loci (except DYS391 and DYS392) investigated among this sub-population were greater than 0.5.

Table 4.4. Allele frequencies for Extended haplotype loci among Maasai males (n = 51)

Allele	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	H	DYS 385a, b
7										11,11	0.0196
8			0.02							12,12	0.1373
9			0.098	0.02		0.02				12,13	0.0196
10	0.02		0.725	0.333	0.039				0.02	13,14	0.0196
11			0.157	0.608	0.549				0.941	13,16	0.0196
12	0.078			0.039	0.314	0.039				14,16	0.0392
13	0.569				0.098	0.451		0.137	0.039	15,16	0.0392
14	0.333					0.333		0.059		16,16	0.0392
15						0.157		0.412		14,17	0.0588
16								0.314		15,17	0.0196
17								0.078		16,17	0.1176
18										17,17	0.2157
19										15,18	0.0392
20										16,18	0.0588
21							0.627			17,18	0.0784
22							0.059			14,19	0.0196
23							0.059			16,19	0.0196
24							0.196			15,21	0.0196
25							0.059			18,21	0.0196
26											
27		0.039									
28		0.039									
29		0.039									
30		0.392									
31		0.431									
32		0.059									
Gene Diversity	0.5702	0.6651	0.4478	0.5278	0.6008	0.6722	0.5686	0.7176	0.1145		0.8094

H = Haplotypes

4.3.5 Allele and haplotype frequencies for Iraqw males

The highest *D* value among the Iraqw males was 0.8223 (DYS385) (Table 4.5). This was the second highest gene diversity observed for any locus among the six populations

evaluated. The highest *D* value for a single-copy locus was 0.7363 (DYS19). The lowest *D* value was 0.5124 (DYS392). Gene diversity values for all the loci in evaluated among this population were greater than 0.5

Table 4.5. Allele frequencies for Extended haplotype loci among Iraqw males (n = 67)

Allele	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	H	DYS 385a, b
7				0.015					0.015	9,11	0.0149
8			0.03							10,11	0.0448
9	0.015		0.045	0.103		0.015			0.045	12,12	0.0448
10	0.119		0.597	0.559	0.209				0.06	12,13	0.0299
11			0.313	0.309	0.388				0.687	11,14	0.0447
12	0.119		0.015	0.015	0.328	0.06			0.119	12,14	0.0149
13	0.328				0.075	0.373		0.403	0.06	15,15	0.0299
14	0.403					0.493		0.09	0.015	14,16	0.0149
15	0.015					0.06		0.239		15,16	0.0149
16								0.209		16,16	0.1791
17								0.06		14,17	0.0597
18										15,17	0.1343
19										16,17	0.1194
20							0.015			17,17	0.0597
21							0.284			15,18	0.0149
22							0.03			16,18	0.0746
23							0.119			17,18	0.0299
24							0.433			14,19	0.0149
25							0.119			16,19	0.0149
26		0.075								17,19	0.0299
27		0.104								16,20	0.0149
28		0.119									
29		0.179									
30		0.239									
31		0.194									
32		0.075									
33		0.015									
Gene Diversity	0.7114	0.8494	0.5504	0.5866	0.7028	0.6201	0.7133	0.7363	0.5124		0.8223

H = Haplotypes

4.3.6 Allele and haplotype frequencies for Alagwa males

The highest *D* value among the Alagwa males was 0.8780 (DYS385) (Table 4.6). This was the highest gene diversity observed for any locus among the six sub-populations evaluated. The highest *D* value for a single-copy locus was 0.7162 (DYS438 and DYS390). The lowest *D* value was 0.2402 (DYS392), indicating that this locus is a least variable in

all populations evaluated in this study. This is to be expected because about 86.5% of the samples shared allele 11 for this locus.

Table 4.6. Allele and haplotype frequencies for Extended haplotype loci among Alagwa males (*n* = 37)

Allele	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	H	DYS 385a, b
9			0.054	0.135						10,11	0.0541
10	0.027		0.784	0.378	0.054					11,11	0.1351
11			0.135	0.351	0.27				0.865	12,12	0.1081
12	0.027		0.027	0.135	0.568					12,14	0.027
13	0.649				0.081	0.649		0.081	0.135	14,14	0.027
14	0.297				0.027	0.297		0.054		14,15	0.0541
15						0.027		0.676		15,15	0.0811
16						0.027		0.135		14,16	0.027
17								0.054		15,16	0.0541
18										16,16	0.0811
19										13,17	0.027
20										15,17	0.027
21							0.432			16,17	0.0541
22							0.081			17,17	0.027
23							0.189			15,18	0.0541
24							0.27			17,18	0.0811
25							0.027			15,19	0.0811
27		0.027									
28		0.054									
29		0.081									
30		0.378									
31		0.459									
Gene Diversity	0.503	0.6532	0.3739	0.7162	0.6111	0.503	0.7162	0.527	0.2402		0.878

H = Haplotypes

Below (Figure 4.2 to figure 4.10) are graphical representations of allele frequency distributions among the populations studied for each locus in the six populations. Across populations, the allele frequency distribution differs for most loci, although the results are not consinstent between loci for example in nearly all population (except Iraqw population which has high number of alleles in nealy all loci) fewer alleles were observed in locus DYS391 and DYS392 resulting in a markedly reduced Gene diversities for these

loci. In contrast locus DYS385 has higher allele range hence greater gene diversity values in most of populations under this study.

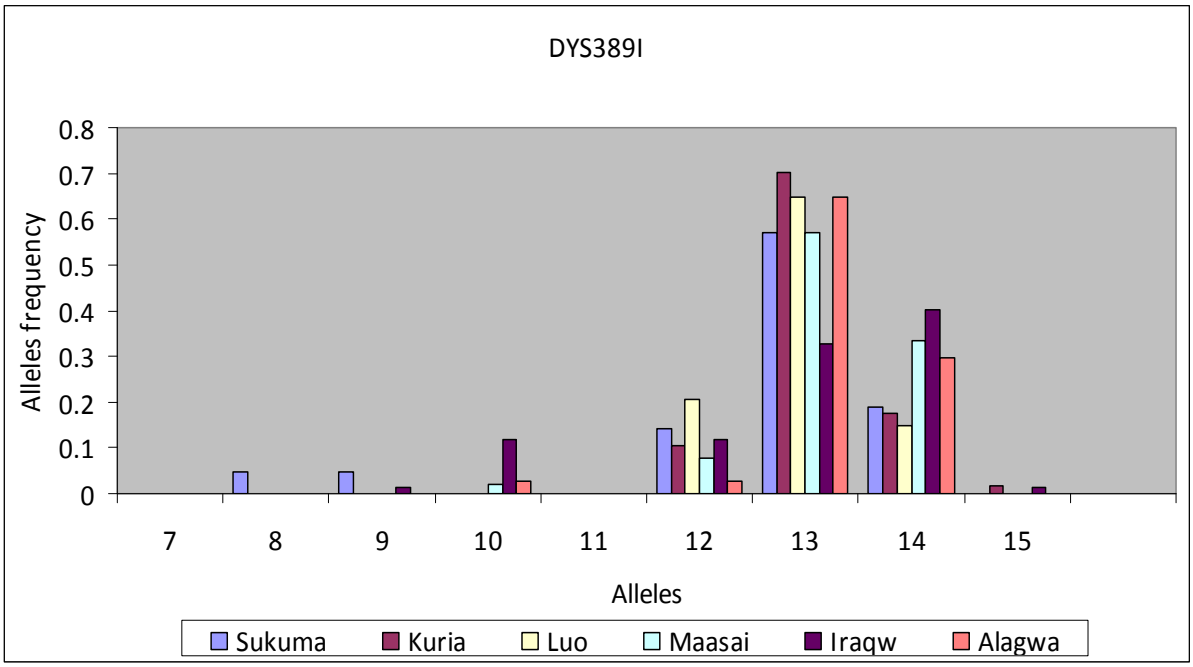

Figure 4.2 Allele frequency of DYS389I

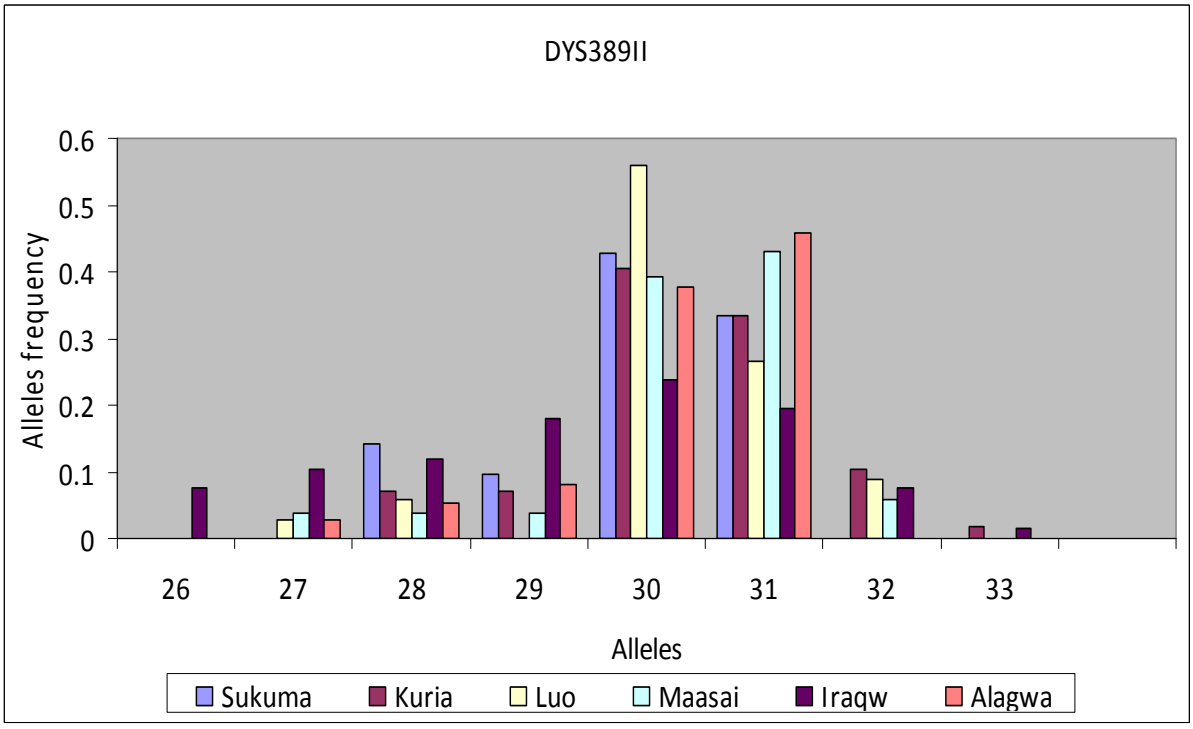

Figure 4.3 Allele frequency of DYS389II

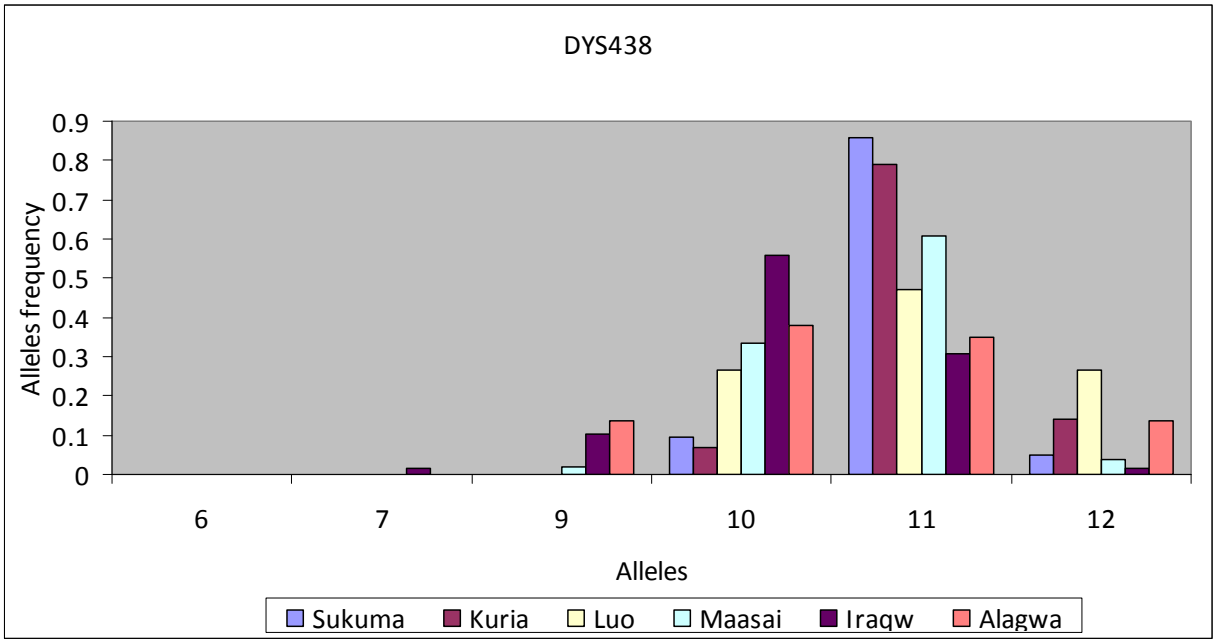

Figure 4.4 Allele frequency of DYS438

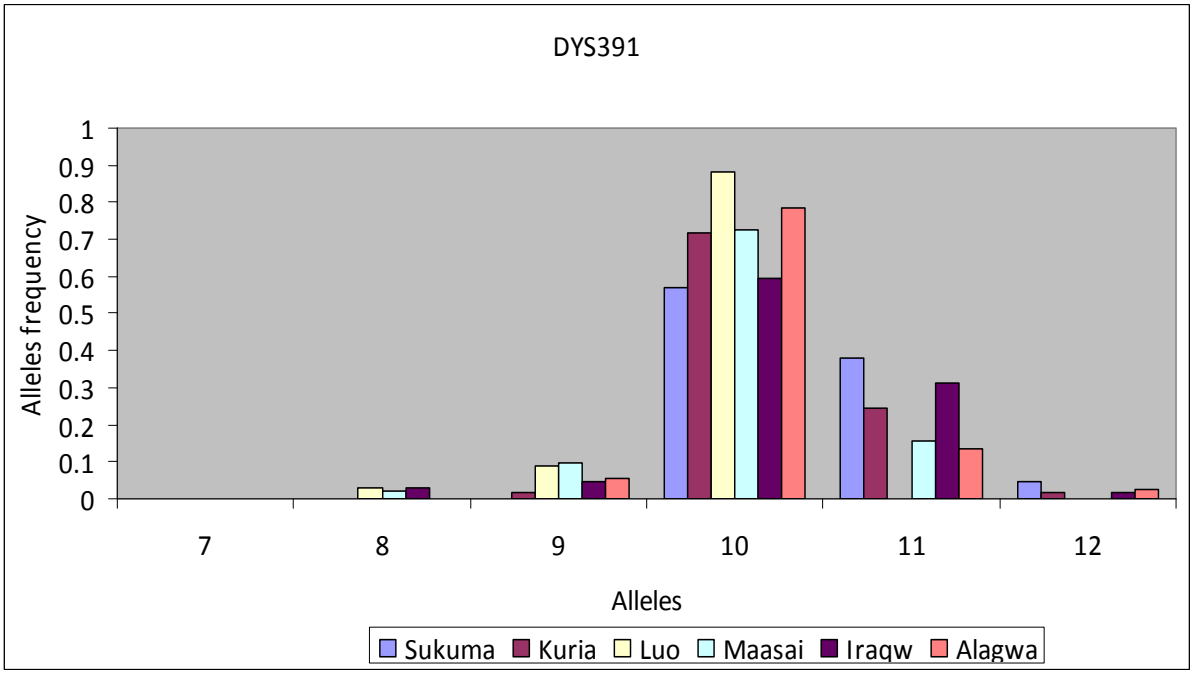

Figure 4.5 Allele frequency of DYS391

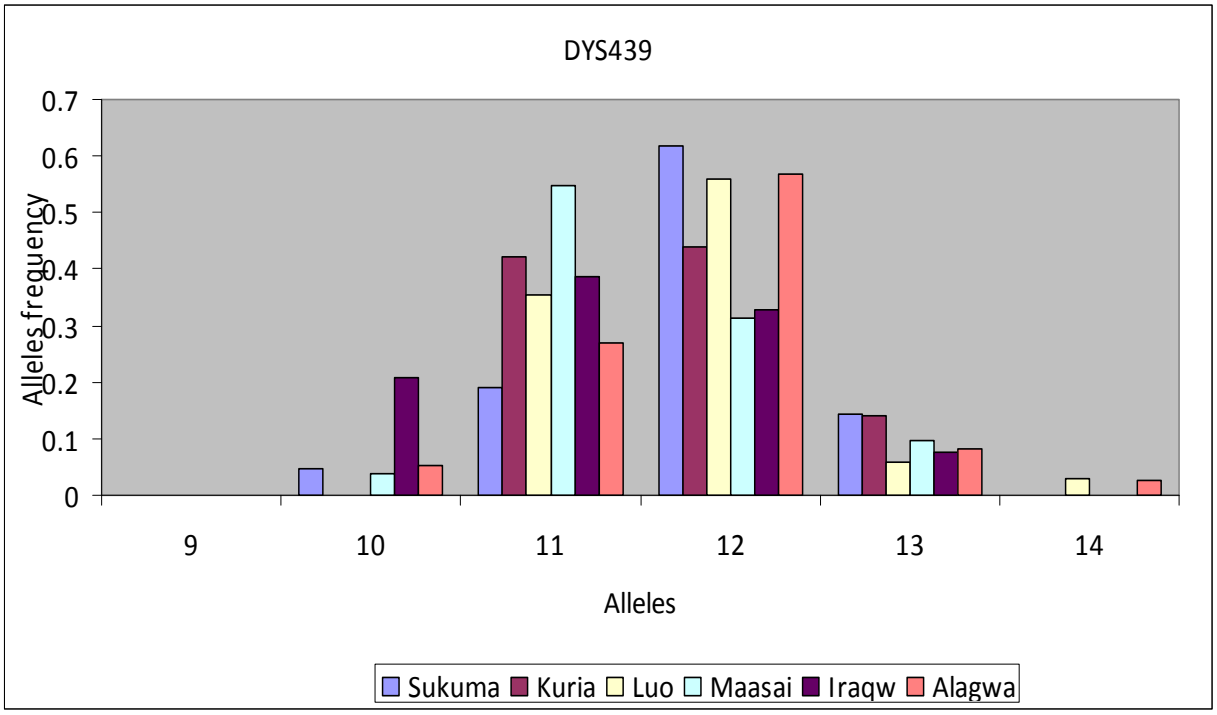

Figure 4.6 Allele frequency of DYS439

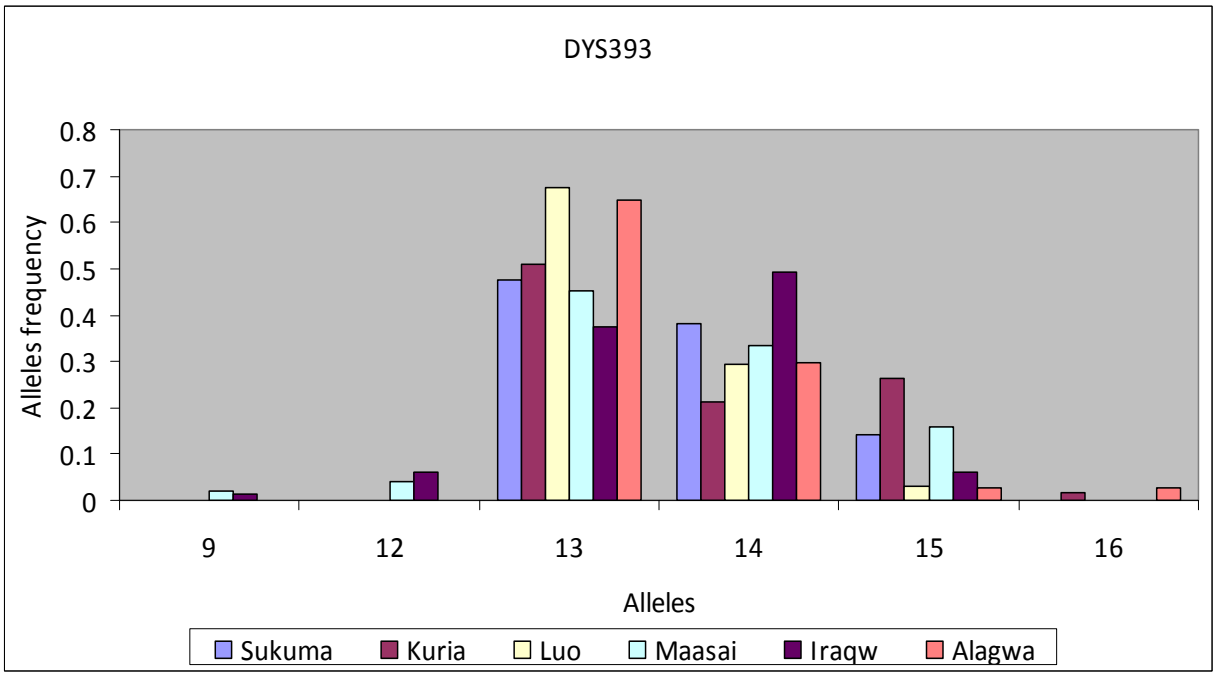

Figure 4.7 Allele frequency of DYS393

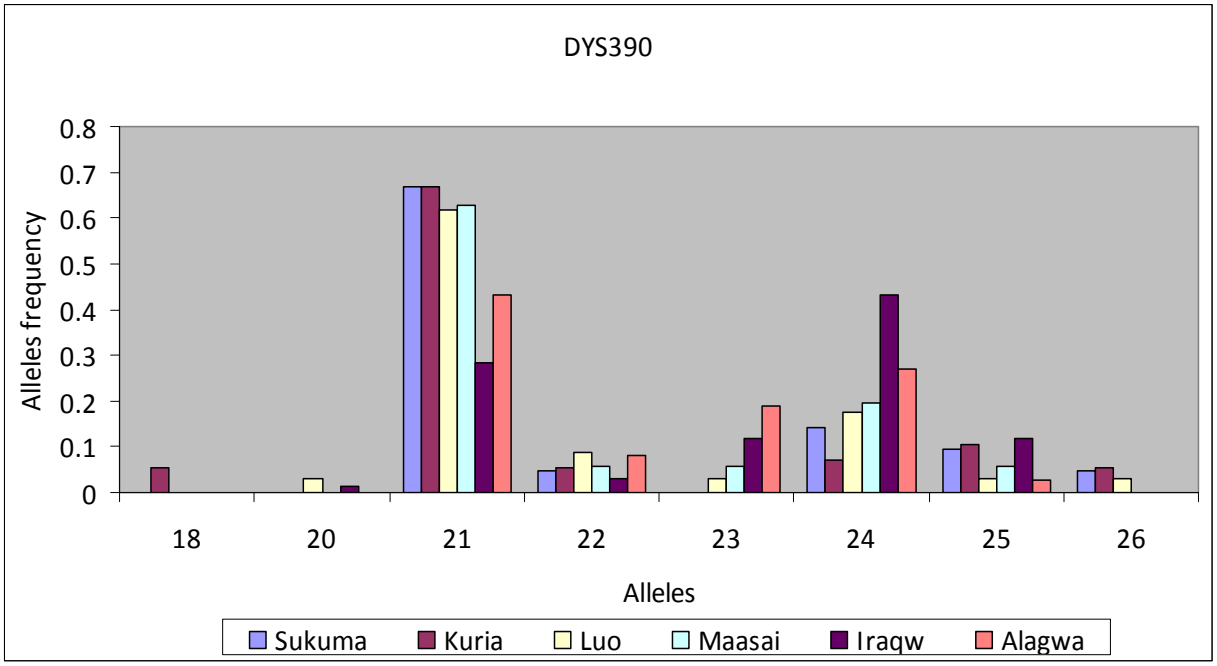

Figure 4.8 Allele frequency of DYS390

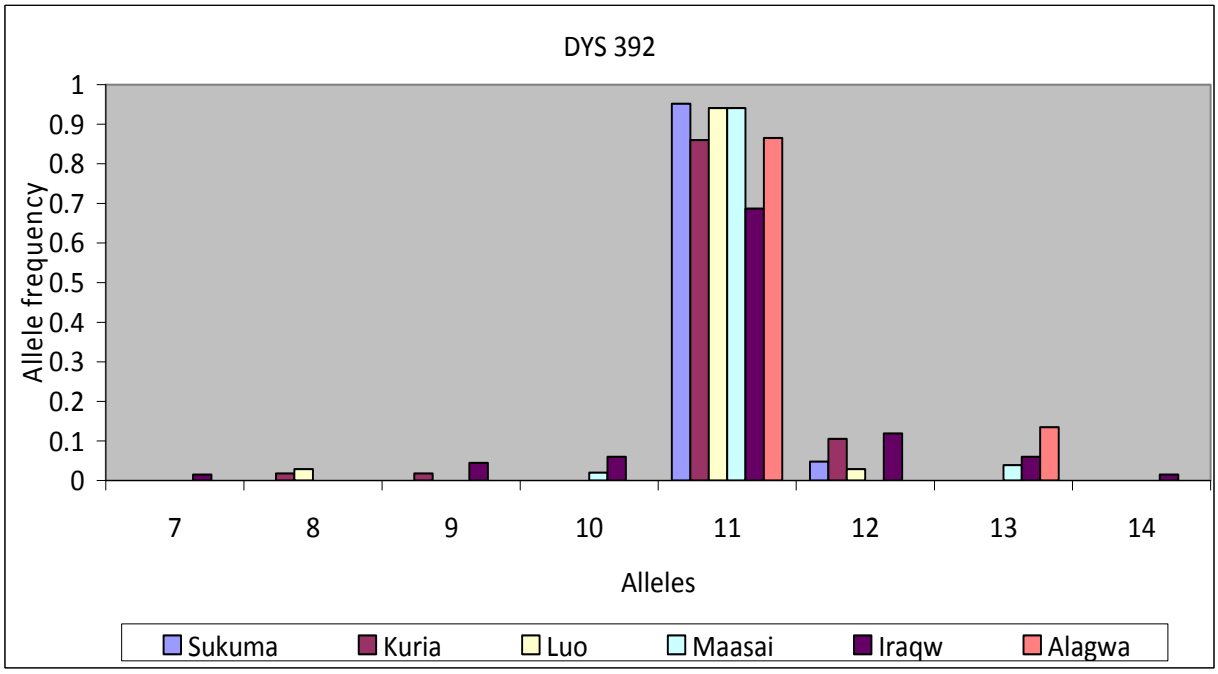

Figure 4.9 Allele frequency of DYS392

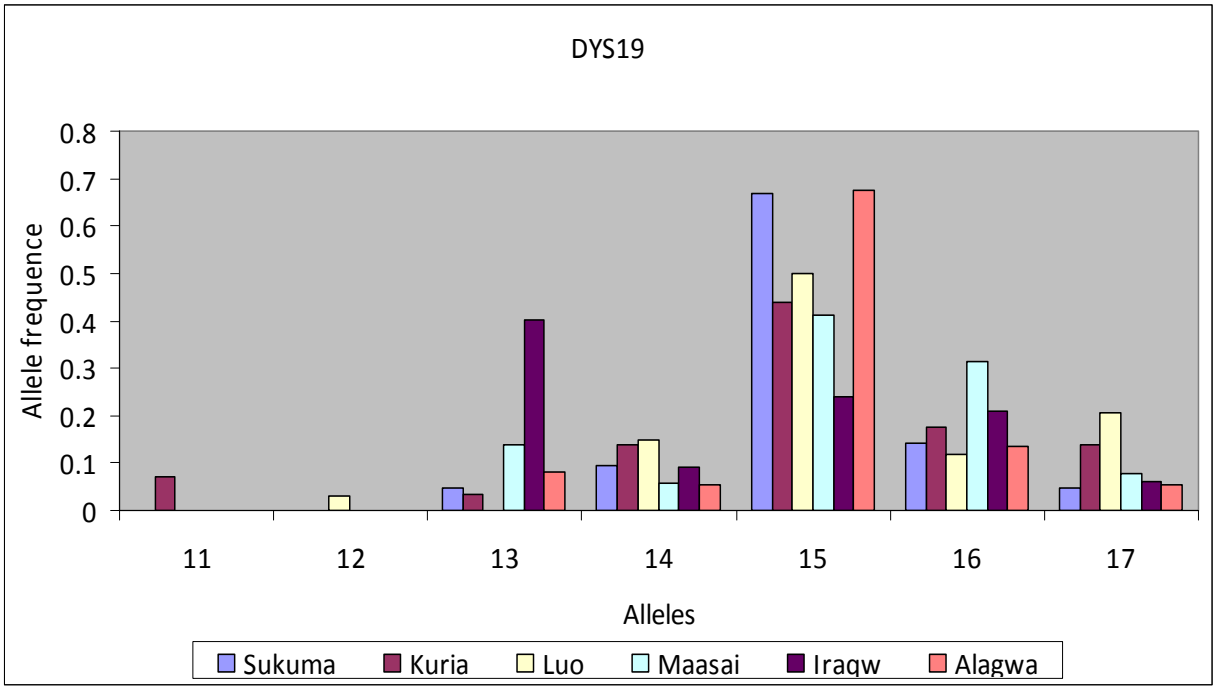

Figure 4.10 Allele frequency of DYS19

4.4 Comparison of Allele frequencies among the six populations of Tanzania

Table 4.7 and figure 4.11 below compares the observed gene diversity (*D*) values of the extended haplotype loci across the six populations of Tanzania. In four out of six sub-populations studied, the highest gene diversity for a single-copy locus was obtained with DYS19. The highest gene diversity for a single-copy locus in the Sukuma males was obtained with DYS393 and Alagwa with DYS438 and DYS390. For Kuria, Luo, Maasai and Alagwa populations, lowest gene diversity values were found with loci DYS392 and DYS391. Lowest gene diversity values in the Sukuma population were obtained with loci DYS392 and DYS438. It was also observed that the Gene diversity D values for Iraqw population were higher in most of the loci compared to other populations. In the overall individuals (Table 4.7 and figure 4.11) the highest gene diversity was found in locus DYS385 and the lowest gene diversity was found in locus DYS392. Similar results with highest level of polymorphism in DYS385 locus and lowest with DYS392 was also observed in Mozambique population (Alves *et al*., 2003).

Table 4.7 Comparison of Gene diversity (D) values for the extended haplotype loci among the six populations in Tanzania

Loci	Populations						
	Sukuma	**Kuria**	**Luo**	**Maasai**	**Iraqw**	**Alagwa**	**Overall**
DYS389I	0.6429	0.4737	0.533	0.5702	0.7114	0.503	0.6007
DYS389II	0.7095	0.7174	0.6239	0.6651	0.8494	0.6532	0.7337
DYS391	0.5524	0.4292	0.2193	0.4478	0.5504	0.3739	0.4534
DYS438	0.2667	0.3584	0.6578	0.5278	0.5866	0.7162	0.6046
DYS439	0.5857	0.6216	0.5758	0.6008	0.7028	0.6111	0.6473
DYS385a/b	0.8690	0.7912	0.8266	0.8094	0.8223	0.8780	0.8251
DYS393	0.6381	0.6385	0.4688	0.6722	0.6201	0.503	0.6193
DYS390	0.5476	0.5407	0.5936	0.5686	0.7133	0.7162	0.6587
DYS19	0.5476	0.7444	0.6916	0.7176	0.7363	0.527	0.7278
DYS392	0.0952	0.2538	0.1159	0.1145	0.5124	0.2402	0.2723

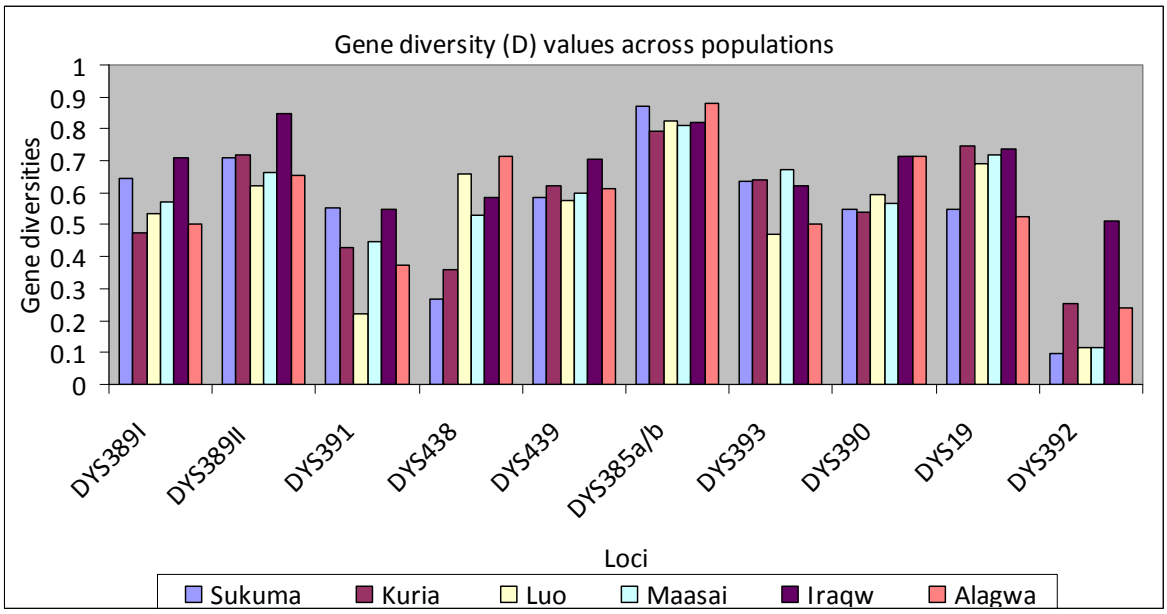

Figure 4.11 A Graphical presentation of gene diversity values across populations

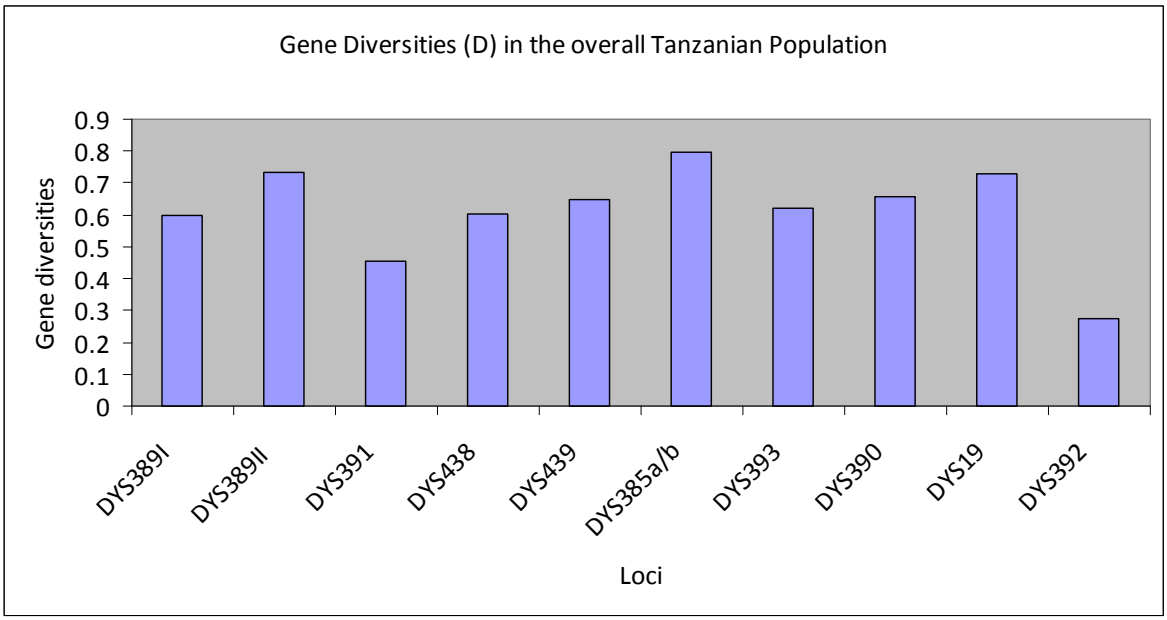

Figure 4.12 A Graphical presentation of gene diversity values across loci in the overall population of Tanzania.

4.5 Haplotype analysis and comparison of common haplotypes with those in the YHRD

Table 4.8 presents the results of the evaluation of population diversities for forensic application, Haplotype Diversity (HD) and Discrimination capacity (DC) of each population and overall population of Tanzania. The haplotype parameters DC and HD of these sets of loci showed the highest values in Sukuma and Iraqw population groups, 100% and 94.7% respectively. The lowest DC was found in Alagwa population (78.9%). A total of 225 haplotypes were obtained from of 267 genotyped individuals, of which 28 were shared, the most frequent haplotype was present within 5 individuals (Table 4.8 and Table 4.15). In the overall population Haplotype Diversity was 0.9984 and DC resulted 84.27% a similar results was found in Angola population with HD 0.9969 and DC 83.13% (Melo *et al*., 2010). The haplotype data were also compared with the data available in YHRD (http://www.yhrd.org), last updated on December 30th, 2010.

Table 4.8 a comparison of Haplotype diversity and Discrimination capacity among the six populations of Tanzania.

Statistics	Population groups						
	Sukuma	Kuria	Luo	Maasai	Iraqw	Alagwa	Overall
N	21	57	34	51	67	37	267
N haplotypes	21	49	32	44	64	29	225
DC	1.0000	0.8596	0.9412	0.8627	0.9552	0.7838	0.8427
HD	1.0000	0.9938	0.9708	0.9772	0.9875	0.9591	0.9984

N = sample size

n haplotypes = number of different haplotypes in the sample

DC = Discrimination Capacity

HD = Haplotype Diversity

4.5.1 Haplotypes from Sukuma and comparisons with those in the YHRD

The overall haplotype diversity for the extended loci among males of the Sukuma population was 1.0000, the highest among the six sub-populations investigated, this was somehow being contributed to the lowest number of sample (n = 21) in this population compared to other population evaluated in this study. A total of 21 unique haplotypes were found out of 21 (100%) males evaluated (Table 4.8). Similar result was observed in males from African population in Guinea Bissau with HD 0.9998 and DC 95.9% (Rosa *et al*., 2007). There was no haplotype that was shared among individuals within this this group, except only one haplotype (Table 4.14 page 54-55) which was shared between Sukuma and Kuria and Luo males, when this haplotype was compared with those in the YHRD, 7 matches were found (figure 4.13) from Angola (African), Venezuela (Mestizo), Brazil (1 admixed Brazilian, 1 European), United Kingdom (Afro caribean), and African American populations.

Figure 4.13. Worldwide distribution of the most common 'extended haplotype' among Sukuma males from Mwanza province (Red dots indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found) (http://www.yhrd.org).

4.5.2 Haplotypes from Kuria and comparisons with those in the YHRD

The overall haplotype diversity for the extended loci among males of the Kuria population was 0.9938, the second highest among the six sub-populations investigated. Distinct haplotypes were found for 49 of the 57 (85.96%) males investigated (Table 4.8). This was unlike the results observed in Bantu population in Mozambique reported with lower diversity using the same markers, HD 0.9921 and DC 77.68% (Alvez *et al*., 2003). Four haplotypes occured at least twice (Table 4.9).The most common haplotypes from the Kuria population were found in three individuals each (5.26%) (Table 4.9). When comparing the first common haplotype to YHRD 16 matches to this haplotype were found in Cote d'lvoire, African American, Botswana, South African (eastern Bantu), French Guiana, Colombia and Brazil populations (figure 4.14). When comparing the second most common haplotype to YHRD no match was found (figure 4.15). Table 4.14 (page 54-55)

indicates haplotypes that were shared between Kuria and other males from Sukuma, Luo, Maasai, Iraqw, and Alagwa populations.

Table 4.9 Haplotypes shared by more than one Kuria males (n = 57)

Haplotype DYS389I, DYS389II, DYS391, DYS438, DYS439, DYS393, DYS390, DYS19, DYS392, DYS385a,b	n	frequency
13-31-10-12-12-13-21-15-11-(16-17)	3	0.05263
13-31-10-11-12-15-18-14-11-(16-18)	3	0.05263
13-30-11-11-11-15-21-17-11-(16-18)	2	0.03509
14-32-10-11-11-13-24-11-12-(16-19)	2	0.03509
12-28-10-11-11-13-25-14-11-(15-20)	2	0.03509
14-30-10-10-11-13-25-13-11-(10-11)	2	0.03509

Haplotypes from duplicated locus DYS385 are presented in parenthesis

Figure 4.14. Worldwide distribution of the most common 'extended haplotype' among Kuria males from Mara province (Red dots indicate haplotype matches, blue dots indicate

populations for which extended haplotype data is available, but in which no matches could be found) (**http://www.yhrd.org**).

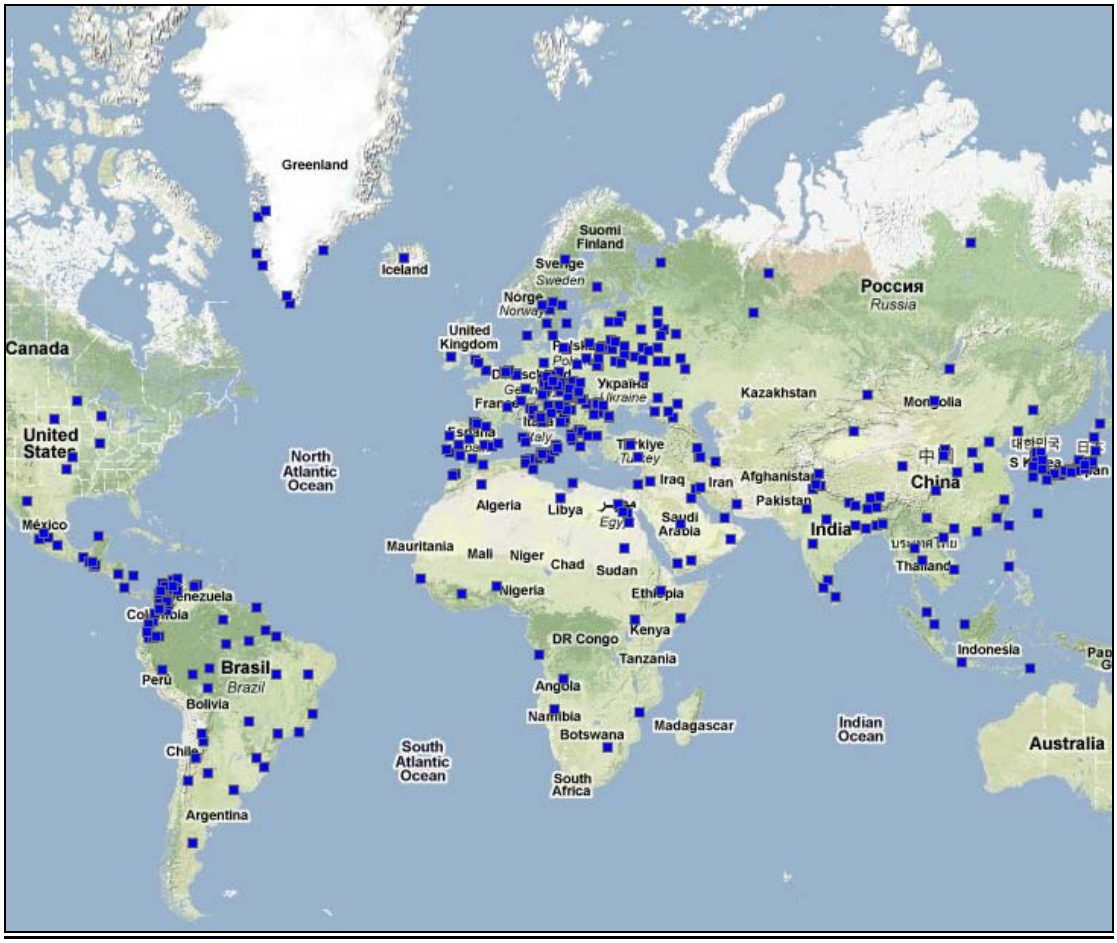

Figure 4.15. Worldwide distribution of the 2nd most common 'extended haplotype' among Kuria males from Mara province (blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found). (**http://www.yhrd.org**).

4.5.3 Haplotypes from Luo and comparisons with those in the YHRD

The overall haplotype diversity for the extended loci among males of the Luo population was 0.9708. Distinct haplotypes were found for 32 of the 34 (94.12%) males investigated (Table 4.8). This was unlike the results obtained from Nilotes population in Uganda that reported lower diversities using the same type of markers i.e. HD 0.9665 and DC 61.86% (Gomes *et al*., 2010). The two most common haplotypes were found in two individuals each (5.88%). When comparing these haplotypes (Table 4.10 – page 47) with those in the YHRD, no match was found with the first most common haplotype (figure 4.16), however

in the second most common haplotype matched 18 other haplotypes in the database, of which were from Mozambique, African American, Afro caribbean, Angola, South Africa (eastern Bantu), Brazil (African) populations (figure 4.17).

Table 4.14 (page 54-55) indicates haplotypes that were shared between Luo and other males from Sukuma, Kuria, Iraqw, and Alagwa populations.

Table 4.10 Haplotypes shared by more than one Luo males (n = 34)

Haplotype DYS389I, DYS389II, DYS391, DYS438, DYS439, DYS393, DYS390, DYS19, DYS392, DYS385a,b	n	frequency
12-30-10-12-12-13-21-15-11-(16-17)	2	0.05882
13-30-10-11-12-14-21-17-11-(17-17)	2	0.05882

Haplotypes from duplicated locus DYS385 are presented in parenthesis

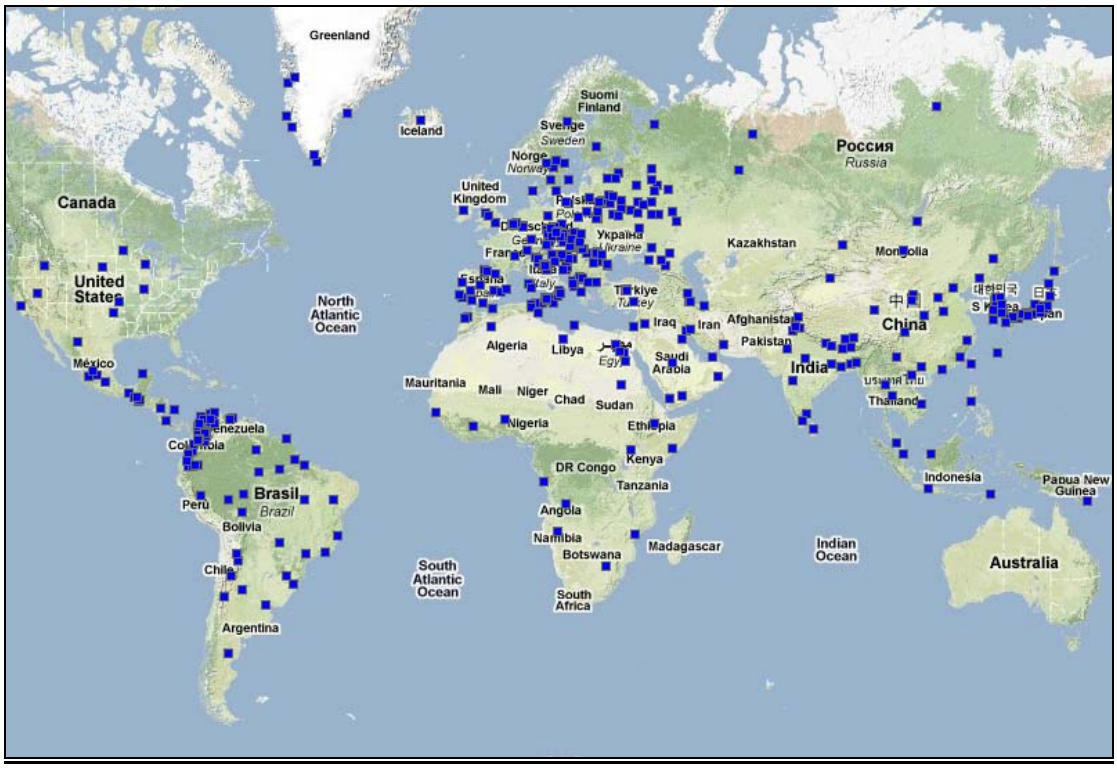

Figure 4.16. Worldwide distribution of the most common 'extended haplotype' among Luo males from Mara province (blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found). (**http://www.yhrd.org**).

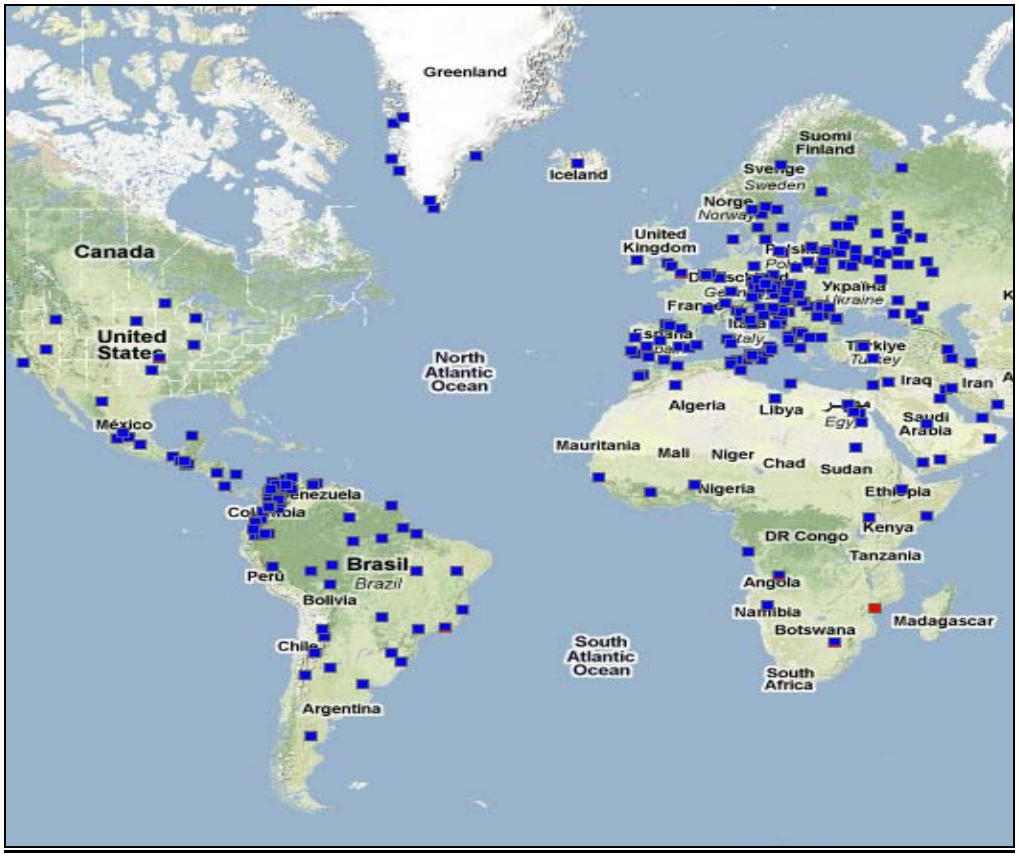

Figure 4.17. Worldwide distribution of the 2nd most common extended haplotype among Luo males from Mara province (Red dots indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found) (**http://www.yhrd.org**).

4.5.4 Haplotypes from Maasai and comparisons with those in the YHRD

The overall haplotype diversity for the extended loci among males of the Maasai population was 0.9772. Distinct haplotypes were found for 44 of the 51 (86.27%) males investigated (Table 4.8). Three haplotypes occurred at least twice (Table 4.11 – page 49). The two most common haplotypes from the Maasai were found in three individuals each (5.88%) (Table 4.11 – page 49), when comparing the two most common haplotypes in Maasai population to those in the YHRD no matches were found (Figure 4.18 and 4.19). Table 4.14 (page 54-55) indicates haplotypes that were shared between Maasai and other males from Kuria, Iraqw, and Alagwa populations.

Table 4.11 Haplotypes shared by more than one Maasai males (n = 51)

Haplotype DYS389I, DYS389II, DYS391, DYS438, DYS439, DYS393, DYS390, DYS19 DYS392, DYS385a,b	n	frequency
14-30-09-10-11-13-21-17-11-(12-12)	3	0.05882
14-31-10-11-11-13-24-16-11-(14-17)	3	0.05882
13-30-10-11-11-14-21-16-11-(17-18)	2	0.03922
14-31-10-11-11-14-21-16-11-(17-17)	2	0.03922
13-30-10-11-13-15-22-15-11-(17-17)	2	0.03922

Haplotypes from duplicated locus DYS385 are presented in parenthesis

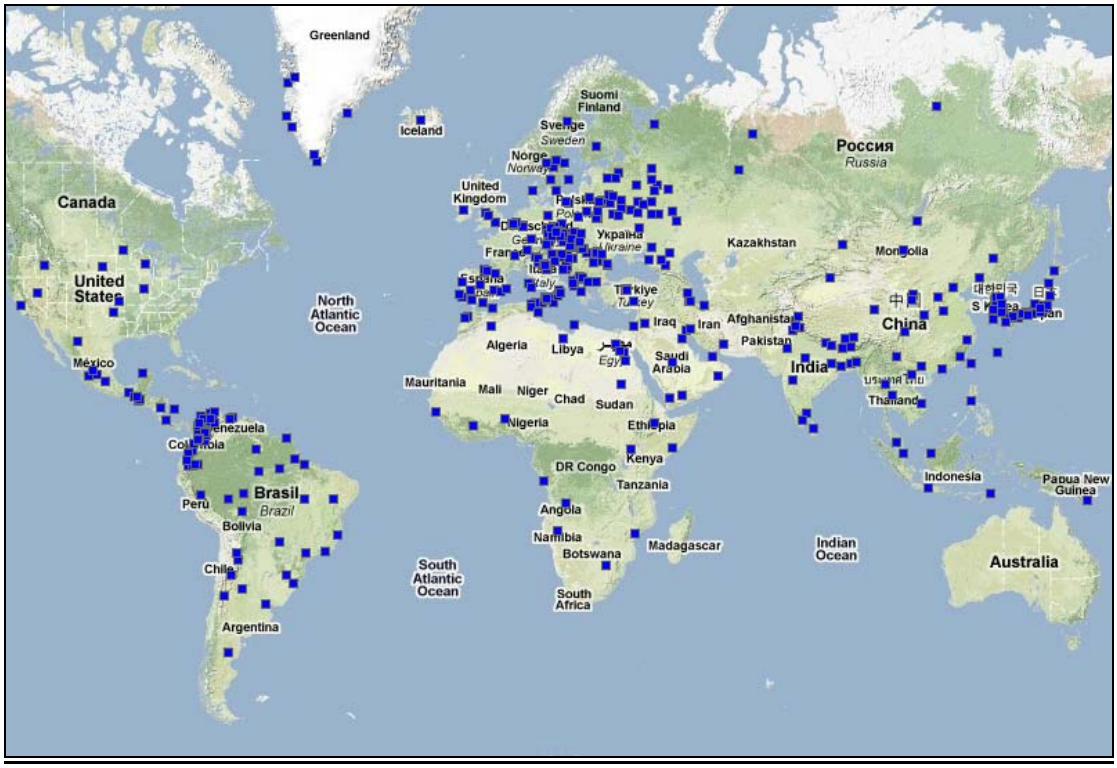

Figure 4.18. Worldwide distribution of the most common 'extended haplotype' among Maasai males from Arusha province (blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found). (**http://www.yhrd.org**).

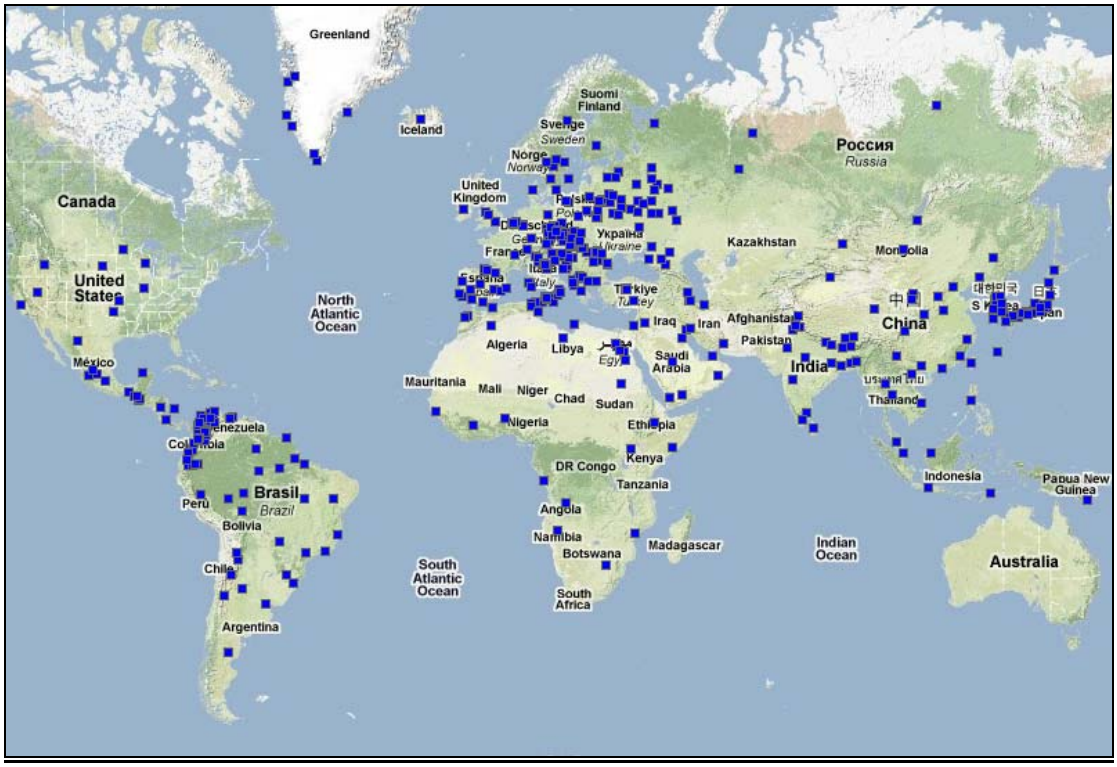

Figure 4.19. Worldwide distribution of the 2nd most common 'extended haplotype'among Maasai males from Arusha province (blue dots indicate populations for which extended haplotype data is available, but in which no match) (**http://www.yhrd.org**).

4.5.5 Haplotypes from Iraqw and comparisons with those in the YHRD

The overall haplotype diversity for the extended loci among males of the Iraqw population was 0.9875, the third highest among the six populations investigated. Distinct haplotypes were found for 64 of the 67 (95.52%) and is the second highest among males in the six populations (Table 4.8). Three haplotypes were shared between males in this population and each occurred twice (29.85%) (Table 4.12 – page 51). When comparing these haplotypes with those in the YHRD, no matches were found in the database (figure 4.20 and 4.21). Table 4.14 (page 54-55) indicates haplotypes that were shared between Iraqw and other males from Kuria, Luo, Maasai and Alagwa populations.

Table 4.12 Haplotypes shared by more than one Iraqw males (n = 67)

Haplotype DYS389I, DYS389II, DYS391, DYS438, DYS439, , DYS393, DYS390, DYS19, DYS392, DYS385ab	n	frequency
14-29-10-11-11-13-24-16-11-(10-11)	2	0.02985
12-28-10-10-11-13-24-16-11-(11-14)	2	0.02985
13-32-11-11-12-14-21-15-11-(16-18)	2	0.02985

Haplotypes from duplicated locus DYS385 are presented in parenthesis

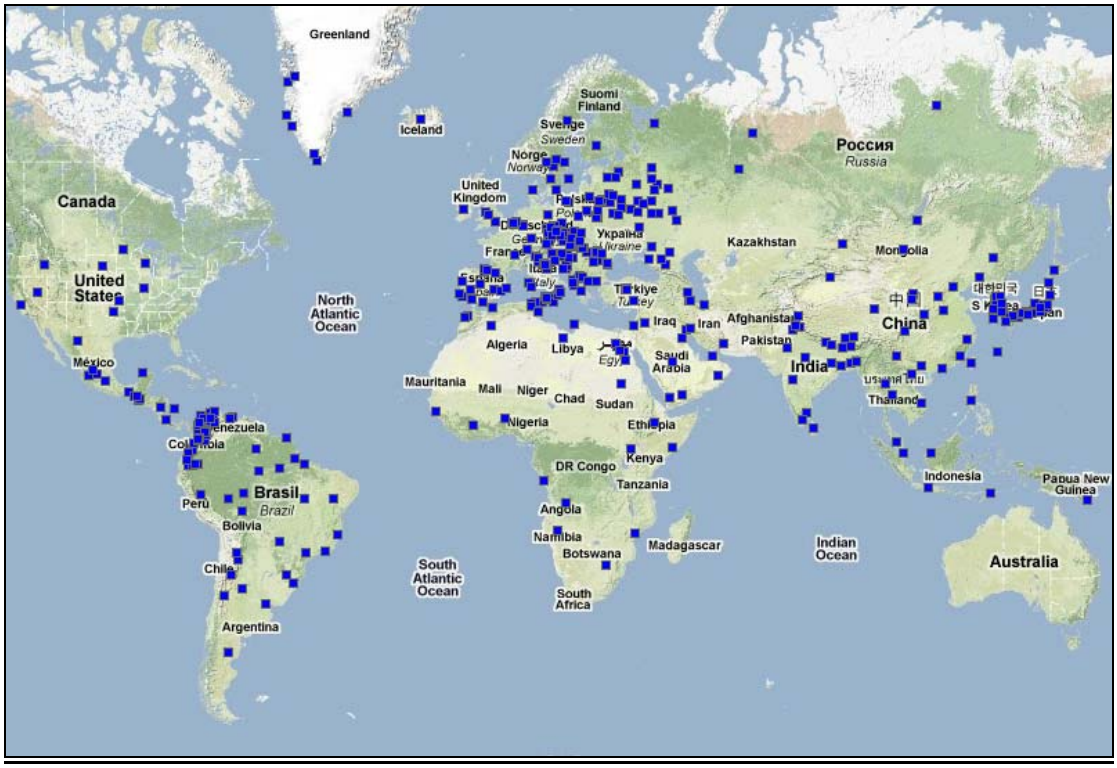

Figure 4.20. Worldwide distribution of the most common 'extended haplotype' among Iraqw males from Arusha province (blue dots indicate populations for which extended haplotype data is available, but in which no matches could be found). (**http://www.yhrd.org**).

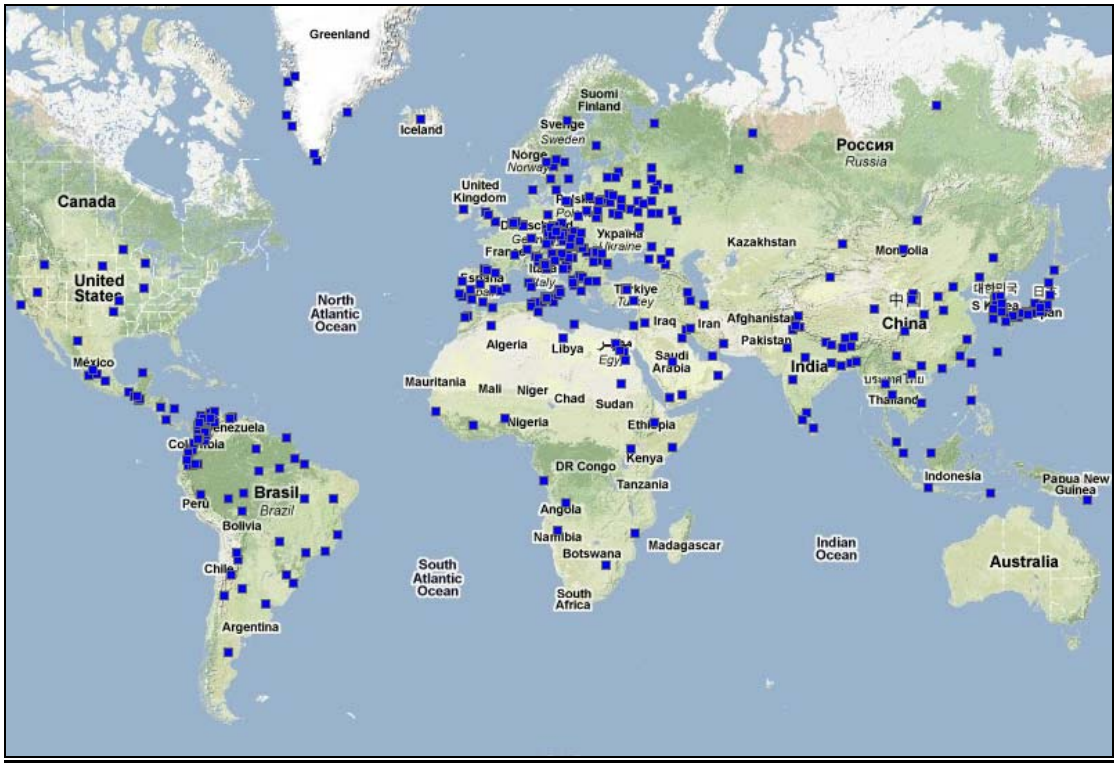

Figure 4.21. Worldwide distribution of the 2nd most common 'extended haplotype' among Iraqw males from Arusha province (blue dots indicate populations for which extended haplotype data is available, but in which no match) ((http://www.yhrd.org).

4.5.6 Haplotypes from Alagwa and comparisons with those in the YHRD

The overall haplotype diversity for the extended loci among males of the Alagwa population was 0.9591, the lowest among the six sub-populations evaluated. Distinct haplotypes were found for 29 of the 37 (78.38%) males investigated (Table 4.8- page 43), similar results were found in Tunisia population with HD 0.9764 and DC 71% (Brandt-Casadevall *et al*., 2003). Three haplotypes occurred at least twice (Table 4.13 – page 53). When comparing these haplotypes with those in the YHRD, only one matched other Africa descent haplotypes in the database, in the two other haplotype no matches were found. The most common haplotype from the Alagwa population were found in four individuals (10.8%). When searched this haplotype to YHRD eight matches were found, 7 of these matches were from Africa (Angola and South Africa, eastern Bantu) and one match was found from Eurasian populations in France (Figure 4.22). The second most common haplotype was found in three individuals (8.12%), one match to this haplotype has been found in YHRD and it was from Africa American population in Brazil (Figure 4.23 – page

54). Table 4.14 (page 54-55) indicates haplotypes that were shared between Alagwa and other males from Kuria, Luo, Maasai and Iraqw populations.

Table 4.13 Haplotypes shared by more than one Alagwa males (n = 37)

Haplotype DYS389I, DYS389II, DYS391, DYS438, DYS439, DYS393, DYS390, DYS19, DYS392, DYS385a,b	N	Frequency
13-31-10-10-12-13-24-15-11-(11-11)	4	0.10811
13-30-10-11-12-13-21-15-11-(15-19)	3	0.08108
13-30-11-11-12-14-21-15-11-(16-16)	2	0.05405
13-31-10-09-10-13-23-15-13-(14-15)	2	0.05405
13-30-10-12-12-13-21-15-11-(15-18)	2	0.05405

Haplotypes from duplicated locus DYS385 are presented in parenthesis

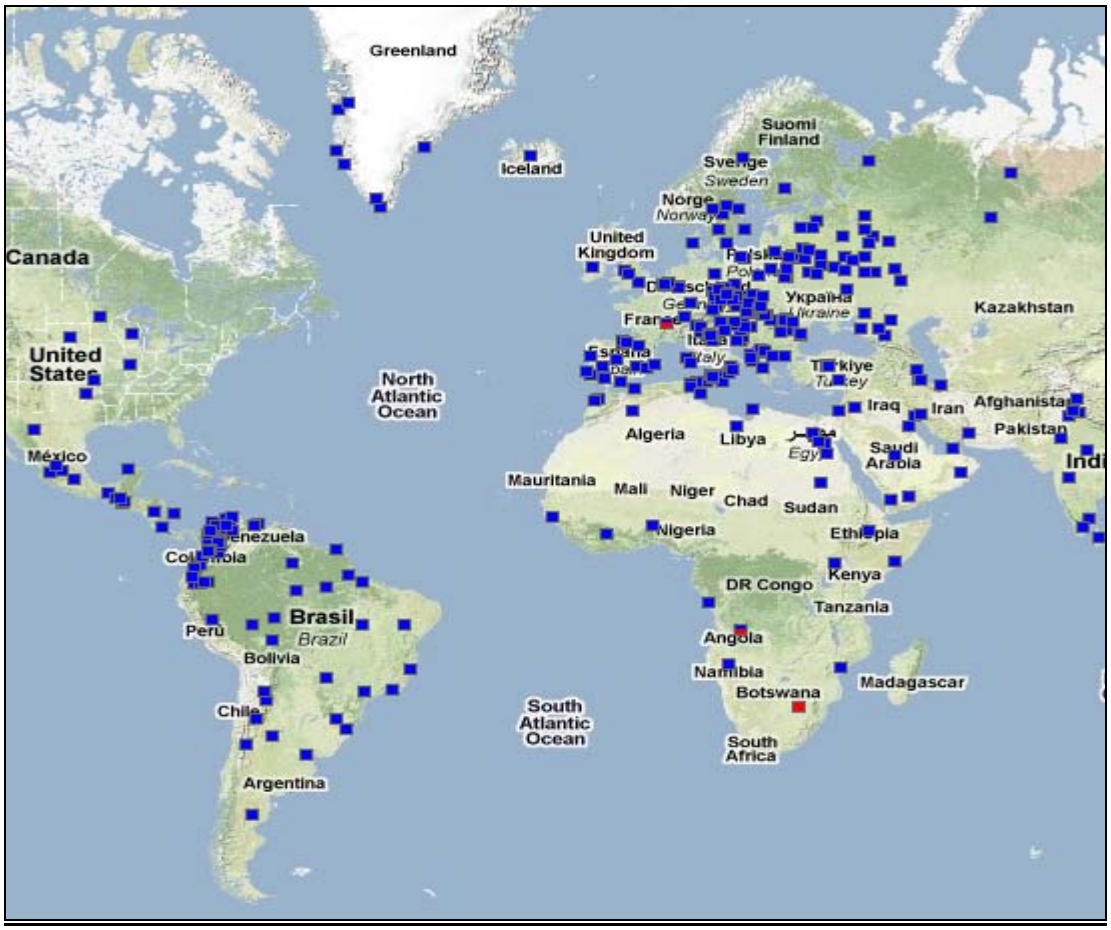

Figure 4.22. Worldwide distribution of the most common 'extended haplotype' among Alagwa males from Dodoma province (Red dots indicate haplotype matches, blue dots

indicate populations for which extended haplotype data is available, but in which no matches could be found) (**http://www.yhrd.org**).

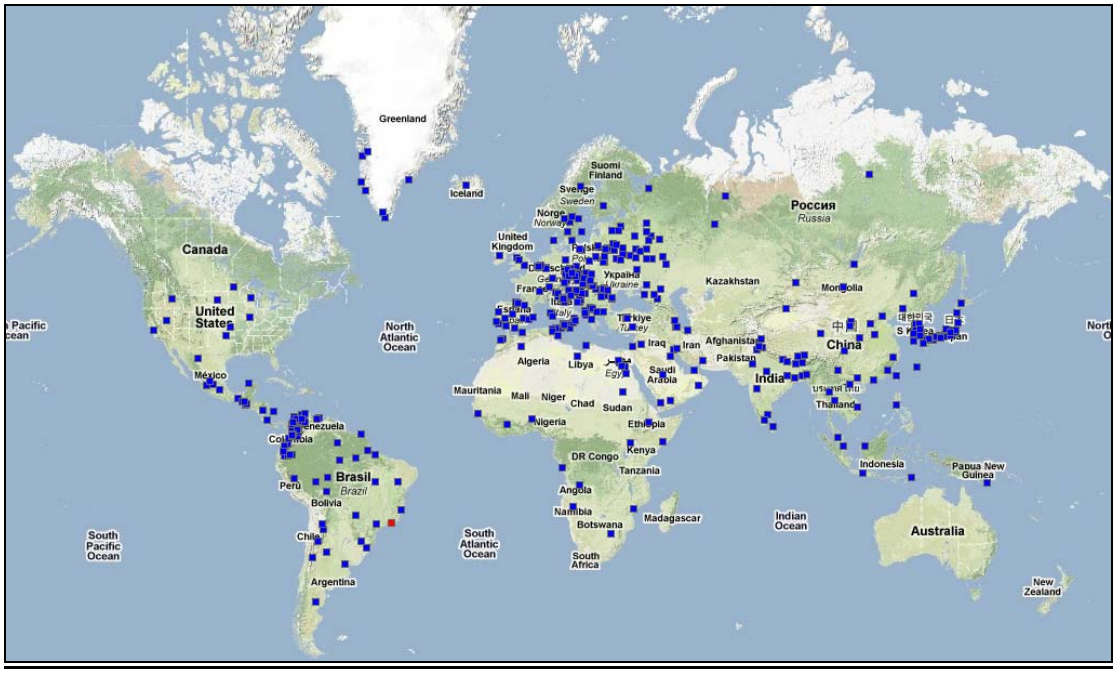

Figure 4.23. Worldwide distribution of the 2nd most common 'extended haplotype'among Alagwa males from Dodoma province (Red dots indicate haplotype matches, blue dots indicate populations for which extended haplotype data is available, but in which no match could be found) (**http://www.yhrd.org**).

Table 4.14 Haplotypes shared among males in the six populations of Tanzania

Haplotype DYS389I, DYS389II, DYS391, DYS438, DYS439, DYS393, DYS390, DYS19, DYS392, DYS385a,b	Population	N	frequency
13-30-10-11-12-14-21-17-11-(16-17)	Sukuma Kuria Luo	1 1 1	0.0476 0.0175 0.0294
13-31-10-12-12-13-21-15-11-(16-17)	Kuria Maasai Alagwa	3 1 1	0.0526 0.0196 0.027
13-30-9-10-11-15-21-15-11-(12-13)	Kuria	1	0.0175

	Luo	1	0.0294
13-30-10-12-12-13-21-15-11-(16-17)	Kuria	1	0.0175
	Luo	1	0.0294
	Iraqw	1	0.0149
13-31-10-11-11-13-21-15-11-(15-17)	Kuria	1	0.0175
	Luo	1	0.0294
13-30-10-11-12-14-21-17-11-(17-18)	Kuria	1	0.0175
	Luo	1	0.0294
	Alagwa	1	0.027
13-30-10-11-12-14-21-17-11-(17-17)	Luo	2	0.0588
	Iraqw	1	0.0149
13-30-10-11-13-15-22-15-11-(17-17)	Maasai	2	0.0393
	Alagwa	1	0.027
13-30-10-11-11-15-21-16-11-(17-17)	Maasai	1	0.0196
	Iraqw	1	0.0149
14-29-10-11-11-13-24-15-11-(10-11)	Iraqw	1	0.0149
	Alagwa	1	0.027

Haplotypes from duplicated locus DYS385 are presented in parenthesis

Table 4.15 List of Y STR extended haplotypes observed in population of Tanzania (n = 267)

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H1	13	30	11	11	12	14	21	15	11	15, 16	0.003745	1
H2	13	30	10	11	12	14	21	15	11	15,17	0.003745	1
H3	14	31	10	11	12	14	21	16	11	17,19	0.003745	1
H4	9	29	12	10	12	13	24	15	11	11,13	0.003745	1
H5	14	30	10	11	13	14	21	15	11	15,18	0.003745	1
H6	14	31	10	11	12	13	21	15	11	15,16	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H7	14	31	10	11	12	15	21	16	11	17,19	0.003745	1
H8	12	28	10	11	11	13	25	15	11	14,20	0.003745	1
H9	13	31	11	11	11	13	21	15	11	15,17	0.003745	1
H10	13	31	10	12	13	13	21	15	11	16,17	0.003745	1
H11	13	30	10	11	12	14	21	17	11	16,17	0.011236	3
H12	13	30	11	11	12	14	21	15	11	15,15	0.003745	1
H13	13	31	11	11	12	14	21	15	11	16,17	0.003745	1
H14	13	30	11	11	11	14	24	13	11	16,18	0.003745	1
H15	13	30	10	11	13	15	21	15	11	15,17	0.003745	1
H16	13	31	10	11	12	13	21	15	11	16,16	0.003745	1
H17	12	28	11	11	12	13	25	14	11	14,20	0.003745	1
H18	13	30	11	11	10	13	21	15	11	15,18	0.003745	1
H19	8	29	11	10	12	13	24	15	11	11,13	0.003745	1
H20	12	28	10	11	11	13	26	14	11	14,21	0.003745	1
H21	13	30	10	11	12	15	22	16	12	17,19	0.003745	1
H22	13	30	11	11	11	15	21	17	11	16,18	0.007491	2
H23	13	30	10	11	12	14	21	16	11	16,16	0.003745	1
H24	15	33	10	11	11	13	24	11	12	16,18	0.003745	1
H25	13	31	11	11	13	13	21	15	11	16,17	0.003745	1
H26	13	30	10	11	12	14	21	16	11	16,17	0.003745	1
H27	13	31	10	12	12	13	21	15	11	16,17	0.018727	5
H28	13	29	10	11	11	15	21	15	11	17,18	0.003745	1
H29	13	31	10	12	13	13	21	15	11	17,17	0.003745	1
H30	14	32	10	11	11	13	24	11	12	16,19	0.007491	2
H31	13	30	10	11	12	14	21	16	11	17,18	0.003745	1
H32	13	31	11	11	13	14	21	14	11	16,17	0.003745	1
H33	13	31	10	12	12	13	21	16	11	16,17	0.003745	1
H34	14	31	10	11	12	15	21	16	11	16,18	0.003745	1
H35	13	30	10	11	13	15	22	16	11	17,18	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H36	12	30	11	11	11	13	26	15	11	14,15	0.003745	1
H37	12	28	10	11	11	13	25	14	11	15,20	0.007491	2
H38	14	30	10	10	11	13	25	13	11	10,11	0.007491	2
H39	14	32	10	10	13	13	21	16	11	14,17	0.003745	1
H40	13	30	10	11	11	14	21	17	11	17,17	0.003745	1
H41	13	31	10	11	12	15	18	14	11	16,18	0.011236	3
H42	13	30	10	11	11	15	21	17	11	18,18	0.003745	1
H43	14	32	10	11	11	13	24	11	12	16,18	0.003745	1
H44	13	30	11	11	12	13	21	15	11	16,17	0.003745	1
H45	13	31	10	11	11	13	21	15	9	15,17	0.003745	1
H46	13	30	10	11	12	13	21	15	11	16,18	0.003745	1
H47	13	30	10	11	12	16	21	16	11	18,18	0.003745	1
H48	13	31	11	11	12	13	21	15	11	16,17	0.003745	1
H49	13	32	10	11	11	13	21	15	11	16,17	0.003745	1
H50	14	31	10	11	11	14	21	17	8	17,17	0.003745	1
H51	13	30	10	11	12	15	22	16	11	17,18	0.003745	1
H52	13	30	9	10	11	15	21	15	11	12,13	0.007491	2
H53	12	29	11	11	11	13	26	15	11	15,15	0.003745	1
H54	12	28	10	11	11	13	25	14	11	14,20	0.003745	1
H55	13	29	11	12	11	14	26	15	12	16,20	0.003745	1
H56	13	31	10	12	13	14	21	15	11	16,17	0.003745	1
H57	13	31	11	11	12	13	21	15	11	15,18	0.003745	1
H58	13	30	10	11	13	15	21	15	11	16,18	0.003745	1
H59	13	30	10	11	12	14	21	16	11	16,19	0.003745	1
H60	13	30	10	11	12	15	21	15	12	17,18	0.003745	1
H61	13	30	10	12	12	13	21	15	11	16,17	0.011236	3
H62	14	32	11	11	11	13	21	15	11	16,16	0.003745	1
H63	13	30	11	11	12	15	21	15	11	17,18	0.003745	1
H64	13	31	12	11	12	13	21	15	11	16,17	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H65	14	31	11	11	13	14	21	17	11	16,17	0.003745	1
H66	12	28	11	11	11	13	25	14	11	15,19	0.003745	1
H67	13	31	10	11	11	13	21	15	11	15,17	0.007491	2
H68	13	30	10	11	12	14	21	17	11	17,18	0.011236	3
H69	13	29	10	11	12	15	22	15	11	16,17	0.003745	1
H70	12	30	10	12	12	13	21	15	11	16,17	0.007491	2
H71	14	32	10	10	13	13	24	16	11	11,11	0.003745	1
H72	14	32	10	11	11	13	21	15	11	16,17	0.003745	1
H73	13	30	10	12	12	13	21	15	11	16,16	0.003745	1
H74	13	30	10	11	12	14	21	17	11	17,17	0.011236	3
H75	13	30	10	11	12	14	22	16	11	17,17	0.003745	1
H76	13	30	10	11	12	14	21	15	11	17,17	0.003745	1
H77	12	28	10	11	11	13	25	14	11	14,18	0.003745	1
H78	14	32	10	11	12	14	26	16	11	11,12	0.003745	1
H79	12	30	10	10	11	13	24	15	11	11,11	0.003745	1
H80	13	30	9	10	12	13	22	14	11	11,13	0.003745	1
H81	13	31	10	10	11	13	24	14	11	16,20	0.003745	1
H82	12	28	10	11	11	13	24	14	11	14,19	0.003745	1
H83	12	27	8	10	11	13	20	14	11	12,16	0.003745	1
H84	13	31	10	12	11	14	21	15	11	16,17	0.003745	1
H85	13	30	10	11	12	13	21	15	11	15,18	0.003745	1
H86	14	30	9	10	14	13	21	15	11	11,12	0.003745	1
H87	14	31	10	10	13	13	24	17	11	11,11	0.003745	1
H88	13	31	10	10	12	13	23	17	11	12,12	0.003745	1
H89	13	31	10	12	12	13	21	15	11	14,17	0.003745	1
H90	13	30	10	11	12	13	21	15	11	15,17	0.003745	1
H91	13	30	10	11	12	14	22	17	11	17,17	0.003745	1
H92	12	30	10	12	12	13	21	15	11	16,16	0.003745	1
H93	13	31	10	11	11	13	21	15	11	15,18	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H94	13	31	10	11	11	13	24	12	12	16,18	0.003745	1
H95	13	30	10	12	11	14	21	16	8	16,17	0.003745	1
H96	13	31	10	12	12	13	21	15	11	16,16	0.003745	1
H97	14	30	9	10	11	13	21	17	11	12,12	0.011236	3
H98	13	30	10	11	11	14	21	16	11	17,18	0.007491	2
H99	14	32	10	10	12	14	24	13	11	16,18	0.003745	1
H100	13	30	10	11	11	15	21	15	11	16,19	0.003745	1
H101	14	31	10	11	11	14	21	16	11	17,17	0.007491	2
H102	14	31	10	11	11	13	24	16	11	14,17	0.011236	3
H103	13	30	10	11	13	15	22	15	11	17,17	0.011236	3
H104	12	28	11	11	11	12	25	14	11	14,19	0.003745	1
H105	12	28	10	11	11	13	25	14	11	15,21	0.003745	1
H106	13	31	10	11	11	13	21	16	11	16,17	0.003745	1
H107	12	29	10	11	12	13	21	15	11	15,18	0.003745	1
H108	14	31	10	10	11	14	25	13	11	16,17	0.003745	1
H109	13	31	10	11	13	13	21	15	11	17,17	0.003745	1
H110	13	31	10	11	12	13	21	15	11	15,16	0.003745	1
H111	13	30	9	10	11	13	22	16	11	12,12	0.003745	1
H112	13	27	8	10	12	13	21	15	11	13,14	0.003745	1
H113	13	31	11	10	11	14	24	13	11	16,16	0.003745	1
H114	14	31	10	10	12	14	24	13	11	16,17	0.003745	1
H115	13	31	10	11	12	15	21	15	11	17,18	0.003745	1
H116	13	31	10	11	11	14	21	16	11	17,17	0.003745	1
H117	13	30	10	11	11	14	21	15	11	16,18	0.003745	1
H118	14	32	10	11	11	13	21	15	11	17,18	0.003745	1
H119	14	31	11	11	12	14	21	15	11	16,17	0.003745	1
H120	13	30	11	10	11	12	23	15	11	12,12	0.003745	1
H121	13	30	10	12	13	13	21	15	11	16,18	0.003745	1
H122	13	30	10	11	11	14	21	16	11	17,17	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H123	14	32	10	11	13	14	21	15	11	14,16	0.003745	1
H124	13	31	10	11	12	13	21	15	11	17,17	0.003745	1
H125	13	31	10	11	11	13	21	15	11	17,17	0.003745	1
H126	13	30	11	10	12	14	24	13	11	16,16	0.003745	1
H127	13	31	10	10	11	13	23	14	11	11,11	0.003745	1
H128	14	31	10	9	11	9	24	16	13	14,16	0.003745	1
H129	12	29	11	11	12	13	24	15	13	13,16	0.003745	1
H130	13	30	10	11	12	15	21	17	10	18,21	0.003745	1
H131	14	30	10	10	12	13	21	15	11	12,13	0.003745	1
H132	14	31	10	11	12	14	21	15	11	15,17	0.003745	1
H133	13	31	10	10	10	13	21	16	11	12,12	0.003745	1
H134	10	27	11	10	12	14	23	13	11	15,18	0.003745	1
H135	13	30	10	11	12	15	21	16	11	17,17	0.003745	1
H136	13	30	10	11	11	14	21	16	11	16,17	0.007491	2
H137	13	30	10	11	11	15	21	16	11	17,17	0.003745	1
H138	13	30	9	10	10	15	21	15	11	12,12	0.003745	1
H139	13	31	11	10	11	13	24	13	11	15,16	0.003745	1
H140	14	29	10	9	12	13	24	15	13	15,18	0.003745	1
H141	13	30	11	10	12	14	23	13	11	16,17	0.003745	1
H142	14	29	10	11	11	13	24	16	11	10,11	0.007491	2
H143	12	28	10	10	11	13	24	16	11	11,14	0.007491	2
H144	12	30	11	11	13	14	25	15	11	15,17	0.003745	1
H145	10	27	10	10	11	14	24	13	12	15,17	0.003745	1
H146	14	31	11	10	11	14	24	13	12	16,16	0.003745	1
H147	15	33	9	10	11	12	21	16	11	12,12	0.003745	1
H148	14	30	11	10	12	14	24	13	12	16,16	0.003745	1
H149	13	31	10	11	11	13	21	15	11	17,19	0.003745	1
H150	10	27	11	10	13	14	24	13	11	16,17	0.003745	1
H151	14	31	10	10	12	14	24	13	11	15,17	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H152	14	31	11	10	11	14	25	13	11	17,17	0.003745	1
H153	10	27	10	10	11	14	24	13	11	15,17	0.003745	1
H154	13	32	10	10	12	12	24	13	11	16,20	0.003745	1
H155	13	30	10	10	12	14	25	14	10	16,19	0.003745	1
H156	14	29	10	9	11	13	24	17	14	14,17	0.003745	1
H157	14	32	11	11	13	14	23	13	11	16,16	0.003745	1
H158	14	31	11	10	12	14	24	13	11	16,16	0.003745	1
H159	14	32	11	10	12	14	22	13	12	16,16	0.003745	1
H160	10	27	10	10	11	14	25	13	11	14,17	0.003745	1
H161	14	29	10	11	11	13	24	15	11	10,11	0.007491	2
H162	14	30	9	10	12	13	21	16	11	12,12	0.003745	1
H163	14	31	8	10	12	13	20	15	11	12,13	0.003745	1
H164	13	29	10	10	10	12	24	13	12	17,17	0.003745	1
H165	12	29	10	10	11	14	25	14	11	16,16	0.003745	1
H166	14	28	10	11	11	13	24	14	9	9,11	0.003745	1
H167	10	26	11	10	10	14	24	13	11	15,17	0.003745	1
H168	9	26	10	10	10	14	24	13	11	15,17	0.003745	1
H169	13	30	12	7	10	13	21	15	11	16,16	0.003745	1
H170	10	26	11	10	10	14	23	13	11	16,17	0.003745	1
H171	13	28	10	9	11	13	24	15	7	15,15	0.003745	1
H172	13	29	10	11	10	13	21	16	9	16,18	0.003745	1
H173	13	28	10	11	11	14	21	17	9	16,16	0.003745	1
H174	14	31	11	11	10	14	21	15	11	16,17	0.003745	1
H175	14	31	10	11	11	14	21	16	11	16,16	0.003745	1
H176	13	30	10	10	11	14	25	13	11	16,16	0.003745	1
H177	14	29	10	9	10	13	22	15	11	15,15	0.003745	1
H178	13	30	10	11	12	15	21	16	11	17,19	0.003745	1
H179	13	30	10	10	11	14	25	13	11	17,18	0.003745	1
H180	10	26	10	10	10	14	23	13	11	15,17	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H181	14	27	11	10	11	9	24	13	10	16,16	0.003745	1
H182	12	28	10	10	10	13	24	16	10	11,14	0.003745	1
H183	12	26	11	10	10	13	23	14	10	12,14	0.003745	1
H184	14	27	10	9	10	13	24	17	12	14,17	0.003745	1
H185	14	31	11	10	12	14	24	13	11	17,17	0.003745	1
H186	12	28	10	11	12	13	24	14	11	14,19	0.003745	1
H187	13	32	11	11	12	14	21	15	11	16,18	0.007491	2
H188	12	29	11	10	11	14	24	13	11	16,16	0.003745	2
H189	14	29	10	9	10	13	23	15	13	14,16	0.003745	1
H190	14	28	10	11	11	13	23	15	13	15,17	0.003745	1
H191	10	27	11	10	12	14	24	13	11	15,17	0.003745	1
H192	13	31	10	11	13	15	21	16	11	16,17	0.003745	1
H193	14	30	11	10	12	14	24	13	11	16,17	0.003745	1
H194	13	31	8	10	12	13	21	14	12	12,13	0.003745	1
H195	14	31	11	10	12	14	25	13	11	15,16	0.003745	1
H196	13	30	10	10	11	12	24	13	12	16,18	0.003745	1
H197	13	31	10	11	11	15	21	16	11	16,18	0.003745	1
H198	13	30	10	10	13	15	21	16	11	17,18	0.003745	1
H199	14	30	9	10	12	13	21	15	11	12,12	0.003745	1
H200	14	29	10	9	10	13	23	15	13	14,17	0.003745	1
H201	13	30	11	11	12	14	21	15	11	16,16	0.007491	2
H202	14	31	10	11	12	14	22	16	11	17,18	0.003745	1
H203	13	28	10	10	12	14	24	16	11	12,14	0.003745	1
H204	14	31	10	10	11	14	23	14	11	15,16	0.003745	1
H205	13	30	11	10	12	14	24	13	11	13,17	0.003745	1
H206	10	27	10	10	11	14	23	13	11	16,16	0.003745	1
H207	14	31	10	9	11	13	23	15	13	14,16	0.003745	1
H208	13	29	10	11	11	13	24	15	11	10,11	0.003745	1
H209	14	30	10	10	11	13	21	15	11	12,12	0.003745	1

H	DYS 389I	DYS 389II	DYS 391	DYS 438	DYS 439	DYS 393	DYS 390	DYS 19	DYS 392	DYS 385a,b	Frequency	N
H210	13	31	10	12	12	14	21	15	11	15,15	0.003745	1
H211	14	30	10	9	13	13	23	16	13	15,15	0.003745	1
H212	12	28	11	11	12	13	25	14	11	14,14	0.003745	1
H213	13	30	11	11	12	14	21	17	11	16,17	0.003745	1
H214	13	31	10	10	12	13	24	15	11	11,11	0.014981	4
H215	14	31	10	10	11	13	22	15	11	12,12	0.003745	1
H216	13	30	10	12	13	16	21	16	11	17,18	0.003745	1
H217	13	31	10	9	10	13	23	15	13	14,15	0.007491	2
H218	13	31	10	10	14	13	24	16	11	11,11	0.003745	1
H219	13	30	10	11	12	13	21	15	11	15,19	0.011236	3
H220	14	31	10	11	12	13	21	15	11	15,17	0.003745	1
H221	13	30	10	12	12	13	21	15	11	15,18	0.007491	2
H222	14	31	10	9	11	13	23	15	13	15,15	0.003745	1
H223	13	29	9	10	11	13	21	15	11	12,12	0.003745	1
H224	14	31	12	10	12	14	24	13	11	15,16	0.003745	1
H225	14	31	9	10	11	13	21	15	11	12,12	0.003745	1

H = Haplotype, n = number of individuals observed in each haplotype

4.6. Population pairwise, Fst results

Table 4.16. Y- STR haplotype pairwise Fst values below diagonal and corresponding *P* values above diagonal (*P* values < 0.05 are bolded) for Six Tanzanian populations (locus DYS385a,b was excluded in the analysis).

Population	Sukuma N = 21	Kuria N = 57	Luo N = 34	Maasai N = 51	Iraqw N = 67	Alagwa N = 37
Sukuma	-	0.17117	**0.04505**	**0.00000**	**0.00000**	**0.02703**
Kuria	0.00864	-	**0.00000**	**0.00000**	**0.00000**	**0.00000**
Luo	0.03885	0.03453	-	**0.03604**	**0.00000**	0.12613

Maasai	0.02946	0.02252	0.02878	-	**0.00000**	**0.00000**
Iraqw	0.09919	0.11248	0.10297	0.05804	-	**0.00000**
Alagwa	0.03942	0.05928	0.01559	0.02730	0.06527	-

4.7 AMOVA results

Table 4.17 a. Non Hierarchical AMOVA, Showing the Percentage of Variance (locus DYS385a,b was excluded in the analysis).

Source of Molecular variation	**Variance components**	**Variance (%)**
Among populations	0.16392 (Va)	6.03862
Within populations	2.55065 (Vb)	93.96138

Fixation index, FST: 0.06039

Significance tests

Va and FST: P (rand. value > obs. value) = 0.00000

P (rand. value = obs. value) = 0.00000

P-value = 0.00000

Table 4.17 b. Hierarchical AMOVA, Showing the Percentage of Variance at three Levels of Population Hierarchy (locus DYS385a,b was excluded in the analysis).

Source of Molecular variation	**Variance components**	**Variance (%)**
Among groups	0.07018 (Va)	2.57
Among populations within groups	0.10684 (Vb)	3.92

Within populations	2.55065 (Vc)	93.51

Fixation indices,

FSC: 0.04020

FST: 0.06490

FCT: 0.02573

Significance tests

Vc and FST: P (rand. value < obs. value) = 0.00000

P (rand. value = obs. value) = 0.00000

P-value = 0.00000+-0.00000

Vb and FSC: P (rand. value > obs. value) = 0.00000

P (rand. value = obs. value) = 0.00000

P-value = 0.00000+-0.00000

Va and FCT: P (rand. value > obs. value) = 0.06061

P (rand. value = obs. value) = 0.07722

P-value = 0.13783+-0.00898

4.8 Factorial Component Analysis (FCA)

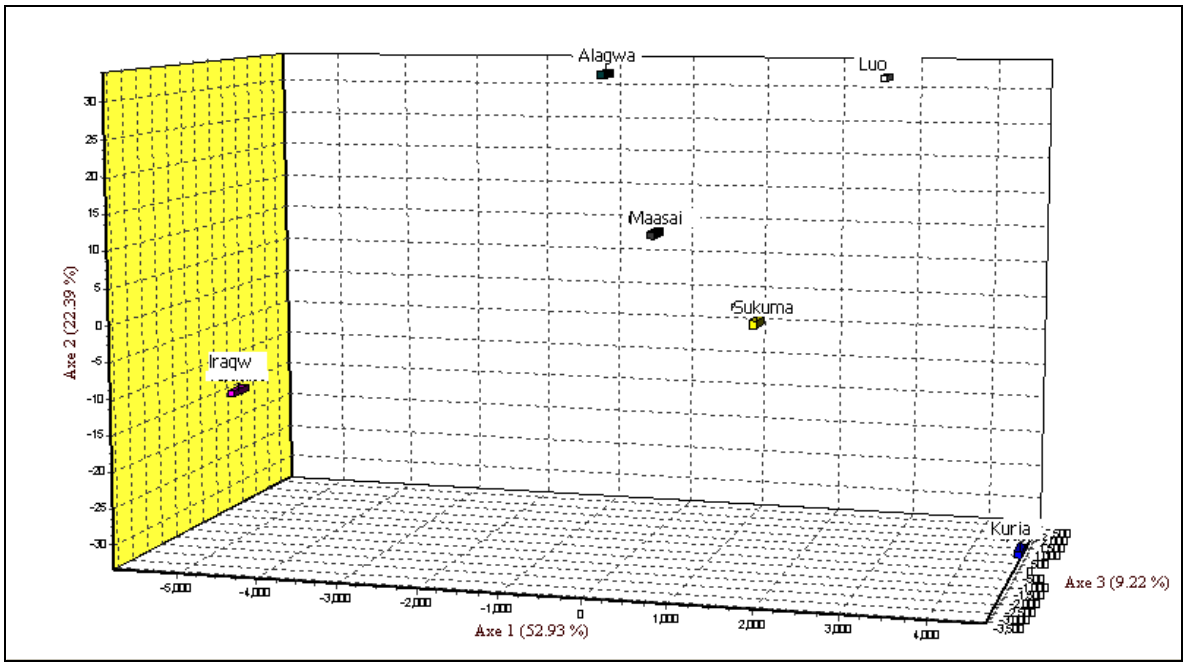

Figure 4.24. Factorial component analysis of the extended haplotypes data for the different samples in the six populations of Tanzania (Locus DYS385a,b was excluded in the analysis).

4.9 Discussion

Genetic differences between populations

Genetic heterogeneity between populations was tested by Fst, AMOVA and multivariate analyses (FCA).

Using non hierchial analysis of Molecular Variance in the six populations of Tanzania, the relative contribution of genetic variance between populations to total observed genetic variance was estimated i.e. (i) genetic variance among populations (Va) and (ii) genetic variance between individuals within populations (Vb) (table 4.17 a). The percentage of variance within population was (93.96%) and among population (6.04%) with (Fixation index Fst = 0.06039, *P* = 0.0000) (Table 4.17 a), indicating genetic heterogeneity among populations. In Hierarchial AMOVA analysis (Table 4.17 b) where populations were subdivided into three groups based on their language; Niger congo (Sukuma and Kuria), Afro Asiatic (Iraqw and Alagwa) and Nilo Saharan (Maasai and Luo) significant results were observed. Variance between individuals within populations (Vc) was 93.51%, Fst = 0.06490, P = 0.0000 and variance among populations within groups (Vb) was 3.92%, FSC =0.04020, P = 0.0000 (Table 4.17 b). Non significant results was observed in variance among groups (Va) when populations were subdivided in three language groups (FCT = 0.02573, P = 0.13783, variance (Va) = 2.57%) (Table 4.17 b). Even though the non significant differences were detected the percentage of variance (2. 57%) among groups is probably being contributed by the haplotypes that are shared by populations of the different language groups (Table 4.14) and probably AMOVA is not sensitive enough to detect differences under these circumstances. The differences among populations that was found by AMOVA analysis is also being supported by population pairwise results (table 4.16) with exceptional in Sukuma and kuria and Alagwa and Luo pairs. Genetic differentiation among Tanzanian populations is supported by the limited level of admixture (gene flow) between them most probably due to differences in social, cultural an ethnical backgrounds. For example the Alagwa through their history they are reserved and hence difficult to receive outsiders, also through the way of life as majority of them

were Muslims and follow Muslim belief systems. Another example is the Kuria with several occurances of inter-clan warfare and cattle raiding which created demographic and environmental stress within their society (Keen, 1996). The Iraqw appears to be genetically distantly related from the rest of population in the scatter plot of FCA (Fig 4.25) and in pairwise population comparison (table 4.16), supported by their population history that during the nineteenth century the Iraqw were confined to a relatively small enclave of the Mbulu highlands (the Iraqw's homeland area) because of the fear of cattle raids and attacks by other ethnic groups, mainly the pastoral Maasai and Datooga, who where in control of large parts of the Mbulu Highlands.

The population pairwise results (Table 4.16) between Sukuma and Kuria populations showed non significant Fst value ($P > 0.05$), which indicates that there is gene flow between Sukuma and kuria populations of the same language group. This result verified that Y-STR extended haplotype data reflect linguistic affiliations in Sukuma and Kuria populations. The possible explanation for their genetic similarities (common ancestry) is through the history that bantu speaking people originated from West Africa, they migrated to East Africa 3000 years ago with the first settlements around the shore of lake Victoria (Newman, 1995). These findings are similar to the study (Wood *et al*., 2005) using Y chromosome SNPs (Single Nucleotide polymorphism) and mtDNA markers that reported genetic and linguistic affiliations between African populations. Wood *et al*. (2005) found the significant association between Y chromosome variation and linguistic differentiation and a marginally significant association between mtDNA variation and linguistic variations in African populations. It was suggested that observed association between Y chromosome and language variation reflected the same co-evolutionary population history and the differing patterns of Y chromosome and mtDNA could be as a result of the greater degree of female than male admixture or the adoption of languages by females is of greater extent than male (Wood., *et al* 2005).

On the otherhand population pairwise results between Alagwa (Afro Asiatic) and Luo (Niger Congo) (Table 4.16) showed non significant Fst value ($P > 0.05$), which indicates that there is gene flow between populations of the different language groups. The genetic relationship between Alagwa and Luo could be explained by the history that the Luo were

able to adapt to many different ecologies and economies as they were spreading, unlike the Maasai, the Luo did not define their ethnicity by their subsitence pattern as herders, hunters or farmers, and hence interacted intimately with people speaking other languages and practicing other cultures thereby adapting new lifestyles and identity. In each place the Luo passed, they had an excessive influence on the people with whom they interacted (Feierman, 1995).

Nevertheless despite of the usage of languages that are in the same language family (Afro Asiatic), the FCA of Iraqw and Alagwa (figure 4.25) suggest that they are distantly related. Also the population pairwise results between Iraqw and Alagwa indicated significant Fst values with $P < 0.01$ (Table 4.16) i.e. there is significant genetic differences between Iraqw and Alagwa both of them classified in Afro Asiatic linguistic family. These results demonstrate that Y-STR haplotype distribution does not reflect genetic and linguistic affiliations in Iraqw and Alagwa populations, but confirms linguistic and genetic affiliations between Sukuma and Kuria (see above discussion). Divergences between genetic and linguistic classifications were also observed in other studies (Tishkoff *et al*., 2009) using different types of markers such as ancestry informative SNPs and microsatellites on African populations. Tishkoff *et al*. (2009) found that click speaking Hadza and Sandawe, were differentiated from South African Khoisan populations all of them classified as khoesan, the afroasiatic Chadic speaking populations from Nothern Cameroon cluster close to Nilo Saharan speaking populations from southern Sudan and Chad, rather than with East African Afroasiatic speakers, and Fulani who speak a West African Niger – Kordofanian language but cluster with Chadic and Southern Saharana speaking populations (Tishkoff *et al*., 2009).

Level of Genetic diversity in Tanzanian populations: Forensic perspective

The levels of genetic diversity for the haplotypes per group as revealed by haplotype diversities (Table 4.8) confirmed that the most diverse group being Sukuma, Kuria, Iraqw, Maasai, Luo and Alagwa being the least diverse.

The Discrimination Capacity of these set of markers showed the highest value in Sukuma population (100%), subsequently followed by Iraqw, Luo, Maasai, Kuria and Alagwa (78.38%) being the lowest (Table 4.8).

In comparison to other African population, the Discrimination Capacity found in Tanzanian population for these sets of markers were higher than those reported for Tunisia (Brandt-Casadevall *et al*., 2003), Uganda (Gomes *et al*., 2010), Angola (Melo *et al*., 2010) and Mozambique (Alvez *et al*., 2003). However the Discrimination Capacity that was found in Tanzanian population was lower than Guinea Bissau population (Rosa *et al*., 2007).

In Sukuma and Kuria locus DYS438 showed low level of polymorphism, while in Luo, Maasai, Iraqw and Alagwa low levels of polymorphism was found in locus DYS391. In all populations the lowest level of polymorphism was found in locus DYS392 (figure 4.12). Similar results were found in Cental African Pygm population (Kayser *et al*., 2001), Mozambican (Alves *et al*., 2003) and South African Xhosa population (Leat *et al*., 2004). Therefore the utility of locus DYS391 and DYS392 is of a limited use in forensics while all other loci studied are likely to be useful (when combined in haplotype) in forensic casework analysis in Tanzania.

It is noteworthy that the levels of genetic diversity for the haplotypes per group (haplotype diversity) (population diversity parameters Table 4.8) are not simililar to the levels of genetic diversity at the individual markers in the corresponding populations (Table 4.7), a typical example is in Iraqw population which has higher gene diversity values in most of the loci compared to other population but its level of genetic diversity is lower compared to other populations such as Sukuma and Kuria.

In comparison to other African population, the level of genetic diversity found in Tanzanian population for these sets of markers were higher than those reported for Tunisia (Brandt-Casadevall *et al*., 2003), Uganda (Gomes *et al*., 2010), Angola (Melo *et al*., 2010) and Mozambique (Alvez *et al*., 2003). However the level of genetic diversity in Tanzanian population was lower than Guinea Bissau population (Rosa *et al*., 2007).

When the most common haplotypes from Sukuma (Niger Congo), Kuria (Niger Congo), where searched in YHRD (figure 4.14 and 4.15) matches where found in African Bantu populations of Angola, Botswana and South Africa indicating a common ancestry or gene

flow in these Bantu populations. On the otherhand haplotypes from Luo (Nilo Saharan) (figure 4.18) and Alagwa (Afro Asiatic) (figure 4.13) were shared by other African Bantu populations in Mozambique, Angola and South Africa indicating admixture which might have occurred during Bantu expansion period approximately 3000 years ago, also during Nilo Saharan expansion which occurred posterior to Bantu expansion .

When the 225 haplotypes obtained form the six populations were searched in the YHRD (as of 30th December, 2010), matches were found in 82 haplotypes, no matches were found in 143 haplotypes, this suggest that still more genetic studies have to be done in African populations and addition of the data in the YHRD database for population and forensic application purposes.

Chapter 5

Conclusion

Y- STR data indicated the existance of genetic differentiation among the six populations of Tanzania which could not be attributed to language barriers. The greater genetic diversity which was found among Tanzanian populations was cont

ributed by a limited level of gene flow between these populations due to different ethnical, social, culturall as well as historical backgrounds between these populations. This was expected through previous genetic studies that have been conducted in Africa and found that there was greater genetic diversity in African populations.

When compared to other African populations using the same type of markers, Tanzanian population seem to be more diverse compared to Uganda (Gomes *et al*., 2010), Tunisia (Brandt-Casadevall *et al*., 2003), Angola (Melo *et al*., 2010) and Mozambique (Alvez *et al*., 2003), but less diverse compared to Guinea Bissau population (Rosa *et al*., 2007).

Genetic variability among the individuals in the six populations of Tanzania was studied and the discrimination potential of the Y chromosome extended haplotype markers in the six populations was evaluated. With exception of DYS391 and DYS92 loci which showed low levels of variability, all other Y chromosome extended haplotype loci are likely to be useful (when combined in haplotype) for forensic application in Tanzania. Furthermore the extended haplotype markers used in this study may be useful in the establishment of the National DNA database following the enactment of the Human DNA Legislation in Tanzania (http://www.parliament.go.tz).

It would also be interesting, future studies to be carried out to further investigate these population groups of Tanzania using additional Y chromosome STRs markers such as UWC 10-Plex (D'Amato et al., 2011) and Y-filer (Mulero et al., 2006) which would increase the discrimination capacity for forensic application, also using mtDNA and ancestry informative Y SNPs markers to provide further information on the ethnic origin and history of the population groups in Tanzania.

References

Ambrose, S.H. (1982). The Archaeological and Linguistic Reconstructions of History in Eastern Africa. In C. Ehret and M. Posnansy (Eds.), *The Archaeological and Linguistic reconstruction of African history*, (pp. 104–157). University of California Press.

Alves, C., Gusmao, L., Barbosa, J. and Amorim, A. (2003). Evaluating the Power of Y – STRs: a Comparative Study using European and new African Haplotype Data, *Forensic Science International*, 134, 126 – 133.

Ayub, Q., Mohyuddin, A., Qamar, R., Mazhar, K., Zerjal T., Mehdi, SQ., and Tyler- Smith, C. (2000). Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res* 28 (2), e8.

Baumann, O. (1894). Durch Massailand zur Nilquelle: Reisen und Forschungen der Massai-Expedition des deutschen Antisklaverei-Komite in den Jarhen 1891-1893. *New York: Johnson Reprint*. German.

Belkhir, K. (1999). GENETIX, logiciel sous WindowsTM pour la ge´ne´tique des populations, Laboratoire Ge´nome et Populations, Universite´ de Montpellier II.

Berntsen, J. (1976). The Maasai and their neighbors: variables of interactioin. *African economic history, 2.*

Blench, R. (2006). *Archaeology language and the African past*, AltaMira press.

Brandt-Casadevall, C., Dhiab, B.M., Taroni, F., Castella, V., Dimo-Simonin, N., Zemni, M., Mangin, P. (2003). Tunisian population data on 10 Y-chromosomal loci. *Forensic Science International* 13, 247–250.

Buckleton, J., Triggs, C.M., Walsh, S.J. (2005). *Forensic DNA evidence interpretation*. CRC press.

Butler, J.M. (2005). *Forensic DNA Typing: Biology, Technology and genetics of STRs markers*, 2nd edition. London: Elisevier Academic Press.

Butler, J.M., Schoske, R., Vallone, P.M., Kline, M.C., Redd, A.J. and Hammer, M.F. (2002). A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Science International* 129, 10-24.

Butler, J., Decker, A., Vallone, P., and Kline M. (2006). Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African–American and Hispanic samples. *Forensic Science International* 156, 250-260.

Butler, J. (2003). Recent Developments in Y-Short Tandem Repeats and single nucleotide polymorphism analysis. *Forensic Science Revised* 15, 91-111.

Butler, J. (2001). *Forensic DNA typing, Biology and Technology behind STR markers*. Academic Press.

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987). Mitochondrial DNA and human evolution. *Nature* 325, 31-36.

Chalcraft, J. (2005). 'Varimu Valale': Rock Art as World Heritage in a Ritual Landscape of Central Tanzania. In W. James and D. Mills (Eds.), *The qualities of time: anthropological approaches.*

Crow, J. F., Berger M.A., Diamond, S.S., Kaye, D.H., Kazazian, H.H., Motulsky, G.A., Nagyalaki, T. A., Nei, M., Sensabaugh, G. F., Siegmund, D.O., and Stigler, M. S. (1996). *The evaluation of forensic DNA evidence*. National Academy Press.

D'Amato, M.E., Ehrenreich, L., Benjeddou, M., Davison, S., and Leat, N. (2008). Ancestry and genetic relationships between groups within the Cape Town metropolitan population inferred using Y-STRs genotyping. *Forensic Science International: Genetics Supplement Series* 1:318-319.

D'Amato, M.E., Bajic, V.B, Davison, S. (2011). Design and Validation of highly discriminatory 10-locus Y - chromosome STR multiplex system. *Forensic Science International: Genetics* 5, 122-125.

De Knijff P. (2000). Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *American Journal of Human Genetics,* 67 1055–1061.

De Knijff, P., Kayser, M., Caglia`, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino,G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P. M., Szibor, R., Teifel-Greding, J., Weichhold, G., and Roewer, L. (1997). Chromosome Y microsatellites: population genetic and evolutionary aspects. *International Journal of Legal Medicine,* 110:134–140

Excoffier, L., Laval, G., and Schneider S. (2006). Arlequin version 3.11: An Integrated Software Package for population genetics data analysis. *Evolutionary Bioinformatics online* 1, 47-50.

Excoffier, L., Smouse, P., and Quattro., J. (1992). Analysis of Molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479 – 491.

Ehret, C. (2000). Language and history. In B. Heine and D. Nurse (Eds.), *African languages: an introduction*. Cambridge: Cambridge University Press. 272–297.

Ehret, C. (1974). Cushites and the Highland and Plains Nilotes to A.D 1800. In B.A. Ogot (ed), *Zamani a survey of East African History*. East African Publishing House and Longman Group Ltd.

Feierman, S. (1995). Economy, society and language in early East Africa. In Philip Curtin, Steven Feierman, Leonard Thompson and Jan Vansina (Eds.), *African History: From Earliest Times to Independence*. New York: Longman, 116–120.

Fleisher, M. L. (2002). 'War is good for thieving!' the symbiosis of crime and warfare among the Kuria of Tanzania. *Africa*, 72, 1.

Fosbrooke, Henry A. (1954). The defensive measures of certain tribes in north-eastern Tanganyika. Part II: Iraqw housing as affected by inter-tribal raiding.*Tanganyika Notes and Records,* 36, 50–57.

Giblin, J. L. (1992). *The Politics of Environmental Control in Northeastern Tanzania, 1840–1940*. University of Pennsylvania press.

Gomes, V., Alves, C., Amorim, A., Carracedo, A., Sanchez – Diz, P. and Gusmao, L. (2010). Nilotes from Karamoja, Uganda: Haplotype Data Defined by 17 Y – Chromosome STRs, *Forensic Science International: Genetics* 4: e83–e86.

Goodwin, W., Linacre, A, and Hadi, S. (2007). *An introduction to forensic genetics*. John Wiley & Sons Ltd.

Guinand, B. (1995). Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biological Journal of Linnean society* 58: 173-195.

Hall, A, and Ballantyne, J. (2003a). The development of an 18-locus Y-STR system for forensic casework. *Anals Bioanalyt Chem* 376: 1234-1246.

Heine, B and Nurse, D. (2000). *African languages: an introduction*. Cambridge: Cambridge University Press.

Henn, B.M., Gignoux, C., Lin, A. A., Oefner, P. J., Shen, P., Scozzari, R., Cruciani, F., Tishkoff, S. A., Mountain, J.L. and Underhill P.A. (2008). Y chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *PNAS* 105, 31(10693 – 10698).

Iliffe, John. (1979). *A Modern History of Tanganyika*, African Studies Series 25. Cambridge: Cambridge University Press.

Jobling, M.A., Bouzekri, N and Taylor, P.G. (1998). Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y specific minisatellite, MS (DYF155S1). *Human Molecular Genetics* 7, 643-653.

Jobling MA, Pandya A. and Tyler-Smith, C. (1997). The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine,* 110: 118–124.

Jobling, M.A and Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics,* 4, 598–612.

Jobling, M.A and Tyler-Smith, C. (2000)**.** New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends in Genetics* 16 (8), 356 -362.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Grazios,i G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C. and Roewer, L. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *International Journal of Legal Medicine* 110, 125–133.

Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L., Jespersen, J., Bakker, E., Roewer, L., De Knijff, P. (2001). An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations, *American* Journal *of Human Genetics* 68, 990–1018.

Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A.C., Mohyuddin, A., Mehdi, S. Q., Rosser, Z., Stoneking, M., Jobling, M. A., Sajantila, A. and Tyler-Smith, C. (2004). A comprehensive survey of human Y chromosomal Microsatellites. *American Journal of Human Genetics* 74: 1183-1197.

Kayser, M., Vermeulen, M., Knoblauch, H., Schuster, H., Krawckzak, M. and Roewer, L. (2007). Relating two deep-rooted pedigrees from Central Germany by high-resolution Y STR haplotyping, *Forensic Science International: Genetics* 125-128.

Keen, D. (1996). 'Organised chaos: not the new world we ordered', *World Today* 52 (1), 14 -17.

Knight, A., Underhill P. A., Mortesen, H. M., Zhivotovsky, A., Lin, A. A., Henn, B. M., Louis, D., Ruhlen, M., and Mountain J. L. (2003). African Y Chromosome and mtDNA Divergence Provides Insight into the History of Click Languages. *Current biology,* 13, 464 – 473.

Kjekshus, H. (1977). *Ecology Control and Economic Development in East African History*: *The Case of Tanganyika 1850–1950*. Berkeley, CA: University of California.

Koponen, J. (1988). *People and Production in Late Precolonial Tanzania*: *History and Structures.* Helsinki: Finnish Society for Development Studies.

Leat, N., Benjeddou, M. and Davison, S. (2004). Nine-locus Y-chromosome STR profiling of Caucasian and Xhosa populations from Cape Town, South Africa. *Forensic Science International* 144, 73 - 75.

Leat, N., Ehrenreich, L., Benjeddou, M., Cloete, K. and Davison, S. (2007). Properties of novel and widely studied Y-STR loci in three South African populations. *Forensic Science International* 168, 154-161.

Medrano, J.F., Aasen, E., and Sharrow, L. (1990). DNA extraction from nucleated red blood cells. *BioTechniques*, 8 (1), 43.

Melo, M.M., Carvalho, M., Lopes, V., João Anjos, M., Serra, A., Vieira, D.N., Sequeiros, J., and Corte-Real, F. (2010). Y- STR haplotypes in three ethnic linguistic groups of Angola population. *Forensic Science International: Genetics*. *doi: 10,1016/j.fsigen.2010.08.002*

Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keys, K.M., Smerick, J.B. and Budowle, B. (2001). Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *Journal of Forensic Science* 46, 647–660.

Mulero, J.J., Chang, C.W., Calandro, L.M., Green, R.L., Li, Y., Johnson, C.L., and Hennessy, L.K. (2006). Development and validation of the AmpFłSTR® Yfiler PCR amplification kit : A male specific, single amplification 17 Y-STR multiplex system. *Journal of Forensic Science* 51 (1), 64-75.

Mullis, K. B., Faloona, F., Scharf, S. J., Saiki, R. K., Horn, G. T. and Erlich, H.A. (1986). Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.,* 51 Pt 1, 263-273.

Nei M. (1987). Molecular Evolutionary Genetics*. New York: Columbia University Press.*

Newman, J. (1995). *The peopling of Africa: a geographic interpretation*. New Haven (CT): Yale University Press.

.

Omran, G.A., Rutty, G. N. and Jobling M. A. (2008). Diversity of 17-locus Y-STR haplotypes in Upper (Southern) Egyptians. *Forensic Science International: Genetics Supplement Series* 1, 230–232.

Pascali, V.L., Dobosz, M. and Brinkmann, B. (1998). Coordinating Y-chromosomal STR research for the Courts. *International Journal of Legal Medicine,* 112, 1.

Posnansky, M. (1974). The prehistory of East Africa. In B.A. Ogot (Ed.), *Zamani a survey of East African History*. East African Publishing House and Longman Group Ltd.

Prinz, M., Boll, K., Baum, H. and Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Science International* 85, 209–218.

Quintana-Murci., L. and Fellous., M. (2001). The Human Y Chromosome: The Biological Role of a 'Functional Wasteland'. *Journal of Biomedicine and Biotechnology* 1, 18-24.

Quintana-Murci, L., Krausz, C., and McElreavey, K. (2001). The human Y-chromosome: function, evolution and disease. *Forensic Science International* 118 (2-3), 169-181.

Redd, A., Agellon, A., Kearney, V., Contreras, V.A., Karafet, T., Park, H., de Knijf, P., Butler, J.M. and Hammer, M.F. (2002). Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Science International* 130, 97-111.

Reed, F. A. and Tishkoff, S. A. (2006). African human diversity, origins and migrations. *Curr. Opin. Genet. Dev.* 16, 597.

Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Cagliá, A., Carracedo, A., Corach, D., Dekairelle, A.F., Dobosz, T., Dupuy, B.M., Füredi, S., Gehrig, C., Gusmaõ, L., Henke, J., Henke, L., Hidding, M., Hohoff, C., Hoste, B., Jobling, M.A., Kärgel, H,J,, de Knijff, P., Lessig, R., Liebeherr, E., Lorente, M., Martínez-Jarreta, B., Nievas, P., Nowak, M., Parson, W., Pascali, V.L., Penacino, G., Ploski, R., Rolf, B., Sala, A., Schmidt, U., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J. and Kayser, M. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Science International* 118, 106-113.

Roewer L, Arnemann J, Spurr N, et al. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Human Genetics* 89, 389-394.

Roewer L., (2000). A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic science international 114, 31-43.*

Roewer, L. (2003). The Y Short Tandem Repeat Haplotype Reference Database (YHRD) and male population stratification in Europe—impact on forensic genetics. *Forensic Sci. Rev.* 15, 164–70.

Roewer, L. (2009). Y chromosome STR typing in crime casework. *Forensic science medical pathology, 5, 77 – 84.*

Rousset, F. (2007). GenePop'007: A complete re-implementation of the GenePop software for Windows and Linux. *Molecular Ecology Resources* 8, 103–106.

Ruel, M. (1991). 'Kuria seers', *Africa: Journal of the international African Institute,* 61 (3). 343 – 352.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Horn, G. T., Mullis, K. B., Erlichet, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermalstable DNA polymerase. *Science* 239, 487-491.

Saad, R. (2005). Discovery, development, and current applications of DNA identity testing. *Proc (Baylor University Medical Center).* 18 (2): 130–133.

Schneider, P.M., D'Aloja, E., Dupuy, B.M, Eriksen, B., Jangblad, A., Kloosterman, A.D., Kratzer, A., Lareu, M.V., Pfitzinger, H., Rand, S., Scheithauer, R., Schmitter, H., Skitsa, I., Syndercombe-Court, D. and Vide, M.C. (1999). Results of a collaborative study regarding the standardization of the Y-linked STR system DYS385 by the European DNA Profiling (EDNAP) group. *Forensic Science International*, 102, 159 – 165.

Scientific Working Group on DNA Analysis Methods (SWGDAM). (2004). Revised validation guidelines. *Forensic Science Communications* 6, 3

Shewale, J., Nasir, H., Schneida, E., Gross, A.M., Budowle, B. and Sinha, S.K. (2004). Y-chromosome STR system, Y-PLEX^{TM} 12, for forensic casework: development and validation. *Journal of Forensic Science* 49, 1-13.

Shetler, Jan Bender. (2007). *Imagining Serengeti: A History of Landscape Memory in Tanzania from Earliest Times to the Present*. Athens, OH: Ohio University.

Spear, T. and Waller, R. (1993). *Being Maasai: Ethnicity and Identity in East Africa*. London: James Currey

Sutton, J. E. P. (1974). The settlement of East Africa. In B.A. Ogot (Ed.), *Zamani a survey of East African History*. East African Publishing House and Longman Group Ltd.

Tanner, R. E. S. (1999). The Sukuma of Tanzania: a traditional social system enabling adaptation to modernity. In G. N. Ratha, G. Pfeffer and D. K. Behera (Eds.), *Contemporary society: Tribal Studies, iv.*

Ten Raa, E. (1969). The moon as a symbol of life and fertility in Sandawe thought. *Africa: Journal of the International African Institute,* 39, 24-53.

Tishkoff, S.A., Williams, S.M. (2002). Genetic analysis of African populations: human evolution and complex disease. *Nature Review Genetics.* 3, 611–621.

Tishkoff S. A., Gonder, M.K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T. B., Ramakrishnan, U., Reed, F. A., and Mountain, J. L. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology Evolution,* 24, 2180-2195.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J. B., Awomoyi, A. A., Bodo, J.M., Doumbo, O., Ibrahim, M., Juma, T. A, Maritha J. K., Lema, G., Moore, H. J., Mortensen, H., Nyambo, T. B., Omar, S.A., Powell, K., Pretorius, G. S., Smith, M. W., Thera, A. M., Wambebe, C., Weber, J. L. and Williams, S. M. (2009). The genetic structure and history of Africans and African Americans. *Science,* 324, 1035-1044.

Watson J. D., and Crick F. H. C. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature,* 171, 737- 738.

Willuweit, S. and Roewer, L. (2007). Y chromosome haplotype reference database YHRD: Update. *Forensic Science International: Genetics*, 1, 83-87.

Wood, E.T., Stover, D.A., Ehret, C., Destro-Bisol, G., Spedini, G., McLeod, H., Louie, L., Bamshad, M., Strassmann, B., Soodyall, H. and Hammer, M.F. (2005). Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes*. European Journal of Human Genetics,* 13, 867–876.

Electronic Supplementary Resources

1. Arlequine: http://cmpg.unibe.ch/software/arlequin3

2. Ethnologue: http://www.ethnologue.com

3. Tanzanian Government Chemist Laboratory: http://www.gcla.go.tz

4. Tanzanian Parliament: http://www.parliament.go.tz

5. Tanzanian Police Force: http://www.policeforce.go.tz

6. Genetix: http://www.univmontp2.fr/_genetix

7. YHRD: http://www.yhrd.org

8. Applied biosystems database: www. Appliedbiosystems.com/yfilerdatabase.

9. Genepop: http://www.cecill.info/index.en.html

Appendix

1. DNA Extraction

1.1. Reagents

5.5 M NaCl

20 mg/ml Proteinase K stored at - 20 °C

Isopropanol

70 % Ethanol stored at -20 °C

Lysis Buffer

- 400 mM NaCl
- 10 mM Tris HCl (pH = 8)
- 2 mM EDTA
- 1 % SDS

1.2. Preparation of Lysis Buffer (Salt lysis)

Lysis Buffer reagents were then added to approximately 80 % of their final volume, of distilled water, SDS was also added and left in an oven set at 60 °C until fully dissolved. The volume was transferred to a volumetric flask and SABAX water was added in order to obtain 100 % of the required volume.

1.3. Procedure

1. The surface of the swab was cut with a clean scalpel/ surgery blade, on a clean petri dish, on top of a tray.
2. 400µL of lysis buffer and 2µL proteinase K (20mg/ml) were added in 2ml Eppendorf tubes.
3. Little pieces of excised swab were added to the Eppendorf tubes containing lysis buffer and proteinase K, the mixture were then vortexed for 30 seconds followed by overnight incubation at 56 ºC.

4. The volume was transferred to a clean tube.
5. As the bud still contained trapped lysis solution, a 0.5 ml tube was perforated with a needle and the bud was placed in perforated tube.
6. The perforated tube was placed in another Eppendorf tube (1.5ml) and spun down, in a micro – centrifuge, for one minute. The volume was collected and added to previous volume.
7. The contents were precipitated by addition of ⅓ of the volume of 5.5 M NaCl. The tubes were then vigorously shaken for 15 seconds and centrifuged at 5000 rpm for 15 minutes.
8. The supernatant, containing DNA, was transferred to another tube and an equal volume of cold isopropanol was added.
9. This was followed by 30 minute incubation at -80 ºC.
10. DNA was pelleted by centrifugation at 14000 rpm for 30 minutes. The pellet was washed with 70 % ethanol, to remove excess salts, and this was followed by centrifugation at 14000 rpm for approximately 10 minutes.
11. The pellet was dried but precautions were implemented as to prevent DNA from becoming too dry. The pellet was dissolved in 30 µl SABAX water and stored at -20 ºC.
12. DNA concentrations were recorded using a Nanodrop spectrophotometer and working dilutions of 5 ng/µl were prepared. Original DNA stock solutions as well as working solutions were stored at – 20 °C.

1.4. Quantifying DNA and working stock dilution

1. 1 µl of the extracted DNA was used for quantification using Nanodrop ND1000 spectrophotometer
2. The DNA concentration and purity was calculated from the Nanodrop spectrophotometer, readings were recorded.
3. Working stock of 2 – 5 ng/µl were made by dilutions calculated from concentrations of all the DNA samples.
4. The original DNA stocks as well as working stock solutions were stored at -20 °C

1.5. DNA Quantification results and purity readings as obtained from Nanodrop ND1000 spectrophotometer

S/n	Sample Number	concentration (ng/uL)	Purity (260/280 ratio)
1	Skm 011	7.1	1.9
2	Skm 012	12.9	1.8
3	Skm 013	5.4	1.69
4	Skm 014	13.8	1.85
5	Skm 015	6.1	1.91
6	Skm 016	20.3	1.81
7	Skm 018	12	2.02
8	Skm 019	18.8	1.87
9	Skm 020	45	1.79
10	Skm 034	8.5	1.75
11	Skm 041	10.4	1.81
12	Skm 042	5.3	2.17
13	Skm 045	14.2	1.97
14	Skm 046	39	1.95
15	Skm 047	100.6	1.92
16	Skm 048	36.3	1.87
17	Skm 049	7.7	1.86
18	Skm 050	17.4	1.85
19	Skm 053	152.5	1.77
20	Skm 054	44.2	1.86
21	Skm 055	10	1.72
22	Kra 056	19.4	2
23	Kra 057	12.7	2
24	Kra 058	6.1	1.92
25	Kra 060	12	1.98
26	Kra 061	16.9	1.7
27	Kra 062	59.3	1.79

28	Kra 063	194.7	1.76
29	Kra 064	60.4	1.71
30	Kra 065	86.6	1.7
31	Kra 066	76.6	1.79
32	Kra 067	71.6	1.69
33	Kra 068	118.5	1.68
34	Kra 069	47.5	1.79
35	Kra 070	129.2	1.72
36	Kra 071	41.95	1.9
37	Kra 072	5.7	2
38	Kra 095	34.9	1.89
39	Kra 096	17.6	1.94
40	Kra 097	43	1.85
41	Kra 098	38.4	1.82
42	Kra 099	19.3	1.89
43	Kra 100	26.3	1.82
44	Kra 101	18.9	1.99
45	Kra 102	14.1	1.76
46	Kra 103	38.9	1.8
47	Kra 104	5.7	1.7
48	Kra 106	51.1	1.78
49	Kra 107	21.4	1.82
50	Kra 108	44.7	1.76
51	Kra 112	13.4	1.86
52	Kra 113	14.1	1.94
53	Kra 114	6	1.79
54	Kra 116	14.9	1.78
55	Kra 117	10.1	1.93
56	Kra 118	40.6	1.81
57	Kra 119	10	1.71
58	Kra 120	17.6	2

59	Kra 122	35.1	1.99
60	Kra 123	19.9	2.01
61	Kra 124	34	1.7
62	Kra 125	7.9	2.01
63	Kra 126	10.3	2
64	Kra 128	17.1	2
65	Kra 129	35.7	2.01
66	Kra 131	25.3	1.85
67	Kra 133	13.9	1.94
68	Kra 135	12	1.8
69	Kra 138	13.4	2
70	Kra 139	28.5	2.09
71	Kra 140	15.6	1.82
72	Kra 141	12.9	2.2
73	Kra 142	18.3	1.97
74	Kra 143	21.9	2.09
75	Kra 144	7.9	2.19
76	Kra 145	40.9	1.98
77	Kra 146	10.4	2
78	Kra 169	44	1.95
79	Luo 151	7.4	2.02
80	Luo 152	10.2	2.19
81	Luo 153	9.3	1.98
82	Luo 154	10.1	2.1
83	Luo 155	20.1	2.12
84	Luo 158	16.7	2.2
85	Luo 159	21.2	2.14
86	Luo 160	22.2	1.98
87	Luo 163	19	1.93
88	Luo 164	17.2	1.83
89	Luo 165	53.6	1.9

90	Luo 170	28	1.99
91	Luo 171	25.6	1.77
92	Luo 172	13.3	1.88
93	Luo 173	8.3	1.7
94	Luo 174	25.5	1.83
95	Luo 175	27	1.84
96	Luo 176	27.5	1.95
97	Luo 177	16.2	1.85
98	Luo 179	16.9	1.65
99	Luo 180	18.6	1.85
100	Luo 182	17.7	1.86
101	Luo 184	30.2	1.97
102	Luo 185	34.1	1.83
103	Luo 187	29.9	1.91
104	Luo 188	29.3	1.88
105	Luo 189	35.8	1.88
106	Luo 190	21.1	1.72
107	Luo 191	16	1.83
108	Luo 192	27	1.68
109	Luo 194	15.1	1.78
110	Luo 195	63	1.9
111	Luo 199	55.4	1.74
112	Luo 200	81	1.8
113	Irq 201	18.7	1.91
114	Irq 202	12.6	1.98
115	Irq 203	11.8	1.78
116	Irq 204	35.2	1.98
117	Irq 206	47.2	1.85
118	Irq 207	26.7	1.98
119	Irq 208	37.5	1.75
120	Irq 209	21.7	1.85

121	Irq 210	92.7	1.88
122	Irq 211	9.5	1.85
123	Irq 213	61	1.83
124	Irq 214	15.3	1.97
125	Irq 215	20.4	2.11
126	Irq 216	36.7	1.78
127	Irq 217	61.3	1.9
128	Irq 218	64.3	1.96
129	Irq 219	36.8	1.83
130	Irq 220	94.9	1.83
131	Irq 221	48	1.9
132	Irq 222	54.2	1.71
133	Irq 223	50.7	1.85
134	Irq 224	21.6	1.89
135	Irq 225	24.2	1.94
136	Irq 226	26.2	1.83
137	Irq 227	18.9	1.78
138	Irq 228	22.3	2.27
139	Irq 230	16.5	2.09
140	Irq 231	66.1	1.75
141	Irq 232	40.5	1.86
142	Irq 233	16.4	1.88
143	Irq 234	10.8	2
144	Irq 235	27.2	1.95
145	Irq 236	106.5	1.84
146	Irq 237	43.2	1.95
147	Irq 238	30.4	1.7
148	Irq 239	29.2	1.87
149	Irq 240	43.6	1.91
150	Irq 241	21.4	1.99
151	Irq 242	35.5	1.71

152	Irq 243	43.5	1.75
153	Irq 244	52.6	1.75
154	Irq 245	188.3	1.81
155	Irq 246	9.7	1.79
156	Irq 247	49.3	1.71
157	Irq 248	58.5	1.67
158	Irq 249	43.4	1.71
159	Irq 250	26.6	1.72
160	Irq 251	15.9	2
161	Irq 252	19.4	1.88
162	Irq 253	11.4	1.85
163	Irq 254	12.9	2
164	Irq 255	15.5	2
165	Irq 257	40.8	1.73
166	Irq 258	25.8	1.89
167	Irq 259	16.1	1.85
168	Irq 260	13.7	2.08
169	Irq 261	70.5	1.78
170	Irq 262	72	1.81
171	Irq 263	42.5	1.7
172	Irq 264	23.1	1.76
173	Irq 265	20.3	1.98
174	Irq 266	1.79	1.79
175	Irq 267	18.1	1.87
176	Irq 268	39.8	1.87
177	Irq 269	25.7	1.84
178	Irq 270	23	1.69
179	Irq 271	26.5	1.67
180	Msi 273	10.3	2
181	Msi 274	15.4	1.94
182	Msi 278	23	2

183	Msi 279	36.9	2.05
184	Msi 280	18.1	1.89
185	Msi 282	48.7	1.82
186	Msi 283	37.5	1.94
187	Msi 285	22.6	1.95
188	Msi 286	10.2	1.9
189	Msi 287	18.9	2
190	Msi 288	44.4	1.95
191	Msi 289	29.1	2
192	Msi 290	55.2	1.93
193	Msi 291	7.4	2.08
194	Msi 292	11.2	2
195	Msi 293	13.1	1.78
196	Msi 294	14.2	2.05
197	Msi 295	84.3	1.86
198	Msi 296	25.7	2
199	Msi 297	27.4	1.95
200	Msi 298	27.9	1.88
201	Msi 299	42.4	1.87
202	Msi 300	20	1.78
203	Msi 301	22.2	1.77
204	Msi 302	17.2	2
205	Msi 304	40.9	1.82
206	Msi 305	32.1	1.8
207	Msi 306	21.7	1.8
208	Msi 308	38	1.78
209	Msi 309	35.2	1.66
210	Msi 310	15.1	1.8
211	Msi 311	20.4	1.67
212	Msi 313	22.4	1.74
213	Msi 316	20.1	1.98

214	Msi 320	27.5	1.78
215	Msi 321	45.3	1.73
216	Msi 322	19.3	1.67
217	Msi 323	18.9	1.87
218	Msi 324	15.8	1.69
219	Msi 325	27.6	1.83
220	Msi 326	28.7	1.69
221	Msi 327	24.6	1.7
222	Msi 328	26.5	1.87
223	Msi 329	25.2	1.84
224	Msi 330	9.2	1.87
225	Msi 331	20.8	1.72
226	Msi 332	16.3	1.98
227	Msi 333	49.7	1.83
228	Msi 334	12.9	1.96
229	Msi 337	92.4	1.67
230	Msi 384	30	1.74
231	Alw 338	21.9	1.84
232	Alw 339	49.6	1.84
233	Alw 340	13.6	1.95
234	Alw 341	31.7	1.73
235	Alw 342	32.8	1.65
236	Alw 343	42.2	1.82
237	Alw 344	23.6	1.92
238	Alw 345	12.4	2.1
239	Alw 347	39.1	1.9
240	Alw 348	37.6	1.65
241	Alw 349	18.7	1.84
242	Alw 350	24.3	1.89
243	Alw 351	10.1	2
244	Alw 352	31.3	1.97

245	Alw 353	39.4	1.6
246	Alw 354	28.9	1.87
247	Alw 356	43.8	1.67
248	Alw 357	32.8	1.8
249	Alw 358	26.9	1.82
250	Alw 359	16.1	1.84
251	Alw 360	31.7	1.98
252	Alw 362	20.5	1.96
253	Alw 363	29.3	1.86
254	Alw 364	46.6	1.91
255	Alw 365	29.6	1.67
256	Alw 366	16.5	1.83
257	Alw 367	37	1.93
258	Alw 369	42.6	1.9
259	Alw 370	33.8	1.88
260	Alw 372	83	1.84
261	Alw 374	12.5	2
262	Alw 376	17.2	1.99
263	Alw 378	21.7	2
264	Alw 379	11.6	1.93
265	Alw 380	38.4	1.86
266	Alw 381	27	1.76
267	Alw 383	15.6	1.85

2.0 Fragment Analysis

2.1 Reagents

10X TBE Buffer (2 Litres)

- Tris (*Merck Laboratory supplies*) 216 gm
- Boric Acid (*Merck Laboratory supplies*) 112 gm
- EDTA (*Merck Laboratory supplies*) 16 gm

- Distilled Water to 2 Litres

4 % Polyacrylamide Gel Mix

- Urea (*Merck Laboratory supplies*) 72 gm
- 40 % 19:1 Polyacrylamide Solution (*Promega*) 20 ml
- 10X TBE Buffer 20 ml
- Distilled Water to 200 ml

TEMED (*Sigma Aldrich*)

10X AMPS (10ml)

- Ammonium Persulfate 1 gm
- Distilled Water 10 ml

Loading Buffer

- Formamide 5 µl
- Loading Dye (*Applied Biosystems*) 2 µl
- LIZ500 (*Applied Biosystems*) 1 µl

3. Preparation of Gel running Plates

3.1 Gel Preparation

1. A 4 % Polyacrylamide gel was prepared
2. 25 ml of gel mix was then transferred to a clean 50 ml beaker.
3. With a 50 ml syringe, gel mix was pulled up and filtered through a 0.22 micron filter directly into a side-arm flask.
4. Gel mix was swirled gently then degassed, for approximately 5 minutes, with intermittent gentle agitation.
5. Gel mix was then transferred from the side-arm flask to a clean 50 ml beaker.
6. 125 µL (AMPS) and 17.5 µL N, N, N', N' Tetramethyl-EthyleneDiamine (TEMED) was added to opposite sides of the beaker, and swirled gently

3.2 Plate Set – Up

1. The plates were cleaned using distilled water

2. The plates were set – up according to the manufacturer's instructions.

3.3 Pouring of Gel

1. Gel mix, containing AMPS and TEMED, was poured into plates which was previously mounted and left for two hours to solidify

2. The plates were cleaned with distilled water and then dried.

3. Plates were placed onto ABI® 377 DNA Sequencer and set – up was followed according to the manufacturer's instructions.

.