KENYA POPULATION FORENSIC DATA ON A SIXTEEN LOCI MICROSATELLITE DNA SYSTEM

KIMANI KAGUNDA JOSEPH (B.Sc)
Reg, No. I56/ 5179/ 03

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (BIOTECHNOLOGY) OF KENYATTA UNIVERSITY

MAY 2009

DECLARATION

I declare that the work presented in this thesis is my original work and has not been presented for a degree in any other university or any other award.

Signature ..... Date 29/05/2009.....

Kimani Kagunda Joseph (B.Sc)

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Supervisors

Signature ..... Date 29/05/2009.....

Prof Eliud N.M. Njagi
Department of Biochemistry and Biotechnology
Kenyatta University
P.O Box 43844-00100
Nairobi, Kenya.

Signature ..... Date 29/5/2009.....

Dr George O. Orinda
Department of Biochemistry and Biotechnology
Kenyatta University
P.O Box 43844-00100
Nairobi, Kenya.

Signature ..... Date 29/05/2009.....

Dr Paul N. Njiru
Department of Health Sciences and Biotechnology
Kenya Polytechnic University College
P.O Box 52428
Nairobi.

DEDICATION

To my wife Nancy Kagunda, Joy Kagunda my daughter, William Kimani my son and the entire Kimani's family.

ACKNOWLEDGEMENTS

I wish to sincerely thank the Almighty God who purposed for me to be in this world and has continually kept me alive and well each day of my life. You have channeled the course of my life and I say thank you. My heart also goes out to my dear wife Nancy N Kagunda, my daughter Joy W Kagunda and son William K Kagunda for their continued support and love throughout the entire time that I was engaged with this work, "I love you". My gratitude also goes to my parents Mr & Mrs William Kimani who kept the fire burning through inspiration, encouragement and financial support to ensure this work is a success, "God bless you".

My sincere thanks also go to my supervisors; Prof Eliud NM Njagi, Dr Paul N Njiru and Dr George O Orinda for the best guidance, criticism and mentoring that you have provided all the way to make this work a success. Prof Njagi you are more like a father. Dr Njiru you took me the hard path of data interpretation and evaluation I am so grateful, Dr Orinda you have always been there with encouragement and support. "May God bless you all".

I cannot forget the support and motivation from the Government Chemist department my employers who paid part of my college fees, provided the money for research and the laboratory to carry out the analysis. I particularly thank Dr Charles K Koskei the former Government Chemist, Mrs Jane A Okado the current Government Chemist, Mr John Mungai, Rose Sikuku and Henry Sang "you are wonderful". I will never forget you people and God bless you.

Prof Chakraborty Ranajit and his student Mr Jianye Ge are not just ordinary folks, we have never met and you agreed to do the data analysis for free. It was a lot of work and many hours of input and I say heartedly a big God bless you.

I cannot forget the serious prayers and moral support from fellow pastors, members of the Lifeline of churches Ministries indeed you are wonderful people. Finally a lot of gratitude for Mr Joseph Onyango (Regional Blood Transfusion Kisumu), Mr Patrick Wambua (Regional Blood Transfusion Mombasa), Mr. Abdi Mohammed (Regional Blood Transfusion Nairobi).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

DNA	-	Deoxyribonucleic acid
VNTR	-	Variable Number of Tandem Repeats
PCR	-	Polymerase Chain Reaction
STR	-	Short Tandem Repeats
AMOVA	-	Analysis of Molecular Variance
AAFS	-	American Academy of Forensic Science
FSS	-	Forensic Science Services
FBI	-	Federal Bureau of Investigation
CODIS	-	Combined DNA Index
U.S.A	-	United States of America
PGM1	-	Phosphoglucomutase1
ESD	-	Esterase D
EAP	-	Erthrocyte Acid Phosphatase
PD	-	Power of Discrimination
HWE	-	Hardy Weinberg Equilibrium
RFLP	-	Restriction Fragment Length Polymorphism
PM	-	Polymarker
HLA	-	Human Leukocyte Antigen
AK	-	Adenylate Kinase
Hp	-	Haptoglobin

Hb	-	Haemoglobin
SNP	-	Single Nucleotide Polymorphism
RAPD	-	Random amplified polymorphic DNA
AFLP	-	Amplified fragment length polymorphism
MHC	-	Major histocompatibity complex
GST	-	G statistic / coefficient of gene differentiation
DST	-	Average gene diversity
FST	-	F statistic
HT	-	Total heterozygosity
HS	-	Subpopulation heterozygosity
LD	-	Linkage disequilibrium
PE	-	Probability of exclusion
TMB	-	Trimethylbenzidine
RFU	-	Relative fluorescence units
PIC	-	Polymorphic information content
AvePI	-	Average paternity index
Het-obs	-	Observed heterozygosity
HE	-	Expected heterozygosity
d.f	-	Degrees of freedom

ABSTRACT

Autosomal micro- satellites due to their hyper- variability have emerged as a potent tool for elucidating human identification. They involve a base motif of one to six base pairs, which are highly polymorphic and uniformly distributed throughout the human genome with high discriminatory power. In forensic analysis for cases of paternity testing, comparative DNA analysis, identification it has been necessary to determine the rarity of a match in the obtained DNA profiles by probability ramifications however lack of a Kenyan based allele frequency database and continued use of the one developed for the African American has made this a setback.. The Kenyan population is diverse in ethnicity with forty two tribes distributed in eight provinces in the country, various cultures and religions. There are three major cities of the country Nairobi (capital city), Mombasa and Kisumu with populations that are cosmopolitan largely due to rural- urban migration. To this end it was necessary to develop allele frequencies drawn from the Kenyan population for which a total of 150 blood samples, 50 from each city were obtained, then pooled together to form a common allele frequency database. The objectives of the study were to develop the Kenyan population forensic allele frequency database, assess the heterozygosity patterns of the Kenyan population, determine the genetic regional disparities among the Kenyan population and compare the Kenyan population forensic allele frequency database with that of the African American population currently in use. Blood was collected from Kisumu (50 samples), Nairobi (50 samples), and Mombasa (50 samples) on clean cotton ear bud ends; DNA was then extracted using the Chelex method then amplified using the Gene Amp PCR System 9700 (Applied Biosystems) and resolved using the 310 ABI Prism Genetic Analyser (Applied Biosystems). The identifiler kit developed by the Applied Biosystems was used to detect the following sixteen loci : D3S1358, VWA, D16S539, Amelogenin (sex chromosome), D2S1338, D8S1179, D21S11, D18S51, D19S433, CSFIPO, D5S818, D13S317, D7S820,TPOX, THOI and FGA using the five dye system for labeling the primers (6- FAM, VIC, NED, PET). Heterozygosity ranged from the highest observed in the pooled population in the STR markers D2S1338 and FGA (0.9) and the lowest observed for the STR marker D7S820 (0.713). The highest power of discrimination was observed with STR markers D2S1338 and FGA (0.96). Accordance with the Hardy Weinberg Equilibrium was established with few deviations in loci: TPOX, D18S51 and D5S818 where $p< 0.05$. However, the deviations were not statistically significant after Bonferroni's correction. AMOVA for the pooled population for the among populations, among individuals and within individuals revealed a variation of 0.20%, 0.32% and 99.87% respectively. The Kenyan population markers have a total of 165 observed alleles with 15 distinct alleles while the African American population dataset has 183 observed alleles with 34 distinct alleles. The two populations revealed significant disparities. The discriminatory potential in the 15 STR loci revealed D2S13338 and FGA to be the most informative loci in the studied Kenyan populations. The overall data supports the utility of the Kenyan population database for estimating STR profile frequencies.

CHAPTER ONE
INTRODUCTION

1.1 Evolution of Forensic Science

The field of forensic science has come a long way since its recorded beginnings in 700'ADs, when the Chinese used fingerprints to establish the identity of documents and clay sculptures. This is one of the few areas of law enforcement where science, technology and crime solving meet. A few significant advances occurred in the year's prior to1800. In 1248, a book, His Duan Yu (The Washing Away of Wrongs) published by the Chinese, described how to distinguish drowning from strangulation. It was the first recorded application of medical knowledge to the solution of crime. In 1609 the first treatise on systematic document examination was published in France. Then in 1784, one of the first documented uses of physical matching saw an English man convicted of murder based on the torn edge of a wad of newspaper in a pistol that matched a piece remaining in his pocket (http://www.quincy.ca/resources.cfm).

Early forensic specialists were self taught. There were no special schools, university courses or formal training. The establishment of forensic science curricula in 1902 by Professor R.A Reiss at the University of Lausanne, Switzerland was one of the first steps towards establishing forensic science as an academic discipline. It wasn't until 1930's that universities began offering courses and degrees in criminalities and police science. In 1950, the University of California at Berkeley established one of the 1st academic departments of

criminology/ criminalistics and American Academy of forensic science (AAFs) was formed in Chicago.

In 1971 Culliford refined the 1965 introduction of elecrophoretic separation of iso enzymes and proteins for forensic biology in the examination and typing of blood stains in the crime laboratory. In 1975 Edwin Southern transferred DNA to a solid support (Southern Blot) for analysis of genetic markers or DNA probes in forensic biology. Kerry Mullis in 1983 created the polymerase Chain reaction (PCR) method for cloning DNA fragments.

The 1980s ended with a few DNA first: Ray White described the first polymorphic RFLP marker in 1980, in 1985 Alec Jeffrey's discovered the multilocus VNTR probes, in 1986 the first paper on PCR on the use of DNA to solve a crime was published and in 1987 the introduction of DNA profiling in the U.S.A criminal court case in which admissibility of DNA was seriously challenged. This set in motion a string of events that culminated in a call for certification, accreditation, standardization and quality control guidelines for both DNA laboratories and the general forensic community. In 1994 the DNA databank legislation was enacted. By the end of the decade, significant progress had been made in the utilization of DNA analysis in case works in the state police laboratory system. In 1995 Forensic Science Services (FSS) started UK DNA database and in 1998 federal Bureau of Investigation (FBI) launched the Combined DNA Index System (CODIS) which has been in use for linking serial

crimes and unsolved cases with repeat offenders. This linked 50 Federal States in the U.S.A with a backlog of 600000 samples (http://www.troopers.state.ny/forensic-science).

1.2 Forensic Serology

Forensic serology is the branch of science that deals with the study of body fluids providing a powerful tool for criminal investigation as well as civil litigation involving adjudication of relationship between individuals. Biological fluids are key elements in the investigation of many kinds of crimes. In sexual assaults cases, the assailant (s) may leave semen in the victim, on clothing or at the location of the assault. In homicide cases the victims blood may be found on the suspects clothing or on a weapon in the suspect's possession. The goal of the forensic serologist is to determine by marker typing whether the biological evidence have originated from a particular individual either victim or suspect. The serologist decides what typing tests should be done based on source, age and condition of the materials choosing those tests that provide the greatest discrimination. If the single genetic type obtained from the evidentiary sample does not match the reference type of the individual in question then the individual can be excluded as the donor of the biological fluid with absolute certainty (Rebecca & Sensabangh 1991).

Previously genetic marker typing and discrimination of individuals was limited to the analysis of blood group markers and soluble polymorphic protein markers. ABO blood grouping and protein profiles have low power of discrimination. Protein profile markers

include Phosphoglucomutase1 (PGM1), Esterase D (ESD) and Erythrocyte acid phosphates (EAP) (Budowle, Jankowski & Corey 1997). The major inherent problem on blood group and protein markers is their failure to analyze poor quality biological material and limited amount of material available for investigations.

These markers are also insensitive as molecular tools. Such constraints necessitated development of DNA based techniques which not only analyses partial degraded biological material and minute samples but are also used for paternity testing, identification and other molecular biology applications. DNA based identification in forensics has been used in courts for many years now. Since 1989 the increasingly widespread use of DNA technology in forensic identification thousands of court cases has led to intense review of all aspects of science. In each court case the challenge has not only been the quality of the results at issue but also the scientific basis for the derivation and interpretation of those results. The definition of a match between samples and the validity of the statistical method used in determining the probability of finding a matching sample at random in a given population have been challenged repeatedly (Gill, Jeffrey & Warret 1985).

1.3 Statement of Problem & Justification

In the current forensic analysis in Kenya, determination of probability of similar match of DNA profiles is through use of the allele frequencies developed for the African American population. This is with the assumption that the African American population allele

frequencies could represent those of the Kenyan population. This is a great set back because the Kenyan population has no reference values for allele frequencies for use in the forensic DNA analysis on routine cases of identifying charred bodies burned beyond any physical recognition, bodies buried in mass graves and decomposed beyond physical recognition, comparative analysis of samples for murder and rape cases and paternity testing. Case studies of the Zimbabwean population on six short tandem repeat loci: CSFIPO, TPOX, THOI, D3S1358, VWA and FGA showed a remarkable difference in population frequency with THOI exhibiting a frequency disparity of $p<10^{-3}$ and FGA exhibiting a frequency disparity of $p<0.001$ between African American and Zimbabwe's population, respectively (Budowle & Nhari 1997).

Woo and Budowle (1995) demonstrated significant differences in allele frequencies for the 7 PCR based loci (LDLR, GYPA, HBGG, D7S8, GC, HLA-DQA1 and D1S80) between the Korean, USA Caucasian and African American populations. However, compared loci per loci, this Korean population data were similar to the Japanese population data. Loci D1S80 of the Korean population data was different when compared to the overall Asian population sample. This disparity served as an eye opener to develop the Kenyan population allele frequencies which would specifically identify with Kenyan population.

The developed Kenyan forensic allele frequencies will be used in cases of paternity testing, identification of remains in holocausts where the bodies are badly charred, mass graves and

murder/ rape cases. Equally, development of this database contributes highly to the global forensic database. This study also revealed the most informative markers amongst the 15 microsatellites used in the Kenyan population.

1.4 Research questions

i.) Are there genetic disparities between the Kisumu, Nairobi and Mombasa regions or do they represent a homogenous population?

ii). Are the allele and genotype frequencies for the Kenyan black population similar to those of the African – American population?

1.5 Hypothesis

The Kenyan population forensic allelic frequencies do not to exhibit significant difference with those of the African American allelic frequency data.

1.6 Objectives

1.6.1 General objective

To determine the forensic genetic diversity of the Kenyan population.

1.6.2 Specific objectives

i) Determine forensic allele and genotype frequencies for the Kenyan population.

ii) Determine heterozygosity patterns exhibited by the Kenyan population.

iii) Compare the homogeneity of loci between the Kenyan population and that of the African American population allele frequencies (in use as the reference frequencies).

iv) Develop allele frequency precision values/data to be used in computations in forensic DNA analysis for routine case works in the Kenyan population.

CHAPTER TWO
LITERATURE REVIEW

2.1 Genetic Marker Typing

In the past decade several key advances in molecular genetics have greatly increased the impact of population genetics on biology. Most important have been the development of polymerase chain reaction (PCR) which amplifies specific stretches of DNA to useable concentrations; the application of evolutionary conserved sets of PCR primers, the advent of hyper variable micro satellite loci and the advent of routine DNA sequencing in biology laboratories. These innovations coupled with the recent explosion of powerful analyses and relatively user friendly computer programs meant that much of the power inherent in molecular genetic data can be tapped. The genetic uniqueness of individuals is a central tenet of human biology. The uniqueness is defined by the combination of genetic markers that an individual inherit from parents. Genomic DNA is present in all nucleated cells in the human body. Since genetic structuring reflects the number of alleles exchanged between populations, it has major consequences on the genetic composition of individuals themselves (Sunnucks 2000).

Understanding gene flow and its effects is central to many fields of research including population genetics, population ecology, conservation biology and epidemiology. Gene flow generates new polymorphism in the populations, and increases local effective population size (the ability to resist random changes in allele frequencies), there by opposing random genetic drift, generating new combinations on which selection can potentially act.

Individual organisms differ in the DNA sequences comprising their genomes. This genetic variation can be considered at the level of individual genes (genic) or genotypes (genotypic). The fate of a given genetic variant will be influenced by the biology and circumstances of individuals through which it passes, including reproductive success, migration, population size and natural selection (Smith 1996).

By measuring genetic variation and by applying population genetic models inferences can be made on the biology of organisms. Processes that affect individual s ultimately accumulate into on population and thus by examining genetic markers with appropriate rates of change and therefore suitable signals, information can be obtained about almost any population and evolutionary process through the hierarchy of life. Selecting suitable genetic analysis is vital to the success of applying molecular genetics in population biology. There is apparently bewildering array of genetic techniques available for population genetic analysis.

Genetic markers are simply heritable characters with multiple states at each character. Typically in a diploid organism each individual can have one or two different states (alleles) per character (loci). All genetic markers reflect differences in DNA sequences. Separate loci can provide independent tests of hypotheses thus using many together can yield extreme sensivity. A population genetic survey must start with a decision regarding appropriate genetic markers. The value of a genetic marker to the forensic serologist depends on several considerations, the marker polymorphism should involve simple qualitative differences: the

marker should be expressed in a tissue of interest (such as blood, semen) throughout the life of the individual, the marker should be robust, that is, should be able to survive drying and other travail experienced by biological evidence and its desirable that the analytical technique consume as little as possible of the sample under test (Rebecca et al. 1991).

The dual goals driving the selection of forensic genetic markers are to enhance discrimination power and to extend the range of evidence samples from which useful information can be obtained. The three criteria relating to the nature of a useful genetic marker are based on the information that can be obtained from genetic studies and the statistical analysis of gene (allele) frequencies. If the genotype frequencies of the marker within a defined population are known, then the probability of two individuals chosen at random from these population will possess the same genotype (P_I) can be calculated.

If the population is in Hardy-Weinberg equilibrium for this marker, then the expected genotype frequencies can be calculated from the observed allele frequencies. The index value known as the discrimination power (P_D) is equal to $1-P_I$ and is the probability that two random individuals will possess different genetic types for the marker being tested. Higher P_D values indicate a greater individualization potential because a larger portion of the population can be excluded. For a single marker system, the greater the number of the alleles (degree of polymorphism) and the more even their distribution in a population (the heterozygosity), the higher the power of discrimination. Genotype frequencies for a

particular marker can vary from population to population. Therefore it becomes important to collect allele frequency data from relevant reproductive populations so the appropriate statistical evaluation of the typing results can be made (Harris 1975).

2.2 Population Genetics

Once a match has been established and the matching criteria satisfied, one must determine the probability of finding this DNA pattern in the population. There is no dispute over the fact that the DNA is different for all individuals except identical twins. However depending on the number of loci examined, a matching DNA pattern may not necessarily be construed as an absolute indictment to the exclusion of other humans, because the loci examined are but a subset of all genetic differences between two human beings. It is the statistical ramifications of a match between the DNA of a suspect and the DNA isolated from an evidentiary sample that has caused DNA identity testing to live such an impact on the legal community. Current serological testing is capable of yielding probabilities for a random match, of one in several thousands. In contrast a DNA pattern from even a moderately resolved VNTR genetic system using as few as four loci can reasonably yield probabilities of one in several 100 million (Balaz, Debbie & Corey 1993).

The frequency of observed alleles or bands in the DNA finger print can be multiplied together to calculate the final probability of the total pattern genotype appearing within an individual. To calculate any allele frequency in the population several criteria must be met:

the alleles have to be distinguishable, either individually or as allelic classes; the inheritable patterns must be identifiable; linkage to other genetic systems must be quantified and resolving power of the assay system must be characterized (Cavalli-Sforza & Bodmer 1971).

2.3 Population Allele Frequency Databases

The ability to type DNA from biological evidence samples has dramatically increased over the past years. Much of this progress has been driven by the discovery of new genetic polymorphisms among human DNA and development of new procedures. The basic approach to characterize patterns and polymorphism differences among populations is by defining the genetic structure of populations. This genetic structure is defined by allele structure (the allele counts or allele frequency distribution for a locus or loci). Theoretically a population is in equilibrium if certain assumptions hold true: a large, randomly mating population, no appreciable mutation, equal fitness to reproduce and no migration.

In the context of Mendelian populations, the Hardy - Weinberg theorem established that the genotype frequencies remain constant and follow a multinomial distribution (binomial for a one locus, two allele model). In reality, these assumptions rarely occur that is, no population is infinitely large. Even in cosmopolitan communities, assortative mating occurs at short distances. The phenomenon generates geographical heterogeneity. Also the mutation rate in multiple loci is enough to produce a high level of diversity in a short time period. Populations also undergo complex patterns of migration because of natural growth and subsequent need

for geographical expansion. The best parametric approach to establish the existence of isolation is to determine the homozygosity - heterozygosity patterns exhibited by a population. Highly inbred populations have increased frequencies of homozygosity and consequently a decreased number of heterozygotes (Slatkin 1995). It is therefore desirable to collect allele/ genotype frequency data from different populations so that the forensic scientist is able to provide an estimate on the rarity of a genetic profile. Huag & Budowle (1995) reporting on the allele frequencies of 5 PCR based loci D1S7, D2S44, D4S139, D5S110 and D17S79 of 126 unrelated Chinese individuals using RFLP analysis observed that similar to the allele frequencies of African American population, loci D17S79 was the least polymorphic and loci DS110 was the most polymorphic.

Woo and Budowle (1995) reporting on the allele frequencies of 7 PCR based loci: LDLR, GYPA, HBGG, D7S8, GC, HLA – DQA1 and D1S80 on 116 unrelated Korean individuals observed significant statistical differences in allele frequencies in all the 7 loci on comparison of Korean database with the US Caucasians or African American population database. A comparison of the allele frequencies for the 7 PCR based loci marker for marker between the Korean database with the Asian population data of the Taiwanese yielded insignificant differences in loci HLA-DQA1 and PM. A comparison of allele frequencies between the Korean database with a Japanese sample and a general Asian sample yielded similarities with the Japanese sample for loci D1S80 data and statistical differences with the general Asian population sample. The allele frequency distributions between the Koreans and

Taiwanese populations had similar trends except for the allele frequencies of marker D1S 80 which were different.

Hayes and Budowle (1995) while reporting on the allele and genotype frequencies of 7 PCR based loci: HLA – DQA1, LDLR, GYPA, HBGG, D7S8, GC and D1S80 allele and genotype frequencies of 94 unrelated Arab populations observed that the Arab HLA- DQA1 data significantly differed from the US Caucasian and African American databases on the bases of the G Statistic ($p<10^{-3}$). Allele distributions exhibited similarities for LDLR and GYPA loci on comparison of the Arabs and Caucasian databases. Arabs and African Americans were not similar at any other loci other than D7S8.

Entrala et al. (1998) reporting on the allele frequencies of 9 STR loci of 171 Caucasians living in Andalucia, southern Spain observed that marker D3S1358 (7 alleles); D5S818 (12 alleles); D7S820 (7 alleles); D8S1179 (10 alleles); D13S317 (7 alleles); D18S51 (12 alleles); D21S11 (15 alleles); FGA (14 alleles); and VWA (8 alleles) had 7, 12, 7, 10, 7, 12, 15, 14, and 8 alleles, respectively. The observed heterozygosity ranged from 0.696 for D5S818 to 0.901 for FGA. The power of exclusion ranged from 0.593 for D13S358 to 0.739 for D18S51 and the combined probability of excluding paternity for the 9 loci exceeded 0.99991. The power of discrimination ranged from 0.875 for D5S818 to 0.964 for D18S51 and the combined power of discrimination for the 9 loci exceeded 0.999997. The most informative loci for this Spanish population were D18S51 and FGA, while the least informative locus

was D5S818. No deviations from Hardy-Weinberg equilibrium were observed except for loci D7S820 (0.003) which was statistically insignificant after Bonferroni correction.

Alves et al. (2001) reporting on allele frequencies of 12 STR loci of 110 unrelated individuals from Mozambique observed that marker CSF1PO; D3S1358; D5S818; D7S820; D8S1179; D13S317; D18S51; D21S11; FGA; THO1; TPO and VWA had 8, 5, 8, 7, 8, 7, 16, 16, 18, 5, 7, and 9 alleles, respectively. The observed heterozygosity ranged from 0.705 for D13S317 to 0.880 for FGA. The power of exclusion ranged from 0.440 for D13S317 to 0.737 for FGA and the power of discrimination ranged from 0.861 for D13S317 to 0.972 for FGA. A comparison of allele frequencies between the Mozambique populations with the African American population of the 12 loci indicated significant differences in loci D21S11, THO1 and TPO. No deviations from Hardy-Weinberg equilibrium were observed in the Mozambique population.

Steinlechner et al. (2002) while reporting on the allele frequencies of 10 STR loci of 108 unrelated Gabon black population observed that marker D3S1358; D8S1179; D18S51; D21S11; FGA; THO1; VWA; D16S539; D2S1338 and D19S433 had 7, 6, 15, 17, 17, 6, 10, 8, 12 and 14 alleles, respectively. The observed heterozygosity ranged from 0.663 for D2S1338 to 0.931 for FGA. The power of exclusion ranged from 0.481 for THO1 to 0.791 for D2S1338 and the combined probability of excluding paternity for the 10 loci exceeded 0.999981. The power of discrimination ranged from 0.888 for D3S1358 to 0.973 for

D2S1338. All the 15 loci were in linkage disequilibrium, that is, there was no association of alleles between the ten STR loci. No deviations from Hardy-Weinberg equilibrium were observed except for three loci D21S11 (0.044); THO1 (0.017) and D16S539 (0.022) which were statistically insignificant after Bonferroni correction. A marker for marker comparison between the Gabon black population allele frequencies with the African American allele frequencies, the two populations were similar at six of the eight loci (THO1, 0.005; D8S1179, 0.008) and with Australian Caucasians allele frequencies, significant differences occurred at all the 10 loci (G statistic test).

Perez et al. (2003) reporting on the allele frequencies of 13 CODIS STR loci of 100 individuals from Peru observed that marker CSF1PO; D3S1358; D5S818; D7S820; D8S1179; D13S317; D18S51; D21S11; FGA; THO1; TPOX; VWA and D16S539 had 6, 6, 7, 7, 9, 8, 12, 12, 12, 6, 6, 8, and 7 alleles, respectively. The observed heterozygosity ranged from 0.6600 for both CSF1PO and VWA to 0.8700 for FGA. The power of exclusion ranged from 0.3691 for both CSF1PO and VWA to 0.7346 for FGA and the power of discrimination ranged from 0.7240 for D5S818 to 0.9660 for FGA. Test of linkage disequilibrium indicated independence of all the 13 loci. No deviations from Hardy-Weinberg equilibrium were observed in the Peru population.

Ang et al. (2005) reporting on the allele frequencies of 13 CODIS loci of 197 unrelated Malays in Singapore observed that CSF1PO (7 alleles); D3S1358 (8 alleles); D5S818 (8

alleles); D7S820 (8 alleles); D8S1179 (9 alleles); D13S317 (8 alleles); D18S51 (13 alleles); D21S11 (12 alleles); FGA (17 alleles); THO1 (7 alleles); TPOX (6 alleles); VWA (8 alleles); and D16S539 (8 alleles) had 7, 8,8, 8, 9, 8, 13, 12, 17, 6, 8, and 8 alleles, respectively. The observed heterozygosity ranged from 0.594 for TPOX to 0.888 for FGA. The power of exclusion ranged from 0.340 for TPOX to 0.723 for FGA and the power of discrimination ranged from 0.780 for TPOX to 0.962 for FGA. Test of linkage disequilibrium indicated independence of all the 13 loci in the Malay database. No deviations from Hardy-Weinberg equilibrium were observed except for loci D7S820 (0.015) which was statistically insignificant after Bonferroni correction.

De Ungria et al. (2005) reporting on the allele frequencies of 19 STR loci of 106 Philippine population generated using AmpFISTR multiplex and ALF singleplex systems noted that marker CSF1PO; D3S1358; D5S818; D7S820; D8S1179; D13S317; D18S51; D21S11; FGA; THO1; TPOX; VWA; D16S539; D2S1338; D19S433; D8S306; DHFRP2; F13A01 and FES/FPS had 8, 6, 7, 6, 8, 7, 13,13, 13, 7, 5, 8, 7, 11, 11, 10, 5, 9, and 7 alleles, respectively. The observed heterozygosity ranged from 0.632 for F13A01 to 0.877 for D21S11. The power of exclusion ranged from 0.331 for F13A01 to 0.749 for D21S11 and the power of discrimination ranged from 0.795 for TPOX to 0.959 for D2S1338. The polymorphic information content ranged from 0.610 for F13A01 to 0.850 for D2S1338. No deviations from Hardy-Weinberg equilibrium were observed except for loci D5S818 (0.00550) which was statistically insignificant after Bonferroni correction.

Zhang et al. (2007) while reporting on the allele frequencies of 15 STR loci of 200 unrelated Chinese population from Sichuan in West China noted that marker CSF1PO; D3S1358; D5S818; D7S820; D8S1179; D13S317; D18S51; D21S11; FGA; THO1; TPOX; VWA; D16S539; Penta E and Penta D had 8, 6, 9, 7, 9, 9, 15, 12, 17, 6, 6, 9, 7, 19, and 10 alleles respectively. The observed heterozygosity ranged from 0.560 for TPOX to 0.915 for Penta E. The power of discrimination ranged from 0.7356 for TPOX to 0.983 for Penta E and the combined power of discrimination for the 15 loci exceeded 0.9999999999999999. The probability of exclusion of paternity ranged from 0.246 for TPOX to 0.826 for Penta E and the combined probability of excluding paternity for the 15 loci exceeded 0.9999994. The polymorphic information content ranged from 0.491 for TPOX to 0.913 for Penta E. No deviations from Hardy-Weinberg equilibrium were observed.

Ota et al. (2007) reported allele frequencies of 15 STR loci in 105 unrelated Sherpa in Namche Bazaar village and 111 unrelated non-Sherpa in Kathmandu from Nepal. In the Sherpa population marker CSF1PO; D3S1358; D5S818; D7S820; D8S1179; D13S317; D18S51; D21S11; FGA; THO1; TPOX; VWA; D16S539; D2S1338; and D19S433 had 6, 6, 8, 6, 8, 9, 12, 12, 12, 6, 5, 7, 7, 10, and 10 alleles, respectively while in the non-Sherpa population, the same markers had 7, 7, 6, 7, 10, 9, 14, 15, 11, 6, 6, 8, 8, 10, and 11 alleles, respectively. The observed heterozygosity ranged from 0.6000 for TPOX to 0.8381 for D2S1358 in the Sherpa population and from 0.6486 for TPOX to 0.8919 for FGA in the non-Sherpa population. For all the 15 STR loci, the power of discrimination ranged from 0.7737

for TPOX to 0.9556 for D2S1338 in the Sherpa population and from 0.8681 for D3S1358 to 0.9647 for D18S51 in the non-Sherpa population. The combined power of discrimination for the 16 loci exceeded 0.999999 in both populations. The combined probability of exclusion for the 15 loci exceeded 0.999993 in the Sherpa and 0.999999 in the non-Sherpa population. The polymorphic information content ranged from 0.5260 for TPOX to 0.8409 for D2S1338 in the Sherpa and from 0.6326 for TPOX to 0.8627 for D18S51. No deviations from Hardy-Weinberg equilibrium were observed except for loci D5S818 in the Sherpa and D7S820 in the non-Sherpa population which were insignificant after Bonferroni correction.

Zheng et al. (2007) while reporting on the genetic data of 9 STR from Henan province in Central China population noted that marker D3S1358; D5S818; D7S820; D8S1179; D13S317; D18S51; D21S11; FGA and VWA had 9, 8, 8, 10, 8, 16, 18, 18, and 10 alleles, respectively. The expected heterozygosity ranged from 0.732 for D3S1358 to 0.855 for VWA. The power of exclusion ranged from 0.480 for D3S1358 to 0.704 for VWA and the power of discrimination ranged from 0.875 for D3S1358 to 0.962 for FGA. The polymorphic information content ranged from 0.68 for D3S1358 to 0.84 for FGA. No deviations from Hardy-Weinberg equilibrium were observed except for loci VWA (p=0.009) which was statistically insignificant after Bonferroni correction.

Sao-Bento et al. (2008) while reporting on STR data for the 15 AmpFISTR Identifiler loci in the Brazilian population of Sao Paulo state noted that marker CSF1PO; D3S1358; D5S818;

D7S820; D8S1179; D13S317; D18S51; D21S11; FGA; THO1; TPOX; VWA; D16S539; D2S1338; and D19S433 had 8, 8, 9, 11, 10, 9, 16, 15, 13, and 14 alleles, respectively. The observed heterozygosity ranged from 0.19100 for TPOX to 0.25500 for FGA. The power of exclusion ranged from 0.42084 for TPOX to 0.74671 for D18S51 and the combined probability of excluding paternity for the 15 loci exceeded 0.99999995. The power of discrimination ranged from 0.66294 for TPOX to 0.87040 for FGA and the combined power of discrimination for the 15 loci exceeded 0.99999999997. No deviations from Hardy-Weinberg equilibrium were observed except for loci CSF1PO (p=0.03693) which was statistically insignificant after Bonferroni correction.

2.4 Markers for Forensic Population Studies

2.4.1 ABO Blood Grouping System

The best known and most venerable of the genetic markers used in forensic testing is the ABO blood group system; the ABO markers are quite stable and can be detected in most types of biological evidence. The other blood group markers such as Rh and MN are rarely used in forensic testing because they exhibit variable stability in blood stains and are not found in semen, saliva, or other biological evidence. However the power of discrimination is quite low, unpublished report indicates 49% of the Kenyan population is blood group O, 26% group A, 22% group B and 4% group AB. Each individual has a pair of allelic genes to make up his genotype, which determines his phenotype for each blood group system. The M-N antigens deteriorate fairly rapidly and the N antigen is more difficult to detect than M. In

addition there is a problem of cross reaction in that there is a residual amount of N activity present in most group M individuals and this can often be detected by absorption elution (Harris 1975).

2.4.2 Protein Markers

During the 1970s the forensic genetic typing tools were expanded to include the electrophoretic detection of protein markers. Of the many protein markers found in the blood, approximately 10 proved to be robust to be typed in bloodstains. Unfortunately only a few protein markers are present in other kinds of biological evidence such as semen. Again evidence samples that have undergone significant environmental deterioration will not yield comparative results. Many enzymes and other blood proteins commonly exist in three or more genetically different forms (polymorphisms). Identification of a particular isoenzyme type in a blood sample can be made using electrophoresis. Examples of protein markers include phosphoglucomutase (PGM), adenylate kinase (AK), Esterase D (ESD), Haptoglobin (Hp) and Haemoglobin (Hb) (Metropolitan Police Forensic Science Laboratory Methods Manual).

2.4.3 Restriction fragment length polymorphisms (RFLP) markers

The technique was developed in the early 1980s. It makes use of bacterial restriction enzymes targeting specific DNA sequences typically four to six base pairs long. Polymorphisms are detected through southern blotting as either presence or absence of

restriction sites and thus fragments are of desired size. Disadvantages are: it's not amenable to automation, slow, use of large samples 50 –500ng of DNA and repeat analysis not always possible (Alastar & Harding 1989).

2.4.4 The HLA DQά Typing system

This system is based on the high degree of polymorphism found in the major histocompatibility gene complex (MHC) proteins. Specifically, a polymorphic coding region of the gene for the ά subunit of the DQ protein is amplified by using flanking primers situated in the regions of constant sequence. The detection assay of the typing system in the kit is a modified dot plot procedure. Limitation is that it involves a laborious procedure that is open to procedural errors (Ambach, Zchethofer & Schcithan 1996).

2.4.5 Single nucleotide polymorphisms (SNP'S) markers

Single nucleotide polymorphisms are caused by point mutations including insertions and deletions on specific coding regions of the genome. They are useful for pedigree, family or population screening within species. Broad application requires an estimate of their frequency and information content as well as a means of automated detection (Chee, Yang & Hubbell 1996).

2.4.6 Random amplified polymorphic DNA (RAPD) markers

This is a PCR based technique that relies on a single arbitrary primer for amplification of genomic DNA (Williams et al. 1990). Polymorphism is detectable as the absence or presence of a band after agarose electrophoresis. It's quick since a prior knowledge of DNA sequence is not required, simple and efficient requiring small amounts of DNA. Results are not reproducible and the mode of generation of the different polymorphisms observed is not understood. Being dominant multilocus markers, they limit the level of hererozygosity assayable and thus are non – informative for population studies (Durand, Sire & Theron 2000).

2.4.7 Amplified fragment length polymorphism (AFLP) markers

AFLP is a technique intermediate between RFLP and RAPDs that involves restriction enzyme digestion of genomic DNA using two specific enzymes and subsequently subjecting the resultant DNA fragments to selective rounds of PCR after ligation of adapters to serve as primers. The products are resolved on a sequencing gel and visualised by X-ray film, if the product is radioactively labeled. The technique is inhibited by large amount of high quality of DNA required as well as being technically demanding and expensive (Vos & Hogers 1995).

2.4.8 Microsatellite markers

These are synonymous to Simple Tandem Repeats (STR'S) and Variable Number of Tandem Repeats (VNTR's) which are relatively short (approximately 100bp) tracts of tandemly

repeated DNA with repeat lengths of six or less base pairs. Tandemly repeated DNA sequences are widespread throughout the human genome and show sufficient variability among individuals in a population that they become important in several fields including genetic mapping, linkage analysis and human identity testing.

These tandemly repeated regions of DNA are typically classified into several groups depending on the size of the repeat region. Minisatellites (variable number of tandem repeats) have core repeats with 9 - 80 bp, while microsatellites (short tandem repeats, STRs) contain 2-5 bp repeats (http://www.sasas.co.za/Sajas.html). The forensic community has moved primarily towards tetranucleotide repeats which may be amplified using the polymerase chain reaction (PCR) with greater fidelity than dinucleotide repeats. The variety of alleles present in a population is such that a high degree of discrimination among individuals in the population may be obtained when multiple STR loci are examined. PCR based STR'S have several advantages over the convectional Southern blotting techniques of the larger variable number tandem repeats (VNTRs).Discrete allele STR systems may be obtained due to their smaller size which puts them in the size range where DNA fragments differing by a single base pair in size may be differentiated. In addition, smaller quantities of DNA including degraded DNA may be typed using STRS. Thus the quantity and integrity of the DNA sample is less of an issue with the PCR based typing methods than with the convectional RFLP methods.

Due to their high polymorphism, microsatellite markers have recently been increasingly used as choice markers for population studies and gene mapping (Bowcock et al. 1994, Montaldo & Meza- Herrera 1998). Because of their ease of scoring and high polymorphic nature, they have become the markers of choice for population genetics studies. Since the advent of polymerase chain reaction in the early eighties, the use of microsatellites has become extremely widespread in biology. Today, a large number of studies rely on these co dominant genetic markers to investigate the genetic structuring of populations (Montaldo & Meza – Herrera 1998). Attributes that make microsatellites markers useful include:

i) Microsatellite loci are found in large numbers and are relatively distributed throughout the genome. Tri and tetra - nucleotide repeats have been shown to occur at a frequency of 1 in every 300 - 500 kilo base (kb) (Primmer et al. 1997).

ii) Due to high mutation rates, majority of microsatellite loci are highly polymorphic in most mammalian cells (Edwards, Hammond & Chakraborty 1992).

iii) Alleles are inherited in a Mendelian fashion with their frequencies often conforming to Hardy - Weinberg equilibrium.

iv) These repeats can be amplified repeatedly using PCR thus enabling precise allele designations in population surveys on the basis of their DNA sequence (Tautz 1989).

2.5 Statistical Analysis

2.5.1 Hardy - Weinberg equilibrium

This is a simple model that relates allele frequencies to genotype frequencies. The Hardy-Weinberg law states that gene frequencies and genotype ratio's in a randomly breeding population remain constant from generation to generation unless acted upon by evolutionary agents other than Mendelian segregation and recombination of alleles.

Therefore a population in Hardy - Weiberg equilibrium (HWE) doesn't show any change in allele frequencies. The genetic change in a population is usually described by the change in gene frequencies. The law predicts that a population given two alleles with frequencies p and q the genotypes will be in the proportions $\mathbf{p^2: 2pq: q^2}$ where p^2 is the frequency of the homozygote genotype for the allele with the frequency p, 2pq is the genotype frequency of the heterozygote and q^2 is the frequency of the homozygote for the allele with gene frequency q. When a polymorphism is in true Hardy - Weinberg equilibrium the gene frequencies remain constant from generation to generation.

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

Usually the HWE holds for many polymorphic loci in out breeding organisms. Many of the assumptions in the HWE do not hold in random mating population but the deviation are usually small and negligible. In situations where one or more of this inherent assumptions is clearly violated in a major way, a variety of major models can be brought to bear on the problem. Violations of these assumptions lead to evolution which is the change in allele

frequencies at one or more loci from generation to generation. Changes in allele frequencies are caused by several factors:

Natural Selection

This takes effect where certain genotypes or phenotypes are better endowed (more fit) to transfer their genes to the next generation compared to others either through differential mortality (survival) and/or differential fecundity(family size) leading to the gradual shift in the gene ratios in favour of the more fit genotypes.

Gene Migration

This is the permanent movement of genes from one population into another. Randomly breeding populations develop a gene pool distinct from other local subpopulations. Many species are made up of local populations whose members tend to breed within the group. Each local group can develop a gene pool distinct from that of other local populations. However, members of one population may breed with occasional immigrants from one adjacent population of the same species. This can introduce new genes or alter existing gene frequencies.

Genetic Drift

An allele may increase or decrease in proportion through chance especially when the population size is small or the marker is neutral. It can lead to loss of rare alleles or fixation of the common allele (homozygosity). If the population is small, drift has little effect.

Mutation

By itself mutation plays a minor role on evolution essentially because mutation rates are low. However, it is a prerequisite so that new alleles are formed upon which natural selection acts to cause evolutionary change. After being shuffled in various combinations with the rest of the gene pool, this provides raw material on which the natural selection can act.

Non – Random Mating

One of the corner stones of HWE is that mating in the population must be random. If individuals are choosy in selection of their mates, the gene frequencies may become altered. Non random is quite common. In breeding – mating between relatives, geographic structuring where proximity limits mating to a certain group - a form of in breeding and assortative mating- between individuals of similar phenotype, predisposes individuals to homozygosity and thereafter selection in case of harmful recessive alleles.

Population Subdivision

More often than not, populations under study include several mating units or sub populations. This increases the frequency of homozygotes relative to the HWE proportions, even though HWE may hold in each individual subpopulation. The genotype frequencies in the entire population will deviate from the expected HWE proportions. This is called the Wahlund effect after the person who first described it (Nei 1987).

2.5.2 Fishers Exact Test

The test is used to detect group differences using frequency (count) data. The Fishers exact test is used for 2x2 tables when members of two independent groups can fall into mutually exclusive categories. The test is used to determine whether the proportions of those falling into each category differ by group. The Chi square test of independence can also be used in such situations but its only an approximation whereas the Fishers exact test returns exact one tailed and two tailed p values for a given frequency table.

Fishers exact test computes the probability given the observed marginal frequencies of obtaining exactly the frequencies observed and any other configuration more extreme (any other configuration with a smaller probability of occurrence in the same direction (one tailed) or in both directions (two tailed) (Guo & Thompson 1992).

2.5.3 Measures of Genetic Diversity

Genetic diversity (average heterozygosity) is the mean percentage of loci heterozygous per individual. Heterozygosities may also be estimated from observed frequencies of alleles assuming the population is in Hardy Weinberg equilibrium.

$$H = 1 - \sum X$$

The average heterozygosity H is the average of the overall loci thus the expected proportion of heterozygous loci in a randomly close individual (Nei 1984).

2.5.4. G-Statistic (G_{ST})

Since natural population are subdivided into subpopulation gene diversities within and between population are studied .The genetic diversity in whole population can be divided into the within and between populations component. The relative magnitude of gene differentiation among subpopulation may be measured by the G statistic also termed as the co – efficient of gene differentiation. The G_{ST}, the coefficient of gene differentiation is an analogue of F_{ST} that doesn't depend on population genotype frequencies or correlation of gene frequencies at higher hierarchical levels. The G statistic (G_{ST}) is a function only of allele frequencies and doesn't incorporate observed proportions of heterozygotes.

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

where D_{ST} is the average gene diversity between subpopulations, including sub populations themselves. For diploid random mating populations: $D_{ST} = H_T - H_S$, thus G_{ST} is equal to F_{ST}.

H_S is the subpopulation heterozygosity. It measures gene frequency variations as gene diversity, the probability that two chosen genes from a population are different (Nei 1973).

2.5.5 F statistic (F_{ST})

The F statistic (F_{ST}) provides a convenient approach for estimating interpopulational gene flow in models that assume selective neutrality. G_{ST} on the other hand, estimates population subdivision without relying on genotype frequencies. However, it is highly dependent on total heterozygosity (H_T) and may be large even when the absolute gene differentiation is small, where $H_{T\ is}$ small (Nei 1984). It is difficult to interpret and give meaning to F_{ST} values. However, values ranging between 0 and 0.05 indicate little differentiation; a value between 0.05 and 0.15 indicate moderate differentiation, and a value between 0.15 and 0.25 signify very great differentiation. Non parametric tests used to compare F_{ST} values are significantly different from zero and thereby indicate the level of structuring. The p values obtained define the significance levels (Balloux & Ligon – Moulin 2002).

2.5.6 Analysis of Molecular Variance (AMOVA)

This is the analysis of variance between gene frequencies, where interclass correlation coefficients do depend on the mutational process. It estimates F statistic through interclass correlations defined as ratios of variance components.

2.5.7 Heterozygosity

Heterozygosity (gene diversity) is the proportion of individual heterozygous at a locus. Allelic diversity is the actual number of alleles present at a locus. The probability of multilocus genotype counts is conditional on allele counts and on allelic independence within and between loci. As the number of loci increases and each locus and each sampled genotype becomes unique, the conditional probability becomes a function of total heterozygosity. In forensic science, multilocus genotype frequencies are often estimated as products of allele frequencies. Heterozygosity is of major interest to students of genetic variation and can tell a lot about the structure and even the history of a population. In Hardy Weinberg equilibrium: $p^2 + 2pq + q^2 = 1$. Heterozygosity is given by 2pq. The rest of the expression ($p^2 + q^2$) is the homozygosity. For many human microsatellites loci H_E expected heterozygosities often >0.85 meaning that one has a >85% chance of being a heterozygote.

2.5.8 Linkage disequilibrium

The term is used in the study of population genetics for non random associations of alleles at two or more loci not necessarily in the chromosome. It is not the same as linkage which describes the association of two or more loci on a chromosome with limited recombination between them. Linkage disequilibrium describes a situation in which some combinations of alleles or genetic markers occur more or less frequently in a population than would be expected from a random formation of haplotypes from alleles based on their frequencies. Non random associations between genes at different loci are measured by the degree of

linkage disequilibrium. Linkage disequilibrium is generally caused by interactions between genes, genetic linkage and the rate of recombination, random drift or non random mating and population structure (Chakraborty & Jin 1993).

2.5.9 Power of Discrimination

The power of discrimination P_D value reflects the probability that two unrelated individuals will have different genotypes at one or more loci.

Combined $P_D = 1 - [(Pi)_1 (Pi)_2 \ldots\ldots (Pi)_n]$ where n is the number of combined markers and Pi is $1 - P_D$ for each number.

2.6 Probability of Exclusion

A genetic marker can be characterized by its ability to exclude an unrelated man from paternity in any situation. The exclusion probability is given by summing the joint probabilities of all mother – child excluded man combinations. These exclusion probabilities depend on allele frequencies for that locus, but do not depend on the genotypes in any particular case. Exclusion probabilities are increased with the use of several loci, since it is sufficient to exclude at any one of several loci when mutation can be ignored. If the Ql is the exclusion probability at locus l, then the overall probability of exclusion follows from being able to exclude from at least one locus (National Research Council 1996). In other words Q is one minus the probability that none allows exclusion. If the Ql are independent

$Q = 1 - I_l I (1 - Ql)$

CHAPTER THREE
MATERIAL AND METHODS

3.1 Sample Collection

Databases of 50 – 150 individuals are deemed as sufficient for forensic calculations according to Hamilton et al. (1996). A total of one hundred and fifty blood samples were collected from Kisumu, Nairobi and Mombasa Blood Transfusion Centers on cotton ear bud's both ends. Simple random sampling was used without replacement where all the samples were put in a row then odd numbers were picked so as to capture a true random sample.

In Kisumu, blood was routinely obtained from blood donation campaigns carried out in Nyanza and the broader Western province by the Nyanza Regional Blood Transfusion Centre. Mombasa samples were routinely obtained from the entire coastal regions by the Coast Regional Blood Transfusion Centre. Nairobi samples represented a sample covering the entire Nairobi region and part of the wider Central Province taken from blood donation campaigns.

The three Regional Blood Transfusion Centers were chosen for their strategic positioning in Kenya and the cosmopolitan nature of their population structure. The ear buds were then dried to preserve the blood sample and then stored under room temperature. Again Kisumu, Mombasa and Nairobi are the three major cities of Kenya.

3.2 DNA Extraction

DNA from whole blood was extracted in the presence of Chelex resin following the procedure of Walsh, Metzger & Higuchi (1991). A small piece of the bloodstain on cotton was cut, and then put in a 1.5ml microcentrifuge tube to which 1 ml of double distilled water was added. This mix was then vortexed (Vortex- Genie model K550- GE scientific industries U.S.A) briefly and left to stand at room temperature for 30 minutes. This mix was then span for 5 minutes at 10000 rpm in a microcentrifuge followed by removal of 950µl of the supernatant to leave about 50µl of the mix in the micro centrifuge.

A volume (100µl) of double distilled water was added, followed by 50µl of 20% Chelex to a final volume of 200µl. The mix was then briefly vortexed for 5 seconds, followed by incubation in a water bath at 56^0C for 30 minutes. This was followed by incubation in a boiling water bath using floating racks for 8 minutes at 95^0C. This was then vortexed for 5 seconds, centrifuged in a microcentrifuge for 5 minutes at 10000rpm. The DNA samples were stored at -15^0C.

3.3 DNA Quantification

The extracted DNA samples were quantified using the Quantiblot method where by 5µl of the extracted DNA was hybridized to 20µl of quantiblot D17Z1 probe then conjugated to an enzyme conjugate (Horse radish peroxidase streptavidin). Colour development was then done using trimethylbenzidine (TMB). The standards were serially diluted from standard A

(Applied Biosystems) whose concentration was 2ng/µl. The intensity of the colour was compared against that of seven standards of concentration ranging from 0.03125ng/µl - 2ng/µl (Quantiblot 1993). The required concentration for amplification is 0.05 – 0.125ng/µl which corresponds to standards designated E and F.

3.4 DNA Amplification

The Amp fl STR reaction mix, Amp fl STR Identifiler plus primer set (containing primers for the 15 loci, D3S1358, VWA, D16S539, Amelogenin, D2S1338, D8S1179, D21S11, D18S51, D19S433, CSF1PO, D5S818, D13S317, D7S820, TPOX, TH01 and FGA (Table 16, Appendix 1) which are dye labeled with the 5' dye system 6-FAM, VIC, NED, PET (Figure 9, Appendix 3) and Amp Taq Gold polymerase were vortexed and span.

The Amp fl STR Identifiler plus master mix was prepared by the addition of the following volumes of reagents to a 1.5 ml micro centrifuge tube: number of samples x 10.5µl of PCR reaction mix, number of samples x 0.5µl of Ampli taq Gold DNA polymerase, and number of samples x 5.5µl of Identifiler plus primer set. The master mix was then vortexed and span for one minute. 15µl of master mix was then dispensed into each PCR tube. This was then followed by the addition of 10µl of DNA sample (0.125ng/µl) in the experimental tubes, 10µl (2ng) of control DNA 9947 to the positive control tubes, and 10µl of distilled water to negative control tubes.

These were then amplified in the Gene Amp PCR 9700 system (Applied Biosystems) PCR profile: the initial incubation step was set at 95^0C for 11 minutes, melting at 94^0C for 1 minute, annealing at 59^0C for 1 minute, extension at 72^0C for 1 minute, final extension at 60^0C for 60 minutes and final step at 25^0C to infinity. After the amplification process, the tubes were removed and stored at 2-6^0C.

3.5 Detection and Resolution

The necessary amount of deionized formamide and gene scan – 500 [LIZ] size standards were combined in a single microcentrifuge tube as: (number of sample + 2) x 12µl deionised formamide and (number of sample + 2) x 0.25µl Gene scan – 500[LIZ] size standard, which were then vortexed and spinned in a microcentrifuge. 9µl of this master mix, was aliquoted into a 0.2 ml Genetic Analyzer sample tube. Then 1µl of PCR product or AmpFl STR Identifiler plus Allelic Ladder per tube was added. The tubes were then sealed with septum. Each sample was then denatured by placing in the Amp PCR 9700 system (Applied Biosystem) for three minutes at 95^0C. The tubes were then chilled for 3 minutes in an ice water bath. Then the sample tray was placed into the autosampler of the ABI Prism 310 Genetic Analyzer (Applied Biosystem) and run.

3.6 Data Analysis

Data collected using the ABI Prism 310 data collection computer software (Applied Biosystem) was transferred to the ABI prism Gene scan analysis software (Applied

Biosystem) which presented the results as electropherograms in relative fluorescence units (RFU) in which allele designations were determined by comparison of the sample fragment with those of allelic ladder (Figure 10, Appendix 4). Genotypes were manually scored using the precision values provided by the ABI Prism 310 Data collection manual. The frequency of each allele for each locus was calculated from the numbers of each observed genotype in the sample set (Gene count method) (Edward et al. 1992). Forensic parameters such as the markers observed heterozygosity (het – obs) and unbiased expected heterozygosity (het – exp), power of discrimination (P_D), probability of exclusion (P_E), polymorphism information content (PIC), random match probability (P_M) and average paternity index (ave P_I) were calculated. The accumulated P_D, P_E and P_I for all the 15 loci were also calculated using methods described by Chakraborty et al. (1999).

Congruence of observed genotypes frequencies with their expectations based on the Hardy Weinberg equilibrium was tested at two levels; by comparing the observed and expected frequencies of all homozygotes and all heterozygotes and second through comparison of frequencies of each possible genotype with their respective expectations under the Hardy Weinberg equilibrium by using the exact test with 2000 shuffling experiments (Guo & Thompson 1992). Linkage disequilibrium for each of these pairs of markers was tested by EM algorithm (Excoffier & Slatkin 1995). The value of F_{ST} was calculated by using Arlequin estimates of the coefficient of gene differentiation (G_{ST}) for each locus and averages of the over all loci were obtained through the apportionment of gene diversity and allele size

variance (Chakraborty, Jin & Deka 2001). The p value was used to define the level of significance as the rejection zone.

CHAPTER FOUR
RESULTS

4.1 Allele Frequencies

The allele frequency distributions of the 15 short tandem repeat loci and the forensic important parameters in the three (Kisumu, Nairobi and Mombasa) cities and the pooled populations are shown in Table 1, Table 2, Table 3 and Table 4. The only exception in the computation of allele frequency was the Amelogenin marker derived from the X chromosomal position Xp22.1 – 3 and Y chromosomal position Xp11.2 used as a sex marker.

In the Kisumu (Table 1) population, the allele frequencies ranged from the lowest 0.01 exhibited by 21 alleles to the highest 0.37 (allele 10 marker D7S820) followed by 0.36 (allele 7 marker THOI). The mean allele frequency ranged from 0.044812 – 0.053866.

For the Nairobi (Table 2) population, the lowest allele frequency was 0.01 for 20 alleles and the highest 0.37 (allele 16 marker D3S1358) while the mean allele frequency ranged from 0.044812 – 0.059635. For the Mombasa (Table 3) population, the lowest allele frequency was 0.01 for 14 alleles and the highest was 0.37 (allele 7 marker THOI). The mean allele frequency ranged from a low of 0.046544 to a high of 0.054566. For the pooled population (Table 4), the lowest allele frequency was 0.01 for 9 alleles and the highest 0.37 (allele 7 marker THOI and allele 12 marker D13S317).

The mean allele frequency ranged from 0.016265 to 0.019853. The Kenyan population forensic allele frequency database (Table 5) when compared marker for marker with the database for the African American population (Table 17; Appendix 2) indicates both consistency and disparity between these two populations.The highest allele frequency for both populations was exhibited by allele 12 marker D13S317. Marker D8S1179 is consistent with that of the African American population though only 10 alleles were detected for this locus as opposed to the 11 alleles detected in the African American population. Allele 9 was present in the African American population with a score of 0.0042. The highest for this loci in the African American population was allele 14 (0.3011) and that for the Kenyan population is allele 15 (0.22).

For the marker D21S11, the number of alleles with a frequency score for the Kenyan population is 21. Allele 35.2 with a frequency score of 0.01333 was absent in the African American data for this marker. The highest frequency for the African American database for the marker D21S11 was allele 28(0.2297) and for the Kenyan population database is allele 29 (0.25333). The number of alleles with a frequency score for the African American population for this marker is 24. Allele 38 has a frequency score of 0.14 for this population and was absent from the Kenyan population database for this marker.

Marker D7S820 has 9 alleles present in the Kenyan population database while in the African American population, this marker had 8 alleles. Allele 6 was present in the Kenya population

database with a frequency score of 0.0033. The highest allele frequency score for the Kenyan population for this marker was exhibited by allele 10 with a frequency score of 0.326667 which is consistent with the African American allele 10 (0.3445).

Marker CSFIPO had 8 alleles for the Kenyan population and 9 alleles in the African American population. Allele 11.3 was present in the African American population with a frequency score of 0.0014. For this marker, allele 11 (0.226667) has the highest frequency in the Kenyan population while in the African American population, allele 10 (0.2787) had the highest frequency.

Marker D3S1358 had 8 alleles for the Kenyan population and 10 alleles in the African American population. For this marker, allele 9 (0.0042) and allele 15.2 (0.0014) are present in the African American database but absent in the Kenyan population database. Marker THOI had 7 alleles for the Kenyan population database and 8 alleles in the African American population database. Allele 5 (0.0028) and allele 13.3 (0.0014) which are present in the African American population are absent in the Kenyan population database; however, allele 11 (0.006667) was only present in the Kenyan population database. Allele 7 has the highest frequency score for both the Kenyan and African American populations for this marker.

Marker D13S317 has 7 alleles in the Kenyan population database and 8 alleles in the African American population database. Allele 15 was present in the African American population

with a frequency score of 0.0014. For this marker, allele 12 which has a frequency score of 0.4622 in the African American population has also the highest frequency score of 0.37 in the Kenyan population. Marker D16S539 had 8 alleles in both the African American and Kenyan population databases. Allele 5 (0.006667) is present in the Kenya population database while allele 15 (0.0014) is present in the African American database. Allele 9 (0.226667) had the highest frequency score in the Kenyan population database while allele 11(0.3151) had the highest frequency score in the African American population database. For marker Marker D2S1338, frequency scores begin at allele 15 both for the Kenyan population and the African American populations with each scoring for 13 alleles. Alleles present for the Kenya population are the same as those for the African American population for the same marker. Allele 22 has the highest frequency score for this marker in both populations.

For marker D19S433, the Kenya population has 13 alleles while the African American population has 15 alleles. Allele 13 (0.256667) has the highest frequency score in the Kenyan population and also the highest frequency score of 0.2983 in the African American population. Allele 9 was present in the Kenyan population with a frequency score of 0.00333 while alleles 11.2, 17.2 and 18.2 were present in the African American population with frequency scores of 0.014, 0.0028 and 0.0014, respectively.

Marker VWA had 11 alleles each in both the Kenyan population and African American population. Allele 16 had the highest frequency score in both the Kenyan and African

American populations of 0.28333 and 0.2675, respectively. Allele 12 was only present in the Kenyan population while allele 2 was only present in the African American population. Marker TPOX has 9 alleles in the Kenyan population and 7 alleles in the African American population. Allele 9 (0.276667) has the highest frequency score in the Kenyan population while allele 8 (0.3613) has the highest frequency score in the African American population. Allele 13 was only present in the Kenyan population with a frequency score of 0.00333.

Marker D18S51 has 15 alleles in the Kenyan population and 18 alleles in the African American population. Allele 16 (0.23333) has the highest frequency score in the Kenyan population and allele 17 (0.1821) has the highest frequency score in the African American population. Allele 9, 10, 11 and 23 are only present in the African American population and allele 25 is only present in the Kenyan population. Marker D5S818 has 9 alleles in both the Kenyan and African American populations. Allele 13 (0.296667) has the highest frequency score in the Kenyan population and allele 12 (0.3641) has the highest frequency score in the African American population.

Marker FGA which begins at allele 17 in the Kenyan population has 17 alleles while in the African American population it begins at allele 16.1 and has 24 alleles. Allele 22 has the highest frequency score in both the Kenyan and African American populations of 0.193333 and 0.2157, respectively. Alleles 16.1, 17.2, 18.2, 19.2, 22.2, 22.3, 23.2, 30.2, 34.2, 44.2,

46.2 and 48.2 are only present in the African American population while alleles 17, 26.2, 30, 31.2 and 42.2 are only present in the Kenyan population.

Table 1: Allele frequencies and statistical parameters of 15 STR loci in the Kisumu population (50 samples)

Sample	D851179	D21511	D75820	CSF1PO	D351358	THOI	D135317	D165539	D251338	D195433	VWA	TPOX	D18551	D55818	FGA	
5	-	-	-	-	-	-	-	0.01	-	-	-	-	-	-	-	
6	-	-	-	-	-	0.12	-	-	-	-	-	0.08	-	-	-	
7	-	-	0.02	0.03	-	0.36	-	-	-	-	-	0.02	-	0.02	-	
8	-	-	0.24	0.08	-	0.26	0.05	0.01	-	-	-	0.2	-	0.07	-	
9	-	-	0.14	0.07	-	0.2	0.01	0.27	-	0.01	-	0.33	-	-	-	
9.3	-	-	-	-	-	0.06	-	-	-	-	-	-	-	-	-	
10	-	-	0.37	0.19	-	-	0.02	0.07	-	0.02	-	0.12	-	0.09	-	
11	0.02	-	0.18	0.24	-	-	0.33	0.28	-	0.06	0.02	0.23	-	0.17	-	
12	0.22	-	0.04	0.27	0.01	-	0.41	0.2	-	0.17	0.01	0.02	0.07	0.32	-	
12.2	-	-	-	-	-	-	-	-	-	0.03	-	-	-	-	-	
13	0.18	-	0.01	0.1	-	-	0.14	0.16	-	0.27	-	-	0.04	0.28	-	
13.2	-	-	-	-	-	-	-	-	-	0.07	-	-	0.02	-	-	
14	0.27	-	-	0.02	0.1	-	0.04	-	-	0.21	0.08	-	0.08	0.04	-	
14.2	-	-	-	-	-	-	-	-	-	0.08	-	-	0.01	-	-	
15	0.24	-	-	-	0.31	-	-	-	0.01	0.03	0.16	-	0.1	0.01	-	
15.2	-	-	-	-	-	-	-	-	-	0.02	-	-	-	-	-	
16	0.04	-	-	-	0.29	-	-	-	0.07	0.02	0.29	-	0.31	-	-	
16.2	-	-	-	-	-	-	-	-	-	0.01	-	-	-	-	-	
17	0.02	-	-	-	0.21	-	-	-	0.11	-	0.2	-	0.14	-	-	
18	0.01	-	-	-	0.06	-	-	-	0.06	-	0.16	-	0.08	-	-	
19	-	-	-	-	0.02	-	-	-	0.16	-	0.07	-	0.07	-	0.06	
20	-	0.01	-	-	-	-	-	-	0.1	-	0.01	-	0.05	-	0.08	
21	-	-	-	-	-	-	-	-	0.09	-	-	-	0.01	-	0.08	
22	-	-	-	-	-	-	-	-	0.17	-	-	-	-	-	0.18	
23	-	-	-	-	-	-	-	-	0.13	-	-	-	-	-	0.2	
24	-	-	-	-	-	-	-	-	0.03	-	-	-	-	-	0.14	
24.2	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
25	-	-	-	-	-	-	-	-	0.05	-	-	-	0.02	-	0.09	
26	-	-	-	-	-	-	-	-	0.02	-	-	-	-	-	0.07	
26.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.01	
27	-	0.07	-	-	-	-	-	-	-	-	-	-	-	-	0.03	
28	-	0.22	-	-	-	-	-	-	-	-	-	-	-	-	0.02	
29	-	0.27	-	-	-	-	-	-	-	-	-	-	-	-	0.02	
30	-	0.13	-	-	-	-	-	-	-	-	-	-	-	-	0.01	

31	-	0.05	-	-	-	-	-	-	-	-	-	-	-	-	-	
31.2	-	0.06	-	-	-	-	-	-	-	-	-	-	-	-	-	
32.2	-	0.07	-	-	-	-	-	-	-	-	-	-	-	-	-	
33.2	-	0.03	-	-	-	-	-	-	-	-	-	-	-	-	-	
34	-	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	
34.2	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
35	-	0.04	-	-	-	-	-	-	-	-	-	-	-	-	-	
35.2	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
36	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
42.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.01	
																Cumulative /average
het_obs	0.76	0.82	0.68	0.8	0.76	0.8	0.56	0.72	0.96	0.9	0.82	0.82	0.8	0.88	0.92	0.8
het_exp	0.794	0.854	0.759	0.819	0.769	0.752	0.706	0.786	0.895	0.844	0.828	0.792	0.861	0.792	0.885	0.80907
P-value of exact	0.252	0.804	0.405	0.261	0.954	0.218	0.043	0.017	0.726	0.143	0.16	0.014	0.118	0.027	0.423	
PD	0.9206065	0.959885	0.899892	0.938508	0.904697	0.892932	0.859067	0.915172	0.97614	0.954382	0.939961	0.916691	0.962795	0.91649	0.972764	1-1.1E-18
PE	0.5784087	0.700504	0.529931	0.628969	0.541178	0.513942	0.455725	0.564256	0.769428	0.680816	0.632574	0.569154	0.70881	0.570538	0.753394	4.18E-07
θ	0.0428212	0.039813	0.104084	0.023199	0.011704	-0.06383	0.206799	0.083969	-0.07263	-0.06635	0.009662	-0.03535	0.070848	-0.11111	-0.03955	0.011206
Min allele freq	0.04923	0.055217	0.048	0.04923	0.048	0.044812	0.048	0.048	0.05311	0.053866	0.05039	0.048	0.053866	0.04923	0.05457	

Het-obs, observed heterozygosity; Het-exp, unbiased expected heterozygosity; PD, power of discrimination; PE, probability of excluding paternity (power of exclusion); p, p values of the exact test for Hardy-Weinberg equilibrium ($p<0.05$); Bonferroni correction (0.05/15=0.003).

Table 2: Allele frequencies and statistical parameters of 15 STR loci in the Nairobi population (50 samples)

Sample	D851179	D21511	D75820	CSF1PO	D351358	THOI	D135317	D165539	D251338	D195433	VWA	TPOX	D18551	D55818	FGA	
5	-	-	-	-	-	-	-	0.01	-	-	-	-	-	-	-	
6	-	-	-	-	-	0.21	-	-	-	-	-	0.06	-	-	-	
7	-	-	0.01	0.02	-	0.38	-	-	-	-	-	0.02	-	-	-	
8	0.01	-	0.2	0.03	-	0.19	0.05	0.03	-	-	-	0.29	-	0.08	-	
9	-	-	0.17	0.09	-	0.12	0.02	0.23	-	-	-	0.26	-	-	-	
9.3	-	-	-	-	-	0.07	-	-	-	-	-	-	-	-	-	
10	-	-	0.28	0.23	-	0.01	0.03	0.09	-	0.01	-	0.09	-	0.05	-	
11	0.05	-	0.22	0.26	-	0.02	0.29	0.3	-	0.03	-	0.25	-	0.19	-	
12	0.16	-	0.11	0.35	-	-	0.36	0.22	-	0.15	-	0.03	0.08	0.34	-	
12.2	-	-	-	-	-	-	-	-	-	0.01	-	-	-	-	-	
13	0.15	-	-	0.02	-	-	0.19	0.11	-	0.27	0.01	-	0.08	0.31	-	
13.2	-	-	-	-	-	-	-	-	-	0.07	-	-	-	-	-	
14	0.35	-	0.01	-	0.03	-	0.06	0.01	-	0.27	0.06	-	0.05	0.03	-	
14.2	-	-	-	-	-	-	-	-	-	0.06	-	-	-	-	-	
15	0.19	-	-	-	0.34	-	-	-	-	0.07	0.2	-	0.15	-	-	
15.2	-	-	-	-	-	-	-	-	-	0.03	-	-	-	-	-	
16	0.07	-	-	-	0.37	-	-	-	0.07	0.02	0.24	-	0.2	-	-	
16.2	-	-	-	-	-	-	-	-	-	0.01	-	-	-	-	-	
17	0.02	-	-	-	0.22	-	-	-	0.11	-	0.22	-	0.21	-	-	
18	-	-	-	-	0.04	-	-	-	0.04	-	0.15	-	0.12	-	0.01	
19	-	-	-	-	-	-	-	-	0.17	-	0.1	-	0.08	-	-	
20	-	-	-	-	-	-	-	-	0.07	-	0.01	-	0.02	-	0.07	
21	-	-	-	-	-	-	-	-	0.09	-	-	-	-	-	0.08	
22	-	-	-	-	-	-	-	-	0.14	-	0.01	-	0.01	-	0.23	
23	-	-	-	-	-	-	-	-	0.08	-	-	-	-	-	0.12	
24	-	-	-	-	-	-	-	-	0.09	-	-	-	-	-	0.15	
25	-	0.01	-	-	-	-	-	-	0.04	-	-	-	-	-	0.12	
26	-	0.01	-	-	-	-	-	-	0.04	-	-	-	-	-	0.12	
26.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.01	
27	-	0.04	-	-	-	-	-	-	0.06	-	-	-	-	-	0.01	
28	-	0.16	-	-	-	-	-	-	-	-	-	-	-	-	0.05	
29	-	0.22	-	-	-	-	-	-	-	-	-	-	-	-	0.03	
30	-	0.2	-	-	-	-	-	-	-	-	-	-	-	-	-	

31	-	0.1	-	-	-	-	-	-	-	-	-	-	-	-	-	
31.2	-	0.05	-	-	-	-	-	-	-	-	-	-	-	-	-	
32	-	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	
32.2	-	0.04	-	-	-	-	-	-	-	-	-	-	-	-	-	
33	-	0.04	-	-	-	-	-	-	-	-	-	-	-	-	-	
33.2	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
34	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
34.2	-	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	
35	-	0.03	-	-	-	-	-	-	-	-	-	-	-	-	-	
35.2	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
37	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
																Cumulative /average
het_obs	0.88	0.9	0.82	0.7	0.68	0.78	0.72	0.88	0.92	0.88	0.78	0.66	0.88	0.78	0.88	0.809333333
het_exp	0.793	0.876	0.802	0.762	0.71	0.767	0.756	0.8	0.909	0.826	0.837	0.797	0.875	0.758	0.869	0.809133333
P-value of exact	0.651	0.477	0.268	0.751	0.88	0.247	0.778	0.207	0.289	0.456	0.045	0.034	0.055	0.182	0.206	
PD	0.925395	0.968908	0.924543	0.894625	0.850348	0.904868	0.892889	0.923056	0.981043	0.943339	0.94134	0.912389	0.963375	0.891234	0.968301	1-1.51E-18
PE	0.591467	0.736959	0.587548	0.519246	0.434368	0.541229	0.517323	0.585485	0.794873	0.645148	0.636613	0.559441	0.713204	0.512839	0.733149	4.81E-07
θ	-0.10971	-0.0274	-0.02244	0.081365	0.042254	-0.01695	0.047619	-0.1	-0.0121	-0.06538	0.0681	0.171895	-0.00571	-0.02902	-0.01266	-0.000247178
Min allele freq	0.049273	0.059635	0.048	0.048	0.044812	0.048	0.048	0.049273	0.05311	0.05311	0.050389	0.048	0.051386	0.046544	0.05311	

Het-obs, observed heterozygosity; Het-exp, unbiased expected heterozygosity; PD, power of discrimination; PE, probability of excluding paternity (power of exclusion); p, p values of the exact test for Hardy-Weinberg equilibrium ($p<0.05$); Bonferroni correction (0.05/15=0.003).

Table 3: Allele frequencies and statistical parameters of 15 STR loci in the Mombasa population (50 samples)

Sample	D851179	D21511	D75820	CSF1PO	D351358	THOI	D135317	D165539	D251338	D195433	VWA	TPOX	D18551	D55818	FGA	
6	-	-	0.01	-	-	0.15	-	-	-	-	-	0.08	-	-	-	
7	-	-	0.02	0.04	-	0.37	-	-	-	-	-	-	-	0.01	-	
8	-	-	0.17	0.03	-	0.24	0.05	0.05	-	-	-	0.3	-	0.14	-	
9	-	-	0.13	0.12	-	0.2	0.04	0.18	-	-	-	0.24	-	0.03	-	
9.3	-	-	-	-	-	0.02	-	-	-	-	-	-	-	-	-	
10	0.01	-	0.33	0.18	-	0.02	0.02	0.17	-	0.03	-	0.07	-	0.11	-	
10.2	-	-	-	-	-	-	-	-	-	-	-	-	0.01	-	-	
11	0.05	-	0.19	0.18	-	-	0.32	0.26	-	0.05	-	0.29	-	0.17	-	
12	0.18	-	0.14	0.34	-	-	0.34	0.22	-	0.17	-	0.01	0.06	0.24	-	
12.2	-	-	-	-	-	-	-	-	-	0.01	-	-	-	-	-	
13	0.15	-	-	0.08	0.02	-	0.15	0.12	-	0.23	-	0.01	0.04	0.3	-	
13.2	-	-	-	-	-	-	-	-	-	0.04	-	-	0.01	-	-	
14	0.31	-	0.01	0.03	0.11	-	0.08	-	-	0.25	0.1	-	0.08	-	-	
14.2	-	-	-	-	-	-	-	-	-	0.11	-	-	0.02	-	-	
15	0.23	-	-	-	0.29	-	-	-	0.03	0.06	0.17	-	0.11	-	-	
15.2	-	-	-	-	-	-	-	-	-	0.04	-	-	-	-	-	
16	0.07	-	-	-	0.35	-	-	-	0.1	-	0.32	-	0.19	-	-	
16.2	-	-	-	-	-	-	-	-	-	0.01	-	-	-	-	-	
17	-	-	-	-	0.18	-	-	-	0.07	-	0.24	-	0.16	-	0.03	
18	-	-	-	-	0.04	-	-	-	0.05	-	0.11	-	0.1	-	0.01	
19	-	-	-	-	0.01	-	-	-	0.08	-	0.06	-	0.1	-	0.06	
20	-	-	-	-	-	-	-	-	0.12	-	-	-	0.06	-	0.04	
21	-	-	-	-	-	-	-	-	0.15	-	-	-	0.03	-	0.15	
22	-	-	-	-	-	-	-	-	0.21	-	-	-	0.03	-	0.17	
23	-	-	-	-	-	-	-	-	0.09	-	-	-	-	-	0.13	
24	-	-	-	-	-	-	-	-	0.02	-	-	-	-	-	0.11	
25	-	-	-	-	-	-	-	-	0.03	-	-	-	-	-	0.12	
26	-	-	-	-	-	-	-	-	0.04	-	-	-	-	-	0.04	
27	-	0.05	-	-	-	-	-	-	0.01	-	-	-	-	-	0.06	
28	-	0.15	-	-	-	-	-	-	-	-	-	-	-	-	0.05	
29	-	0.27	-	-	-	-	-	-	-	-	-	-	-	-	0.02	
30	-	0.16	-	-	-	-	-	-	-	-	-	-	-	-	-	
30.2	-	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	
31	-	0.06	-	-	-	-	-	-	-	-	-	-	-	-	-	

31.2	-	0.06	-	-	-	-	-	-	-	-	-	-	-	-	0.01	
32	-	0.03	-	-	-	-	-	-	-	-	-	-	-	-	-	
32.2	-	0.09	-	-	-	-	-	-	-	-	-	-	-	-	-	
33	-	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	
33.2	-	0.03	-	-	-	-	-	-	-	-	-	-	-	-	-	
34.2	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
35	-	0.03	-	-	-	-	-	-	-	-	-	-	-	-	-	
35.2	-	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	
																Cumulative /average
het_obs	0.76	0.84	0.64	0.78	0.84	0.76	0.8	0.86	0.82	0.8	0.88	0.84	0.88	0.82	0.9	0.8146667
het_exp	0.797	0.861	0.801	0.816	0.772	0.772	0.761	0.831	0.9	0.848	0.823	0.8	0.913	0.815	0.9	0.8273333
P-value of exact	0.159	0.025	0.037	0.35	0.376	0.275	0.762	0.501	0.008	0.148	0.857	0.896	0.381	0.013	0.03	
PD	0.92424	0.965793	0.925626	0.931993	0.895057	0.891244	0.898001	0.933584	0.975615	0.952067	0.922792	0.900263	0.977556	0.925255	0.97817	1-3.34E-19
PE	0.58812	0.722674	0.591105	0.609717	0.5205	0.509975	0.529264	0.612621	0.76655	0.672654	0.584379	0.530846	0.776442	0.590445	0.779834	2.35E-07
θ	0.046424	0.02439	0.200999	0.044118	-0.08808	0.015544	-0.05125	-0.0349	0.088889	0.056604	-0.06926	-0.05	0.036145	-0.00613	0	0.0153102
Min allele freq	0.048	0.054566	0.049273	0.049273	0.048	0.046544	0.048	0.046544	0.053866	0.052288	0.046544	0.048	0.054566	0.048	0.054566	

Het-obs, observed heterozygosity; Het-exp, unbiased expected heterozygosity; PD, power of discrimination; PE, probability of excluding paternity (power of exclusion); p, p values of the exact test for Hardy-Weinberg equilibrium ($p<0.05$); Bonferroni correction (0.05/15=0.003).

Table 4: Allele frequencies and statistical parameters of 15 STR loci in the pooled Kenyan population (150 samples)

Sample	D851179	D21511	D75820	CSF1PO	D351358	THOI	D135317	D165539	D251338	D195433	VWA	TPOX	D18551	D55818	FGA	
5	-	-	-	-	-	-	-	0.006667	-	-	-	-	-	-	-	
6	-	-	0.003333	-	-	0.16	-	-	-	-	-	0.073333	-	-	-	
7	-	-	0.016667	0.03	-	0.37	-	-	-	-	-	0.013333	-	0.01	-	
8	0.003333	-	0.203333	0.046667	-	0.23	0.05	0.03	-	-	-	0.263333	-	0.096667	-	
9	-	-	0.146667	0.093333	-	0.173333	0.023333	0.226667	-	0.003333	-	0.276667	-	0.01	-	
9.3	-	-	-	-	-	0.05	-	-	-	-	-	-	-	-	-	
10	0.003333	-	0.326667	0.2	-	0.01	0.023333	0.11	-	0.02	-	0.093333	-	0.083333	-	
10.2	-	-	-	-	-	-	-	-	-	-	-	-	0.003333	-	-	
11	0.04	-	0.196667	0.226667	-	0.006667	0.313333	0.28	-	0.046667	0.006667	0.256667	-	0.176667	-	
12	0.186667	-	0.096667	0.32	0.003333	-	0.37	0.213333	-	0.163333	0.003333	0.02	0.07	0.3	-	
12.2	-	-	-	-	-	-	-	-	-	0.016667	-	-	-	-	-	
13	0.16	-	0.003333	0.066667	0.006667	-	0.16	0.13	-	0.256667	0.003333	0.003333	0.053333	0.296667	-	
13.2	-	-	-	-	-	-	-	-	-	0.06	-	-	0.01	-	-	
14	0.31	-	0.006667	0.016667	0.08	-	0.06	0.003333	-	0.243333	0.08	-	0.07	0.023333	-	
14.2	-	-	-	-	-	-	-	-	-	0.083333	-	-	0.01	-	-	
15	0.22	-	-	-	0.313333	-	-	-	0.013333	0.053333	0.176667	-	0.12	0.003333	-	
15.2	-	-	-	-	-	-	-	-	-	0.03	-	-	-	-	-	
16	0.06	-	-	-	0.336667	-	-	-	0.08	0.013333	0.283333	-	0.233333	-	-	
16.2	-	-	-	-	-	-	-	-	-	0.01	-	-	-	-	-	
17	0.013333	-	-	-	0.203333	-	-	-	0.096667	-	0.22	-	0.17	-	0.01	
18	0.003333	-	-	-	0.046667	-	-	-	0.05	-	0.14	-	0.1	-	0.006667	
19	-	-	-	-	0.01	-	-	-	0.136667	-	0.076667	-	0.083333	-	0.04	
20	-	0.003333	-	-	-	-	-	-	0.096667	-	0.006667	-	0.043333	-	0.063333	
21	-	-	-	-	-	-	-	-	0.11	-	-	-	0.013333	-	0.103333	
22	-	-	-	-	-	-	-	-	0.173333	-	0.003333	-	0.013333	-	0.193333	
23	-	-	-	-	-	-	-	-	0.1	-	-	-	-	-	0.15	
24	-	-	-	-	-	-	-	-	0.046667	-	-	-	-	-	0.133333	
24.2	-	0.006667	-	-	-	-	-	-	-	-	-	-	-	-	-	
25	-	0.003333	-	-	-	-	-	-	0.04	-	-	-	0.006667	-	0.11	
26	-	0.003333	-	-	-	-	-	-	0.033333	-	-	-	-	-	0.076667	
26.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.006667	
27	-	0.053333	-	-	-	-	-	-	0.023333	-	-	-	-	-	0.033333	
28	-	0.176667	-	-	-	-	-	-	-	-	-	-	-	-	0.04	
29	-	0.253333	-	-	-	-	-	-	-	-	-	-	-	-	0.023333	

30	-	0.16	-	-	-	-	-	-	-	-	-	-	-	-	0.003333	
30.2	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
31	-	0.07	-	-	-	-	-	-	-	-	-	-	-	-	-	
31.2	-	0.056667	-	-	-	-	-	-	-	-	-	-	-	-	0.003333	
32	-	0.016667	-	-	-	-	-	-	-	-	-	-	-	-	-	
32.2	-	0.066667	-	-	-	-	-	-	-	-	-	-	-	-	-	
33	-	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	
33.2	-	0.023333	-	-	-	-	-	-	-	-	-	-	-	-	-	
34	-	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	
34.2	-	0.013333	-	-	-	-	-	-	-	-	-	-	-	-	-	
35	-	0.033333	-	-	-	-	-	-	-	-	-	-	-	-	-	
35.2	-	0.013333	-	-	-	-	-	-	-	-	-	-	-	-	-	
36	-	0.003333	-	-	-	-	-	-	-	-	-	-	-	-	-	
37	-	0.003333	-	-	-	-	-	-	-	-	-	-	-	-	-	
42.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.003333	
																Cumulative /average
het_obs	0.8	0.853	0.713	0.76	0.76	0.78	0.693	0.82	0.9	0.86	0.827	0.773	0.853	0.827	0.9	0.807933
het_exp	0.792	0.863	0.788	0.799	0.749	0.763	0.738	0.805	0.902	0.837	0.826	0.795	0.883	0.788	0.886	0.814267
P-value of exact	0.331	0.412	0.0669	0.312	0.729	0.311	0.417	0.648	0.111	0.14	0.306	0.035	0.032	0.003	0.099	
PD	0.907777	0.950839	0.90296	0.91068	0.869552	0.882617	0.868511	0.911162	0.969348	0.936131	0.919446	0.896367	0.956797	0.89744	0.962745	1-1.9E-17
PE	0.716137	1.123243	0.67581	0.560187	0.576784	0.554635	0.475552	0.643429	0.753492	0.717314	0.821404	0.584199	0.875031	0.625443	1.013961	5.95E-10
θ	-0.0101	0.011587	0.095178	0.048811	-0.01469	-0.02228	0.060976	-0.01863	0.002217	-0.02748	-0.00121	0.027673	0.033975	-0.04949	-0.0158	-0.007778
Min allele freq	0.017431	0.019853	0.017087	0.016702	0.016702	0.016265	0.016265	0.016702	0.018288	0.018288	0.017742	0.016702	0.018755	0.017087	0.019164	

Het-obs, observed heterozygosity; Het-exp, unbiased expected heterozygosity; PD, power of discrimination; PE, probability of excluding paternity (power of exclusion); p, p values of the exact test for Hardy-Weinberg equilibrium ($p<0.05$); Bonferroni correction (0.05/15=0.003).

4.2 Hardy Weinberg Equilibrium (HWE)

Hardy Weinberg equilibrium expectations was exhibited in most of the fifteen short tandem repeat loci (based on the Fishers Exact test) with exceptions of some markers significantly deviating from HWE ($p< 0.05$) as shown in Table 5. However, the departures were not significant after the application of the Bonferroni's correction which involves performing multiple tests on the data set.

Table 5: Significant Exact Test for Hardy Weinberg Equilibrium ($p< 0.05$)

Kisumu	Mombasa	Nairobi	Pooled
D135317 (0.043)	D21511 (0.025)	VWA (0.045)	TPOX (0.035)
D165539 (0.017)	D75820 (0.037)	TPOX (0.034)	D18551 (0.032)
TPOX (0.014)	D251338 (0.008)		D55818 (0.003)
D55818 (0.027)	D55818 (0.013)		
	FGA (0.03)		

4.3 Heterozygosity

Heterozygosity is the proportion of individual heterozygous at a locus – gene diversity. For many humans' microsatellite loci, the expected heterozygosity (H_E) is often > 0.85 meaning that one has a > 85% chance of being a heterozygote. In Kisumu population, the observed heterozygosity (het- obs) ranged from the highest shown by marker D2S1338 (0.96) followed by that shown by marker FGA (0.92) and the lowest was shown by marker THOI (0.56). For the Nairobi region population (Table 2), the highest observed heterozygosity (het – obs) was observed in marker D2S1338 (0.92) followed by markers D21S11 (0.9) and FGA (0.88) and the lowest was shown by marker TPOX (0.66).

In Mombasa population, the highest het- obs was shown by marker FGA (0.9) followed by that shown by markers VWA and D18S51 (0.88) and the lowest observed heterozygosity exhibited by marker D7S820 (0.64) (Table 3). For the pooled population for all the three cities, the observed heterozygosity was the highest for markers FGA (0.9) and D2S1338 (0.9) followed by marker D19S433 (0.86) and the lowest was shown by marker D7S820 (0.713) (Table 4).

4.4 Linkage Disequilibrium

Linkage disequilibrium was tested using the EM algorithm where an interclass correlation cliterion for two loci was used. This was performed at 105 pairwise comparisons. The departures for the three cities and the pooled population are shown in Table 6. Some STR markers showed association between them ($p < 0.05$). The numbers of departures are slightly higher than expected but after Bonferroni's correction they were insignificant.

Table 6: Exact Test for Linkage Disequilibrium ($p< 0.05$)

Kisumu	Mombasa	Nairobi	Pooled
D21S11-D13S317 (0.004)	D21S11-D3S1358 (0.041)	CSF1PO-D2S1338 (0.045)	D8S1179-THOI (0.034)
D7S820- THOI (0.044)	D3S1358-D135317 (0.041)	D16S539-D18S11 (0.037)	D21S11-D3S1358 (0.048)
D8S119-D18S51 (0.004)	TPOX-D5S818 (0.023)	D2S1338-D18S11 (0.005)	D21S11-D2S1338 (0.018)
THOI-D5S818 (0.022)		VWA-FGA (0.038)	D3S1358-D18S51 (0.029)
D19S433-D1S511 (0.008)			VWA-TPOX (0.021)
			TPOX-FGA (0.015)

4.5 Power of Exclusion (P_E)

The power of exclusion was computed for each of the three cities populations (Tables 1, 2 and 3) and the pooled population (Table 4). As shown in Table 1 for the Kisumu population sample, the highest power of exclusion (P_E) was for STR marker D2S1338 (0.769) followed by marker FGA (0.753394) and the lowest P_E was for marker THOI (0.455725). For the Mombasa population sample, the highest P_E was for the marker FGA (0.97817) followed by marker D2S1338 (0.975615). The lowest P_E was observed for marker THOI (0.509975) (Table 3). For the Nairobi population sample, the highest P_E was observed for the STR marker D2S1338 (0.794873) and the lowest P_E was for marker D3S1358 (0.434368) (Table 2). The highest P_E for the three cities was for the STR marker D2S1338 for the Mombasa population.

4.6 Power of Discrimination (P_D)

This was computed for each of the three regions as shown in Table 1, 2 and 3. Considering the P_D observed in Kisumu region data, the highest was for marker D2S1338 (0.9761) followed by FGA (0.972764) and the lowest was for marker THOI (0.859067). For the data collected from the Nairobi region, the highest was observed for marker D2S1338 (0.981043), followed markers D21S11 (0.968908) and FGA (0.968301). The lowest P_D was for marker D3S1358 (0.850348).

For the Mombasa region data, the highest P_D was for marker FGA (0.97817), followed by marker D2S1338 (0.975615) and the lowest was marker D3S1353 (0.895057). For the pooled population (Table 4) of 150 samples, the highest P_D was observed for marker D2S1338

(0.969348) followed by marker FGA (0.962745) and the lowest was for marker THOI (0.882617). The combined P_D was greater than 0.99999999.

4.7 Polymorphic Information Content (PIC)

For the Kisumu region data (Table 7), the highest PIC was observed for marker D2S1338 (0.875136), and the lowest marker was marker D13S317 (0.648589). For the Mombasa region data, the highest PIC was observed for marker FGA (0.880894), followed by marker D2S1338 (0.8247) and the lowest was exhibited by marker THOI (0.699905) (Table 8). For the Nairobi region data (Table 9), the highest PIC was observed with marker D2S1338 (0.889925) and the lowest was marker D3S1358 (0.638999). Therefore marker D2S1338 is the most polymorphic marker as exhibited in the PIC value from the three cities.

Table 7: Polymorphic information content values Kisumu

Sample	D8S1179	D21S11	D7S820	CSF1PO	D3S138	THO1	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S818	D5S818	FGA
2N	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Het-obs	0.76	0.82	0.68	0.8	0.76	0.8	0.56	0.72	0.96	0.9	0.82	0.82	0.8	0.88	0.92
PIC	0.75252	0.82891	0.71309	0.78511	0.72313	0.70286	0.648589	0.742456	0.87514	0.817278	0.78781	0.7432	0.83293	0.74223	0.86459

Table 8: Polymorphic Information Content values Mombasa

Sample	D8S1179	D21S11	D7S820	CSF1PO	D3S138	THO1	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S818	D5S818	FGA
2N	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Het-obs	0.76	0.84	0.64	0.78	0.84	0.76	0.8	0.86	0.82	0.8	0.88	0.84	0.88	0.82	0.9
PIC	0.75753	0.84339	0.75915	0.76925	0.70597	0.69991	0.7098	0.7771	0.87246	0.81309	0.75425	0.71621	0.87857	0.75982	0.88089

Table 9: Polymorphic Information Content values Nairobi

Sample	D8S1179	D21S11	D7S820	CSF1PO	D3S138	THO1	D13S317	D16S539	D2S1338	D19S433	VWA	TPOX	D18S818	D5S818	FGA
2N	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Het-obs	0.88	0.9	0.82	0.7	0.68	0.78	0.72	0.88	0.92	0.88	0.78	0.66	0.88	0.78	0.88
PIC	0.75685	0.85301	0.75981	0.70573	0.639	0.7202	0.70184	0.75566	0.88993	0.79307	0.79221	0.73681	0.84068	0.69999	0.85255

4.8 Allele and Genotype distribution

Histograms for both the number of shared alleles and shared genotypes for each of the three populations and the pooled populations are shown in Figures 1 – 8.

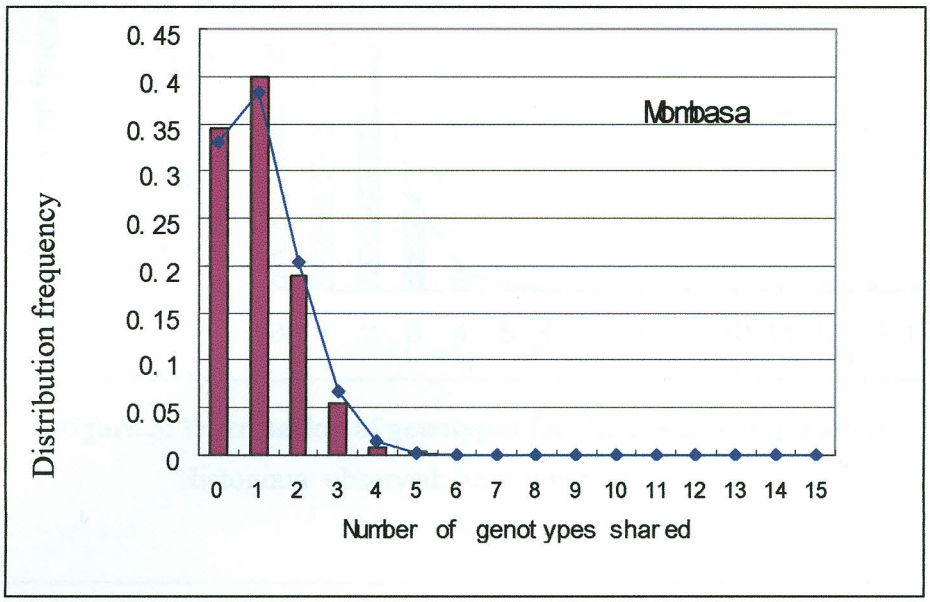


Figure 1: Distribution of genotypes for the Mombasa population.

Histogram: observed; line: expected

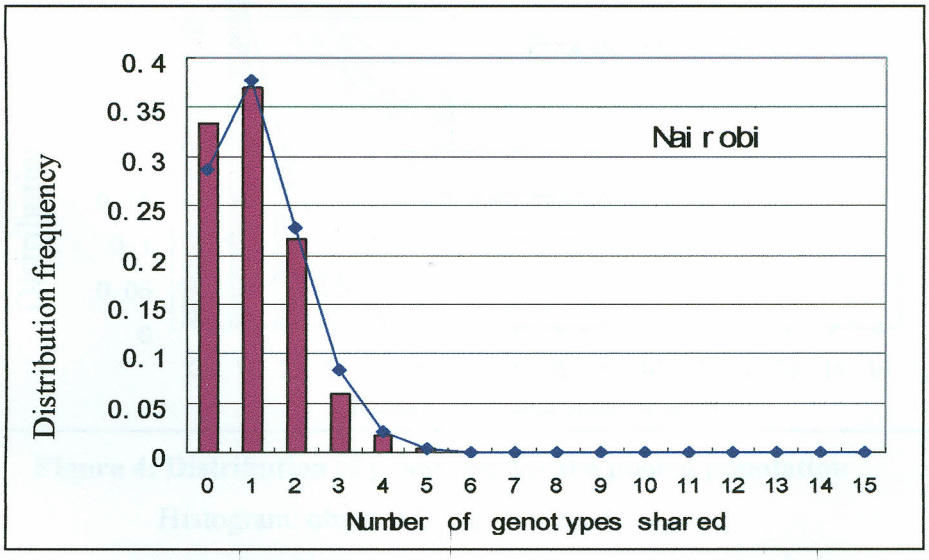


Figure 2: Distribution of génotypes for the Nairobi population

Histogram: observed; line: expected

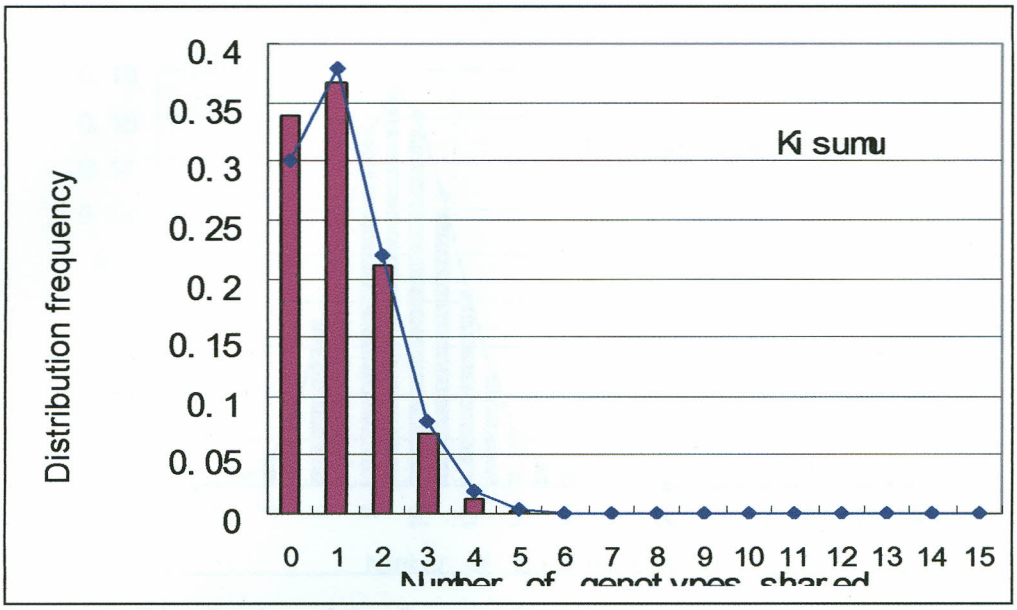


Figure 3: Distribution of genotypes for the Kisumu population

Histogram: observed; line: expected

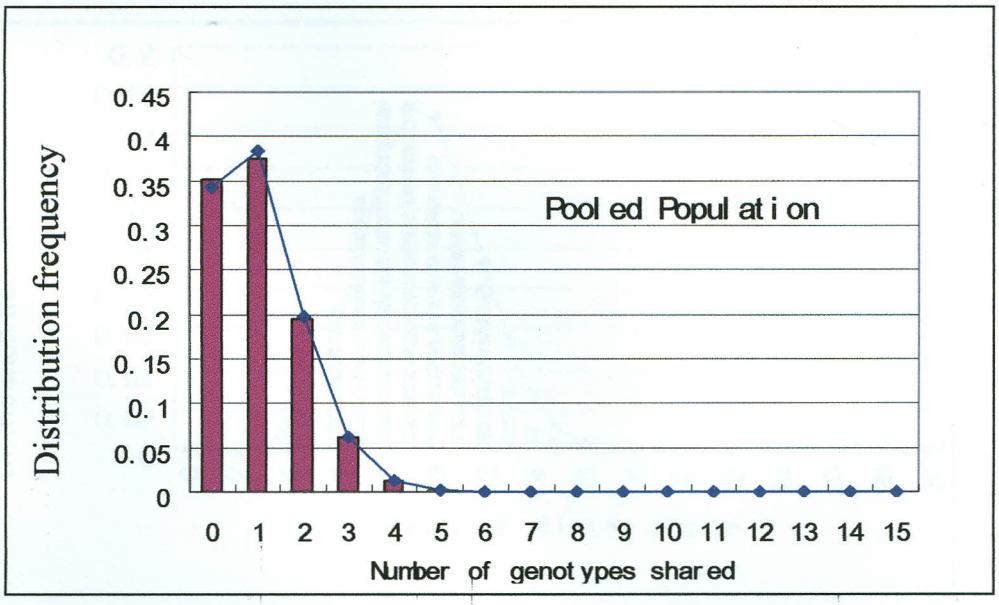


Figure 4: Distribution of genotypes for the pooled population

Histogram: observed; line: expected

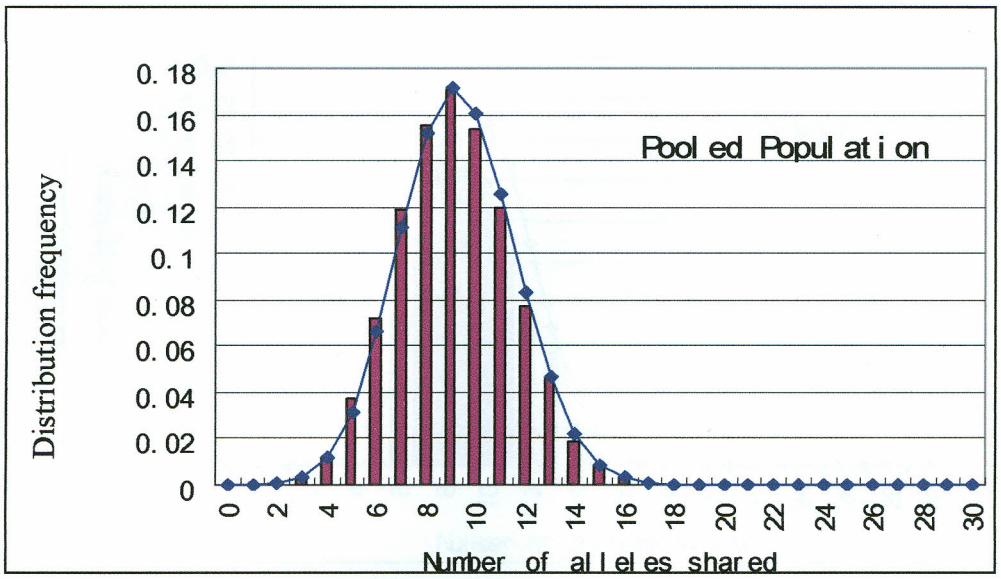


Figure 5: Distribution of the shared alleles for the pooled population

Histogram: observed; line: expected

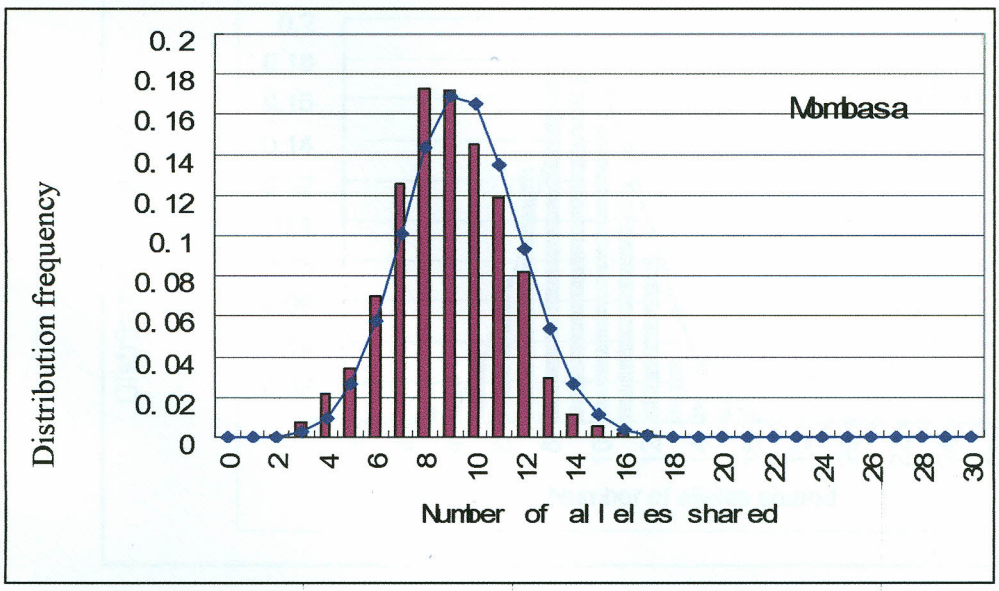


Figure 6: Distribution of the shared alleles for the Mombasa population

Histogram: observed; line: expected

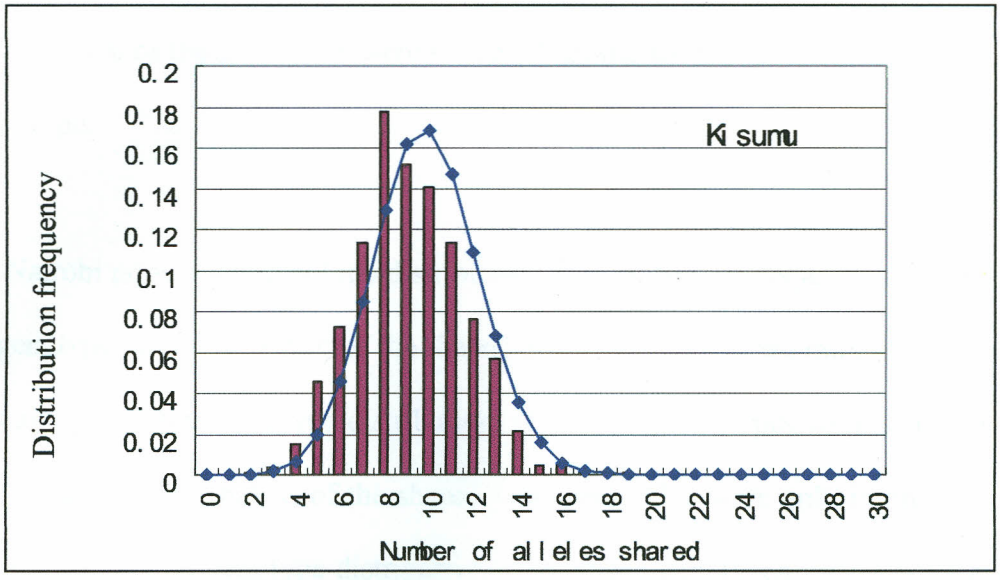


Figure 7: Distribution of shared alleles for the Kisumu population

Histogram: observed; line: expected

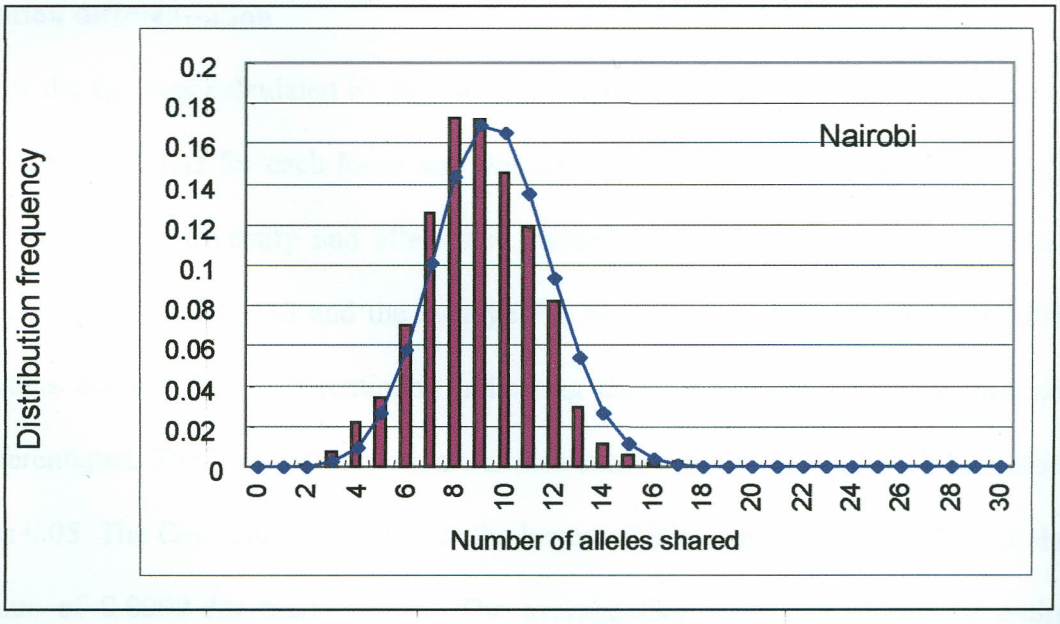


Figure 8: Distribution of shared allele for the Nairobi population

Histogram: observed; line: expected

In all the four shared genotype distributions, the observed do not deviate significantly from the expected. For the Mombasa population genotype distribution, 66% of the genotypes were shared and 34% of the genotypes were unshared; of the shared genotypes, 40% were only in one genotype.

In the Nairobi population genotype distribution, 67% of the genotypes were shared and 33% of the genotypes were unshared; of the shared genotypes, 37% were only in one genotype. In the Kisumu population genotype distribution, 66% of the genotypes were shared and 34% of the genotypes were unshared; of the shared genotypes, 37% were only in one genotype. For the pooled population genotype distribution, 65% of the genotypes were shared and 35% of the genotyped were unshared; of the shared genotypes, 37% were only in one genotype.

4.9 Population differentiation

The value of the F_{ST} was calculated by the Arlequin method. Estimates of the coefficient of gene differentiation (G_{ST}) for each locus and averaged over all loci were obtained through apportionment of gene diversity and allele size variance (Table 10, 11). The F_{ST} values ranged from 0.00007 to 0.00583 and the average F_{ST} value for all the loci was -0.001905. This F_{ST} value depicts little differentiation indicating that the pooled population had not greatly differentiated. The low differentiation was also confirmed by the p values which were greater than 0.05. The G_{ST} values ranged from the lowest -0.00039 for marker D18S51 to the highest value of 0.0009 for marker FGA. The average G_{ST} was -0.0013 for the entire population meaning that the differentiation was very low (0.13%).

Table 10: Population Specific F_{ST} Indices per Polymorphic Locus (absolute values)

Locus	Kisumu	Mombasa	Nairobi	Average F_{ST}	p-value
D8S1179	-0.00452	-0.00455	-0.00451	-0.00453	0.83773
D21S11	-0.00209	-0.00223	-0.00234	-0.00222	0.71652
D7S820	-0.00088	-0.00136	-0.00140	-0.00122	0.59238
CSF1PO	0.00041	-0.00021	0.00040	-0.00007	0.48192
D3S1358	-0.00350	-0.00330	-0.00262	-0.00314	0.67253
THO1	-0.00127	-0.00123	-0.00141	-0.00131	0.49951
D13S317	-0.00540	-0.00609	-0.00601	-0.00583	0.90518
D16S539	-0.00166	-0.00202	-0.00178	-0.00182	0.57185
D2S1338	0.00124	0.00127	0.00110	0.00121	0.31085
D19S433	-0.00570	-0.00567	-0.00546	-0.00561	0.94135
VWA	-0.00487	-0.00453	-0.00493	-0.00478	0.86804
TPOX	-0.00079	-0.00053	-0.00074	-0.00069	0.50244
D18S51	0.00082	0.00033	0.00070	0.00062	0.40665
D5S818	-0.00184	-0.00205	-0.00142	-0.00177	0.54154
FGA	0.00129	0.00113	0.00141	0.00127	0.28055

Table 11: Coefficient of Gene Differentiation (G_{ST}) values

Locus	H_T	Hs	G_{ST}
D8S1179	0.7923	0.7947	-0.003
D21S11	0.8635	0.8648	-0.0015
D7S820	0.7847	0.7853	-0.0008
CSF1PO	0.7924	0.7924	0
D3S1358	0.7409	0.7424	-0.0021
THOI	0.7544	0.7551	-0.0009
D13S317	0.7346	0.7374	-0.0039
D16S539	0.7974	0.7984	-0.0012
D2S1338	0.899	0.8982	0.0008
D19S433	0.8336	0.8367	-0.0037
VWA	0.8108	0.8134	-0.0032
TPOX	0.7761	0.7765	-0.0005
D18S51	0.8731	0.8727	0.0004
D5S818	0.7763	0.7772	-0.0012
FGA	0.8876	0.8868	0.0009
Average	0.8078	0.8088	-0.0013
SE	0.0135	0.0134	0.0004

H_T– Total heterozygosity
H_S– Subpopulation heterozygosity

4.9.1 Matrix of F_{ST} Pairwise Comparison of the Three Cities Population and their F_{ST} p values

The p values for the F_{ST} pair wise comparisons between the three cities population data was greater than 0.05. For the F_{ST} pairwise comparison between the African American population and each of Kisumu and Mombasa population, the p value was 0.0000 which was lower than the confidence level. This difference is statistically significant.

The F_{ST} pairwise comparison between the Nairobi population and the African American population yielded a p value of 0.22523 which was lower than the p values of the F_{ST} pairwise comparisons of the three cities. The p value for the F_{ST} pairwise comparison between the pooled Kenyan population (150 samples) and the African American population was 0.003. The two estimators of genetic differentiation were used with the following observations; the values of F_{ST} and G_{ST} were almost similar but the F_{ST} was a little higher in magnitude than the G_{ST}. The F_{ST} values ranged from 0.00007 to 0.00583.

Table 12: F_{ST} Pairwise Comparisons of the Three Cities Population

Population	Kisumu	Mombasa	Nairobi	African-American
Kisumu	*			
Mombasa	-0.00192	*		
Nairobi	-0.00098	-0.00278	*	
African-American	0.00286	0.00220	0.00058	*

Table 13: F_{ST} p values of the Pairwise Comparisons of the Three Cities Population.

Population	Kisumu	Mombasa	Nairobi	African-American
Kisumu	*			
Mombasa	0.84685±0.0365	*		
Nairobi	0.76577±0.0434	0.92793±0.0274	*	
African-American	0.00000±0.0000	0.00000±0.0000	0.22523±0.0365	*

4.9.2 Analysis of Molecular Variance (AMOVA)

The analysis of molecular variance was done for the Kenyan population three regions (Kisumu, Nairobi and Mombasa) among populations, among individuals within populations and within individuals and the percentage of variation calculated. -0.20 % variation was observed for the among populations at 2 df, 0.32 % variation was observed among individuals within populations at 147 df, 99.87 % within individuals at 150 df (Table 15).

AMOVA was also done to determine the variance between the Kenyan pooled population and the African American population (Table 16). Variation of individuals within the two populations is very high (100.19%) meaning two individuals in the populations would be characterized at this percentage.

Table 14: AMOVA for the Kenyan Population

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	2	9.810	-0.01181 Va	-0.20
Among individuals within populations	147	894.620	0.01959 Vb	0.32
Within Individuals	150	907.000	6.04667 Vc	99.87
Total	299	1811.430	6.05445	

Fixation Indices
FIS: 0.00323, FST: -0.00195, FIT: 0.00129

FIS and FIT represent correlation of alleles to the local and total populations, respectively and their values describe departures from the expected HWE genotypic frequencies.

Table 15: AMOVA values between the Kenyan and the African American populations

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	2	8.433	-0.00982 Va	-0.19
Within populations	297	1543.980	5.19859 Vb	100.19
Total	299	1552.413	5.18877	

Fixation Index FST: -0.00189

CHAPTER FIVE
DISCUSSION, CONCLUSION AND RECOMMENDATION

5.1 DISCUSSION

Kenya as a country has a population of about 33 million people according to the 1999 National Population Census. Its divided into eight provinces (Figure 12; Appendix 6) and has forty two different tribes distributed in these provinces with particular tribes concentrated in specific places of the country (Figure 11; Appendix 5). There are three major cities recognized in the country which are located strategically, Nairobi is the capital city positioned centrally, Kisumu is in the western region of the country while Mombasa covers the coastal zones as an island of the Indian Ocean spanning to the Eastern region of Kenya. Nairobi city is largely cosmopolitan with an admix of almost all the Major tribes of the country. Kisumu has a high concentration of the Luo and Lugha tribes while Mombasa has a high concentration of the Swahili (cross breed of the Arab and ancient black coastal tribes), Mijikenda and Kamba tribes although it's generally assumed that both cities are largely cosmopolitan.

Due to this, therefore it was generally assumed that samples drawn from each of three cities would form a representative sample of the entire Kenyan population. The Kenyan population markers have a total of 165 observed alleles with 15 alleles appearing in the Kenyan population only for the identifiler profiling system. The African American markers for the identifiler profiling system have a total of 183 observed alleles with 34 alleles present in this population but totally absent in the Kenyan population database. This presence of common alleles and population specific alleles in STR markers has also been observed between the Kenyan population databases and other population databases including Mozambique,

Venezuela, Japanese, Asian, West China, Scottish, Brazilian, Central China, South Poland, Northern Spain, Philippine, Singapore Malays, Singapore Chinese, Spanish and Tibetan populations (Alves et al. 2001, Fernandez et al. 2005, Asamura et al. 2007, Yamamoto et al. 2003, Zheng et al. 2007, Zhang et al. 2007, Goodwin, Scoular & Linacre 2001, Sao-Bento et al. 2008, Steinlecher et al. 2002, Wolanska-Nowak, Branicki & Kupiec 2001, Miozzo et al. 2003, Zarrabeitia & Riancho 2001, De Ungria et al. 2005, Ang et al. 2005, Syn et al. 2005, Lim et al. 2005, Entrala et al. 1998, Ota et al. 2007)

According to the results, the highest heterozygosity was observed for the STR marker D2S1338 for both the Nairobi and Kisumu cities while the highest for Mombasa was FGA which is equally as high in the other two cities. The slight difference in Mombasa which is also evident in the other forensic parameters could be attributed to gene flow as a result of cross breeding of the coastal black population with the Arab population which dates way back to the time prior to the onset of the ancient slave trade. The results of the heterozygosity for the pooled Kenyan population indicate that the STR markers D2S1338 and FGA are the most prevalent and widely distributed markers in the Kenyan population. That FGA was the most prevalent marker has also been reported by Entrala et al. (1998), Perez et al. (2000), Alves et al. (2001), Steinlechner et al. (2002), Miozzo et al. (2003), Ang et al. (2005), Zhang et al. (2007), Ota et al. (2007), and Sao-Bento et al. (2008) for the Spanish (FGA, 0.901), Peru (FGA, 0.870), Mozambique (FGA, 0.880), Gabon (FGA, 0.931), Argentina (FGA, 0.884), Singapore Malays (FGA, 0.888), West China (FGA, 0.880), Nepal Kathmandu (FGA, 0.892), and Brazilian (FGA, 0.255) population, respectively.

All the markers were observed to be highly polymorphic with D2S1338 having the highest 88.99% and the lowest D3S1358 (63.9%). This is consistent with other databases where the polymorphism of markers was quite high (Entrala et al. (1998), Perez et al. (2000), Alves et al. (2001), Steinlechner et al. (2002), Miozzo et al. (2003), Ang et al. (2005), Zhang et al. (2007), Ota et al. (2007), and Sao-Bento et al. (2008) for the Spanish (0.696-0.901), Peru (0.660-0.870), Mozambique (0.705-0.880), Gabon (0.663-0.931), Argentina (0.728-0.884), Singapore Malays (0.594-0.888), West China (0.560-0.880) and Nepal Kathmandu (0.649-0.892) population, respectively.

Departures from the Hardy Weinberg Equilibrium in some loci in the three cities could suggest the existence of some population heterogeneity and Wahlund effect (Zarrabeitia & Riancho 2001). However, the disappearance of statistical significance after Bonferroni correction could widely be attributed to chance or inbreeding due to assortative mating. Kenya, a highly ethnic country could be practicing intramarriages with a high preference to ethnicity. Lack of significance after Bonferroni correction of some previously significant loci have also been reported by Entrala et al. (1998), Zarrabeitia & Riancho (2001), Goodwin, Scoular & Linacre (2001), Steinlechner et al. (2002), Ang et al. (2005), De Ungria et al. (2005), Ota et al. (2007), Zheng et al. (2007), and Sao-Bento et al. (2008) for the Spanish, Northern Spanish, Scottish, Gabon, Singapore Malays, Philippine, Nepal Sherpa and Kathmandu, Central China, and Brazilian populations. In the Coastal region, gene flow due to migration could also be a factor. The coastal region (Mombasa) where the highest heterozygosity was observed has a high population of descendants of Arab origin. Therefore

a Kenyan Arab, Asian and Caucasian populations would be an appropriate indicator of the level of cross breeding in the Kenyan population sample.

For the Power of Discrimination (P_D) and the Power of Exclusion (P_E), higher values for any given population indicate a greater individualization potential. The higher the P_D values, the larger the proportion of the population that can be excluded. According to the obtained results the greatest potential of individualization can be obtained with the markers D2S1338 and FGA because these two markers have the highest P_D and P_E values for all the three cities. The high values of these parameters were also reported by Entrala et al. (1998), Perez et al. (2000), Alves et al. (2001), Goodwin, Scoular & Linacre (2001), Zarrabeitia & Riancho (2001), Steinlechner et al. (2002), Miozzo et al. (2003), Ang et al. (2005), Syn et al. (2005), De Ungria et al. (2005), Lim et al. (2005), Zhang et al. (2007), Zheng et al. (2007), Ota et al. (2007), and Sao-Bento et al. (2008) for the Spanish (PE, 0.593-0.739; PD, 0.875-0.964), Peru (PE, 0.369-0.735; PD, 0.724-0.966), Mozambique (PE, 0.440-0.737; PD, 0.861-0.972), Scottish (PE, 0.472-0.752; PD, 0.862-0.959), Northern Spain (PE, 0.373-0.725; PD, 0.873-0.967), Gabon (PE, 0.481-0.791; PD, 0.888-0.973), Argentina (PE, 0.480-0.754; PD, 0.864-0.971), Singapore Malays (PE, 0.341-0.723; PD, 0.780-0.962), Singapore Chinese (PE, 0.341-0.732; PD, 0.782-0.967), Philippine (PE, 0.331-0.749; PD, 0.795-0.959), Singapore Indian (PE, 0.433-0.727; PD, 0.856-0.962), West China (PE, 0.246-0.755; PD, 0.736-0.966), Central China (PE, 0.480-0.704; PD, 0.875-0.962), Nepal Kathmandu (PE, 0.353-0.778; PD, 0.863-0.965), and Brazilian (PE, 0.421-0.747; PD, 0.663-0.870) population, respectively.

Although in forensic science using molecular markers it's not adequate to discriminate individuals based on one single marker system the two markers D2S1338 and FGA would actually appear to be the choicest and most informative markers. The marker THOI would be the least informative STR marker amongst individuals in the general Kenyan population based on the low values obtained for both P_D and P_E. The greater the number of alleles (degree of polymorphism) and the more even the distribution of these alleles in a population (the heterozygosity), the higher the power of discrimination and exclusion.

The D2S1338 and FGA in each of the three cities are quite high in terms of P_D with the pooled population P_D being 96.93% (D2S1338) and 96.27% (FGA). Therefore although individual power of exclusion estimates varied at different loci, the combined exclusion value is greater than 0.99999973 in all the three populations and the combined power of discrimination is greater than 0.99999999. The combined power of discrimination and combined probability of exclusion is similar to that reported by Entrala et al. (1998), Steinlechner et al. (2002), Miozzo et al. (2003), Syn et al. (2005), Ota et al;. (2007), Zhang et al. (2007), and Sao-Bento et al. (2008) for the Spanish (PE, 0.99991; PD, 0.999997), Gabon (PE, 0.999981), Cordoba (Argentina) (PE, 0.9999074; PD, 0.999999999991), Singapore Chinese (PE, 0.9999899), Nepal (Sherpa and Kathmandu) (PE, 0.999993, 0.99991; PD, 0.999999, 0.999997), West China (PE, 0.9999994; PD, 0.999999999999999), and Brazilian (PE, 0.99999995; PD, 0.99999999997) population.

The distribution of both the observed and expected shared allele and genotypes were almost equal in the three regions and exhibited similar patterns indicating an even distribution of

these alleles and genotypes implying that the populations in the three regions are homogenous. This argument is also supported similar small F_{ST} (0.00007-0.00583) and mean overall loci G_{ST} (0.0013) values in the three Kenyan populations. The small values of FST indicate little population differentiation and the small GST indicates that only 0.13% of the total gene diversity exists in the Kenyan population data. F_{ST} estimates interpopulational gene flow and G_{ST} estimates population subdivision. This is also supported AMOVA which indicated that both among population and among individuals within populations variations were negligible (0.20% & 0.32%, respectively) and variations between two individuals in the population was high (99.87%). Thus the three populations can therefore be taken as subsets of the same general population. This would actually mean that two individuals taken at random in any of the three cities in the entire Kenyan population would vary with a very high probability of 99.87 %, a very high level of individualization. The findings of this study compare with the F_{ST} values observed in Japanese subpopulations by Uchichi (2007) which ranged from 0.0020 to 0.0118.

The F_{ST} p values comparing the pooled Kenyan population from the three cities and the African American population data (p=0.003) indicated that these populations are entirely different. This shows that though it largely assumed that the African American descent has its roots in Africa and therefore considered to be closer to the black population than the Caucasians, the years of separation have led to great differentiation. These population differences have also been reported between Sichuan population in West China and each of Argentina population, West African population, Japanese population, Korean population and Thai population (Zhang et al. 2007). The similarities in frequency histograms could be

explained by their relative common origin (Hamilton et al. 1996). Phylogenetic and genetic distance studies on the Kenyan population data could reveal whether the population has a relatively recent common origin. Databasing of the major tribes of Kenya would also reveal the level of population subdivision and structure which was not covered by this study.

The results on linkage disequilibrium indicated the presence of some correlation between some loci which disappeared after the application of Bonferroni correction. The data showed little evidence of association between all 15 loci, therefore, the assumption of independence is valid and a multiple-locus profile frequency can be estimated using the product rule in this population. The entire battery of markers could then be rightly described to be in linkage disequilibrium, that is, they are in different haplotypes and thus independent of one another. The frequency of one is not dependent on the other. This is a characteristic of a good forensic molecular marker. These results are in line with genetic data reported by Entrala et al. (1998), Perez et al. (2000), Alves et al. (2001), Goodwin, Scoular & Linacre (2001), Zarrabeitia & Riancho (2001), Steinlechner et al. (2002), Miozzo et al. (2003), Ang et al. (2005), Syn et al. (2005), De Ungria et al. (2005), Lim et al. (2005), Zhang et al. (2007), Zheng et al. (2007), Ota et al. (2007), and Sao-Bento et al. (2008) for the Spanish, Peru, Mozambique, Scottish, Northern Spain, Gabon, Argentina, Singapore Malays, Singapore Chinese, Philippine, Singapore Indian, West China, Central China, Nepal Kathmandu, and Brazilian population, respectively.

5.2 Conclusion

In conclusion,

- A Kenyan population database has been established for the loci D8S1179, D21S11, D7S820, CSF1PO, D3S1358, THO1, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818 and FGA. The allelic frequencies of these PCR based loci can be used to estimate the frequency of a multiple locus DNA profile in the Kenyan population for forensic individual identification, paternity investigation and population genetic studies.. The data demonstrates that a high degree of discrimination can be obtained when all the fifteen loci are used to characterize forensic biological evidence.

- A common Kenyan population database was produced by pooling together all the three Kenyan populations since they were similar in the observed heterozygosity, power of discrimination, probability of exclusion of paternity (power of exclusion), polymorphic information content, and allele and genotype frequency histograms and were in Hardy-Weinberg Equilibrium. The most informative markers for the three Kenyan populations were FGA and D2S1338. However, for the Mombasa population, the most informative marker was FGA..

- Because the 15 loci Kenya STR based database was different from the African American population database, each population should use a separate database for forensic individual identification, paternity investigation and population genetic studies. Compared marker for marker, the study revealed 15 alleles unique in the Kenyan population and 34 alleles unique in the African American population.

5.3 Recommendations

- The data generated for the 15 STR Identifier system is suitable for use in forensic DNA analysis for the Kenyan population.

- To establish the Asian and Arab population Kenyan database in order to assess the level of gene flow into the Kenyan population due to cross breeding as suggested in the Mombasa population database.

- Study of the major ethnic groups in the Kenyan population in order to assess the level of population sub structuring and sub division.

REFERENCES

Alastar, M.R & Harding, H.N. 1989, 'DNA typing', Forensic Science International, vol. 41, pp. 197-203.

Alves, C., Gusmao, L & Amorim A. 2000, 'STR data (AmpFlSTR Proflier Plus and Gene Print CTTv) from Mozambique', Forensic Science International, vol. 119, pp. 131 – 133.

Ambach, E., Zchethofer K & Schcithan, R. 1996, 'HLA – DQ alpha genotype and allele frequencies in Austrian', American Journal of Human Genetics, vol. 46, pp. 71-75.

Ang, H.C., Sornarajah, R., Lim, S.E.S., Syn, C.K.C., Tan-Siew, W.F., Chow, S.T & Budowle, B. 2005, 'STR data for the 13 CODIS loci in Singapore Malays', Forensic Science International, 148, pp. 243-245.

Asamura, H., Fujimori, S., Ota, M & Fukushima, H. 2007, 'MiniSTR multiplex systems based on non-CODIS loci for analysis of degraded DNA samples', Forensic science International, vol. 173, pp.7-15.

Balaz, I., Debbie, V.F & Kevin, C.M. 1993, 'DNA-based Identity testing in Forensic Science', Bioscience, vol. 43, pp. 153-156.

Balloux, F & Lugon-Moulin, N. 2002, 'The estimation of population differentiation with micro-satellite markers', Molecular Ecology, vol.11, pp. 155-165.

Bowcock, A. M., Ruiz-Linares, A., Tomfohdre, J., Minch, E., Kidd, J. R & Carvalli-Sforza, L. L. 1994, 'High resolution of human evolutionary trees with polymorphic micro- satellites' Nature, vol. 368, pp. 455-457.

Budowle, B., Jankowski, L.B & Corey, H. W. 1997, 'Evaluation of independence. Assumption for PCR Based and Protein based Genetic markers, New Jersey Caucasians', Journal of Forensic Science, vol. 42, pp. 223-225.

Budowle, B & Nhari, L. T. 1997, 'Zimbabwe Data on six short tandem repeat loci CSF1PO, TPOX, THOI, D3S1358, VWA and FGA', Journal of Forensic Science, vol. 42, pp. 278 – 280.

Cavalli-Sforza, L & Bodmer, W. 1971, 'The genetics of human population', Focus on population genetics.

Chakraborty, R & Jin, L. 1993, 'Determination of relatedness between individuals using DNA fingerprinting', Human Biology, vol. 65, pp. 875-895.

Chakraborty, R. D., Stivers, B., Zhong, Y. & Budowle, B. 1999, 'The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems', Electrophoresis, vol. 20, pp. 1682-1696.

Chakraborty, R., Jin, L & Deka, R. 2001, 'Extent and pattern of gene diversity at micro satellite loci: Implications for disease-gene association studies', American Journal of Human Genetics, vol. 69, pp. 262.

Chee, M., Yang, R & Hubbell, E. 1996, 'Accessing genetic information with High Density DNA Arrays', Science, vol. 274 (5287), pp. 610-614.

De Ungria, M.C.A., Roy, R.K., Tabbada, K.A., Rao-Coticone, S., Tan, M.M.M & Hernandez, K.N. 2005, 'Allele frequencies of 19 STR loci in a Philippine population generated using AmpFISTR multiplex and ALF singleplex systems', Forensic Science International, vol. 152, pp. 281-284.

Durand, P., Sire, C & Theron, A. 2000, 'Isolation of microsatellite markers in the digenetic trematode *Schistosoma mansoni* from Guadalupe Island', Molecular Ecology, vol. 9, pp. 993-1011.

Edwards, A., Hammond, H. A & Chakraborty, R. 1992, 'Genetic Variation at five trimeric and tetrameric tandem repeat loci in four human population groups', Genomics , vol. 12, pp. 241-253.

Entrala, C., Lorente, M., Lorente, J.A., Alvarez, J.C., Moretti, T., Budowle, B & Villanueva, E. 1998, 'Fluorescent multiplex analysis of nine STR loci: Spanish population data' Forensic Science International, vol. 98, pp. 179-183.

Excoffier, L and Slatkin, M. 1995, 'Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population', Molecular Biology and Evolution, vol. 12, pp. 921-927

Fernandez, W.M.Z., Borjas-Fajardo, L., Salgado, E.F., Castillo, C., Socca, L., Portillo, M.G., Sanchez, M.A., Delgado, W., Morales-Machin, A., Layrisse, Z & Bernal, L.P. 2005, 'Use of short tandem repeats loci to study the genetic structure of several populations from Zulia state, Venezuela', American Journal of Human Biology, vol. 17, pp. 451-459.

Gamero, J.J., Romero, J.L., Gonzalez, J.L., Arufe, M.I., Cuesta, M.I., Corte-Real, F., Carvalho, M., Anjos, M.J., Vieira, D.N & Vide, M.C. 2000, 'A study on ten tandem repeat systems: African immigrant and Spanish population data', Forensic Science International, 110, pp. 167-177.

Gill, P., Jeffrey's, A S & Warret, D.J. 1985, 'Forensic application of DNA fingerprints', Nature, vol. 318, pp. 577-579.

Goodwin, W., Scoular, C & Linacre, A. 2001, '13 STR loci frequency data from a Scottish population', Forensic science International, vol. 116, pp. 187-188.

Guo, S. W & Thompson, E.A. 1992, 'Performing the exact test of Hardy-Weinberg proportion for Multi alleles', Biometrics, vol. 48, pp. 361-372.

Hamilton, J.F., Starling, L., Cordiner, S.J & Monahan, D. L. 1996, 'New Zealand Population data at five VNTR loci: Validation as databases for forensic identity testing', Science and Justice, vol. 36(2), pp. 109 – 117.

Harris, H. 1975, 'Principles of Human Biochemical Genetics', New York Academic Press, edition 2, pp. 286-295.

Hayes, J. M., Budowle, B & Freund, M. 1995, 'Arab Population Data on the PCR- Based loci-HLA-DQA1, LDLR, GYPA, HBGG, D758, GC and DIS80', Journal of Forensic science , vol. 40, pp. 888-892.

Helene, P., Bertrand, L., Pascal, K., Antoine, T & Patrice, M. 1995, 'French Caucasian population data for HUMTHOI and HUMFES/ FPS short tandem Repeat (STR) systems', Journal of Forensic Sciences, vol. 40, pp. 270-274.

Huang, N.E & Budowle, B. 1995, 'Fixed Bin Population Data for the VNTR Loci D157, D2S44, D4SI39, D5S110 and D17S79 in Chinese from Taiwan', Journal of Forensic Sciences, vol. 40, pp. 287-290.

Jeffery's, A. J., Wilson, V & Thein, S. L. 1985, 'Hyper variable "Minisatellite" region in Human DNA', Nature, vol. 314, pp. 67-73.

Lim, S.E.S., Tain-Siew, W.F., Syn, C.K.C., Ang, H.C., Chow, S.T & Budowle, B. 2004, 'Genetic data for the 13 CODIS STR loci in Singapore Indians', Forensic Science International, vol. 148, pp. 65 – 67.

Ota, M., Droma, Y., Basnyat, B., Katsuyama, Y., Asamura, H., Sakai, H & Fukuhsima, H. 2006, 'Allele frequencies for 15 STR loci in Tibetan populations from Nepal', Forensic Science International, vol. 169, pp. 234 – 238.

Metropolitan police forensic science laboratory methods manual.

Miozzo, M.C., Pacharoni, C.M., Mutai, S.A., Maxzud, M.K., Casanova, A.E & Modesti, N.M. 2003, 'STRs data for the loci D3S1385, VWA, FGA, D8S1179, D21S11, D5S818, D13S317 and D7S820 from Cordoba (Argentina)', Forensic science International, vol. 131, pp. 214-217.

Montaldo, H & Meza-Herrera, C. A. 1998, 'Use of molecular markers and major genes in the genetic improvement of livestock', Electronic Journal of Biotechnology, vol. 1(2), pp. 83-89.

National Research Council 1996, 'The evaluation of forensic DNA evidence', National Academy Press, Washington DC.

Nei, M. 1984, 'Molecular evolutionary Genetics', Columbia University Press, New York

Nei, M. 1987, 'Molecular evolutionary Genetics', Columbia University Press, New York.

Ota, M., Droma, Y., Basnyat, B., Katsuyama, Y., Asamura, H., Sakai, H & Fukuhsima, H. 2007, 'Allele frequencies for 15 STR loci in Tibetan population from Nepal', Forensic Science International, vol. 169, pp. 234-238.

Perez, L., Hau, J., Izarra, F., Ochoa, O., Zubiate, U & Garcia, O. 2003, 'Allele frequencies for the 13 CODIS STR loci in Peru', Forensic Science International, vol. 132, pp. 164-165.

Pfitzinger, H & Ludes, B. 1995, 'French Caucasian Population Data for HUMTHOI and HUMFES/FPS short Tandem Repeat (STR) systems', Journal of Forensic Sciences , vol. 40, pp. 270-274.

Primmer, C. R., Raudsepp, T., Chowdhary, B. P., Moller, A. P & Ellegren, H. 1997, 'Low frequency of micro-satellite in the avian genome', Genome Research, vol. 7, pp. 471-481.

Quantiblot 1993, 'Human DNA quantitation kit, package insert', Roche molecular systems, Inc, Alameda, CA.

Rebecca, R & Sensabangh, G. 1991, 'Analysis of Genetic marker in Forensic DNA samples using the Polymerase chain Reaction', Analytical Biotechnology, vol. 63, pp. 2-15

Sao-Bento, M., Carvalho, M., Andrade, L., Lopes, V., Serra, A., Gamero, J.J., Corte-Real, F., Vide, M.C & Anjos, M.J. 2008, 'STR data for the 15 AmpFISTR identifiler loci in the Brazilian population of Sao Paulo state', Forensic Science International: Genetics Supplement Series, vol. 1, pp. 367-369.

Slatkin, M. 1995, 'A measure of population subdivision based on micro satellite allele frequencies', Genetics, vol. 139, pp. 457-462.

Smith, T. B. 1994, 'Molecular genetic approaches in conservation', Oxford University Press.

Steinlencher, M., Schmidt, K., Kraft, H. G., Uttermann, G & Parson, W. 2001, 'Gabon black population data on the ten short tandem repeat loci', International Journal of Legal Medicine, vol. 116, pp. 176 – 178.

Sunnucks, P. 2000, 'Efficient Genetic Markers for Population Biology', Tree, vol. 15, pp. 199-201.

Syn, C.K.C., Chuah, S.Y., Ang, H.C., Lim, S.E.S., Tan-Siew, W.F., Chow, S.T & Budowle, B. 2005, 'Genetic data for the 13 CODIS STR loci in Singapore Chinese', Forensic science International, vol. 152, pp. 285-288.

Tautz, D. 1989, 'Hyper variability of simple sequences as a general source of polymorphic DNA markers', Nucleic Acids Research, vol. 17, pp. 6463-6471.

Tsutsumi, H., Komuro, T., Mukoyama, R & Nogami, H. 2008, 'Polymorphism of the VWA locus in the center of the three VWA STRs in the Japanese population', Legal Medicine, vol. 10, pp. 31-34.

Uchihi, R., Yamamoto, T., Yoshimoto, T., Inoue, C., Kishida, T., Yoshioka, N & Katsumata, Y. 2006, 'Population data of six STR loci using hexaplex PCR amplification and typing system, "Midi-6" in four regional populations in Japan', Forensic Science International, vol. 169, pp. 255–259.

Vos, P & Hogers, R. 1995, 'AFLP, a new technique for DNA fingerprinting', Nucleic Acid Research, vol. 23, pp. 4407-4414.

Walsh, S., Metzger, D.A & Higuchi, R. 1991, 'Chelex 100 as a medium for simple extraction of DNA for PCR- based typing from forensic material', Bio Techniques, vol. 10, pp. 506-513.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A & Tingey, S.V. 1990, 'DNA polymorphisms amplified by arbitrary primers are useful as genetic markers', Nucleic Acid Research, vol. 18(22), pp. 6531-6535

Wolanska-Nowak, P., Branicki, W & Kupiec, T. 2001, 'STR data for AmpF/STR profile plus loci in south Poland', Forensic Science International, vol. 122, pp. 173-174.

Woo, K.M & Budowle, B. 1995, 'Korean Population Data on the PCR-based Loci LDLR, GYPA, HBGG, D7S8, GC, HLA – DQA1 and D1S80', Journal of Forensic Sciences, vol. 40, pp. 645-648.

Yamamoto, T., Mizutani, M., Uchihi, R., Tanaka, M., Yoshimoto, T., Misawa, S., Saitou, N., Katsumata, Y. 2003, 'Allele distributions and genetic relationship with 13 CODIS core STR loci in various Asian populations in or near Japan', International Congress Series, vol. 1239, pp. 117-120.

Zarrabeitia, M.T & Riancho, J.A. 2001, 'Population data on nine STRs from Cantabria, a mountainous region in northern Spain', Forensic Science International, vol. 122, pp. 175-177.

Zheng, X., Zeng, Z., Meng, X., Xiang, Z., Zhu, Y & Yan, H. 2007, 'Genetic data of 9 STR loci Henan province (Central China)', Forensic Science International, vol. 169, pp. 244-246.

Zhang, H., Li, Y., Jiang, J., Zhang, J., Wu, J., Du, H., Yan, J., Shen, Y., Hou, Y. 2007, 'Analysis of 15 STR loci in Chinese population from Sichuan in West China', Forensic Science International, vol. 171, pp. 222-225.

Websites

http://quincy.ca/resources.cfm - History of forensic science

http://www.troopers.stae.ny/forensic-science - Forensic science time line

http://www.gov/hq/lab/fsc/backissu/april2005 - India forensic database

http://www.sasas.co.za/Sajas.html –Micro satellite Markers

APPENDICES

Appendix 1: Chromosomal locations for the Identifiler STR markers and their corresponding dye labels.

Locus designation	Chromosome location	Alleles included in the Identifiler allelic ladder	Dye label	Control DNA 9947A
D8S1179		8 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	6 FAM	13, 13
D21S11	21q11.2 – q21	24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2,		30, 30
D7S820	7q11.21 - 22	6, 7, 8, 9, 10,11, 12, 13, 14, 15		10, 11
CSF1PO	5q33.3 - 34	6, 7, 8, 9, 10,11, 12, 13, 14, 15		10, 12
D3S1358	3p	12, 13, 14, 15, 16, 17, 18, 19	VIC	14, 15
THO1	11p15.5	7, 8, 9, 9.3, 10,11, 13, 13.3,		8, 9.3
D13S317	13q22 - 31	8, 9, 10,11, 12, 13, 14, 15		11, 11
D16S539	16q24 - qter	5, 8, 9, 10,11, 12, 13, 14, 15		11, 12
D2S1338	2q35 – 37.1	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28		19, 23
D19S433	19q12 – 13.1	9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 17, 17.2	NED	14, 15
vWA	12q12 - pter	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24		17, 18
TPOX	2q23 – 2per	6, 7, 8, 9, 10, 11, 12, 13	NED	8, 8
D18S51	18q21.3	7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27		15, 19
Amelogenin	X: p22.1 – 22.3 Y: p11.2	X, Y	PET	X, X
D5S818	5q21 - 31	7, 8, 9, 10, 11, 12, 13, 14, 15, 16,		11, 11
FGA	4q28	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2		23, 24

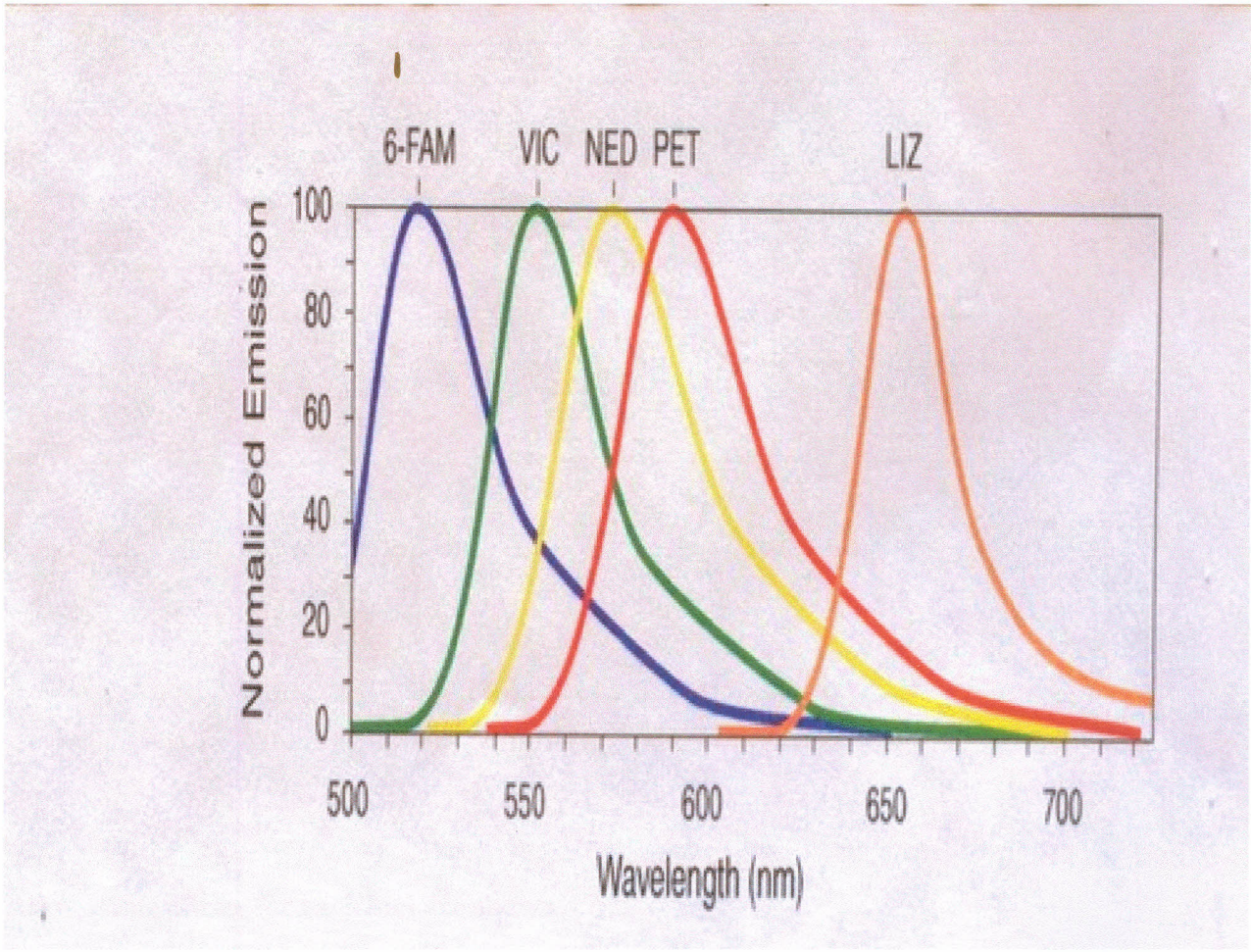
Appendix 2: 16AmpFlSTR Identifiler kit allele frequencies for African American Population (n = 357)

Sample	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	THO1	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
5	-	-	-	-	-	0.0028	-	-	-	-	-	-	-	-	-
6	-	-	-		-	0.1106	-	-	-	-	-	0.0672	-	-	-
7	-	-	0.0042	0.0462	-	0.4286	-	-	-	-	-	0.0224	-	0.0014	-
8	0.0042	-	0.1877	0.0756	-	0.2073	0.0308	0.0322	-	-	-	0.3613	-	0.5460	-
8.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	0.0042	-	0.1373	0.0378	-	0.1232	0.0252	0.1905	-	-	-	0.2115	0.0014	0.0168	-
9.3	-	-	-	-	0.0042	0.1162	-	-	-	-	-	-	-	-	-
10	0.0238	-	0.3445	0.2787	-	0.0098	0.0378	0.1092	-	0.0154	-	0.0924	0.0028	0.0672	-
10.2	-	-	-	-	-	-	-	-	-	-	-	-	0.0014	-	-
11	0.0392	-	0.1989	0.2059	-	-	0.2451	0.3151	-	0.0714	0.0028	0.2115	0.0028	0.2549	-
11.2	-	-	-	-	-	-	-	-	-	0.0014	-	-	-	-	-
11.3	-	-	-	0.0014	-	-	-	-	-	-	-	-	-	-	-
12	0.1331	-	0.1078	0.2913	0.0056	-	0.4622	0.1877	-	0.1078	-	0.0308	0.0700	0.3641	--
12.2	-	-	-	-	-	-	-	-	-	0.0630	-	-	-	-	-
13	0.2325	-	0.0154	0.0532	0.0070	-	0.1541	0.1485	-	0.2983	0.0126	-	0.0434	0.2157	-
13.2	-	-	-	-	-	-	-	-	-	0.0574	-	-	0.0042	-	-
13.3	-	-	-	-	-	0.0014	-	-	-	-	-	-	-	-	-
14	0.3011	-	0.0042	0.0098	0.1204	-	0.0434	0.0154	-	0.2101	0.0714	-	0.0686	0.0238	-
14.2	-	-	-	-	-	-	-	-	-	0.0420	-	-	0.0028	-	-
15	0.2017	-	-	-	0.3053	-	0.0014	0.0014	0.0014	0.0476	0.2003	-	0.1947	-	-
15.2	-	-	-	-	0.0014	-	-	-	-	0.0336	-	-	-	-	-
16	0.0462	-	-	-	0.2857	-	-	-	0.0532	0.0238	0.2675	-	0.1653	-	-
16.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0014
16.2	-	-	-	-	-	-	-	-	-	0.0238	-	-	-	-	-
17	0.0112	-	-	-	0.1947	-	-	-	0.1078	-	0.2059	-	0.1821	0.0014	-
17.2	-	-	-	-	-	-	-	-	-	0.0028	-	-	-	-	0.0014
18	0.0028	-	-	-	0.0672	-	-	-	0.0560	-	0.1471	-	0.1190	-	0.0070
18.2	-	-	-	-	-	-	-	-	-	0.0014	-	-	-	-	0.0140
19	-	-	-	-	0.0084	-	-	-	0.1415	-	0.0672	-	0.0602	-	0.0672
19.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0028
20	-	-	-	-	-	-	-	-	0.0602	-	0.0196	-	0.0490	-	0.0700
20.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-	0.1401	-	0.0028	-	0.0210	-	0.1289
21.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	0.1317	-	0.0028	-	0.0070	-	0.2157
22.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0028
22.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0014
23	-	-	-	-	-	-	-	-	0.1078	-	-	-	0.0042	-	0.1499
															0.0014

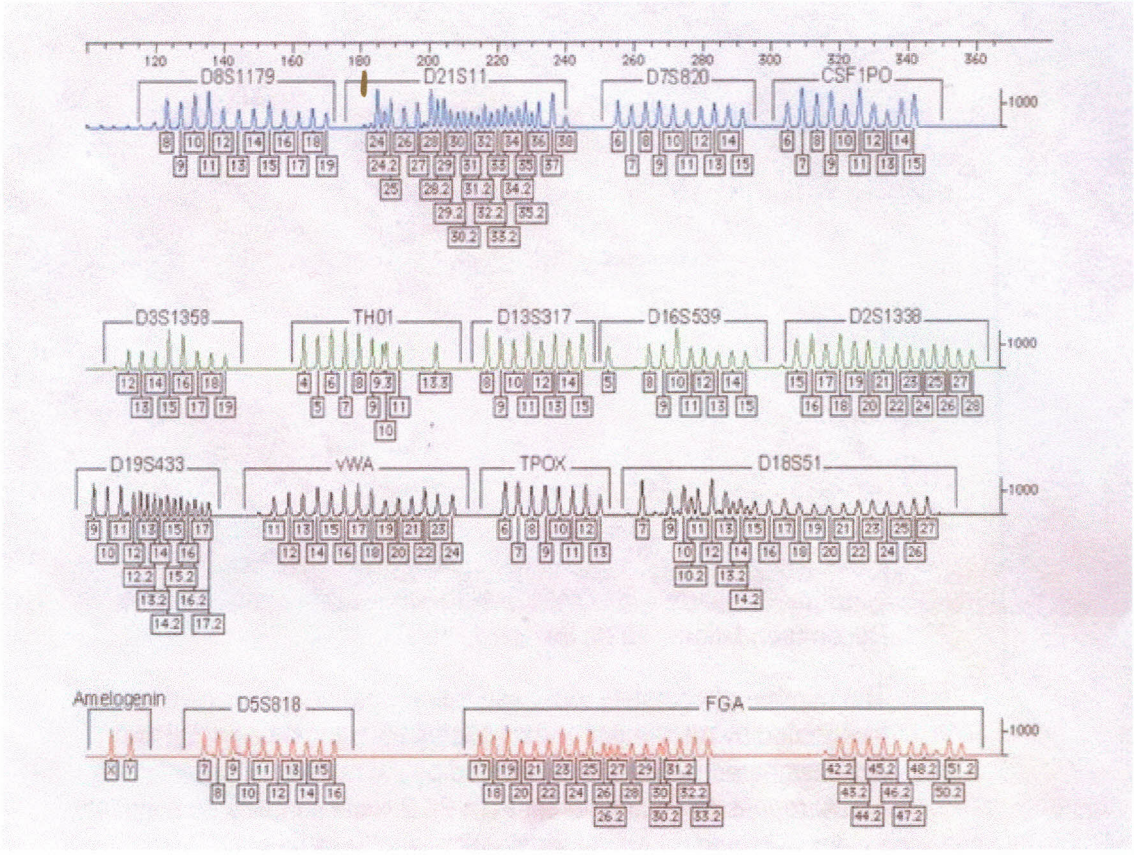
24	-	-	-	-	-	-	-	-	0.0980	-	-	-	-	-	0.1751
24.2	-	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-
24.3	-	0.0028	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	-	-	-	-	-	-	0.0812	-	-	-	-	-	0.0798
25.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	-	0.0014	-	-	-	-	-	-	0.0196	-	-	-	-	-	0.0350
26.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27	-	0.00504	-	-	-	-	-	-	0.0014	-	-	-	-	-	0.0182
28	-	0.2297	-	-	-	-	-	-	-	-	-	-	-	-	0.0140
28.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	0.1933	-	-	-	-	-	-	-	-	-	-	-	-	0.0056
29.2	-	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-
29.3	-	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	0.1723	-	-	-	-	-	-	-	-	-	-	-	-	-
30.2	-	0.0140	-	-	-	-	-	-	-	-	-	-	-	-	0.0014
31	-	0.0798	-	-	-	-	-	-	-	-	-	-	-	-	-
31.2	-	0.0798	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	0.0112	-	-	-	-	-	-	-	-	-	-	-	-	-
32.2	-	0.588	-	-	-	-	-	-	-	-	-	-	-	-	-
33	-	0.0056	-	-	-	-	-	-	-	-	-	-	-	-	-
33.2	-	0.0378	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	0.0126	-	-	-	-	-	-	-	-	-	-	-	-	-
34.1	-	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-
34.2	-	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	0.0014
35	-	0.0294	-	-	-	-	-	-	-	-	-	-	-	-	-
35.1	-	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-
35.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
36	-	0.0084	-	-	-	-	-	-	-	-	-	-	-	-	-
37	-	0.0028	-	-	-	-	-	-	-	-	-	-	-	-	-
38	-	0.0014	-	-	-	-	-	-	-	-	-	-	-	-	-
39	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
42.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
43.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
44	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
44.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0028
45	-	-	-	--	-	-	-	-	-	-	-	-	-	-	-
45.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0014

47	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
48	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
48.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0014
49	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
50.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
51.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Appendix 3: Emission spectra of five dye labels used in the AmpF1STR Identifiler Primers.

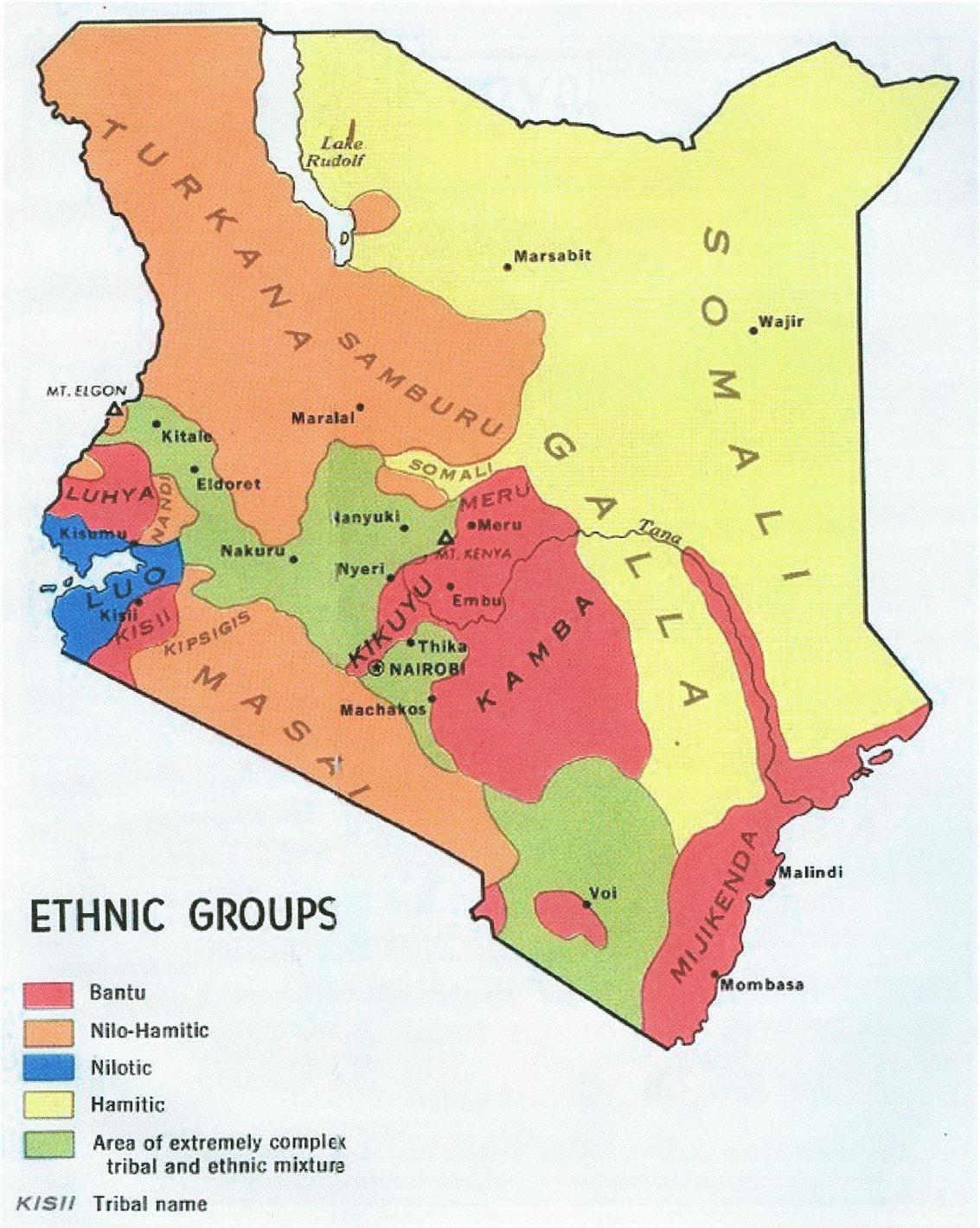


Appendix 4: Genotyper software plot of AmpF1STR Identifiler Allele ladder



All possible alleles for each loci are shown.

Appendix 5: Map of Kenya, Ethnic distribution



Appendix 6: Map of Kenya, Provinces