STUDY OF MITOCHONDRIAL DNA VARIABILITYIN FOUR ETHNIC GROUPS WITHIN THE SOUTHERN PART OF KADUNA STATE NIGERIA.

BY

JAMES ABRAK TIMBUAK

DEPARTMENT OF ANATOMY
AHMADU BELLO UNUIVERSITY, ZARIA
NIGERIA.

JUNE, 2015

STUDY OF MITOCHONDRIAL DNA VARIABILITY IN FOUR ETHNIC GROUPS WITHIN THE SOUTHERN PART OF KADUNA STATE, NIGERIA.

BY

James AbrakTIMBUAK, B.Sc.2000(A.B.U.), M.Sc.2008(A.B.U.)

Ph.D/MED/01123/2008-2009

ADISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATESTUDIES, AHMADU BELLO UNIVERSITY, ZARIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D.) IN HUMAN ANATOMY.

DEPARTMENT OF HUMAN ANATOMY,
FACULTY OF MEDICINE
AHMADU BELLO UNIVERSITY, ZARIA
NIGERIA

JUNE, 2015

DECLARATION

I declare that the work in this dissertation entitled, "STUDY OF MITOCHONDRIAL DNA VARIABILITYIN FOUR ETHNIC GROUPS WITHIN THE SOUTHERN PART OF KADUNA STATE, NIGERIA"has been carried out by me in the Department of Human Anatomy, Faculty of Medicine.

The information derived from the literature has been duly acknowledged in the text and a list of references provided. No part of this Dissertation was previously presented for another degree or diploma at this or any other Institution.

__________________________________ _______________________ _______________

Name of StudentSignature Date

CERTIFICATION

This dissertation entitled "STUDY OF MITOCHONDRIAL DNA VARIABILITY IN FOUR ETHNIC GROUPS WITHIN THE SOUTHERN PART OF KADUNA STATE, NIGERIA" by James Abrak TIMBUAK, meets the regulations governing the award of the of Doctor of Philosophy (Ph.D)degree of the Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

Prof.S.S.Adebisi*B.Sc., M.Sc. ,Ph.D*______________________________
Chairman Supervisory CommitteeSignature Date
Department of Anatomy
Faculty of Medicine
Ahmadu Bello Univeristy, Zaria

Dr. B. Danborno*B.Sc., M.Sc.,Ph.D* ______________________________
Member Supervisory Committee Signature Date
Department of Anatomy
Faculty of Medicine
Ahmadu Bello University Zaria

Prof. A.J. Nok *B.Sc., M.Sc., Ph.D., MFR*______________________________
Member Supervisory Committee Signature Date
Department of Biochemistry
Faculty of Medicine
Ahmadu Bello University, Zaria

Prof. S.S.Adebisi*B.Sc., M.Sc. Ph.D*____________________________
Head of Department Signature Date
Department of Anatomy
Faculty of Medicine
Ahmadu Bello University, Zaria

Prof. A.Z. HASSAN *DVM, M.Sc. Ph.D*______________________________
Dean, School of Postgraduate studies SignatureDate
Ahmadu Bello University, Zaria

DEDICATION

I dedicate this work to the loving memory of our son Ethan Bisan, rest on in the loving bosom of our Lord.

ACKNOWLEDGMENTS

I give thanks to God Almighty who made all this possible. I also wish to thank all the participants who provided samples for this research, without you this research would not have been possible.

I would like to thank all the members of my supervisory committee namely, Prof. SS Adebisi, Dr. B. Danborno and Prof. A.J. Nok for their commitment, guidance, encouragement and patience.

Much appreciation goes to the Tertiary Education Trust Fund (TETFUND) and the management of Ahmadu Bello University for providing part of the funds for this research. I would like to thank Prof. M.H. Crawford of the Laboratory of Biological Anthropology (LBA), Kansas, through the connection established by Dr B. Danbornofor giving me the opportunity to experience molecular genetics on hands on basis. I also wish to thank other members (past and present) of the LBA Dr. A. Justice, Stephen Johnson, Kristie Beaty and Orion Graf, you all contributed to my rich laboratory experience.

I wish to appreciate Dr. Pat Williams and Torrey Parish, of Evogen Inc., who made their facility available to me without charge. I am also extending my gratitude to Dr. Sarah Tishkoff and JibrilHirbo for their advice and I wish to thank my sister Mrs. Alheri Timbuak Bala and lateMr. Tauna for their invaluable assistance throughout the sample collection period.

I also wish to thank my dearest wife Mrs. Rose Sharon A. Timbuakwho not only brought my long search to an end but also went the “whole nine yards” withme through the nights and days of proof reading this work. I will not forget my lovely son Ethan B. Timbuak who has broughtme so much joy and hope.

I am also expressing my gratitude to Dr. (Mrs)Danborno and her children for providing me with a support system far from home and for giving me the rare experience of a tiny bit of family responsibility. Be assured all the lessons learnt are safely tucked away for the inevitable future.

I also wish to thank my senior colleaguesDr. U.E Umana and Dr. W.O.Hammanfor their friendship and support.Many thanks go to my Parents, Mr. and Mrs. Timbuakfor the joy and travails of parenting me. My siblingsAlheri, John, Yamai, Dinatu, Timo and Priscilla, for always being there and knowing I just cannot ask for a finer set of siblings. I wish to thank my colleaguestoo for the contributions they towards the completion of this study. I will not forget to mention the staff of the department of Human Anatomy and all those who assisted me in one way or the other urged me on in this endeavor.

I really do appreciate you all.

ABSTRACT

The middle-belt of Nigeria is known for its ethnic diversity and linguistic complexity. Studies on biological variation within the region have been mostly based on anthropometry but with recent advances in anthropological genetics, newer tools now provide greater resolution on human variation. This study examined the efficacy of the mitochondrial DNA (mtDNA) as a genetic marker to characterize the genetic structure of four ethnic groups of Benue-Congo affiliations from Kaduna State. Column and propriety salting based methods were used to extract mtDNA hypervariable segment-I (HVS-I) sequences from samples belonging to four ethnic groups, the Bajju, Chawai, Atyap and Kagoro. Sequences were amplified and amplicons purified using ExoSap. Sequencing for the light strand was done followed by sequence alignment, restriction fragment length polymorphism (RFLP) andsingle nucleotide polymorphism (SNP) analysis. Nucleotide positions 16050-16460 were compared to the revised Cambridge Reference Sequence (rCRS) and 91 haplotypes were observed. A total of 107 polymorphic sites characterized the haplotypes. The African specific *HpaI*cut site at 3592 defined the L1 and L2 haplotypes which were most frequent but absent for the L0 and L3 haplotypes. Subclade L3e had the highest frequency while other sub clades of the sub Saharan haplogroups werealso present across the study populations in appreciable frequencies, indicative of substantial gene flow between them and other neighbouring populations.A fewsamples, however failed to cluster with the majority as they lacked SNPs belonging to the region and were merely identified as Non-L haplogroups.The Nucleotide diversities (π) were 0.019, 0.026, 0.025 and 0.020 for the Atyap, Bajju, Chawai and Kagoro respectively. The haplotype diversities (HD) were

high and consistent with the regional and overall African values, with the Atyap having the leastdiverse value (0.960) while the Bajju had the most diverse haplotypes(0.992). Sub-clade analysis based on L0, L1, L2 and L3, for haplotypediversity (HD) and nucleotide diversities, π, exhibited greater diversity for L3 and L2 while L0 had the least diversity. Mismatch distributions for the major haplogroups showed stable demographic patternsfor L0, L1 and L2 but the more recent L3 clade exhibited an expansion pattern as expected. The same expanding demographic was observed for the Atyap, Chawai and Kagoro with the Bajju having a more stable population. Partitioning the genetic variation using the linguistic group model using the analysis of molecular variance (AMOVA) revealed little variation among the populations (3.66%) but showed a high level of variation (94.6 %) within each population.This study has revealed the presence of a shared genetic structure among the Atyap, Bajju, Chawai and Kagorousing molecular markers, which is indicative of close genetic relationship due to common history, substantial gene flowand geographical proximity.

TABLE OF CONTENTS

Cover Page i

Title page iii

Declaration iv

Certification iii

Dedication v

Acknowledgements vii

Abstract viii

Table of Contents .... ix-xx

List of Abbreviations xxi

CHAPTER ONE

1.0 Introduction 1

1.1 Background of Study 1

1.2 Statement of research the Problem 4

1.3 Justification of the Study 4

1.4 Aims of the Study 5

1.5 Objectives of the Study 5

1.6 Hypotheses of the Study 6

CHAPTER TWO

2.0 Literature Review 7

2.1 Origins of Man and Anthropological Genetics 7

2.2 Biological Sources of Human Variation..11

2.2.1 Base Substitutions..12

2.2.2 Insertions and Deletions 13

2.3 Linguistic Affiliations 13

2.3.1 The Niger-Kordofanian family 16

2.3.2 Benue-Congo 19

2.4 Historical Background of the Bajju, Atyap, Chawai and Kagoro 21

2.4.1The Chawai (Atsam)...23

2.4.2 The Bajju 25

2.4.3 The Atyap (Kataf)...25

2.4.4 Kagoro (Oegworok) 27

2.5 Genetic Markers 29

2.5.1ClassicalMarkers 29

2.5.2 Molecular Markers..31

2.5.3 Autosomal Markers 33

2.6 Y-Chromosome 33

2.7 Mitochondrial DNA 34

2.8 History of Mitochondria..39

2.9 Biology of the Mitochondria 41

2.9.1 Replication..41

2.9.2 Energy Generation..43

2.10 mtDNA Variation in Humans...44

2.11 DNA sequencing..47

2.12 Polymerase Chain Reaction (PCR) 49

2.13 Gel Electrophoresis (Agarose) 50

2.14 Restriction Fragment Length Polymorphisms (RFLPS) 51

2.15 Phylogenetics 53

2.16 Networks...57

CHAPTERTHREE

3.0 Materials and Method 60

3.1 Subject Sampling 60

3.1.1 Comparative Populations...61

3.1.2 Ethical Approval..61

3.1.3 Comparative Populations...61

3.2 Buccal Cell Collection....63

3.3 DNA Extraction and Polymerization....63

3.4 Amplification (Polymerase Chain Reaction)....67

3.5 Electrophoresis Procedure....70

3.5.1 Preparing the Gel....70

3.5.2 Running the Gel....71

3.6 Purification of PCR Product and Sequencing....71

3.7 HVS-I Sequencing....73

3.8 Restriction Fragment Length Polymorphism (RFLP)....73

3.9 Precautionary Measures....76

3.10 Haplogroup Assignment....77

3.11 Data Analysis....77

3.11.1 Genetic Diversity....77

3.11.2 Fst Distance Matrix....78

3.11.3 Intra-population Diversity Measures....78

3.11.4 Neutrality Test....78

3.11.5 Interpopulation Diversity Measures (Networks)....78

3.11.6 Mismatch Distribution....79

3.11.7 Analysis of Molecular Variance(AMOVA)....79

3.11.8 Phylogenetic Analysis 80

3.11.9 Principal Component Analysis (PCA) 80

CHAPTER FOUR

4.0 Result...82

4.1 DNA Extraction...82

4.2 HVS-I Sequencing and Gel Electrophoresis...84

4.3 HVS-I Sequences and Haplogroup Characterization...87

4.4 Diversity Indices..97

4.5 Analysis of Molecular Variance (AMOVA) 102

4.6 Neighbour-Joining (NJ) Tree...107

4.7 Mismatch Analysis...108

4.8 Principal Component Analysis 110

4.9 Median-Joining network 120

CHAPTER FIVE

5.0 Discussions...122

CHAPTER SIX

6.0 Summary, Conclusion and Recommendation...134

6.1 Summary...134

6.2 Conclusions......135

6.3 Recommendations......135

References...... 137

LIST OF TABLES

Table 3.1 Restriction Enzymes, Recognition Sequences and their Cut Sites 75

Table 4.1 Method Comparison for Extraction time, Final volume and DNA yield 83

Table 4.2 Haplotypes and Sequence Segregating Sites 90

Table 4.3 Sequence Polymorphisms 92

Table 4.4: Major African mtDNA L clades' frequencies 94

Table 4.5 Descriptive data of the HVS-I Polymorphisms 98

Table 4.6 HVS I Diversity Indices...99

Table 4.7 HVS-I diversity Indices for the major L sub-haplogroups 100

Table 4.8 Diversity Indices of Comparative Populations 101

Table 4.9: Results of the AMOVA Analysis with Populations 103

Table 4.10: Genetic distance matrix 104

Table 4.11: Neutrality test Scores for mtDNA HVS-I Sequence Data............................105

LIST OF FIGURES

Figure 2.1: Models of Human Evolution 10

Figure 2.2: Distribution of the Major African Language Families and Migrations 15

Figure 2.3: The Niger-Kordofanian Language Tree 18

Figure 2.4: Revised Classification of the Niger-Congo Subfamily 20

Figure 2.5: Map of Kauru LGA Showing Geographic Region of the Chawai 24

Figure 2.6: Map of ZangonKataf LGA Showing Bajju and Atyap Settlements 26

Figure 2.7: Map of Kaura LGA Showing the Geographic Spread of the Kagoro 28

Figure 2.8: Phlogenetic mtDNA tree from the World's Geographic Regions 32

Figure 2.9: Y Chromosome 35

Figure 2.10: The Human Mitochondrium 36

Figure 2.11: Mitochondrial DNA 38

Figure 2.12: D-Loop Formation in mtDNA 42

Figure 2.13: Sanger's Sequencing Method 48

Figure 2.14: Sample Sequence Electropherogram 49

Figure 2.15: A Phylogenetic Tree with a Common Ancestor 55

Figure2.16: Representative Phylogenetic Tree 57

Figure 3.1: Map of Kaduna Showing Study Area 62

Figure 3.2: Steps in QIAamp (Qiagen) Extraction 65

Figure 3.3: Quantitation of DNA 66

Figure 3.4: Thermal Profile for HVS-I Amplification 69

Figure 3.5: Electrophoresis Tank with the Loaded Gel 72

Figure 3.6: ABI® PRISM™ 3730 Genetic Analyzer 74

Figure 4.1: A260/A280 Absorbance Ratio 85

Figure 4.2: Representative mtDNA Sequence Chromatogram 88

Figure 4.3: BioEdit Window of Aligned HVS-I Sequences 89

Figure 4.4: Haplogroup Assignment of Individuals 96

Figure 4.5: Neighbour-Joining tree (NJ) 106

Figure 4.6: L0 Haplotype Mismatch Distribution 110

Figure 4.7: L1 Haplotype Mismatch Distribution 111

Figure4.8: L2.Haplotype.Mismatch.Distribution 112

Figure.4.9: L3.Haplotype.Mismatch.Distribution 113

Figure 4.10: Atyap Population Mismatch Distribution 114

Figure4.11: Bajju Population Mismatch Distribution 115

Figure 4.12: Chawai Population Mismatch Distribution 116

Figure 4.13: Kagoro Population Mismatch Distribution 117

Figure 4.14: Principal Component Analysis 118

Figure 4.15 Median-Network of Haplogroups Based on HVS-I Sequences 120

LIST OF PLATES

Plate I: PCR Amplicons Visualized on a 2% Agarose Gel..86

LIST OF APPENDICES

Appendix I- Consent form 158

Appendix II-ABUTH Ethical committee clearance form 159

Appendix III-Human Subject committee Lawrence Campus, University of Kansas.
USA. ethical clearance 160

Appendix IV - PCR Amplification (HVS1_mini DNA) protocol 161

Appendix V - Gel (RFLP) Digest protocol...162

Appendix VI - Exosappurification protocol 163

LIST OF ABBREVIATIONS

(r)CRS (revised) Cambridge Reference Sequence

(T)MRCA (time to the) Most Recent Common Ancestor

AMH Anatomically Modern Human

AMOVA, Analysis of Molecular Variance.

bp/kb(p)/Mb(p) base pair / thousand (kilo) base pairs / mega (million) base pairs

ca. circa, about

D-loop displacement loop/control region of mtDNA

DNA deoxyribonucleic acid

Dpn*Diplococcuspneumoniae*

haplotype a sequence type that comprises all identical sequences

haplogroup a group of haplotypes that share a common ancestor defined by specific polymorphisms

HpaI*Haemophilus parainfluinzae*

HVS-I/HVS-II first/second hypervariable segment of mtDNA

Indel(s) polymorphism of insertion-deletion

ky(a) thousand/kilo years (ago)

MboI*Moraxella bovis*

mtDNA, mitochondrial DNA;

my(a) million years (ago)

np(s) nucleotide position(s)

PCA, Principal Components Analysis;

PCR Polymerase Chain Reaction

RFLP Restriction Fragment Length Polymorphism

SNP(s) Single Nucleotide Polymorphism(s)

STR(s) Short Tandem Repeat(s)

CHAPTER ONE

1.0 INTRODUCTION

1.1 BACKGROUND OF STUDY

Linguistic inclinations, anthropometric records and geographical background of human populations are known to provide a historical basis for human evolution and variation, as well as the reasons underlying such changes (Tishkoff *et al.,* 2007; Adebisi, 2008). The diversity exhibited by *Homo sapiens* arose during their processes of dispersal into their present regions, thus, the subsets of variation tend to be associated with particular geographic areas and populations (Rosa *et al.,* 2004). Human variation has been measured using simple visual characters like size, form and skin colour leading to the conclusion that sharing one or more features is an indication of common descent (Molnar, 1998), but advances in the science of genetics have revealed greater distinctions (Relethford, 1990). Thus, anthropological genetics, a comparatively new discipline makes attempt at answering questions that concern human origin and variation using methods and theories of genetics (Crawford, 2007).

The mapping of the human genome and the emergence of technologies has made it possible to identify variation at the level of the individual (Underhill *et al.,* 2000) and more recently, the haploid characteristics and of some genetic markers allow the successful application of phylogenetic and phylogeographic approaches to population genetics (Ennafaa *et al.,* 2009). This has moved the focus to more detailed methods like the genome wide characterization of natural variation (Thangaraj *et al.,* 2005). These advances in

human genetics have supplied researchers with tools for the investigation of the variation between language groups and haploid markers such as the mitochondrial DNA (mtDNA) and non-recombining region of the Y-chromosome (NRY) (Hurles *et al.,* 2002; Wood *et al.,* 2005). The focus of these studies has ranged from explorations of disease factors to historically focused research on the genetic relations between African populations (MacEachern, 2006). A striking factor in many of these researches is the search for relationship between populations within Africa with that elsewhere. Some of these studies are germane to the considerations of ancient relations and migrations which indicate strong correlations between genetic and linguistic relationships among globally distributed human populations (Chen, 1995).

From an evolutionary view point, such relationships spread across the world's genetic map have led to efforts on illuminating the origin and dispersal of anatomically modern man across the world (Saccone *et al.,* 1992). Postulations based on developed models have shown different origins for man, out of which the Recent African origin (RAO) also known as the "out of Africa" model asserts a common descent for all populations from an anatomically modern Homo sapiens ancestor (Ramachandran *et al.,* 2005; Relethford 2008). This makes the African continent, the ancestral home of all humans today.

Reconstructing the history of the West African population is considered complex (Rosa *et al*., 2004), this is due to short and long migration events within the region (Tishkoff and Williams, 2002). One of the earliest indications of West Atlantic occupation by modern humans goes back to about 40 KYA (Wood *et al*., 2005), but subsequent changes in the climatic conditions resulted in the significant movement of these occupants (Aumassip *et al*., 1994).

The dearth of archeological evidence (attributable to differences in sea level which may have buried such artifacts) for reconstructing the past has also contributed to this situation, as such language groups were used to genetically classify populations into groups of common descent (Brown and Ogilvie, 2009). This is evident in the fact that cultures may spread without attendant spread of genes, but languages are not easily acquired in later life than other cultural transformations (Bellwood, 2001). The recent development of genetic tools has proven that linguistic groups within Africa share common gene pools (Excoffier *etal.,* 1992)which have become useful in probing phylogenies. The genetic variation in modern man occurred during the events of early migration into new territories, with concomitant localization of these variations to particular regions (Ingman *et al.,* 2000; Atkinson *et al.,* 2009). Studies are ongoing to understand the past events involving population expansion, contraction, genetic drift and substructure. Some of these studies employ genetic methods to probe the human genome to investigate and analyse single nucleotide polymorphisms (SNPs) in conjunction with restriction fragment length polymorphism (RFLP) techniques obtained from the hypervariable region of the d-loop (Chen *et al.,* 1995, 2000; Salas *et al.*, 2002, 2004). These studies have demonstrated that human mtDNA is geographically structured and may be classified into groups of related haplotypes (Chen *et al.,* 1995; Wallace *et al.,* 2007).

All these approaches attempt to provide the historical perspectives of genetic lineages by using various human population groups as the focus of their investigations. It is estimated that more than 2,000 distinct ethnic groups and languages are spoken in Africa (www.ethnologue.com, 2013); however, they belong to comparatively few language families. Studies have shown extensive genetic diversity among geographically close

African populations. Some 50 African languages have more than half a million speakers each, but many others are spoken by relatively few people. Yet, many studies rely on very few populations within Africa as a representation of the diversity within it (Tishkoff and Williams, 2007).

1.2 STATEMENT OF THE PROBLEM

Human variation based on features under selection, often lead to spurious inferences. The push is now towards understanding variability founded on available non-recombining genetic systems, which represent a fundamental part of a population's evolutionary history. Thus the focus on the Atyap, Bajju, Chawai and Kagoro located within the ethnically diverse but little studied Middle-belt region. These populations are described as a homogenous population due to shared oral traditions and proximity in linguistics, culture and geography. But this nondescript lumping together has been rejected by these groups as they see themselves as ethnically distinct. This study, however, is undertaken to apply mtDNA markers in testing the nature of relationship within and among these groups and to also establish if there is substantial gene flow between them.

1.3 JUSTIFICATION OF THE STUDY

Africa has played a principal role in the origin of diverse human populations. Therefore, understanding the patterns of genetic variation and the demographic history of populations within Africa is important for understanding the demographic history of global human populations. Out of a vast array of population genetic studies, only a handful of studies

have been carried out within Africa. Relatively, Nigeria has recorded only a handful of studies on genetic variation at the level of its populations. Thus, there is the need to fill up the gap by carrying out more studies on the available subsets of ethnic populations. The present study aims at providing such needed information to add to existing data on the diverse genetic landscape in Nigeria, and also the African region as a whole.

1.4 AIMS OF STUDY

The study is aimed at the identification of the composition and distribution of inherited mtDNA haplogroups of four ethnic groups from Southern Kaduna area. It also seeks to corroborate their origins and to determine the basis for the shared ethnocultural similarities between them as well as neighbouring populations within the Niger-Kordofanian construct.

1.5 OBJECTIVES OF STUDY

The objectives of the study are to

i. Isolate mitochondrial hypervariable segment I (HVS-I) sequences from sampled populations of the Bajju, Atyap, Chawai and Gworok ethnic groups.

ii. Sequence the hypervariable segment I (HVS-I) of each individual with respect to their ethnic group.

iii. Identify and compare polymorphisms in these ethnic groups using the revised Cambridge reference sequence (rCRS).

iv. Assign each individual to a specific haplogroup based on confirmation from characteristic restriction fragment length polymorphisms (RFLP) and single nucleotide polymorphism (SNP) markers

v. Characterize the genetic structure within each study ethnic population using the haplogroup assignments.

vi. Investigate the genetic variability of the study populations which may explain their evolution into different ethnic groups.

vii. Compare the obtained sequences from this study to those in previously published works on Africans.

1.6 HYPOTHESES OF STUDY

The following hypothesis would be tested using statistiical variability measures.

i. mtDNA haplogroups markers will indicate genetic structure within each ethnic populations.

ii. There is shared genetic structure between Bajju, Atyap, Chawai and Gworok due to geographical proximity

iii. There is a statistically significant relationship between language family, territory and genetics across the study populations.

iv. There is gene flow between the study populations and other comparative African populations.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 ORIGINS OF MAN AND ANTHROPOLOGICAL GENETICS

While interactions between biology and culture defines the field of biological anthropology, anthropological genetics is an amalgamated discipline with its origins drawn from developments in evolutionary biology, population genetics and biological anthropology (Crawford, 2007; Gonder *et al.,* 2007; Omkar *et al.,* 2009). It is concerned with understanding human evolution through genetic and biological variation, from which it also provides tools for reconstructing population history (Marks, 2012). The discipline today is a dynamic one that has benefited from developments in molecular biology and the more recent field of bioinformatics (Lander *et al.,* 2001; Mardis 2008). This has made possible the availability of a vast array of genetic markers to aid genomic analysis in contrasts to the earlier approach where research in the field utilized only a limited number of classical polymorphisms.

Endeavours of the early biologists like Carl Linnaeus led to the discovery of evolutionary processes and man's descent as elucidated by Charles Darwin, Alfred Russell Wallace and Thomas Huxley and others (Stoneking, 2008). Based on fossil evidence they postulated the divergence of man from apes circa 5-7 million years ago (MYA). Subsequent evidence inferred from molecular structure by Goodman (1963) using qualitative immunological methods to compare proteins revealed a close relationship

between African apes and humans. Quantitative immunological methods were later used by Sarich and Wilson (1967) which also corroborated this relationship.

The first in-depth study of human mtDNA variation was carried out by Cann *et al.* (1987), where they used mtDNA polymorphisms to produce a tree portraying human ancestral relationships and how long it would take to accumulate these mutations as observed in populations today. The outcome implied an African source of all extant mtDNA variation which dated back to 150,000 YBP.This brought into limelight the concept of a common ancestor widely reported as the "mitochondrial Eve." The idea of an African origin for mitochondrial Eve which is the source of all present diversity, was not accepted readily until the successful extraction of mtDNA sequence from a Neanderthal fossil in 1997 and its subsequent comparison to those of modern humans which showed dissimilarity as predicted by the African origin hypothesis (Krause *et al.,* 2007; Stoneking 2008). Consequently, the mtDNA has been shown to be useful in tracing origins through its distinct geographical diversity and frequency in Africa (Rosa *et al.,* 2004; Reed and Tishkoff 2006; Campbell and Tishkoff 2008; Jakobsson *et al.,* 2008).

It is reasoned that the array of genetic variation in modern humans have been influenced by the history of our ancestors (Tishkoff and Gonder, 2007). This history has been dominated by debates with the proponents presenting alternate views on the origin of mankind. Based on archaeological, anthropological and genetic data three model hypothesis have been proposed (Stringer, 2002; Mellars, 2006).

The multiregional model as seen in Figure 2.1 holds it that after the migration of *Homo erectus* or *Homo ergaster* from Africa into areas of the “old world” such as the

European and Asian continents over one million years ago (MYA) (Nei 1995). There has been a gradual and continuous conversion through gene flow among regional populations from the indigenous *Homo erectus* to *Homo sapiens* or anatomically modern humans leading to the production of diverse modern human characteristics (Wolpoff *et al.,* 2000). A point of weakness in the assertions of this model is the requirement for an initial large effective population to sustain gene flow amongst the populations with an attendant high degree of diversity as against the known restricted population size of our ancestors (Harpending *et al.,* 1993; Tishkoff and Gonder, 2007). In contrast, the replacement theory popularly known as the Out of Africa (OOA) model depicted in Figure 2.1 has it that anatomically modern humans originated in Africa and then spread to the rest of the world within the past ~100,000 years (Stringer, 2002; Tishkoff and Verrelli, 2003). These humans replaced the hominids already existing in those continents. The assimilation or hybridization model (Figure 2.1) bridges the aforementioned models; it accepts the "Out of Africa theory" but further suggests the integration of archaic African genes and non-African populations to the modern gene pool (Stringer, 2002; Tishkoff and Verrelli, 2003). Thus modern humans may have evolved due to combination of characteristics.

Throughout time, the main molecular differentiation of anatomically modern humans occurred during their dispersal into different continents and regions, and therefore the subsets of variation tend to be associated to particular geographic areas and populations (Tishkoff *et al.,* 2009). The phylogeographic approach is applied to clarify past demographic phenomena such as range expansion, genetic drift (founder-effects and bottlenecks)

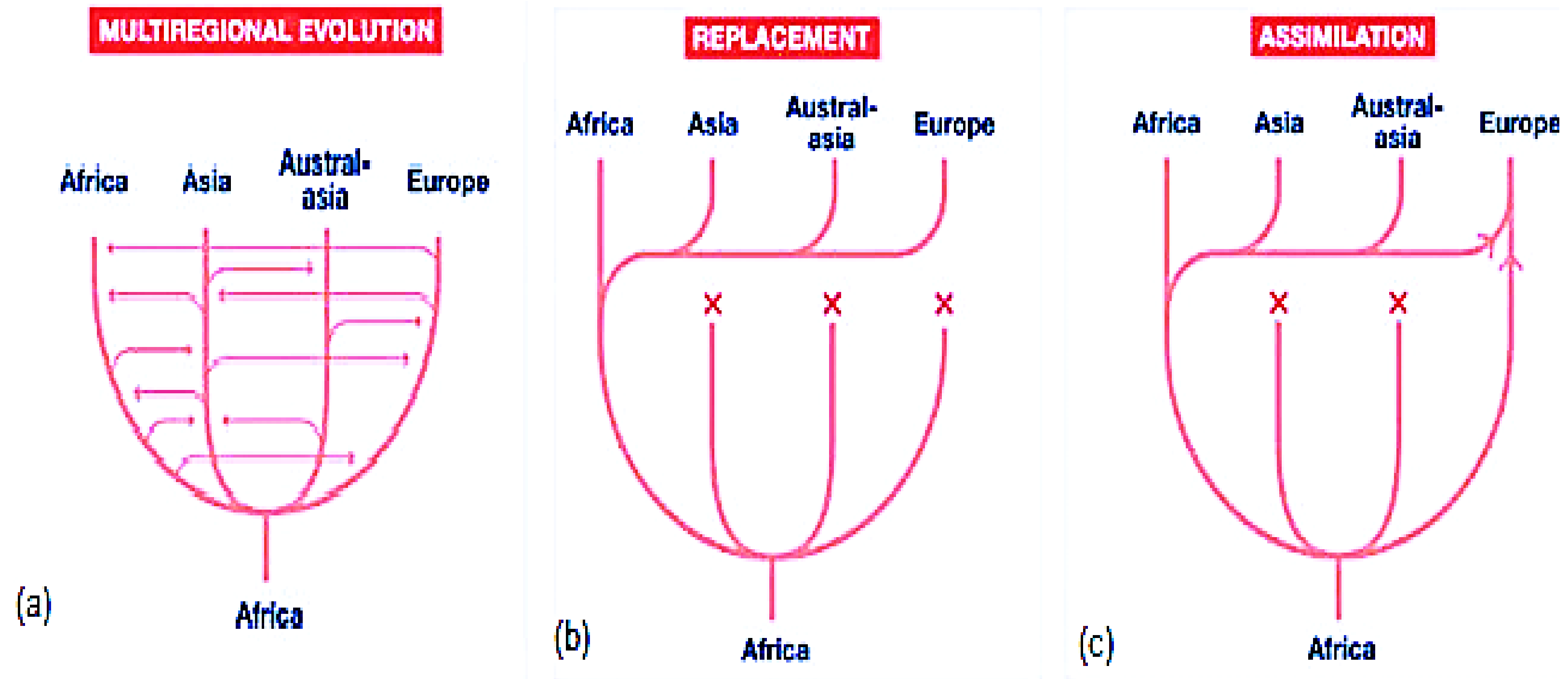


Figure 2.1: Models of human evolution. Adapted from Stoneking (2008)

and population subdivision. The lineage based approach attempts to unravel the history of genetic lineages of shared ancestry while the population-based approach focuses on the prehistory of individual populations, geographical regions or the changing pattern of population settlements (Pakendorf and Stoneking, 2005).

2.2 BIOLOGICAL SOURCES OF HUMAN VARIATION

Reproduction across the mammalian order shows similarity with that that in other mammals. It results in inheritance of characters from parents through the shuffling of genetic material. Within the mechanism through which this inheritance is acquired lies the source of the diversity exhibited by organisms (Molnar, 1998). The deoxyribonucleic acid (DNA) is the molecule through which progenies inherit genetic information from their parents. The success of this transmission lies in the ability of the DNA to reproduce itself through a process of cell division, known as mitosis. In humans this process does not result in the production of identical cells beyond the 16-cell stage because after this stage, the cells assumed heterogeneity and commit to differentiated embryonic stem cell lines also termed cellular differentiation (Mielke *et al.,* 2011).

In the contribution of gametes by the father and mother in the production of the diploid zygote, it is required that the germ cells go through a reduction division known as meiosis, that takes them from the diploid to the haploid state. Thus, meiosis does not produce identical copies because the parental nucleus is diploid, while the daughter nuclei must be haploid. The shuffling of maternal and paternal variation leads to additional genetic variation in offspring. The source of these variations which produce different alleles

at the same genetic locus is mutation. A mutation is defined as any random change in DNA sequence and it involves a range of events with different rates and different molecular mechanisms (Jobling *et al.,* 2004). Mutational events could arise as a result insertions, replacement, deletion or inversion of genetic systems. But genetic anthropology is mainly concerned with the followings: Base substitutions, Insertions and Deletions.

2.2.1 Base Substitutions

In a substitution mutation, a DNA nucleotide is replaced with another. When a purine is exchanged for another purine (Adenine (A) for Guanine (G) and vice versa) or a pyrimidine for another pyrimidine (Cytosine (C) for Thymine (T) and vice versa) it is called a transition. But when a purine is replaced with a pyrimidine or vice versa, it is called a transversion. If these changes involve a coding sequence, the base substitutions may result in a change in the amino acid. Consequently it may lead to a function change for the amino acid involved (Mielke *et al.,* 2011). Other mutations can change a nucleotide while having no effect on the coded amino acid because of redundancy in the genetic code, such mutations are referred to as "silent" mutations. Base substitutions arise from mis-incorporation of nucleotides during DNA replication and chemical modification of bases or physical damage from ionizing radiation (Jobling *et al.,* 2004). Single base-pair substitutions are among the most frequently encountered mutations in the human genome (Antonarakis *et al.,* 2005).

2.2.2 Insertions and Deletions

Insertions and deletions popularly known as indels, refer to the addition or deletion of a single base. Together with base substitutions they are collectively known as single nucleotide polymorphisms (SNPs). Another class of indels are those made up of short variable stretches of DNA sequences that may be inserted or deleted from a locus such as the “Alu insert” on the CD4 gene (Tishkoff *et al.,* 1996). If this occurs in a coding sequence it may result in the entire sequence being mistranslated. This is also referred to as a reading frame shift. Reading frame shift mutations are rare because they usually lead to a nonviable embryo (Mielke *et al.,* 2011).

2.3 LINGUISTIC AFFILIATIONS

Very little is known about primeval accounts of West Africa, but it is known that agriculture came to the area around the third millennium BC and as a consequence led to rapid population expansion (Gonzalez *et al.,* 2006). Studies have linked geographical affiliations to long-range migration of languages and the spread of agriculture (Williamson and Blench, 2000). Records of Sub-Saharan linguistic groups were dated between the tenth and twelfth centuries according to Arabic documents, though ancestral relationships of languages spoken in Africa are probably beyond linguistic reconstruction as languages evolve much faster than genes. Languages may even be lost and can, be replaced by newer ones. In the recent past so much had been put into the study and classification of the African languages, but it has been comprehensively worked out only in this century

(Ruhlen, 1991). Brown and Ogilvie (2009) proposed linguistic areas based on the following characteristics:

i. The number of languages spoken in one and the same general area.

ii. The sharing by these languages of a set of linguistic features whose presence can be explained with reference to neither genetic relationship, drift, universal constraints on language structure or language development, nor to chance.

iii. This set of features is exclusively found within languages in a specific area but not in languages outside the area.

iv. On account of (ii), the presence of these features must be the result of language contact.

The main language families of Africa as seen in Figure 2.2 are grouped into four phyla. They are the Afro-Asiatic; Niger-Kordofanian (including Niger-Congo); Nilo-Saharan; and Khoisan, or Click. Niger-Kordofanian and Nilo-Saharan are two large language families. The Niger-Kordofanian languages are found in most parts of southern Sub-Saharan Africa, while the Nilo-Saharan languages are distributed in Central and East-Central Africa except for Songhai which is spoken in Western Africa (Welmers, 1956). The Niger-Kordofanian and Nilo-Saharan families are thought to have a common descent. The Afro-Asiatic languages are spoken across most of the northern part of Africa and extend up to the Middle East. The Khoisan language groups are found in South Africa, Namibia, Botswana, Angola and northern Tanzania (Greenberg, 1954, 1963). This study population are derived from branches of the Niger-Kordofanian.

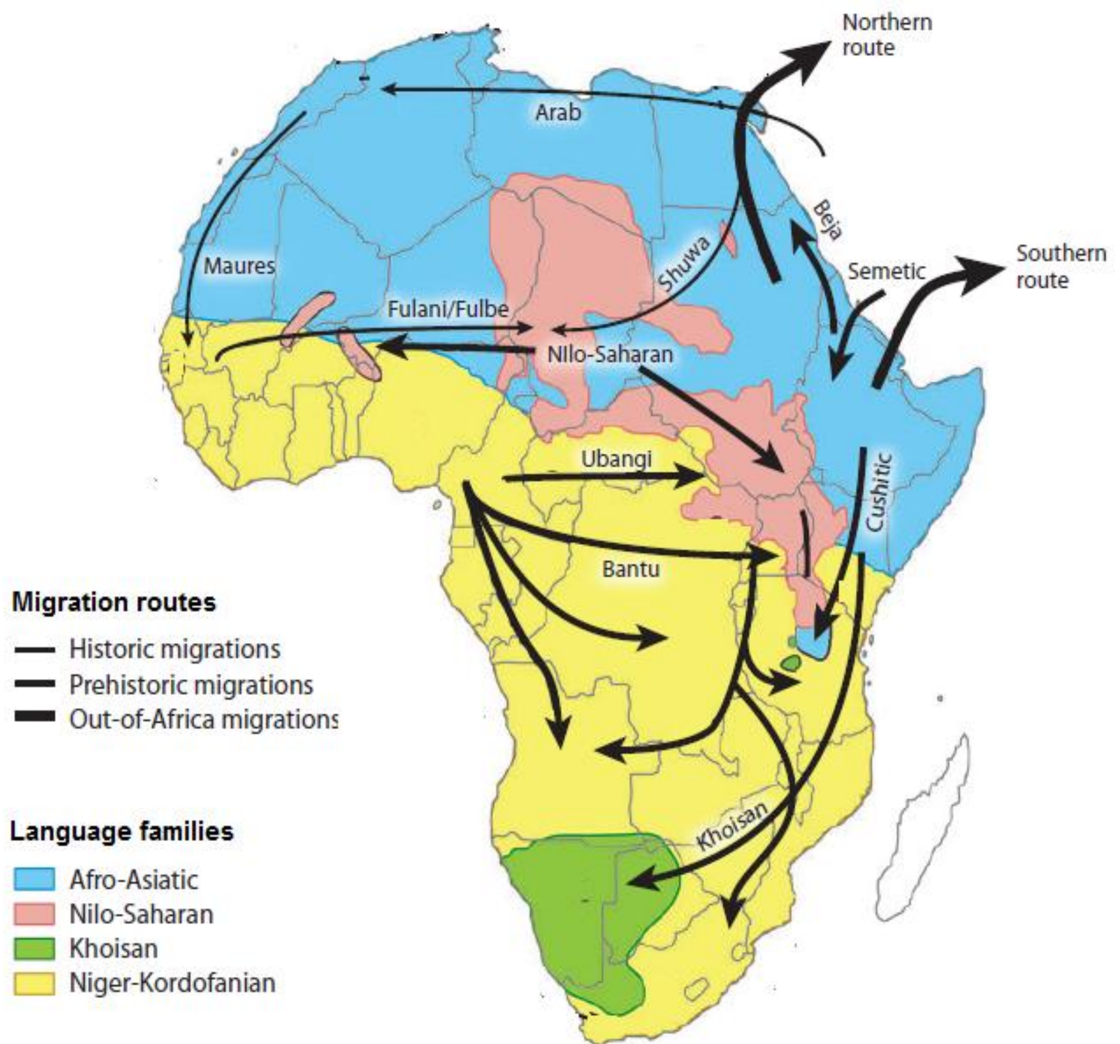


Figure 2.2: Distribution of the Major African Language Families and Migrations

2.3.1 The Niger-Kordofanian family

The Niger-Kordofanian family is split into two branches, the Niger-Congo and Kordofanian. Greenberg (1963) even suggested the similarity between these two families. Kordofanian is divided into four main groups: Heiban, Talodi, Rashad, and Katla. The Kordofanian group is made up of about thirty languages spoken in Sudan. The Niger-Congo is a large branch with more than a thousand languages and over 180 million speakers. Its languages are dispersed throughout South and Central Africa and in most of Sub Saharan Africa (Crabb 1974). Greenberg (1963) categorised the Niger-Congo family into six subfamilies as shown in Figure 2.3. These consist of:

a. West Atlantic (now known as Atlantic) which is scattered between Senegal to Sudan but mainly found along the Atlantic coast from the Senegal River to Liberia;

b. Mande, found from Senegal to Burkina Faso and Ivory Coast;

c. Gur, or Voltaic located in West Africa in a stretch running from Senegal to Cameroon;

d. Kwa stretching across Southern Ivory Coast, Ghana, Togo, Benin, and into the Southwest;

e. Benue-Congo which stretches from the Benin-Nigeria border across Nigeria Eastward to Kenya and Southward to the Cape; and

f. Adamawa-Eastern that lies north of the Bantu family in a belt across eastern Nigeria into the Sudan,

But Samarin (1971) suggested the use of the name 'Ubangi' instead of 'Eastern.'The Adamawa languages are found in Northern Nigeria, Cameroon, and Chad, whereas the Ubangi languages are spoken in the Central African Republic (CAR), Northern Zaire, and Southwestern Sudan (Brown and Ogilvie, 2009).

The West Atlantic branch includes languages such as Wolof (in Senegal), Temne (in Sierra Leone), and Fulani, widely spoken by people located in an area from Senegal to Lake Chad. The Mande group consists of languages prevalent in the Niger valley, Liberia, and Sierra Leone, such as Mende in Liberia and Malinke in Mali. Gur, or Voltaic, is made up of several language groups and includes Mossi, the dominant tongue of Burkina Faso, as well as the Dagomba and Mamprusi of Nothern Ghana (Bendor-Samuel, 1971). The Kwa languages, found mainly in Ghana, Côte d'Ivoire, Benin, Nigeria, and Liberia, include Ewe, Yoruba, Igbo, Nupe, Bini, Ashanti, and perhaps Ijo. Benue-Congo includes the huge Bantu group of hundreds of tongues found throughout Central and Southern Africa as well as such non-Bantu languages as Tiv, Jukun, and Efik, which are spoken in Nigeria and Cameroon. The Adamawa-Eastern branch, to which Banda, Zande, and Sango belong, is composed of a number of languages spoken in Nigeria, Cameroon, and an area north of the Bantu territory to Sudan (Gregersen, 1977).

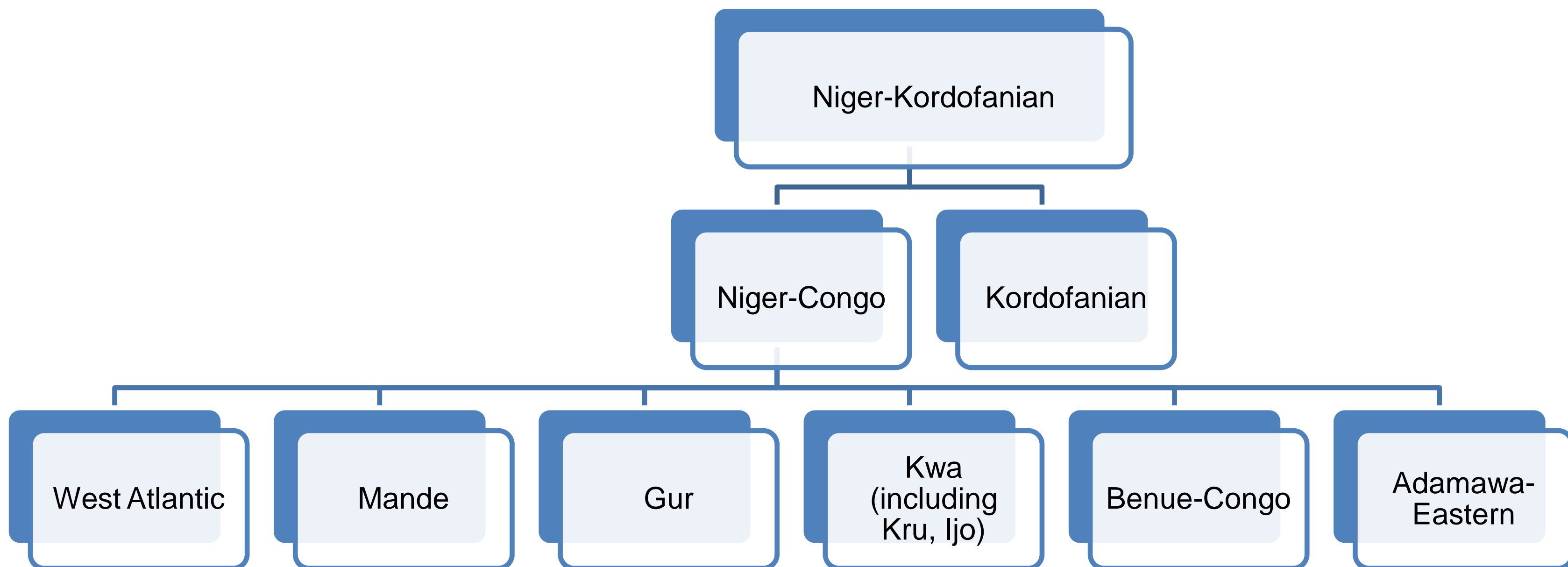


Figure: 2.3: The Niger-Kordofanian language tree (Greenberg, 1963)

2.3.2 Benue-Congo

Benue-Congo is the largest of the subfamilies within Niger-Congo in terms of the number of languages, speakers, and geographical extent. The Benue–Congo languages form a very large group in Africa and include the well-known Bantu languages. Greenberg (1963) introduced the term 'Benue–Congo' to designate one of the six branches of his Niger–Congo family. Previously, the Bantu languages had been treated as a separate family and the similarity of the other Benue–Congo languages to Bantu had been recognized by referring to them as 'Semi–Bantu' or 'Bantoid' (Guthrie, 1948). It stretches from the Benin-Nigeria border across Nigeria Eastward to Kenya and Southward to the Cape. Thus, it covers over half of the Continent and a similar percentage of the population. Benue-Congo is divided into 11 groups that can be arranged on an approximately West-to-East basis as shown in Figure 2.4. All these groups, with the exception of Bantoid, are found primarily in Nigeria. The principal languages of each group are as follows: Defoid: Yoruba and Igala, Edoid: Edo and Urhobo, Nupoid: Nupe, Ibira (Ebira) and Gwari (Gbagyi), Idomoid: Idoma and Igede, Igboid: Igbo, Cross River: Efik, Ibibio, and Ogoni, Kainji: Kambari, Platoid: Berom, Tarok and Jukun.

The Bantoid group forms the largest branch of the Niger-Congo, comprising many languages covering most of the area Southeast of Nigeria and Chad. Bantoid is divided into a small Northern group of languages spoken in Eastern Nigeria and Western Cameroon, and the very much larger Southern group, which includes all the Bantu languages.

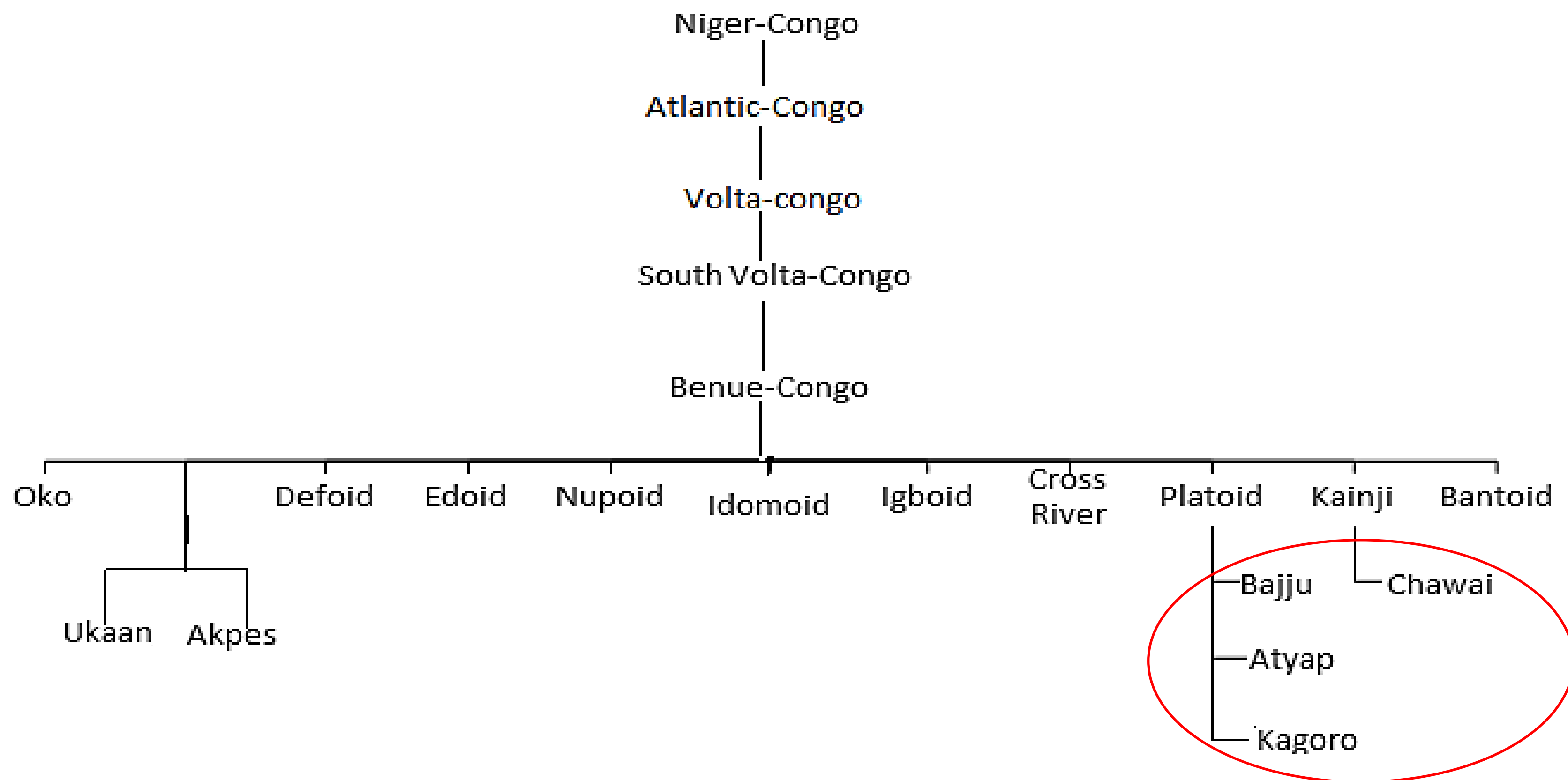


Figure 2.4**:** Revised classification of the Niger-Congo Subfamily with study populations encircled in red. (Adapted from Greenberg, 1963)

2.4 HISTORICAL BACKGROUND OF THE BAJJU, ATYAP, CHAWAI AND KAGORO

Kaduna State, found in Northern Nigeria is located just to the North of Abuja. The state shares the same name with its capital, a city established by the British along the Kaduna River and from which the former Northern Region was governed (Habu, 1993). Within its boundaries is the Southern part of Kaduna State located between latitudes 9° and 11° north and longitudes 7° and 9° East (Habu, 1993), which lies within the geographical entity known as "Middle belt". It is roughly sandwiched between larger ethnic groups, the Hausa to the North and combination of the Igbo to the South East and Yoruba to the South West. According to how broadly the region is defined, it includes between 250 and 400 distinct languages (Blench, 1998) roughly about 80 percent of Nigeria's over 500 languages. Very little is known about this area as there are no written records and even when oral records are available they do not carry a depth of more than two hundred years (Blench, 2003). The reasons for the Middle Belt's linguisticdiversity could be attributed to migratory events of which the Bantu expansion is notable (Vansina, 1995).

The Bajju, Atyap, Chawai (Atsam) and Kagoro are found in the Southern part of Kaduna state. The Bajju and Atyap within the Zangon Kataf Local Government Area and environs, while the Gworok and Chawai are found in Kaura and Kauru Local Government Areas respectively. They are classified as Benue-Congo languages, particularly of the Plateau variety excepting the Chawai that belong to the Kainji branch (Williamson and Blench 2000). These groups are extant within the area of Nigeria's oldest known civilization called Nok, after a village in Southern Kaduna near where a terracotta head was found in 1943 and brought to the attention of archaeologist Bernard Fagg and eventually,

the world. Although not definitively proven, Nok may well have influenced later Nigerian civilizations (Blench, 2003). The Nok civilization flourished from the Southern part of Kaduna and South of Jos plateau to far away Tiv settelement of Katsina Ala, South of the Benue river. Traces of the Nok culture were found as far as the Bachama people of Numan of Adamawa and beyond. The Nok culture flourished extensively in northern Nigeria from 500 BC into the early centuries AD. Furnaces found by archaeologists at Taruga which were older than 5,000 years served as proof that the Nok were the first known workers of iron in this part of the world. Apart from smelting iron, they were also into production of sculptures of the Terracotta type (http://zitt.sourceforge.net, 2013). The Bajju, Atyap, Chawai (Atsam) and Kagoro all have several traditions ascribed to their origins. Apart from the Chawai that claim an autochthonous origin in one of the three accounts of their historical origins, they all, including the Chawai share similar migration histories. They collectively have it in their traditional history tales of long distance migrations, and that they had common ancestors who migrated from an area around the present day Bauchi State to the Jos-Plateau before moving on to their present settlements (Nengel *et. al.,* 2002). Thus they can be said to be remnants of a probable single migration which in part may explain the proximity in culture and tradition. The latter account is preferred for the purpose of this study since their oral histories places them together in a particular region as their common or separate origins (Nengel *et al.,* 2002) while disregarding timing. Put together they constitute about 1.5-2 million of the Nigerian population as reported by the 2006 Census (NPC, 2006). Except for Chawai, the remaining tribes have a high level of intelligibility among their speakers. Therefore, from a population genetics perspective it is pertinent to determine if the levels of gene flow that may exist between these groups were

sufficient to homogenize the interpopulation genetic variation accumulated by geographic isolation (Gonzalez *et al.,* 2006).

2.4.1 The Chawai (Atsam)

The Chawai occupy two ecological zones of hills and plains. The Eastern area lies on the slopes of the Western escarpment of the Jos Plateau while the Western part lies in the upper part of the River Kaduna basin (Nengel *et al.,* 2002). The choice of the settlement was probably influenced by the economy, the need for security and religious inclinations (Chawai History Project Transcripts, 1993). The Chawai had close relationship with their neighbours; they participated in cultural events like the annual communal hunting with nearby tribes such as the Atyap, Piti, Kurama, Miango, Amo and Ribang in their traditional hunting grounds which are the lowlands of Mambo, Warsa and Tsararin Mata. Apart from the Hausa and Irigwe, the Chawai did not intermarry with any of their neighbours for reasons of a cultural nature (Nengel *et al.,* 2002).

Available sources show that there are seven founding settlements of "great antiquity", they are Zambina, Maizanko, Fadan Chawai, Pari, Kizakoro, Kidundun and Badurum. There are over 43 villages associated with these centres (Nengel *et al.,* 2002) as shown in Figure 2.5.

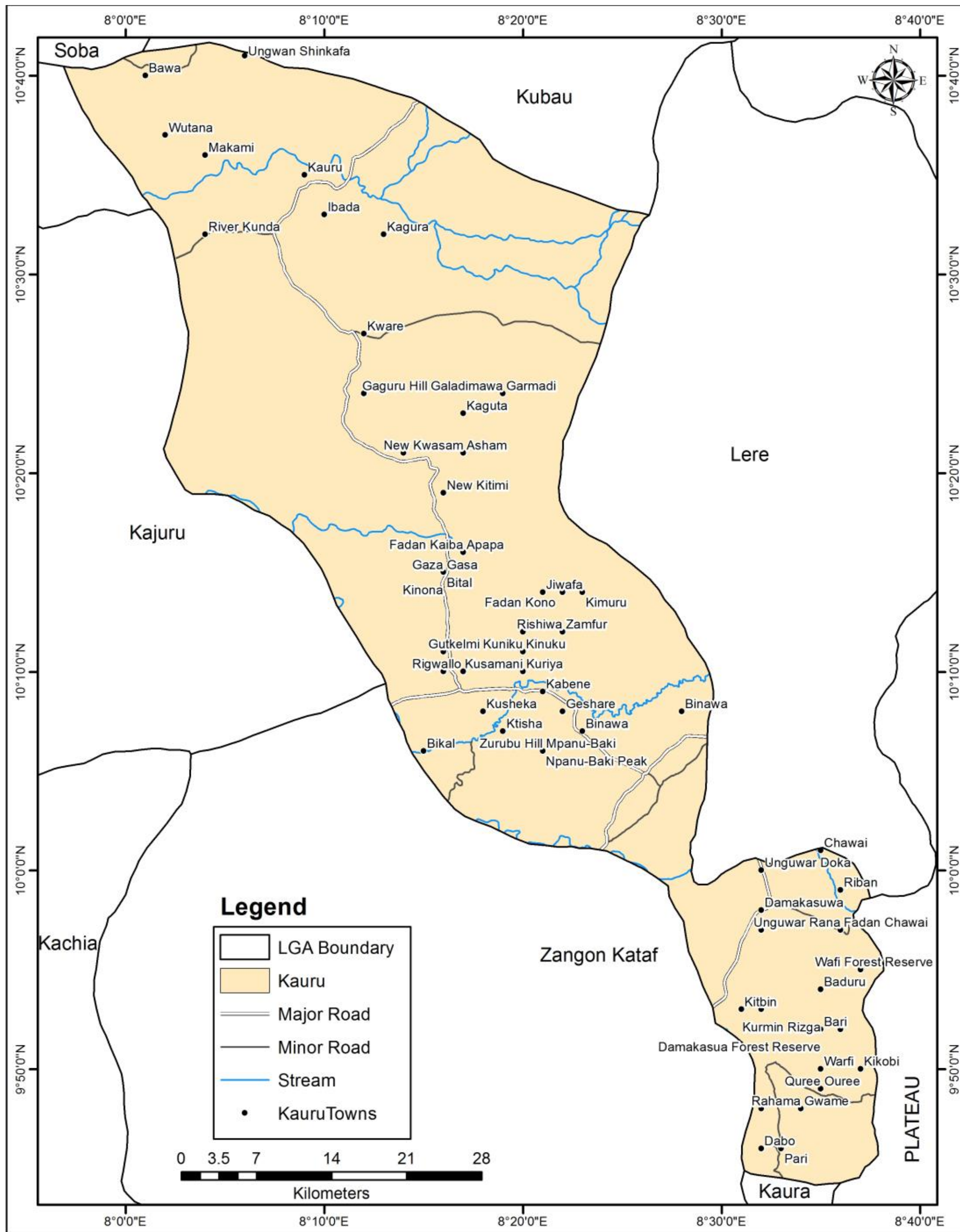


Figure 2.5: Map of Kauru LGA showing geographic region of the Chawai. Adapted from the administrative map of Kaduna State.

2.4.2 The Bajju

Based on oral traditions, the origin of the Bajju can be traced to an area as far as the present Bauchi State. The Bajju claim to have left a settlement close to the ancestral home of the Jarawa from where they migrated to Miango (an Irigwe settlement on the Western edge of the Jos Plateau) (Danbo, 2011). Other sources have it that their ancestor had his origin from Niger and/or Cameroon. It was believed that their migration was for the search of better hunting grounds. They migrated from Bauchi State to Plateau State and settled on a hill called 'Hurruang'. The hill was already occupied by a tribe called the Jarawa, but the Jarawa people left and lived on another hill called 'Tsok-kwon' (Danbo, 2011). The Bajju moved further down to their present settlement in the Southern part of Kaduna (Figure 2.6) and have remained since about the early 1800's. Baranzan the son of Zamfara is regarded as the founding ancestor of the Bajju and according to oral traditions the Atyap and Chawai may be his descendants. The Bajju are predominantly subsistence agriculturists (McKinney, 1985).

2.4.3 The Atyap (Kataf)

The Atyap derive their name from their language, the Tyap description generally means "the people who speak Tyap" (www.atyap.net). The Atyap people are also referred to as "Katab" or "Kataf" from Katambari which is the Hausa word for camwood because they were known for trading with camwood (Katambari in Hausa). Tyap is usually used to refer to a group of similar dialects comprised of the Atyap – tyap; the Agworok – Gworok; the Asholio – Sholio and the Atakad – Takad (http://www.atyap.net,www.ethnologue.com).

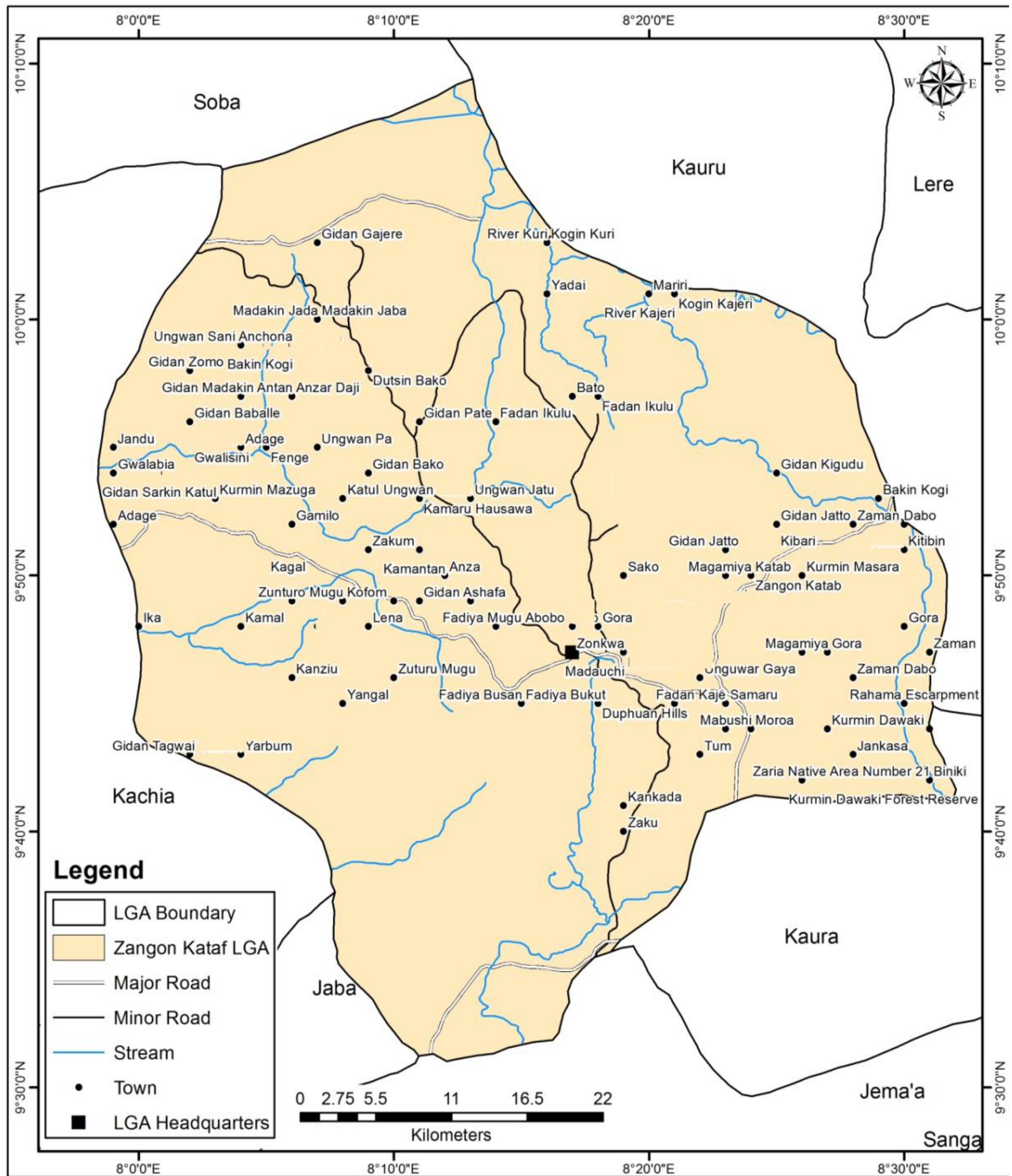


Figure 2.6: Map of Zangon Kataf LGA showing Bajju and Atyap settlements. Adapted from the administrative map of Kaduna State.

The Atyap have four clans, namely, Agbaat, Aminyam, Aku and Shokwa. Some of these clans have sub-clans. The Agbaat has three sub-clans: Akpaisa, Akwak and Nje. Aminyam has two sub-clans: Aswon and Fakan. Aku and Shokwa do not have sub-clans. There are no distinct settlements for specific clans or sub-clans today because clans are highly mixed because of the mass movement of people occasioned by need for land, the 19th century raids and British colonial policies aimed at effective exploitation of the people (www.atyap.net). The Atyap villages are closely situated to those of the Bajju (Figure 2.6)

2.4.4 Kagoro (Oegworok)

Kagoro is a shared name between the language as well as the people who regard themselves as Oegworok. They are found in Kaura Local Government Area (Figure 2.7). Their traditions have it that they arose in the East precisely from Sudan (Joshua, 2013) as part of the Nubian empire (http://kagoro.online.cm). They emigrated with other populations with push factors attributable to wars and sundry reasons. The migratory path was Southwest through South of Chad into Borno eventually reaching Bauchi and from there to Assop presently in Plateau state, chronologically estimated around 1600-1700. They kept on with the Westward path to Numbio (Nimbia) near the Kagoro Hills and finally settled on Tsok-Busa (meaning a plain land on top of a mountain) on top of the hills (Joshua, 2013) and environs. Some oral sources maintain that the present population today may have origins among Atyap migrants from the Agbaat clan and from surrounding populations such as the Miango of Plateau state.

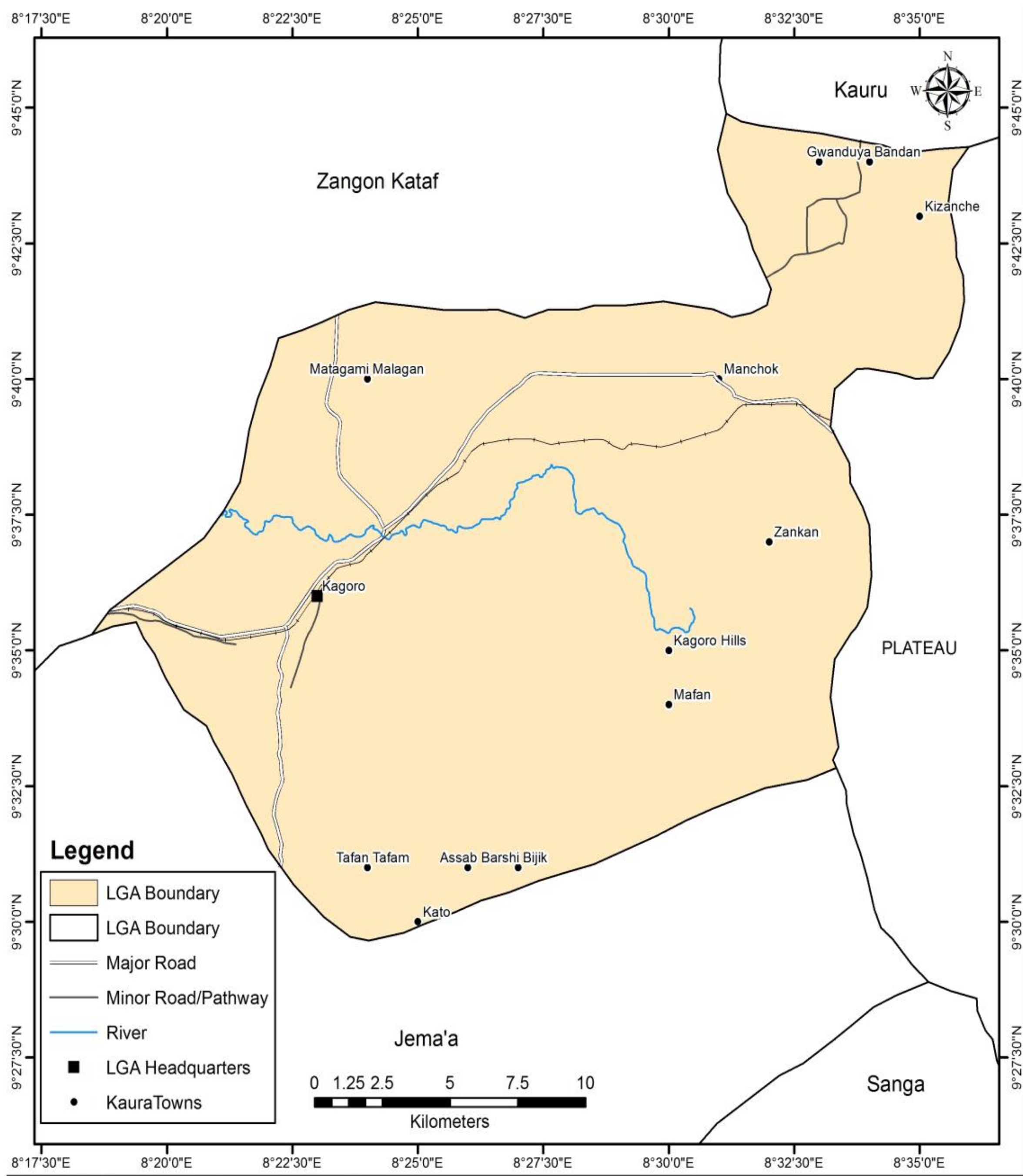


Figure 2.7: Map of Kaura LGA showing the geographic spread of the Kagoro. Adapted from the administrative map of Kaduna State.

2.5 GENETIC MARKERS

The early use of genetic markers in evolutionary studies of human populations was summarized by Crawford (1973), defined as "discrete segregating, genetic traits which can be used to characterise populations by virtue of their presence, absence, or high frequency in some populations and low frequency in others' (Crawford, 2007). In other words they may be variable allelic state of genes, gene products (proteins) or unique DNA polymorphisms, which produce measurable phenotypes (physical or biochemical) in a given population. In answering anthropological questions, these marker types are preferred as they are usually not affected by the environment and hence have largely replaced phenotypic and quantitative traits like those used in anthropometry (Rogers and Harpending 1983; Crawford, 2007). In population studies, two types of genetic systems have been identified, classical (non DNA) and molecular (DNA) markers.

2.5.1 CLASSICAL MARKERS

Long before the recombinant DNA revolution, studies on human diversity, population structure, group affinities, migration patterns, and microevolution were done basically at the level of plasma proteins and red blood cell enzymes because of the presence of polymorphic states (Mielke *et al.,* 2011). Since the introduction of zone electrophoresis for proteins by Oliver Smithies in 1955, available technologies such as electrophoretic techniques were used in the detection of polymorphisms. The first of these polymorphisms discovered by Landsteiner (1900, 1901) was the ABO blood group system. This marker is located on chromosome 9 and codes for two antigens A and B, giving rise to three alleles

which are the A, B and O (which contains antigen H also known as FUT1) (Agre and Cartron, 1992). The Duffy system is another blood marker discovered for its involvement with malarial infection by *Plasmodium vivax* and *Plasmodium knowlesi*. This system consists of two principal codominant alleles, $Fy^{a}$and Fy^{b}, that differ by a single amino acid. The Fy^{o}, is the third allele which represents a lack of the Fy antigen. Therefore the antisera (anti-Fy^{a} and anti-Fy^{b}) give rise to four main phenotypes. Another five duffy alleles viz Fy3, Fy4, Fy5, Fy6 and Fy^{x} have also been discovered (Pogo and Chaudhuri 1997; Reid and Lomas-Francis 1997). Other blood group markers such the Lutheran blood group, Kidd, Kell, Diego, Cartwright, Scianna, Dombrock, Colton, Xg, and Landsteiner-Wiener have also been studied. These "classic" genetic markers were used erstwhile for establishing paternity and for forensics but anthropologist have been studying them to ascertain their relative occurrence along with their subtypes in populations leading to understanding of population structure (Mielke *et al.,* 2011). Some examples of the use of these markers in anthropology include the study by North *et al.* (2000), in which was demonstrated the use of 10 classical markers from blood systems consisting of the Duffy, ABO, Kell, Rhesus, MNS, P, transferrin, phosphoglucomutase, adenylate kinnse, and haptoglobin to ratify to the relationship of Irish Travelers/Tinkers to the main Irish population as earlier suggested by Crawford (1975). The association between genes and language in Europe had also been reported using 120 alleles from the classical genetic systems by Cavalli-Sforza *et al.* (1988), paving way to attempt the synthesis of the classical genetic marker data of different populations around the world by Cavalli-Sforza *et al.* (1994). A shortcoming on the use of this marker type is the uncertainty of influence by selective environmental pressures, as one trait could have resulted from different polymorphism at the DNA level as in the case of Glucose-6-phosphate dehydrogenase

deficiency (Tripathy and Reddy 2007). With the dawn of molecular techniques, the use of classical markers in studies of population origins, expansion, regional diversity, and replacement has largely waned.

2.5.2 MOLECULAR MARKERS

With advanced and rapid development in molecular technologies, analysis of DNA found outside or within of the nucleus of human cells (nuclear, mitochondrial and ancient) and Y chromosome analysis have been used in studies such as Stone and Stoneking (1998) which involved the use of mtDNA from 147 ethnic groups. The result as seen in Figure 2.8, showed a split between the African population and the rest of the world, represented as two main phylogenetic branches. Sykes (1999), Barbujani and Bertorelle (2001) and also Richards *et al.,* (2003) using similar techniques tried to explain population origins, genetic variation, migration and gene flow. These genetic systems are characterized by polymorphisms that serve as markers for genetic analysis. Some of these are the substitution of nucleotide bases known as the single nucleotide polymorphism (SNP), insertions and deletions of single nucleotide bases (Indels) and short tandem repeats (STRs) which are tandemnly repeating sequences also known as satellites (mini and micro) and retroelements (Lander *et al.,* 2001; Crawford, 2007). From other studies (Rosenberg *et al.,* 2002; Watkins *et al.,* 2003; Jakobsson *et al.,* 2008; Li *et al.,* 2008; Tishkoff *et al.,* 2009), it has been shown that autosomal chromosome data have exhibited considerable homogeneity among Niger-Congo–speaking groups (including Bantu) and adeep structure among hunter-gatherer communities in sub-Saharan Africa, although the data are still scanty and contradictory.

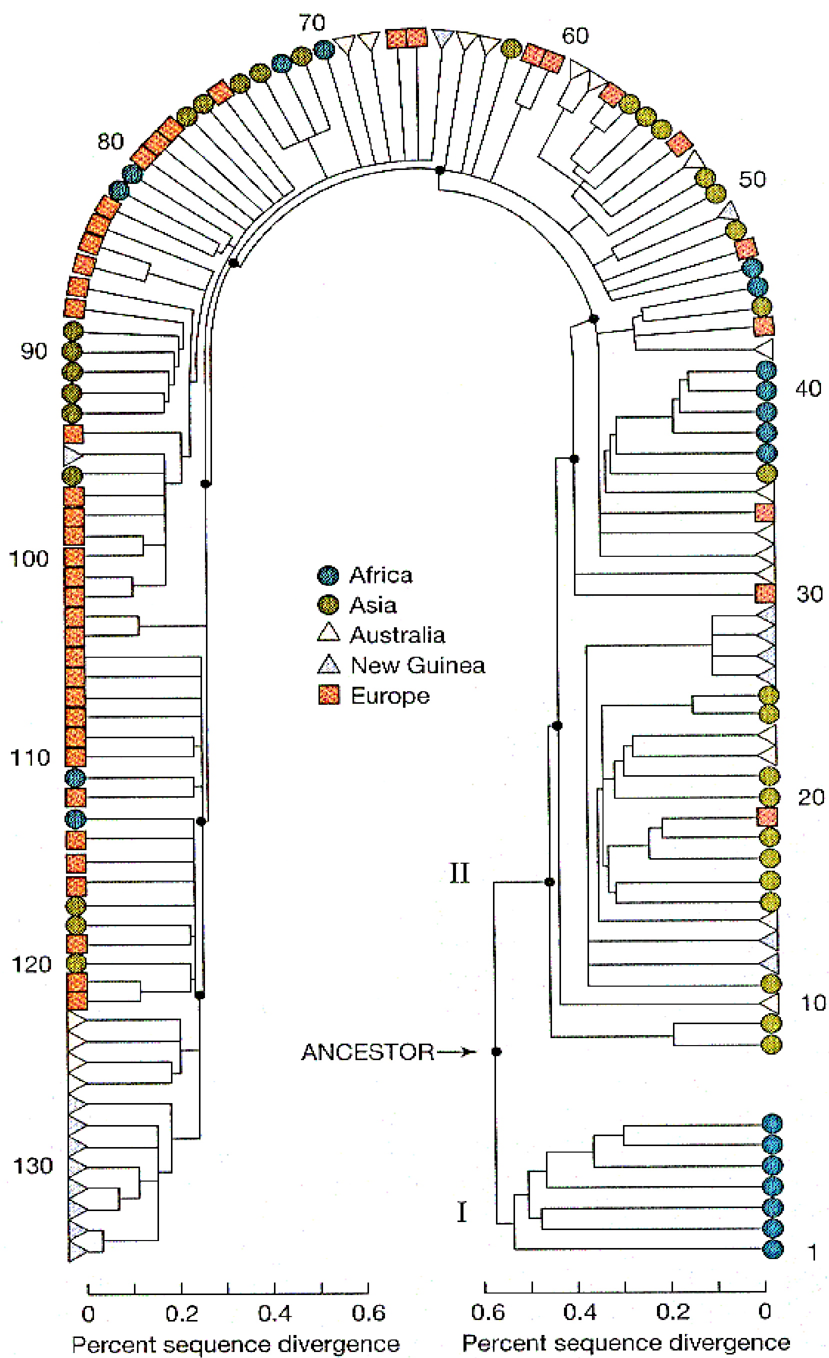


Figure 2.8: Phylogenetic mtDNA tree from the world's geographic regions based on genetic markers (Adapted from Cann *et al.,* 1987)

The components of the human genome used in these phylogenetic studies are the autosomal markers, the non-recombining portion of the Y-chromosome and mtDNA. In general they provide a unique set of tools that can be used to reconstruct sex specific population history.

2.5.3 AUTOSOMAL MARKERS

The utility of autosomal markers in evolutionary studies has the advantage of being able to capture genome wide polymorphisms in individuals. The autosomal polymorphic traits are inherited biparentally as against the markers inherited from single parents (mtDNA and Y-chromosome) (Rubicz *et al.,* 2007). The autosomal genome can be regarded as a series of large blocks of low recombination, interrupted by small blocks of high recombination. Alleles at these different loci may undergo recombination and assort independently as they are inherited from one generation to another (Darby, 2004). The recombination may occur within the large blocks which may vary between populations (Tishkoff and Verrelli, 2003). This reshuffling produces new genetic combinations and may sometimes not be suitable in answering certain research questions (Crawford, 2007).

2.6 Y-CHROMOSOME

The Y-chromosome (Figure 2.9) is considered poor in genes mainly relating to male sex determination and coding only for 27 proteins (Skaletsky *et al.,* 2003). The gene loss is as a result of the inability of Y-chromosome to eliminate mutant alleles by recombining with a non-mutant homolog leading to the degeneration of genes caused by forces such as Muller's ratchet (Jobling *et al.,* 2004).

The small region of pseudoautosomal chromatin recombines with X-chromosome leaving approximately 95% of the Y-chromosome known as the male specific portion of the Y (MSY), free of recombination (Rubicz *et al.,* 2007). The absence of recombination means that any accumulated mutations will be passed from a father down to his male lineage only, which makes the Y-chromosome useful in phylogenetic and population studies. The polymorphisms used are considered as unique mutational events (UMEs) in anthropological genetics are short tandem repeats and binary markers mainly single nucleotide polymorphisms and indels (Brion *et al.,* 2003; Quintans *et al.,* 2003; Crawford, 2007).

2.7 MITOCHONDRIAL DNA

Mitochondria (see Figure 2.10) are organelles found outside the nucleus of the cell and are concerned with energy production via oxidative phosphorylation and ATP synthesis for metabolic processes. Mitochondria possess their own genome that encodes genes needed for mitochondrial protein synthesis. However, most mitochondrial proteins are encoded by nuclear genes and are synthesized on cytoplasmic ribosomes before being imported into the mitochondria. Unlike the simple bacterial genome which is made up a single DNA molecule, the human genome is complex consisting of 25 different DNA molecules (one mitochondrial DNA molecule and 24 nuclear molecules). It is made up of a nuclear genome containing 26,000 genes, and a simple mitochondrial genome with only 37 genes (Strachan and Andrews, 2011). The mitochondrial full sequence was produced by Anderson *et al.* (1981) from a human tumour line now known as “Cambridge reference sequence”, and revised by Andrews *et al.* (1999).

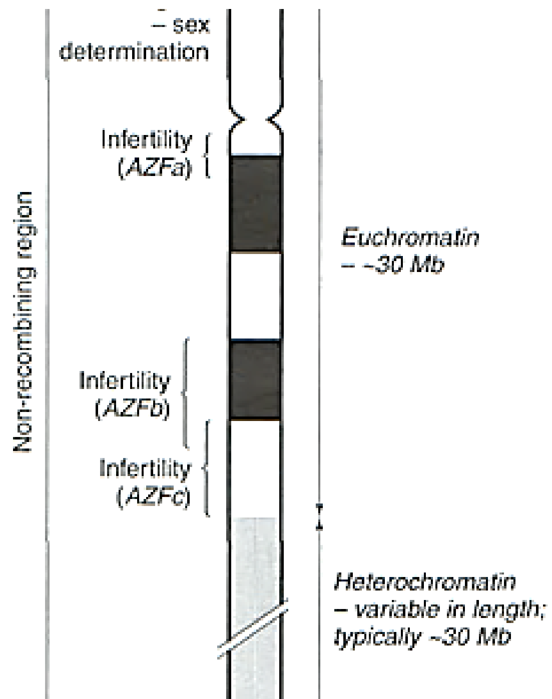


Figure 2.9: The Y chromosome. The non-recombining region forms the locus used in studies of genetic variation. (Adapted from Jobling *et al.,* 2004)

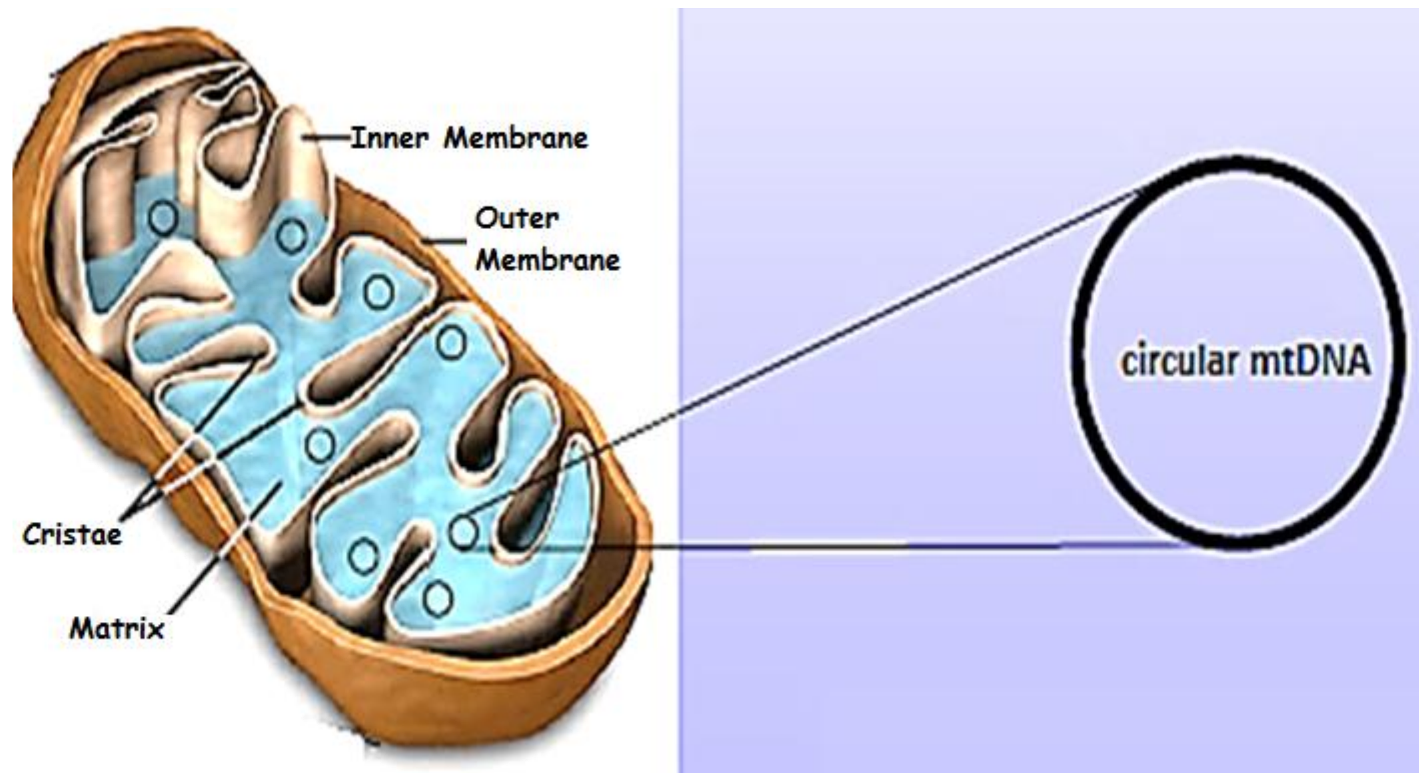


Figure 2.10: The Human Mitochondrium. The circular mtDNA are shown located within the matrix of the organelle.

mtDNA is a short sequence of circular double stranded DNA of about 16,569 base pairs (figure 2.11), found only in the mitochondria of higher animals (Ebner *et al.,* 2011).

In man, each cell holds as much as 100-1000 mitochondria per cell and each mitochondrion has 2-10 copies of the molecule (Tang *et al.,* 2000; Jobling *et al.,* 2004). This high copy number makes it a rich source of genomic DNA (Torroni *et al.,* 1994). The mtDNA is organized into coding and non-coding parts with the coding parts lackingintrons. The transmission of mtDNA is typically from the mother to her offspring with usually no contribution from the father (Butler, 2005). The explanation to this phenomenon is the presence of fewer mitochondria in a sperm cell when compared to an oocyte, so paternal mitochondria would make up less than a tenth of a percent of the total mitochondria in the newly formed zygote. Second, there appears to be some mechanism by which paternal mitochondria that enter the oocyte are inactivated.

So far, there has been only one documented case of paternal mtDNA transmission, where there was recombination between the maternal and paternal mtDNA (Schwartz and Vissing, 2002; Kraytsberg *et al.,* 2004). mtDNA provides genetic markers for inferring genealogical and evolutionary relationships among and within populations. Since it is maternally inherited and transmitted down the lineage with little or no recombination of genetic material, all members of a maternal line will share identical mtDNA sequences (Wallace, 2007).

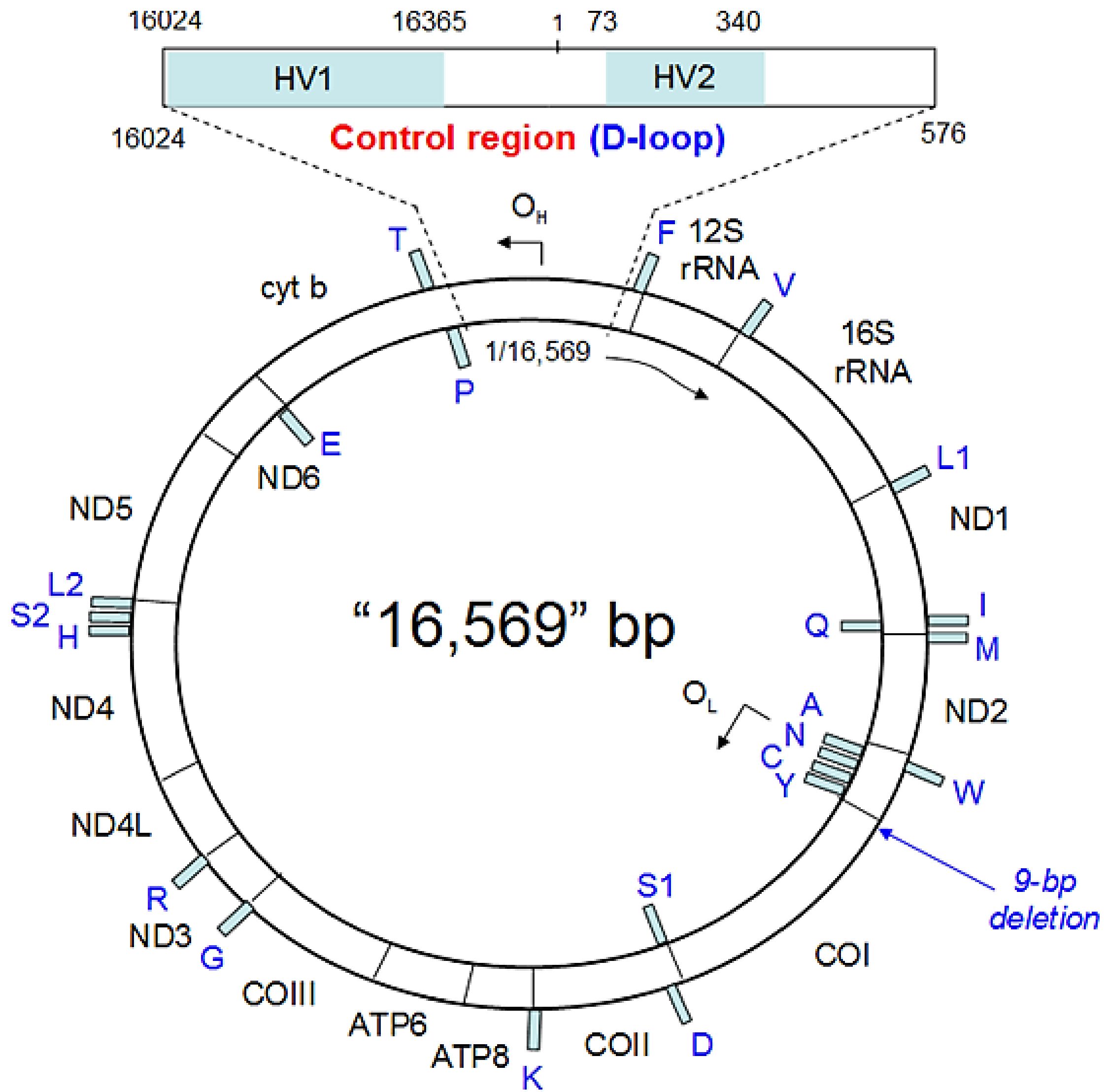


Figure 2.11: Mitochondrial DNA. The non-coding HVS I and HVS II portions of the control region and the rest of the coding region represent the loci used for variation studies. Adapted from Jobling *et al.* (2004)

Even though mtDNA inheritance does not reflect the nuclear pattern, it does form a fundamental part of an organism's heritage and evolutionary history (Rubinoff and Holland 2005).

mtDNA is now widely used to describe human lineages and migration patterns (Cavalli-Sforza and Cavalli-Sforza 1995), due to its low rate of recombination (Piganeau *et al.,* 2004), maternal inheritance, simple genetic structure, reduced effective population size (Ne), and relatively rapid rates of evolution (Avise *et al.,* 1983; Moritz *et al.,* 1987; Chen *et al.,* 2005). This is also aided by characteristic mutations arising within the mtDNA (Maca-Meyer *et al.,* 2001). The high mutation rate of mtDNA may be due to the fact that the mitochondrial genome is closely apposed to the respiratory structures of the cell. The respiratory machinery is a major source of mutagens known as oxygen free radicals, which are natural by-products of respiration (Bermisheva *et al.,* 2003). The resulting mutations provide the basis for the different haplotypes have been defined as independent lineages and relationship between these lineages are estimated according to phylogenetic analysis (Bandelt *et al.,* 1995). These have been constituted into major haplogroups used to describe human populations (Torroni *et al.,* 1996; Pereira *et al.,* 2005).

2.8 HISTORY OF MITOCHONDRIA

From the theory put forward by Margulis, (1981), it is now generally established that the mitochondrion originated as an endosymbiotic bacteria over 1.5 billion years ago. The choice of a safer environment in return for energy delivery led to its choice of a proto-eukaryotic cell as a host (Margulis 1970a, 1975; Mishmar *et al.,* 2003; Wallace *et al.,* 2003). This relationship is thought to be a fall out of a global catastrophe billions of years

ago, which saw to an increase in levels of atmospheric oxygen from activities of microorganisms (Holland, 1994).

The need for protection in such toxic climate caused anaerobic organisms to find refuge within higher forms of life. Mitochondrial DNA represents an extranuclear genome whose content and size varies in different living organisms. When compared to the conserved organization in metazoan organisms (Saccone *et al.,* 1999) most of mtDNA genes have been lost in mammals. Its prokaryotic history is evidenced by the possession of features similar to modern bacteria besides having similar proportions, such as the non-linear (circular) DNA, the circular structure is thought to provide protection from exonucleases, which digest free ends of linear DNA molecules (linear chromosomes possess "end caps" known as the telomeres for protection). Other common features between the mitochondrial and bacterial genomes, is the presence of only little noncoding DNA (Margulis, 1970b). Genes are usually tightly packaged on the chromosome, with few intergenic regions between genes and few introns within genes. This presents a different conformation from that of eukaryotic genes, which are widely spaced on chromosomes and have numerous introns. There is also absence of histones and the discrete origin of replication support the hypothesis of an endosymbiosis (Vellai *et al.,* 1998). Another strong suggestion of its exogenous source is the difference in the genetic code of the mtDNA when compared to the nuclear genome. While the codon UGA codes for the STOP signal in Nuclear DNA, it codes for the amino acid Tryptophan (Trp). Likewise AGA and AGG code for Arginine (Arg) in the nuclear DNA while they represent the STOP signal in mtDNA. AUA and AUU code for Isoleucine (Ile) and Methionine in nuclear DNA and mtDNA respectively (Venter *et al.,* 2001). Molecular phylogenies have also provided evidence of a single and monophyletic bacterial origin for these modern cellular organelles

with mtDNA genes having being closely related to those of the present-day alpha-proteobacteria (Karlberg *et al.,* 2000; Andersson *et al.,* 2003). Thus in the process of co-evolution, some of the mitochondrial genes have been transferred to the nucleus but not as functional genes (Lopez *et al.,* 1994). This is illustrated in the reports that there are more than 600 nuclear inserts of mtDNA (numts) of different lengths within the human nuclear genome (Shoubridge, 2001; Mishmar *et al.,* 2004; Riccheti *et al.,* 2004; Hazkani-Covo and Graur 2007).

2.9 BIOLOGY OF THE MITOCHONDRIA

2.9.1 Replication

The replication of both the Heavy (H) and Light (L) strands is unidirectional and starts at specific origins. Although the mitochondrial DNA is principally double-stranded, repeat synthesis of a small segment of the H-strand DNA produces a short third DNA strand called 7S DNA (Mitomap, 2013). As seen in Figure 2.12, the 7S DNA strand can base-pair with the L-strand and displace the H-strand, this leads to the formation of a triple-stranded structure (Strachan and Read, 2011). The overall base composition is 44% guanine and cytosine (G+C), but the two mtDNA strands have significantly different base compositions: the heavy (H) strand is rich in guanines, but the light (L) strand is rich in cytosine.

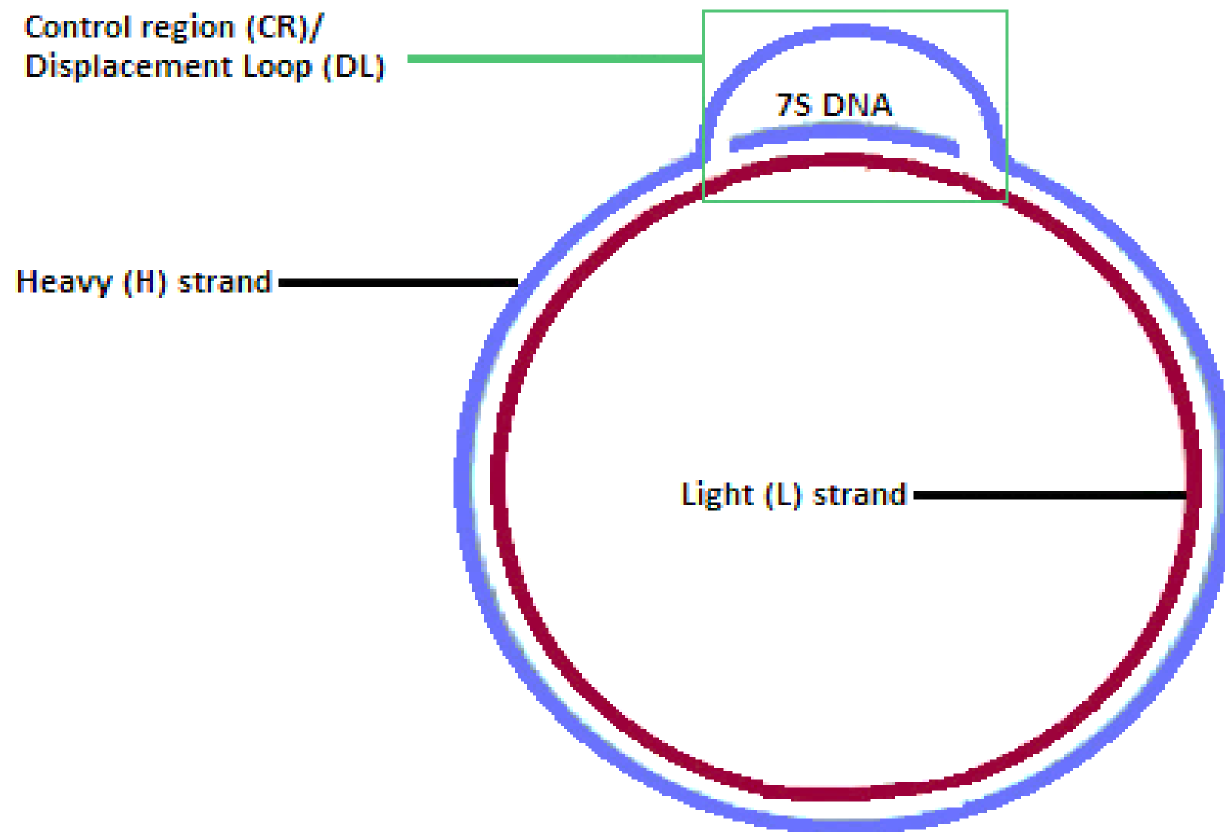


Figure 2.12: D-Loop formation in mtDNA. Adapted from Strachan and Read (2011).

This area contains many of the mtDNA control sequences comprising the major promoter regions that contain the signals to control RNA and DNA synthesis. Thus, it is known as the Control/Displacement-loop (CR/D) region or hypervariable region due to accumulation of point mutations at approximately 10 times the rate of nuclear DNA (Falkenberg *et al.,* 2005). It is approximately 1,200 base pairs in length. The origin of replication for the H-strand lies in the CR/D-Ioop region, and that of the L-strand is sandwiched between two tRNA genes. A single promoter on each DNA strand initiates transcription in each direction, and a single origin initiates replication of each strand. When about two-thirds of the daughter H strand has been synthesized using the L-strand as a template and displacing the old H strand, the origin for L-strand replication then become exposed. Subsequent replication of the L-strand continues in the opposite direction with the H strand serving as the template (Strachan and Read, 2011).

2.9.2 Energy Generation

Mitochondria are the source of energy within the cells and this function is achieved through the energy-generating oxidative phosphorylation (OXPHOS) pathway (Tranah *et al.,* 2011). Other concomitant processes include the tricarboxylic acid cycles, the breakdown of amino acids, lipids, cholesterol, steroids and nucleotides intracellular signalling and apoptosis (Chinnery, 2006). Most of its length comprises contiguous coding regions, encoding for 13 polypeptides involved in OXPHOS electron transport system (ETS), 22 tRNAs and 2 rRNAs, essential to protein synthesis. The population of mitochondria in human tissues vary depending on the cell type and size and function. Thus thousands of mitochondria are found in tissues that are in dire need of energy. While

somatic cells have about 1000 to 10000 (Lightowlers *et al.,* 1997), the oocytes have around 200,000 molecules (Reynier *et al.,* 2001; Santos *et al.,* 2006) and sperm cells mid-piece averaging about 50 to 75 mitochondria (Diez-Sanchez *et al.,* 2003) for their motility (Ruiz-Pesini *et al.,* 2000). During the production of energy, protons are created from electron transfer and these protons then react with the oxygen to yield water. This process may come with errors such as points of faulty transfer of electrons within the electron transport chain result in the formation of the superoxide free radical. To reduce the occurrence of mutations as a result of the abundance of these free radicals, the superoxide free radical is dismutated, this reduction and oxidization converts the free radicals into hydrogen peroxide and oxygen. The hydrogen peroxide is also converted to water and oxygen. Consequently, mutagenic hydroxyl ions which are more reactive than superoxide produced. They produce free radicals which react with essential biological substances such as amino acids, nucleic acids, lipids and sugars that eventually impede DNA and RNA processes such as cell replication and transcription.

2.10 MtDNA VARIATION IN HUMANS

mtDNA studies concerning human variation began with a pioneering study into human mtDNA as a molecular marker by Wesley Brown and Douglas Wallace during the late 1970s (Brown 1980). It involved the use of single to several restriction enzymes to digest the entire mtDNA molecule of 21 humans from various ethnicities and geographical backgrounds (Richards and Macaulay 2001). The result showed that restriction-enzyme fragment length polymorphism (RFLP) was useful in unravelling human genetic history

and coalescence time (Torroni *et al.,* 2006). Following that, studies on the discovery of origins for anatomically modern humans (*Homo sapiens*) commenced in the early 1990s. The strategy was localizing the studies to individual continents using RFLP sites on the coding regions of the mtDNA molecule or sequencing the first hypervariable segment (HVS-I) of the non-coding control region.

The first large-scale studies were carried out within Native American populations to answer questions on the origin, the time and numbers of ancestral migrations from Asia (Torroni *et al.,* 1993). They discovered that the molecular differentiation which had taken place after the process of human radiation into the different continents and their subsequent colonization, led to the restriction of the mtDNA haplotype/haplogroup types to specific geographical areas and populations (Achili *et al.,* 2004; Achilli *et al.,* 2005). The haplotypes are defined by specific array of mutations along the mtDNA molecule compared to a reference sequence, the Cambridge Reference Sequence (CRS) (Anderson *et al.,* 1981) and revised by Andrews *et al.* (1999).

Two systems of classification were put forward, one used the Roman numerals (Horai *et al.,* 1993) and the other utilized the English system of alphabets (Torroni *et al.,* 1992). The latter quickly became widely applied and it now serves as the nomenclature of choice. Together with archaeological findings, themitochondrial genome has become one of the key tools for not only understanding human origins (Atkinson, 2009) but also plays a role in understanding genetic diseases whether sex-linked or among ethnicities (Price and Price2004). It also provides for the appraisal of the degree of admixture in any given population (Salas *et al.,* 2002).

The use of mtDNA polymorphisms using HVS-I data in combination with partial RFLPs in defining the ethnic origins of populations is based on recognisable haplogroups (Rando *et al.,* 1998, 2000; Kivisild*et al.,* 2004). Torroni *et al.* (2000) defined a mtDNA haplogroup, as a monophyletic clade of the mtDNA genealogy. Haplogroups consist of different mtDNA haplotypes of common ancestral descent and sharing a characteristic mutational event within the mtDNA structure. Arising from the study on Native Americans, the first haplogroups were identified and named in an alphabetic sequence A, B, C, D and later X (Torroni *et al.,* 1993; Schurr and Sherry 2004; Torroni *et al.,* 2006). Subsequently, haplogroups of other populations based within the extant continents were characterized using up all the letters of the alphabet excepting O (vanOven and Kayser, 2008). The convention for mtDNA haplogroup nomenclature was proposed by Richards *et al.* (1998). The major mtDNA haplogroups reported for the sub-Saharan African populations are mainly L0, L1, L2 and L3. Found also are rare haplogroups such as L4, L5, L6 and L7 (Gonder *et al.,* 2007). Apart from these, very small frequencies of haplogroups from Europe, Asia and North Africa (H, I, J, K, T, U, V and W) have been described in sub-Saharan Africa (Lum *et al.,* 1989; Torroni *et al.,* 1994).

The distribution of these haplogroups within the African continent portrays a picture of human prehistory. Haplogroup L0 is localized within eastern and southern Africa (Salas *et al.,* 2002), L1 is found within west and central Africa while L2 which is more frequent than any other haplogroup is widely spread within the western and south eastern parts of Africa. Haplogroup L3 is the most recent in origin, is commonly distributed within the western, eastern and south eastern parts of Africa (Watson *et al.,* 1997; Salas *et al.,* 2002). With the expansion events, African-specific L3 haplogroup yielded descendants

which are the Eurasian macro-haplogroups M and N found outside Africa (Mishmar *et al.,*2003).

2.11 DNA SEQUENCING

DNA sequencing was based on an enzymatic sequencing method first developed in the 1970s, in which a DNA polymerase was used to synthesize new DNA chains by using a cloned single-stranded DNA template, consisting of millions of identical copies of a specific DNA sequence (Mielke *et al.,* 2011). The Sanger dideoxy method introduced by Fred Sanger in 1970s involved the chain termination reaction for defining sequence composition. The method begins with many copies of single-stranded DNA obtained either by asymmetric PCR, or by tagging one of the primers so that after the DNA is denatured the targeted strand can be isolated (Mielke *et al.,* 2011). There is random inhibition of chain elongation, creating newly synthesized DNA strands of various lengths that can be separated by size. Sequencing is carried out in four parallel reactions as depicted in Figure 2.13.

Each reaction uses the PCR method to grow DNA strands in one direction but uses only one primer and a mixture of ddNTPs (dideoxy-ATP, -GTP, -CTP, and –TTP) that serve as chain terminators and the dNTPs (deoxy- ATP, -GTP, -CTP, and –TTP). These dideoxy molecules lack the 3' hydroxyl (OH) group and also at the 2’ position of the regular dNTPs ribose sugar.

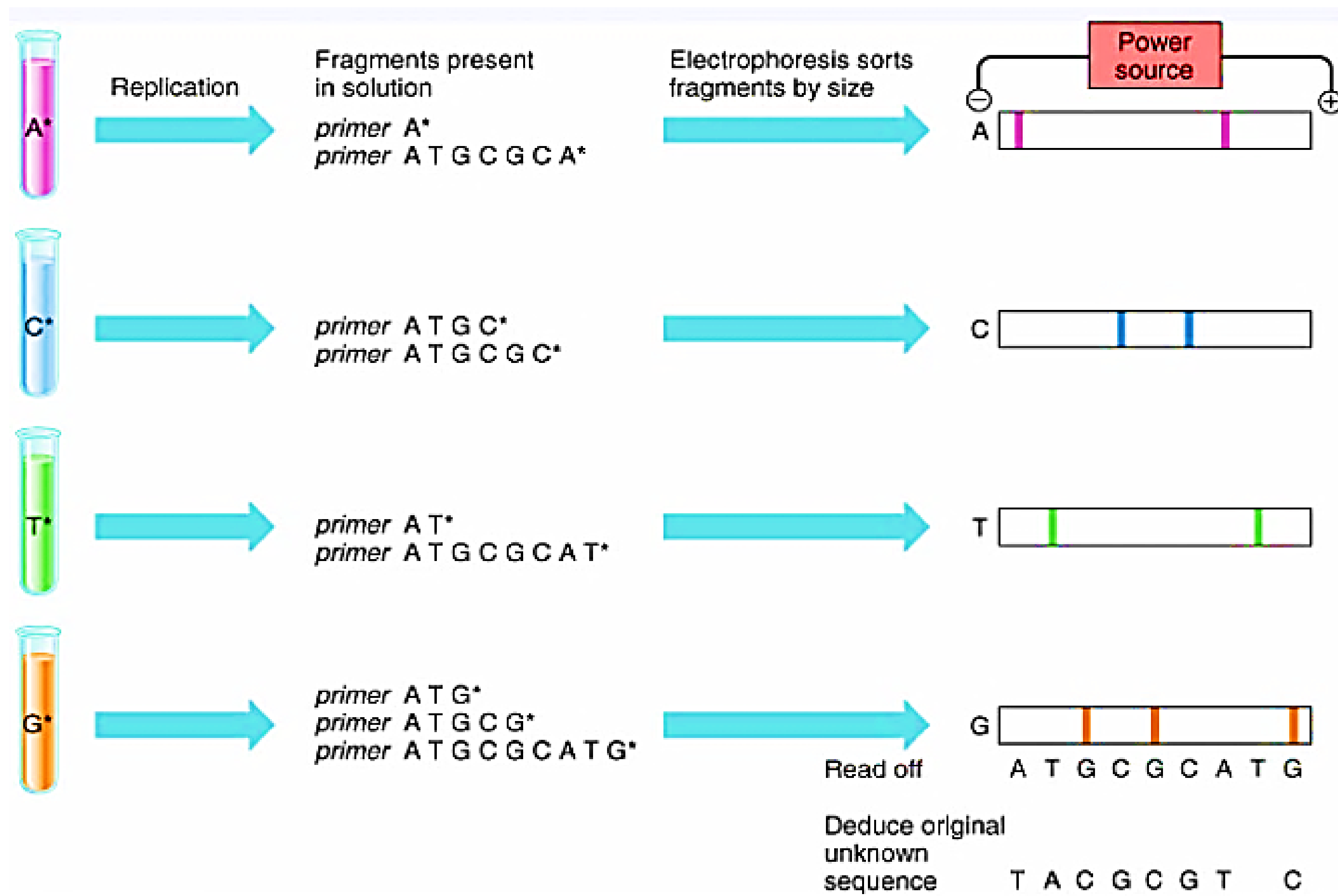


Figure 2.13: Sanger's sequencing method

When incorporated into a growing chain to form phosphodiester bond between the 5' and 3' carbon of the last incorporated nucleotide during extension, they abruptly terminate the formation of the phosphodiester bond thus stopping the growth of the DNA chain producing amplicons of different lengths (Strachan and Read 2011). The ddNTPs are fluorescently colour labelled so they can be differentiated with the aid of an electrophoretic automated sequencer which uses a laser to detect the wavelengths and colour of the dye. This results in an electropherogram or chromatogram (Fig 2.14) from which the sequence can be easily read.

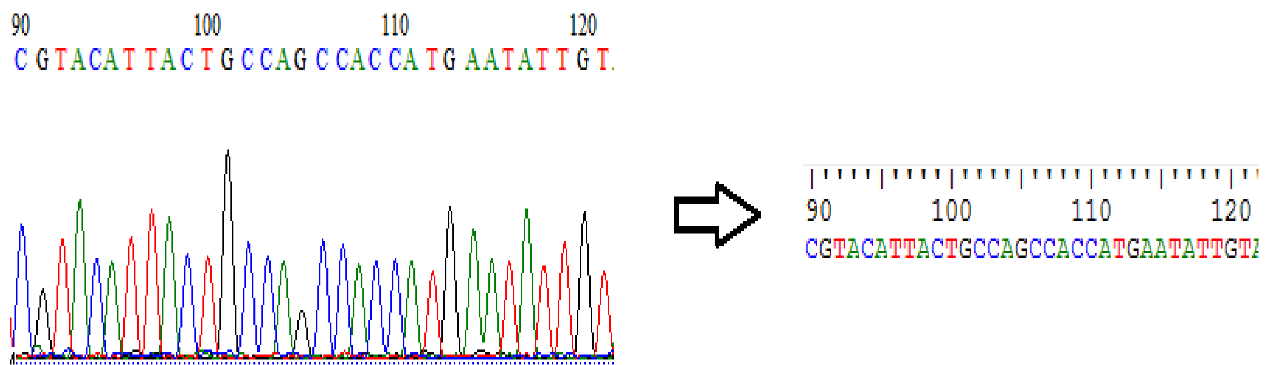


Figure 2.14: Sample sequence electropherogram. Coloured peaks correspond to nucleotide bases

2.12 POLYMERASE CHAIN REACTION (PCR)

The polymerase chain reaction is a method of cloning DNA *in vitro* developed in1984 by Kary Mullis. Using the enzyme DNA polymerase, it is designed to synthesize new DNA by using a short single-stranded DNA template known as a primer (Saiki *et al.,* 1985; Mullis and Faloona 1987; Mielke *et al.,* 2011). It typically produces millions of copies of

little quantities of DNA (Bartlett and Stirling 2003). Its importance comes from the fact that a specific sequence of interest usually forms a very tiny portion of the entire genome, masking its signal. As such this technique provides a direct, cheap and rapid tool for the amplification of that particular sequence (Jobling *et al.,* 2004). Thus, this makes the PCR a popular technique with a wide range of applications in basic and clinical research. It is also robust allowing the possible amplification of DNA from bad damaged or degraded tissues or cells for forensic or anthropologic purposes.

The procedure involves the denaturing of a DNA template by heating in a 'thermal cycler' which is a programmable heating block. The DNA template and reactants (oligonucleotides, primers, a thermostable *Taq* DNA polymerase, magnesium ions, and buffer) are further cooled to a specific temperature so primers can anneal to their target sequence. The temperature is again raised in a phase known as the extension phase so that the *Taq* (*Thermus aquaticus*) adds complementary nucleotide bases producing complementary strands to the genomic template (Mullis *et al.,* 1992; Jobling *et al.,* 2004). The three process namely: denaturation, annealing and extension are repeated for a pre-set number of cycles with each cycle yielding double the amount of target sequence. The outcome is the production of enough DNA for downstream applications.

2.13 GEL ELECTROPHORESIS (AGAROSE)

The ability of different sized molecules with specific charges to migrate in an electric field has led to the identification of nucleic acids based on size. This enables the separation of nucleic acid fragments with the rate of migration being dependent on

fragment size. The most common type of gel electrophoresis uses agarose gels for analysing DNA fragments less than 25 kb in size. Fragments greater than 25 kb and up to 10,000 kb are usually subjected to pulse-field gel electrophoresis.

Tris is a common buffer used in electrophoresis because it works well to withstand pH changes under an electric current. TBE (Tris-borate-EDTA) and TAE (Tris-acetate-EDTA) are two buffers usually used in agarose gel electrophoresis with EDTA serving as a preservative. DNA is negatively charged, so if it is placed in an electrical field, it will migrate away from the negative pole and toward the positive pole. The sample is added to the gel at the negative pole and a voltage gradient is applied. The gel acts as a sieve through which the negative DNA travels within the gel toward the positive pole causing the migration to be inversely proportional to the fragment size.

2.14 RESTRICTION FRAGMENT LENGTH POLYMORPHISMS

Restriction fragment length polymorphisms (RFLPs) are a class of DNA variants that have been widely applied in human genetics research. The technology underlying RFLP analysis is based on bacterial enzymes referred to as restriction endonucleases. Restriction endonucleases are proteins that recognize a specific sequence in target DNA and will cleave the molecule across the sugar-phosphate backbone, leaving two fragments. The intimidating name for these enzymes refers to both their function and how they operate. They are "restriction" enzymes in that they first recognize viral DNA that could otherwise run rampant in the bacterial cell and then cleave it to deactivate it (Mielke *et al.,* 2011). Thus, they restrict the growth of viral DNA. A *nuclease* is an enzyme that

cleaves nucleotides, while the *endo* prefix refers to the fact that these particular nucleases cleave the DNA internally (away from the ends). Endonucleases function by cutting across the sugar-phosphate backbones of DNA molecules whenever they "read" a particular sequence of bases. The enzymes are named after the bacteria in which they are produced (e.g., *AluI* is made by *Arthrobacter luteus).* Consequently, when human DNA is confronted with these bacterial enzymes, it will be cleaved wherever the particular recognition sequence occurs. These enzymes therefore provide a powerful tool for manipulating human DNA. In particular, the enzymes can be used to cut human DNA into pieces and then insert these pieces into host chromosomes or other "cloning vectors." This is the basis for recombinant DNA ("cloning") technology, which has been a cornerstone for the Human Genome Project. In addition to their role in recombinant DNA technology, restriction endonucleases are essential for identification of RFLPs in human population genetics research. Let us presume that we have used PCR to make many copies of a particular stretch of DNA from an autosomal chromosome. If this sequence contains one *restriction site* (i.e., the recognition sequence for a particular restriction endonuclease), then when we digest the DNA using the enzyme, there will be two fragments formed from each DNA molecule.

However, if there has been a mutation within the restriction site, then the enzyme will not digest the DNA and we will have one long piece of DNA between the two primer sites. When some individuals have a particular restriction site and others lack it, this is an RFLP. RFLPs act as codominant systems in that the heterozygote, which has the restriction site on one chromosome but not on the homolog, is distinguishable from both homozygotes. This raises the issue of how RFLPs are assayed (visualized).

2.15 PHYLOGENETICS

A phylogeny is the evolutionary history of a group of entities while phylogenetics entails the study or estimation of the evolutionary history underlying biological diversity. It provides a means of the deducing existingrelationships within and among species (Harrison and Langdale 2006). These relationships are usually represented as branching, treelike diagrams with branches joined by nodes known as Phylogenetic trees or Dendrograms (Figure 2.14). The trees are either rooted, denoting common descent for all the groups from an ancestor or unrooted trees meaning there is no known common ancestor (Hall, 2004). The tips of the branches are the taxa or sequences under study representing species, families, orders, populations, they are also called Operational Taxonomic Units(OTUs). The nodes in the tree are points of bifurcation representing ancestral states or an ancestor and all branches arising from a node constitutes a clade or a monophyletic group (Figure 2.15) in which all members share a common descent from a unique ancestor and a set of common traits.

Some clades may exclude some of its descendants and are known as a paraphyletic group (Figure 2.16), such as a grouping of animals excluding man. The connecting branchesdepict the relationship between the clades while the branch length suggests changes due to evolutionary events over time. Monophyletic and paraphyletic groups have a single evolutionary origin. In contrast, polyphyletic groups result from convergent evolution, and the characters that support the group are absent in the most recent common ancestor (Kitching *et al.,* 1998). In gene families these principles approximate to orthology and paralogy (Fitch, 1970). Orthology refers to groups of genes that reveal species phylogeny. Thus, within a monophyletic gene group each species is represented by a single orthologue. In contrast, paralogues reveal the history of a gene

family. Thus, within a gene group each species may be represented by a number of paralogues. The term Cladistics being a derivative of the word ‘‘clade’’ coined from “Klados” the Greek word for branch or twig, may be applied to Phylogenetics indicating a descent from a single ancestor (Brinkman and Leipe 2001). Evolutionary studies of the genetic variety involve the comparison of sequences that have common origins (homologs). Thus, sequences that share some level of similarity are termed homologous.

According to Brinkman and Lieppe (2000), homologs are classified as orthologs, paralogs, or xenologs.

i. Orthologsare homologs produced by a speciation event. They represent genes derived from a common ancestor that diverged due to divergence of the organisms they are associated with. They possess similar functions.

ii. Paralogsare homologs produced by gene duplication. They represent genes derived from a common ancestral gene that duplicated within an organism and then subsequently diverged. They exhibit different functions.

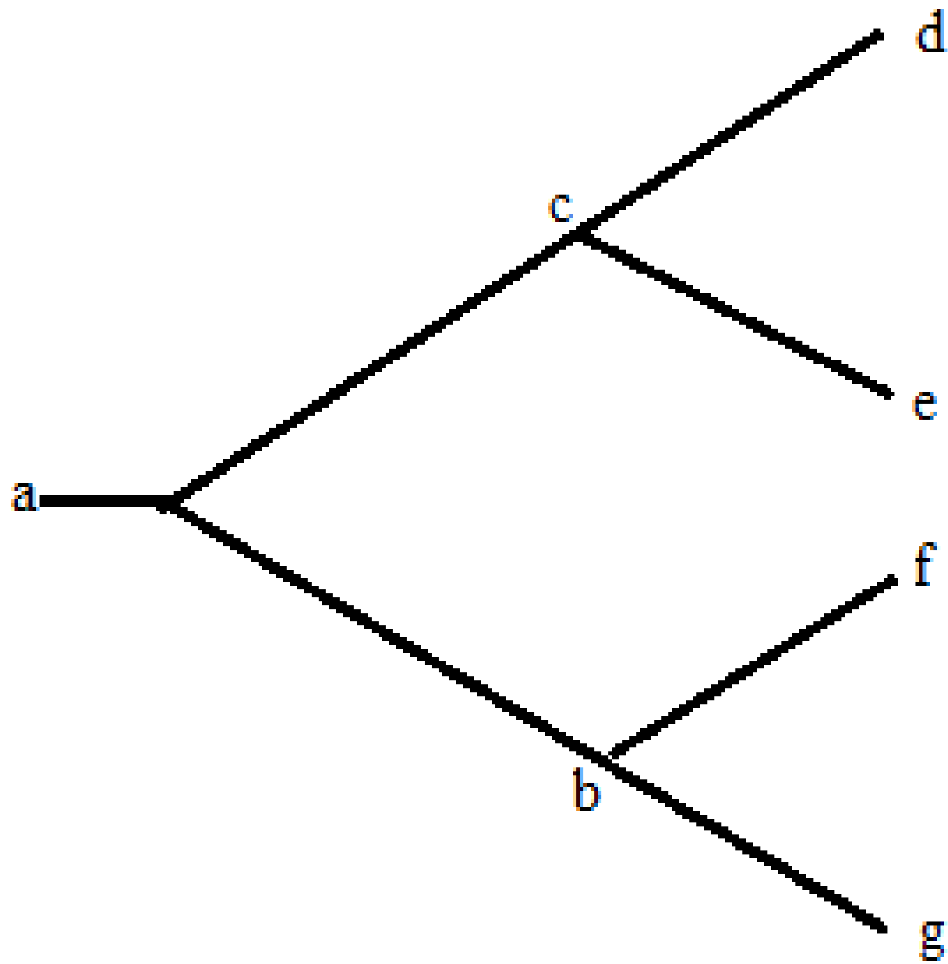


Figure 2.15: A Phylogenetic tree with a common ancestor a at the root, two other ancestral internal nodes c and b, and branches whose tips represent four taxa d, e, f and g also known as Operational Taxonomic Units (OTUs).

iii. Xenologsare homologs resulting from horizontal gene transfer between two organisms. The determination of whether a gene of interest was recently transferred into the current host by horizontal gene transfer is often difficult. Occasionally, the % (G-C) content may be so vastly different from the average gene in the current host that a conclusion of external origin is nearly inescapable, however often it is unclear whether a gene has horizontal origins. Function of xenologs can be variable depending on how significant the change in context was for the horizontally moving gene; however, in general, the function tends to be similar (Brinkman and Leipe, 2001; Hall, 2004,).

A difficulty arises when some biological processes are not represented by typical phylogenetic trees where taxa are continuously splitting without reticulations (Jobling *et al.,* 2004). These reticulations are due to recombination and gene flow between previously divergent populations. These processes result in the generation of loops within trees and these types of phylogenies are known as Networks (Bandelt *et al.,* 1995).

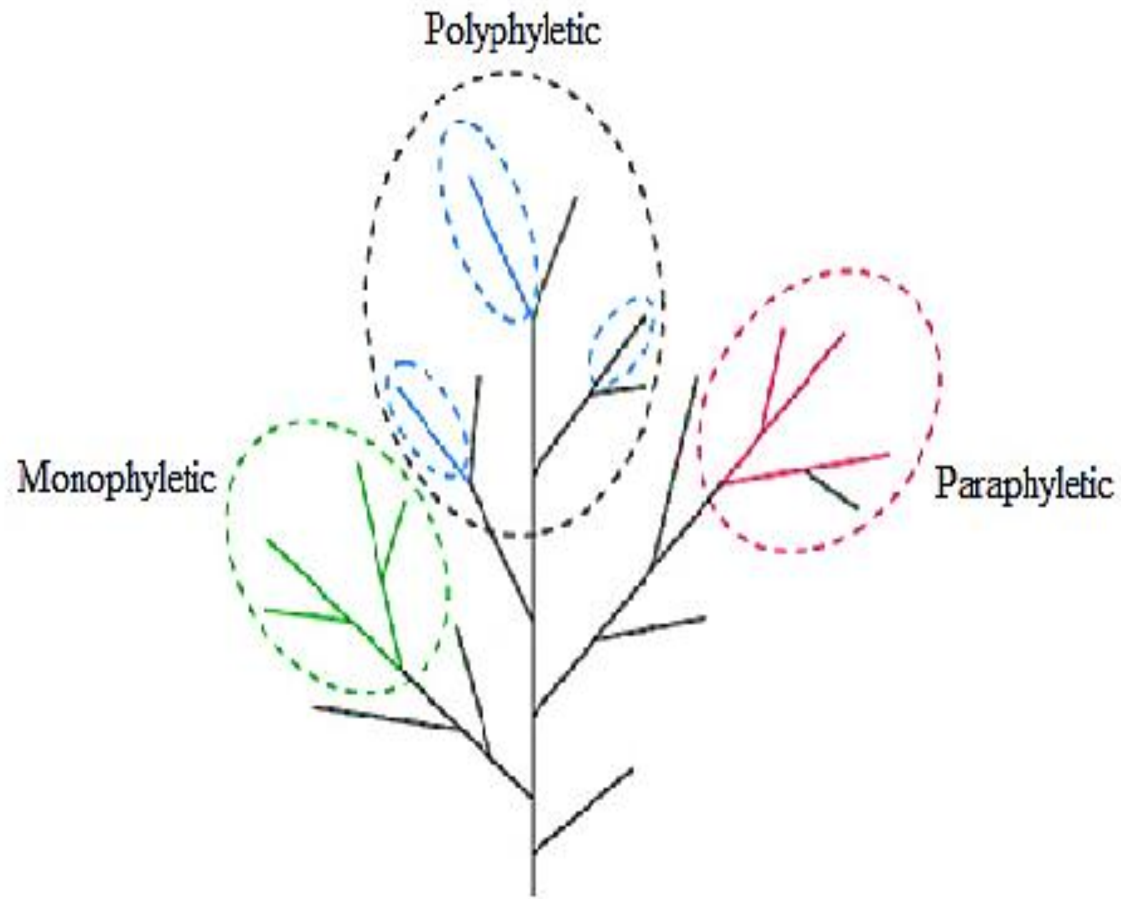


Figure 2.16: Representative phylogenetic tree with monophyletic, polyphyletic and paraphyletic clades. Adapted from Baldauf (2003).

Building phylogenetic trees using molecular data depend on the efficiency, consistency, robustness, powerful and the ability to validate the tested assumptions (Jobling *et al.,* 2004). It also generally follows two approaches, the distance based methods and character based methods (Page and Holmes 1998). The distance methods, which are iterative clustering methods such as the Unweighted-Pair Group Method with Arithmetic Mean (UPGMA), Neighbour-Joining (NJ), Fitch–Margoliash, measure percentage sequence diversity for all pairwise alignments of OTUs as a proportion of overall sequence length, then the distances are organized into an ultrametric (additive) tree (Jobling *et al.,* 2004; Hall, 2004).

The discrete data methods, also known as tree searching or character based methods such as maximum parsimony (MP), maximum likelihood (ML), Bayesian methods scrutinize each column of the alignment separately and look for the tree that parsimoniously accommodates all of this information (Baldauf, 2003; Hall, 2004). Branch lengths of character based methods represent the number of evolutionary changes along the branch. Reconstructed phylogenies can be tested statistically to assess levels of confidence using the Bootstrap method. This method investigates a phylogenetic finding using random subsets of the data to support that result. Simply put, to check for concordance between the overall tree and trees produced using about 100-1000 subsets of the data (Efron, 1982).

2.16 NETWORKS

mtDNA and a few other biological data (microsatellite) that are not well represented by a tree structure in which the branches continuously split without joining are better represented by networks. These data type produce reticulations in the form of loops thus

reuniting two previously divergent haplotypes. These types of phylogenies are known as Networks. Their advantage over traditional trees is that they are able to resolve conflicts that may produce reticulations that represent homoplasy, recombination, or sequencing errors, thus producing a parsimonious tree (Bandelt *et al.,* 1995). The method of network construction is based on the state of the data (character based or distance based). Split decomposition is employed when using genetic distance data while Minimum spanning networks and Median networks are used to define genetic relationships where character based data are used (Jobling *et al.,* 2004).

When a data set containing parallel mutations or reversions results in homoplasies, median-joining networks are preferably used. This method yields multi- dimensional reticulation with large sample sets leading to difficulty in reading such networks. This situation is managed by reducing the network through the removal of unlikely links (Bandelt *et al.,* 1995). Networks with fewer reticulations can be produced using the Median-joining method which uses smaller numbers of the most likely ancestral sequences into a phylogeny of the observed sequences (Bandelt *et al.,* 1999; Jobling *et al.,* 2004).

CHAPTER THREE

3.0 MATERIALS AND METHOD

This chapter documents the details of collection of biological specimen from the field, laboratory extraction of mtDNA, sequencing, sequence alignment and analytical aspects of this study.

3.1 SUBJECT SAMPLING

mtDNA yielding specimen (buccal cells) for this study were obtained from the Bajju, Atyap, Kagoro and Chawai found within the Zangon Kataf, Kaura and Kauru Local Government Areas of Kaduna State (Figure 3.1). Collection of specimen was done in randomly selected Secondary Schools from villages where each of the ethnic groups predominates. The villages sampled were: Madakiya, Zonkwa and Wadon where the Bajju subjects were obtained. Zangon-Kataf and Jan Kasa for the Atyap, Fadan Chawai for the Chawai and Malagun and Kagoro for the Kagoro subjects. 400 unrelated males (100 from each ethnic group) were enrolled for this study. Certification of ethnicities was based on information given by the participants on both parental sides, whose ancestors were known to belong exclusively to a specific ethnic group for at least the last two generations. This information was corroborated by community elders from each of the villages where the participants were drawn.

3.1.1 Inclusion criteria

The inclusion criteria for the sampled subjects was based on those who by definition are descended from any of the following four ethnic groups Bajju, Atyap, Kagoro and Chawai

at the level of the parent, grandparents and great grandparents and only for subjects who had no challenge with their recall ability and ascertainment from the elders. All relationships were biological. Prior to the commencement of sample taking the purpose and the significance of the research were carefully explained first, to the School management and later the participants. The participants were recruited by through verbal communication with the help of the School authorities. Specimen collection was carried out in the respective Schools over a period of one month. It was further clarified that participation was voluntary, non-invasive and the biological specimens obtained would be used solely to achieve the aims of this research. Written consents were then obtained (appendix I).

3.1.2 Ethical Approval

The Ethical Board of the Faculty of Medicine Ahmadu Bello University, Zaria; the Zonkwa Education Inspectorate, Kaduna State, approved the use of the selected schools as collection sites and the Human Subject Committee of Lawrence (HSCL) University of Kansas, USA. (appendix II), reviewed and gave approval for the protocol.

3.1.3 Comparative Populations

For interpopulation genetic analyses the obtained HSV-I data from this study were compared to some available previously published reports from sub Saharan African populations. Sequences from the African populations were extracted from the National Center for Biotechnology Information (NCBI) GenBank database (www.ncbi.nih.gov/Genbank/). The samples comprised of West-African and Eastern African HVS-I sequences. From the West are 20 Hausa, 14 Kanuri, 60 Fulbe, 10 Songhai,

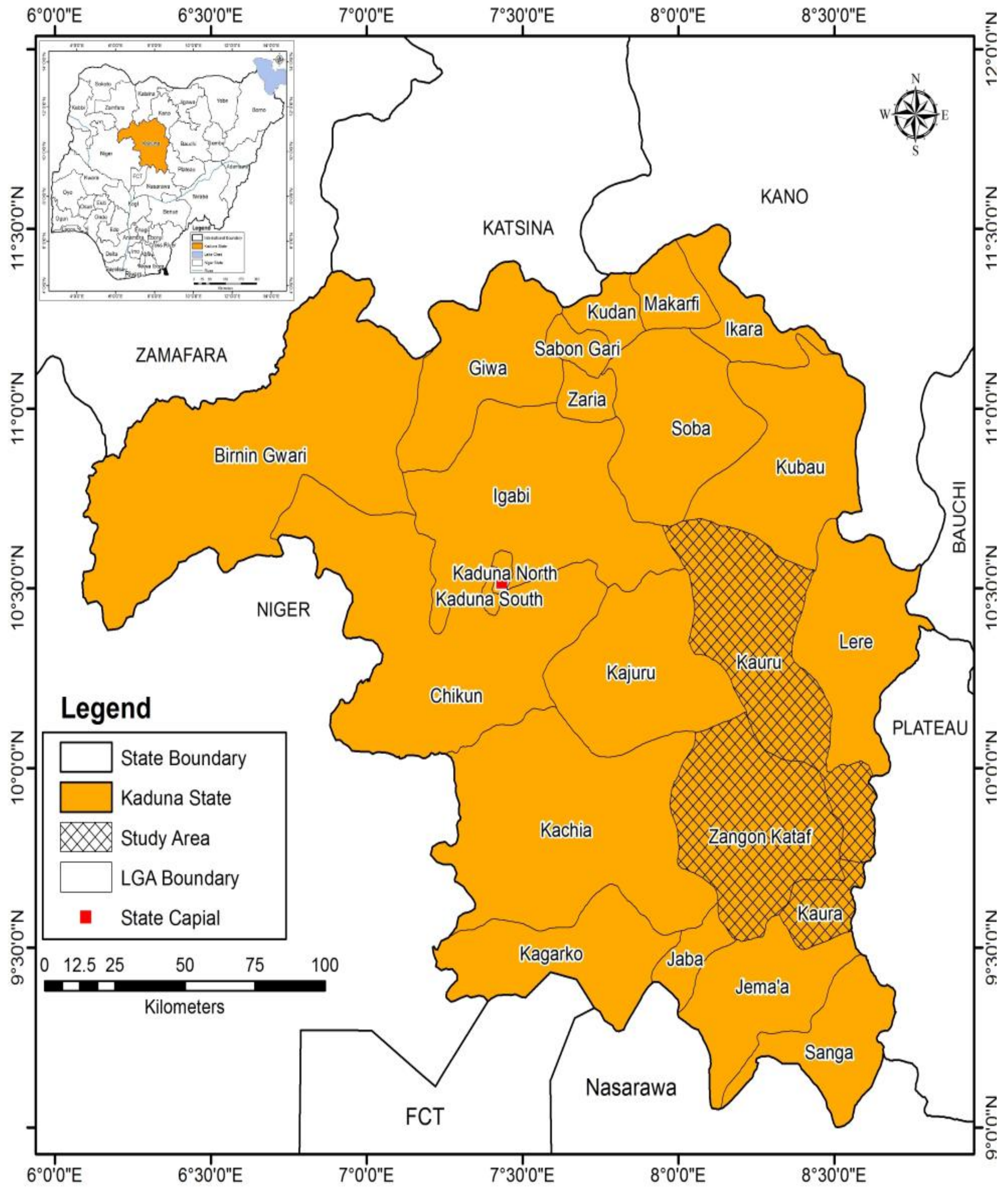


Figure 3.1: Map of Kaduna showing study area covering Kauru, Zangon Kataf and Kaura LGAs (hatched). Adapted from Administrative map of Kaduna state.

23 Tuareg, and 32 Yoruba, while the eastern populations included 37 Turkana, 27 Somali, and 24 Kikuyu (Watson *et al.,* 1996).

3.2 BUCCAL CELL COLLECTION

Buccal cells collection was done according to the protocol employed by Freeman *et al.* (1997). For each participant, sterile cotton tipped applicators (swabs) were used to scrape off the cells of the oral mucosa, the ends of the swabs were then cut, air-dried at room temperature and placed in 15 ml plastic tubes. After the swab, each subject was asked to rinse his mouth with the provided 10ml of clean water. The water was swished vigorously in the mouth of the subjects for one minute and then poured into 50 ml collection tubes. The tubes were transferred to portable cooling chambers (flasks) containing ice so that the samples were preserved at 4 °C prior to transportation and subsequent extraction.

3.3 DNA EXTRACTION AND POLYMERIZATION

The following extraction, purification and amplification of genomic DNA were carried out at the Laboratory of Biological Anthropology, University of Kansas, and Evogen Inc, Kansas City, Kansas and Lenexa, Kansas in the United States of America.

Prior to the laboratory procedures, all necessary equipment and laboratory wares (tips and tubes) were autoclaved to prevent contamination. DNA from the cells on the cotton tips of the swabs and that from the collected rinses were extracted using a commercial kit, QiaAmp® DNA Mini Kit from Qiagen®, Valencia, CA., following the manufacturer's instruction.This is a spin-column method of extraction which involves lysing the cells with

the provided Qiagen Proteinase K and lysis buffer. The lysis solution functions in that it disrupts the hydrophilic bonds within the cell membrane, effectively breaking it open so that the DNA is released into the aqueous solution. The DNA containing solution is then loaded onto spin columns with a silica-gel membrane (for DNA-binding), where the nucleic acids are attracted to the silica bead under high chaotropic salt concentrations. Impurities such as the proteins and divalent cations are removed using multiple buffer washes and centrifugation steps as illustrated in Figure 3.2. The need to remove cations for example Mg^{2+}, is due to the fact that they are needed by enzymes such as nucleases to function. As such their removal inhibits these enzymes from degrading the DNA. The pure DNA is then washed to remove impurities and eluted from the membrane into 100 µl TE buffer. In principle the formulation of the kit is based on the mechanism of affinity which results in the direct removal of DNA from solution (Cattaneo *et al.,* 2000). Another extraction method was utilized. This method employed the “one tube, one reagent, one step” Evogen One^{TM} kit formulated by Evogen Inc, which uses a proprietary salt, detergent and heat method to lyse the buccal cells samples. The cut swab tips are immersed in 200 µL of Evogen One^{TM} within an eppendorf tube and vortexed for 30 seconds to loosen the cells from the cotton tips. The tubes were then placed on a heating block at 95 °C for 10 minutes. The lysate was further centrifuged at 13,200 revolutions per minute (rpm) to pellet the cellular debris which go down to settle at the bottom of the tubes. The supernatant containing PCR-ready DNA was then decanted into fresh tubes.

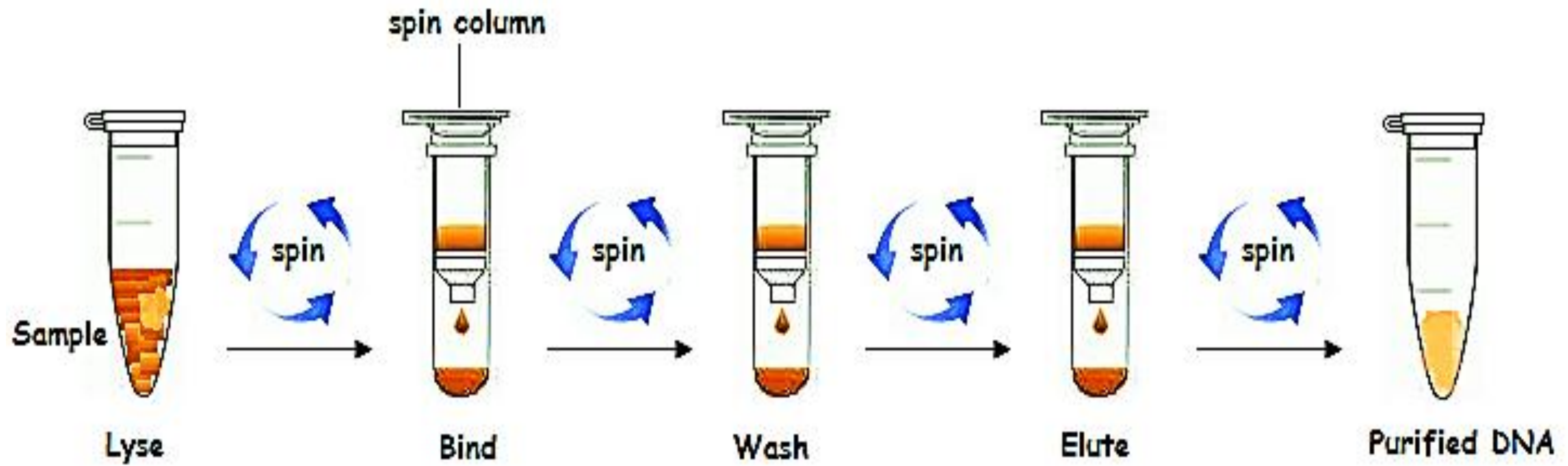


Figure 3.2: Steps in QIAamp (Qiagen) Extraction. Sample cells were lysed using Proteinase K and lysis buffer and DNA was transferred to spin columns where the DNA binds to silica-gel membrane. The cellular contaminants were removed by wash steps. The purified DNA was then eluted in an elution buffer.

Figure 3.3: Quantitation of DNA. Using photo-spectrophotometry and absorbance curves to determine the quantity and quality of DNA with the NanoDrop^{TM} 2000c (Thermo Scientific).

Sample DNA purity and quantitation were obtained through photo-spectrophotometry absorbance curves with the aid of the Thermo Scientific NanoDrop^{TM} 2000c (Figure 3.3) and adjusted for volume to obtain total yield. An absorbance ratio greater than 1.7-2.0 indicates that the absorption in the UV range is due to nucleic acids and that the DNA is sufficiently pure for further downstream application like the PCR, given that DNA absorbs light at 260 nm while impurities (Dnase, RNA and cell matter) absorb light at 280 nm. Therefore all isolated DNA with A260/A280 absorbance ratio greater than 1.75 were incorporated for analytical purposes as proposed by Hoisington *et al.* (1994); Oswald (2007).

The DNA was quantified as calculated using the formula:

$$\text{DNA concentration } (\mu g/ml) = \frac{OD_{260}\ x\ 100\ (dilution\ factor)\ x\ 33\ \mu g/ml}{1000}$$

- OD= Absorbance at 260 nm
- 33 ug = constant for single stranded DNA

3.4 AMPLIFICATION (POLYMERASE CHAIN REACTION)

Within the control region of the mitochondrial DNA (mtDNA) molecule the Hypervariable Region I (HVS1) precisely between nucleotide positions (np) 15976 to 16498 was amplified for each sample. The region was targeted using two primers, a forward primer L-15976 (5'- CCA CCA TTA GCA CCC AAA GCT AAG -3'), and a reverse primer H-16498 (5'- CCT GAA GTA GGA ACC AGA TG – 3') designed by the Integrated DNA technologies (IDT). Amplification was done using the polymerase chain reaction (PCR). The PCR reaction mix consists of:

5 µL 5X buffer,

100 mM $MgCl_2$,

1 µL purified BSA (New England Biolabs Inc., Beverly, Massachusetts),

1 unit (U) Taq DNA Polymerase,

20 mM dNTPs (New England Biolabs Inc., Beverly, Massachusetts),

10 pM forward and reverse primers,

1 ng DNA template,

and molecular grade water to make it up to 25 µL volume.

The PCR was carried out on a 24-well 2400 and a 96-well 9600 Perkin Elmer® models DNA thermocycler. Temperature for the reaction profile was an initial denaturation step of95 °C for 11 minutes, followed by 40 amplification cycles for denaturation at 94 °C for 30 seconds, annealing at 55 °C for 30 seconds, and extension at 72 °C for 30 seconds with a final extension step of 72 °C for 10 minutes with a final soak or hold at 74 °C for 5 minutes as represented in Fig 3.4.

Primer annealing temperature were set 2 °C to 5 °C below the lowest corresponding melting temperatures of the individual primers according to Devor (2004) and calculated by the given formula:

$$Tm = [2(A + T) + 4(G + C)]\ ^{\circ}C$$

Where A, C, G and T represent the number of each type of nucleotide exhibited in a primer sequence.

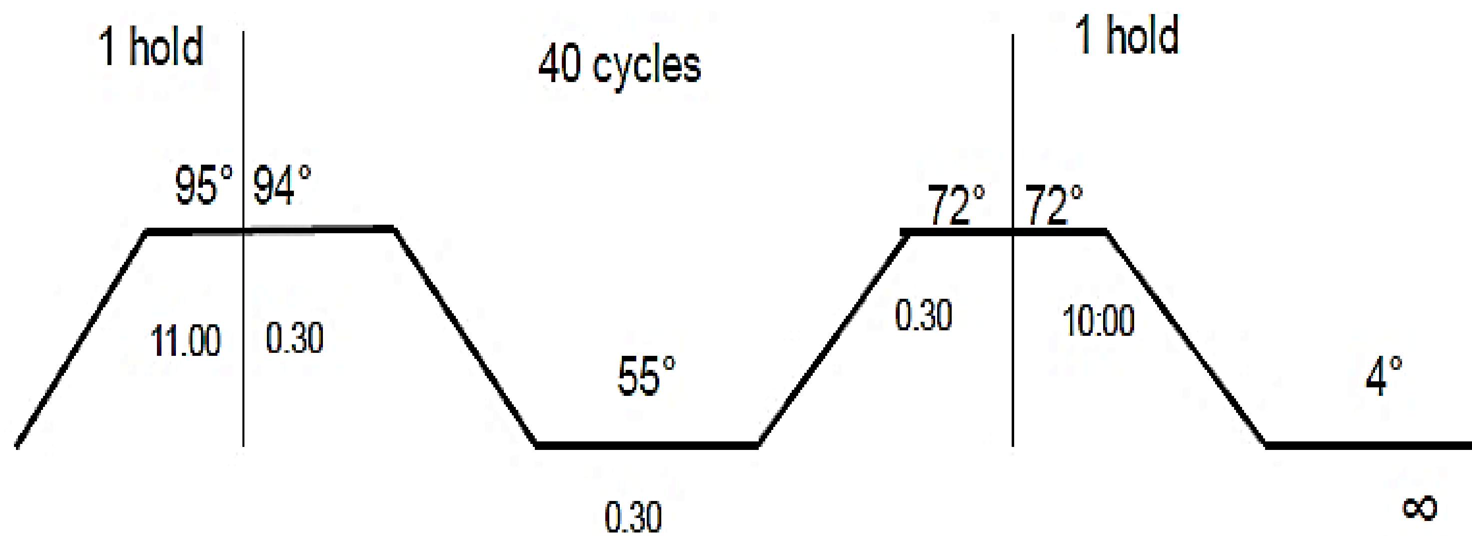


Figure 3.4: Thermal profile for HVS-I amplification. An initial denaturation step of 95 °C for 11 minutes was followed by 40 amplification cycles. Primer annealing was set at 55 °C for 30 seconds, and extension at 72 °C for 30 seconds with a final extension step of 72 °C for 10 minutes with a final hold at 4 °C.

3.5 ELECTROPHORESIS PROCEDURE

3.5.1 Preparing the Gel

The gel used in this study was from 1.5 % Agarose. To make 1.5 % gel, a 2.25 g SeaKem® LE Agarose (Cambrex, East Rutherford, NJ) was measured and poured into 150 ml of 1xTBE within a conical flask. The top of the flask was covered with foil paper with a tiny opening to prevent excessive build-up of pressure during heating. The mixture was shaken slightly and microwaved to dissolve the agarose. After bringing it to a boil, it was removed and allowed to cool slightly and then re-weighed to make sure the volume of liquid had not reduced considerably during heating due to evaporation. In case of any reduction the volume was made up to the original volume with distilled water. This ensures the right agarose concentration and same buffer composition for the gel and electrophoresis buffer. 8 µl of ethidium bromide (acting as a dye by intercalating between the DNA strands to fluoresce) were added and the flask was shaken to mix thoroughly while taken necessary precautionary measures to prevent spillage as ethidium bromide is a known carcinogen. The mixture was cooled to 55 °C using a water bath and a magnetic stirrer. The cooled mixture was then poured into the gel casting plate that had been tightly sealed at the sides with making tape and two 24-well combs were inserted into the molten solution at the top and middle holder of the casting plate to create the wells while the gel was left to set for about 30mins. After the solidification of the gel, it was wrapped in foil paper and kept in the refrigerator until needed.

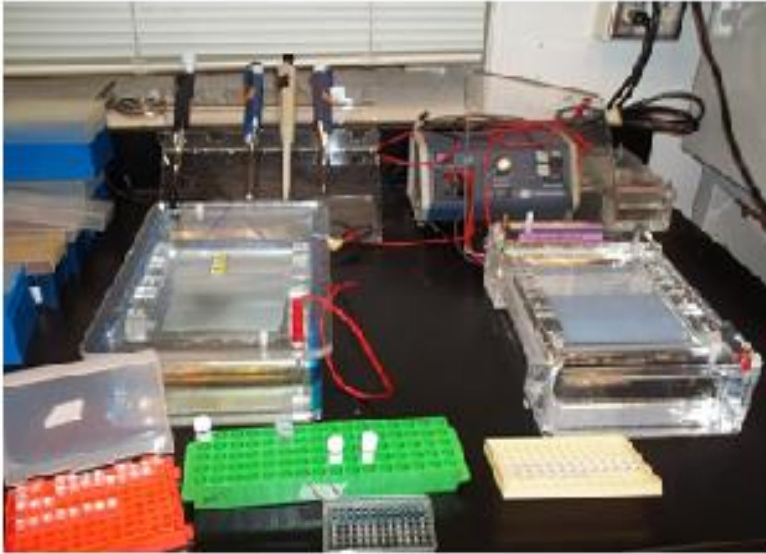
3.5.2 Running the gel

To check for successful amplification, the TBE buffer was added to the electrophoresis tank and the gel within the casting plate was placed immersed in the buffer making sure the top of the gel lay at least 1mm below the surface of the buffer. Then to 5 µl of each sample,2 µl of the loading dye, bromophenol blue (New England BioLabs) was added and the sample placed in each well. A molecular marker to track the migration of the samples through the gel was placed in the outermost well, made up of a mixture of 5 µl of the 50 bp DNA ladder (New England BioLabs) and 2 µl of the loading dye. The tank (Figure 3.5) was covered to prevent evaporation and the gel was allowed to run at 97 V for about 1 hour for maximum resolution. Visualization of DNA bands was done by placing the gel on a UV illuminator (254-366 nm) and subsequently photographed on Polaroid.

3.6 PURIFICATION OF PCR PRODUCT AND SEQUENCING

After verifying the presence of DNA through the gel, the amplicons were purified to remove excess dNTPs and unbound primers using ExoSAP DNAse. The ExoSAP purification procedure requires diluting 30µl of stock into 120 µl of molecular grade water. For every 15 µl of PCR product, 1.2 µL of ExoSAP was added and ran in the thermocycler at 37 °C for 30 min, followed by 80 °C for 15 min, ending with a hold temp of 4 °C. After purification, samples were re-labeled and sent for sequencing at the Museum of Natural History, University of Kansas.

The sequence chromatograms were viewed and aligned with BioEdit ver 7.2.0 software (Ibis Therapeutics) (Hall, 1999) against a standard mtDNA sequence that was



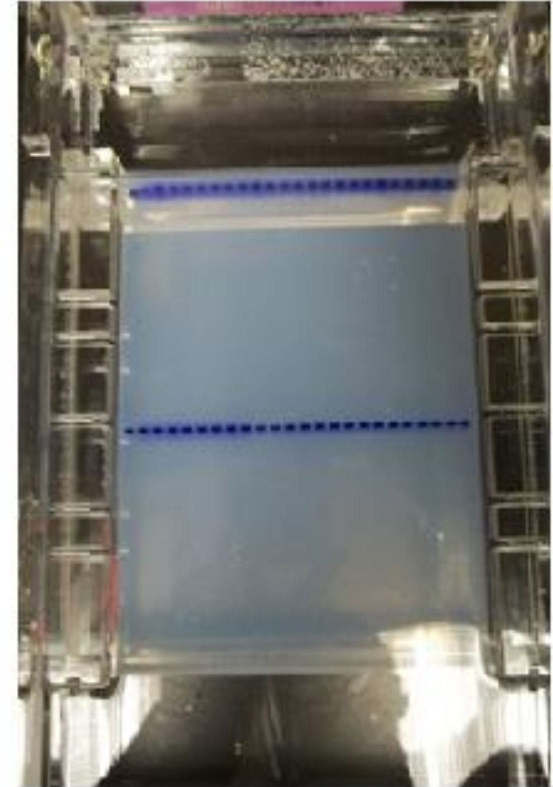


Figure 3.5 Electrophoresis tank with the loaded gel. The results are analyzed quantitatively to visualize extracted DNA with Ultraviolet light

obtained from the revised Cambridge Reference Sequence (rCRS) (Anderson *et al.,* 1981; Andrews *et al.,* 1999).

3.7 HVS-I SEQUENCING

Nucleotide positions (nps) 16,050 to 16,460 were amplified by PCR at the LBA (primer pair 15976 FOR and 16498 REV) and sequenced for the light strand using the Big Dye Terminator cycle sequencing kit on an ABI® PRISM™ 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA) as shown in Figure 3.6. This sequencer uses the Sanger dideoxy method (Sanger *et al.,* 1977). It employs an automated capillary electrophoresis system involving a PCR reaction to polymerize DNA strands using a mixture of dNTPs and dideoxynucleotides (ddNTPs). The reaction stops when a ddNTP is incorporated to the end of a growing strand.

3.8 RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)

In case of ambiguity in defining mtDNA haplogroups on the basis of the HVS-I haplotype, additional data was gathered from restriction fragment length polymorphisms (RFLPs) of diagnostic sites. All restriction digests were made according to the manufacturer's instructions (Fermentas and New England BioLabs).

The PCR amplicons were digested for 10-18 hours at 37 °C with the appropriate restriction enzymes (*Mbo1*and *HpaI*) shown in Table 3.1 to distinguish the subclades of the haplogroup L.

Figure 3.6: ABI® PRISM™ 3730 Genetic Analyzer (Applied Biosystems Foster City, CA). This 48-capillary tube analyzer is used for DNA fragment analysis and traditional DNA sequencing.

Table 3.1 Restriction enzymes, recognition sequences and their cut sites

Restriction Enzyme	Recognition sequence	Cut site
Haemophilus parainfluinzae (*HpaI*)	5' GTTAAC 3' CAATTG	5' ---GTT AAC--- 3' 3' ---CAA TTG--- 5'
Moraxella bovis (*MboI*)	5' GATC 3' CTAG	5' --- GATC--- 3' 3' ---CTAG --- 5'

Restriction sites characterising the clades of L3 were tested in this study, L3d was defined by the −8616 *Dpn*II site, L3e by +2349 *Dpn*II (Rando *et al.,* 1998) and L3b by 10084 *Taq*I (Chen *et al.*, 2000). The restriction reagents per sample are as follows: 2.0 µL of 10X buffer (provided by the manufacturer and varies according to restriction enzyme); 1.0 µL of 100X bovine serum albumin (BSA); 0.5 µL of enzyme, 9.0 µL of ddH_2O and 7.5 µL of PCR product for L3888 FOR/L3745 REV. The digestion reactions were stopped by the addition of 5 µL of 3X loading dye. The restricted DNA was loaded unto 3 % 3:1 NuSieve® agarose gel (ISC BioExpress, Kaysville, Utah) stained with ethidium bromide and then run at 100 volts for about 2 hours. The gels were visualized using a UV illuminator to check for the presence or absence of restriction sites.

3.9 PRECAUTIONARY MEASURES

To prevent the risk of contamination between samples or from personal laboratory handling, the following were done: all samples were labelled with care, gloves were worn throughout, all tubes, tips, holders and related wares were autoclaved before use to ensure they were sterile and there was no talking over the samples to avoid saliva spills. PCRs were performed in a different room under a flow hood. All samples were compared to positive and negative controls to further ensure the quality of extraction.

3.10 HAPLOGROUP ASSIGNMENT

HVS1 sequences from nucleotide positions 16,050 and 16,460 were compared to the revised Cambridge Reference Sequence (rCRS) to determine key diagnostic SNPs that could be used for mtDNA haplogroup assignment. Position numbers were relative to the rCRS and the scoring method was as specified by Bandelt and Parson (2008). Mutations were identified by the three last digits of their positions in the reference sequence (Anderson *et al.,* 1981), as for transversions the variant base was specified by an additional letter. For haplogroup assignment the SNPs were sorted following the nomenclature put forward by Salas *et al.* (2002) based on the control HVS- I mutations with some modifications and verified by entry into the National Geographic's Genographic Project haplogroup predictor (http://nnhgtool.nationalgeographic.com/classify).

3.11 DATA ANALYSIS

3.11.1 Genetic Diversity

Differences between the sequences were determined by comparing nucleotide positions relative to the rCRS. The differences were categorized as substitutions (transitions and transversions) and length alterations (insertions and deletions) following which the average number of nucleotide pairwise differences and the polymorphic positions known as the number of segregating sites (S) within the sequences were evaluated. All intra population diversity analyses were done using DNAsp 5.10 (Librador and Rozas 2009), XLSTAT 4.06 v13 andArlequin 3.513 package (Schneider *et al.,* 2000; Excoffier *et al.,* 2010).

3.11.2 F_{st} Distance Matrix

This distance matrix was used to measure genetic distances of the study populations. Its values range from 0 to 1. Values from 0 - 0.05 represent little genetic differentiation among populations, values from 0.05 - 0.15 indicate moderate genetic differentiation, values from 0.15 - 0.25 reflect high differentiation, and values above 0.25 indicate great differentiation.

3.11.3 Intra-population Diversity Measures

The amount of variation within each study population was determined using the Nei's gene/haplotype diversity (H) or expected heterozygosity and nucleotide diversity (π), or the number of nucleotide diversity per site (Nei, 1987). It measures the probability that two alleles drawn at random from the population are different from each other.

3.11.4 Neutrality Test

Tajima's D (Tajima, 1989) and Fu's F were used to test for any departure from neutrality. The expectation is that for a panmictic population under the neutral mutation model, there should be no difference between the average number of nucleotide differences and the number of segregating sites.

3.11.5 Interpopulation Diversity Measures (Networks)

mtDNA genealogies were constructed using median-joining network approaches (available in the Network 4.6 program, from the Fluxus Engineering Web site) (Bandelt *et al.*,

1995).Phylogenetic relationships among HVSI and genomic mtDNA sequences were established using the reduced median network algorithm (Bandelt *et al.,* 1995).

3.11.6 Mismatch Distribution

This is a common method of representing diversity known as the distribution of pairwise differences between sequences, mainly known as mismatch distributions (Rogers and Harpending, 1992). Apart from providing a description of the diversity that exists between populations, the shape of the distribution allows populations expansions to be detected. A raggedness index (r) distinguishes between the unimodal and multimodal distribution and was calculated as the sum of the squared difference between neighbouring peaks (Harpending, 1994). DnaSP 5.10 was used for mismatch analysis.

3.11.7 Analysis of Molecular Variance (AMOVA)

This is a variant of the Analysis of variance (ANOVA). It tests the hypothesis that for any two populations, their genetic diversity is not significantly different from that which would result from merging the two populations (Excoffier, 1992). The statistic partitions the total genetic variation into two additive parts, a part between populations and a part within populations (Mielke *et al.,* 2011). The outcome is the Φ_{ST} which is comparable to Wright's F_{ST}, which is the ratio of the between-group mean square to the total mean square (Cockerham, 1973). Kimura two parameter distances with a gamma correction were used for the sequence data following the methods of Kimura 1980; Excoffier and Yang, 1999; Meyer *et al.,* 1999. The values of Φ_{ST} range between 0 for panmictic populations and 1 for

genetically divergent or dissimilar populations. The significance of the test was performed using a non-parametric permutation approach as specified in Excoffier *et al.* (1992).

The output from the reduced median network analysis was used for input in constructing the median joining network in order reduce any large phylogenetic unrealistic reticulations in the network. For the relationship between the genetic diversity and language, the populations were analysed based on the major linguistic groups and a p-value less than 0.05 was considered significant. Arlequin 3.5.1.2 (Excoffier and Lischer, 2010) was employed for the AMOVA. The significance of test was determined by bootstrapping the molecular data 1000 times.

3.11.8 Phylogenetic Analysis

Phylogenetic analysis was performed utilising the nucleotide sequence data generated in the study. Sequences were analysed via the neighbour joining (NJ) tree using the Molecular Evolutionary Genetics Analysis (MEGA) v2.1 software programme (Kumar *et al.,* 2001). The neighbour-joining method uses genetic distances and the resultant tree is produced by successively inserting branches between a pair of neighbours and the branch tip in the tree.

3.11.9 Principal Component Analysis (PCA)

Graphical displays of genetic variation provide a comprehensible method of presenting such without much loss in information. PCA is a data reduction method to compress variance data into a few axes or variables known as the principal components (PCs). Simply put, it is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. PCs are extracted from multivariate data

frequencies with each successive PC containing a fraction of the overall variance. PCA was plotted using the excel addin XLSTAT v4.06, 2013.

CHAPTER FOUR

4.0 RESULT

This chapter presents the outcome of laboratory procedures such as isolation, amplification of mtDNA and analytical procedures.

4.1 DNA EXTRACTION

The DNA extraction was done using two methods, the spin column based extraction method (QiaAmp®) and the Evogen One™ extraction which is a novel and rapid method using a propriety salt, heat and detergent. From a total of 400 collected samples, mtDNA was successfully extracted from only 265 individuals. The loss or degradation of biological material was attributable to factors relating to storage and transportation. From Evogen One™, 231 samples were extracted while 34 samples were extracted using QiaAmp® following the manufacturers' protocols. For the final downstream analyses only 126 viable samples (Atyap 29, Bajju 35, Chawai 30 and Kagoro 32) of the total extracted were useable. For quality assurance and the need of standardization the QiaAmp® and Evogen One™ methods and purity of the extract were compared and the results obtained are shown in Table 4.1 of which the total extraction time was less than 25 minutes per 24 samples for Evogen One™ and ranged from approximately 90-120 minutes for QiaAmp® including lysis steps.The average reaction volume for Evogen One™ was 61.81 ± 52.93 µL while QiaAmp® had a higher volume at 136.32±23.28 µL. DNA quantification using the NanoDrop, showed an average total DNA yield three times higher at 16.48±24.50 µg for Evogen One™ than at 4.32 ±1.84µg for QiaAmp® ($p<0.001$).

Table 4.1: Method comparison for extraction time, final volume and DNA yield

Extraction method	Total extraction time (mins)	Final volume (µL)	total DNA yield (µg)
QIAamp ®	90-120	136.32±23.28	4.32 ±1.84*
Evogen One ™	< 25	61.81 ± 52.93	16.48±24.57*

*$p < 0.001$

To determine sample purity, the absorbance ratio (A260/A280) was calculated. It estimates how free a sample is from the presence of RNA, sheared DNA and other contaminants in the form of unreacted reagents. The presence of impurities could interfere with the evaluation of total yield of high molecular weight DNA. Figure 4.1 provides the absorbance ratios of the resulting DNA samples from the two methods both having ratios above 1.7. This indicates that the sample is viable for downstream applications.

4.2 HVS-I SEQUENCING AND GEL ELECTROPHORESIS

HVS-I segments of the control regions of 126 individuals were successfully amplified from the total sample population. The process utilised primer sets that generated 410 bp fragments when separated on gel as illustrated in plate 4.1. However, the sequencing reactions of certain samples failed due to unforeseen reasons not restricted to preservation of samples or PCR optimization.

The viable PCR amplicons were visualized on low resolution agarose gel in comparison with samples of known size acting as a positive control and a 50 bp molecular weight marker for the estimation of the amplified product size. The electropherograms or chromatograms (Figure 4.2) produced were further read using the BioEdit software to check for the presence of background noise from overlapping nucleotide peaks and incidences of “dropped” sequences. Signal intensity numbers of the electropherograms were checked to ensure high quality; this is to reduce identification errors due to mis-call of the nucleotides.

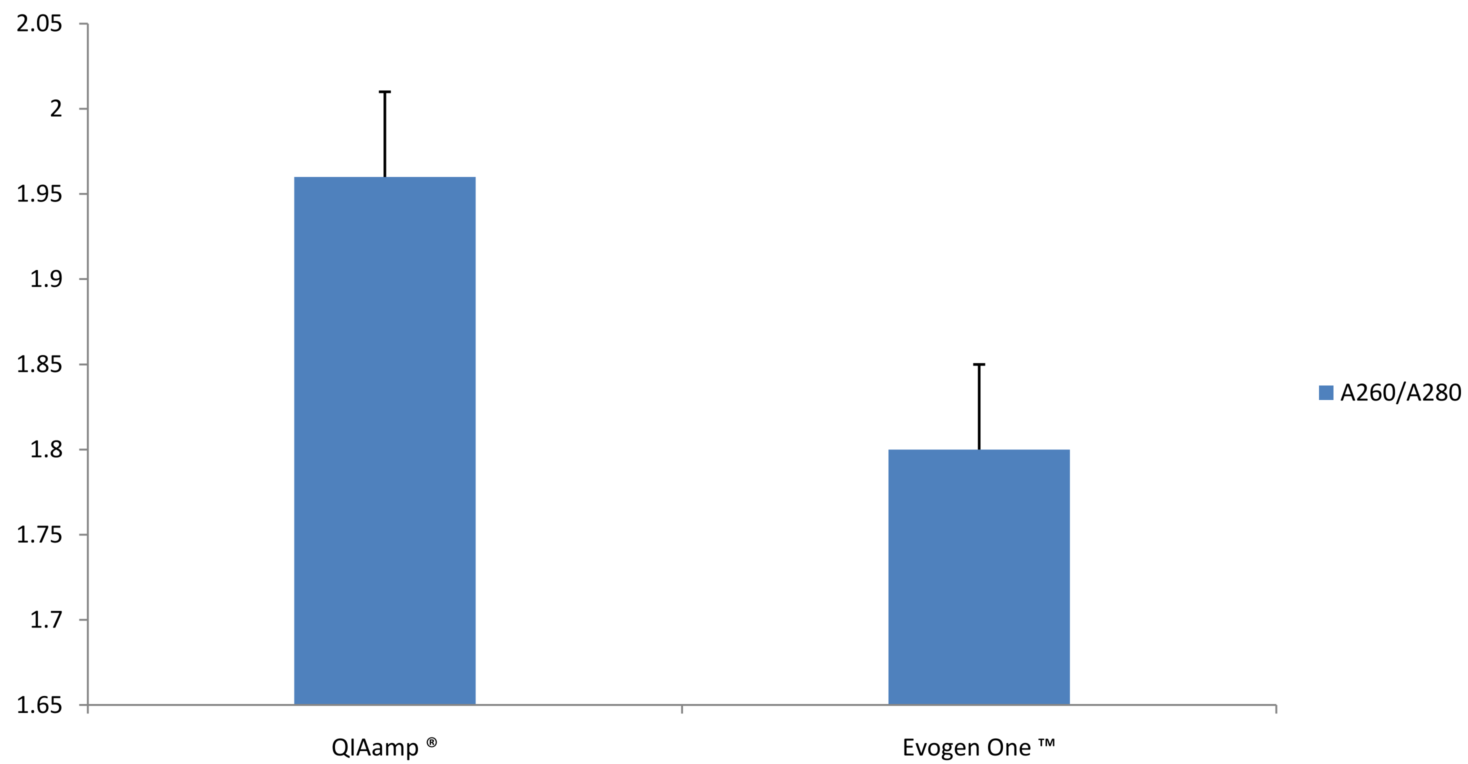


Figure 4.1: A260/A280 absorbance ratio between QIAamp® and Evogen One™ extracted DNA.
A260/A280 > than 1.7 indicates sample purity

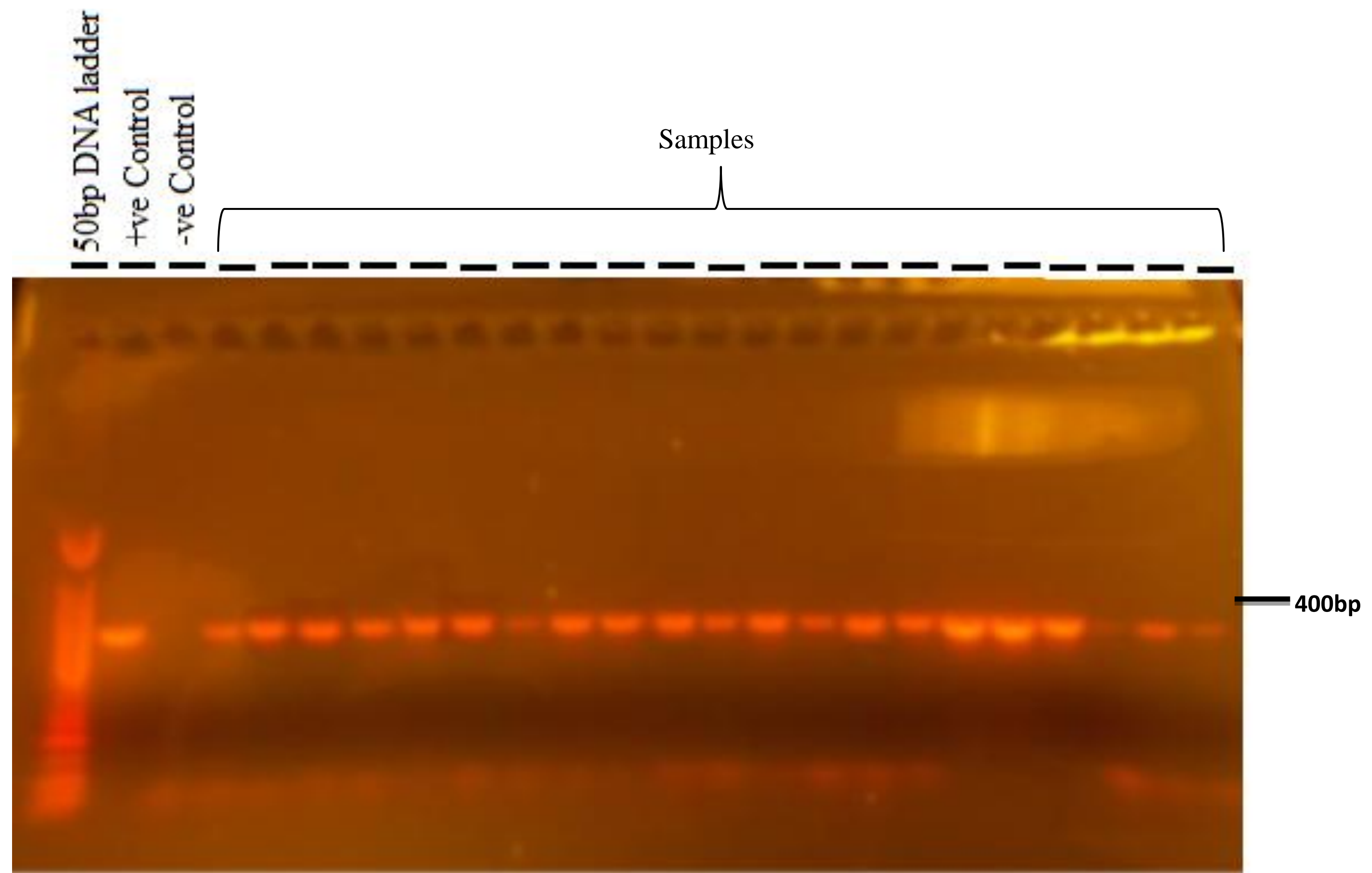


Plate I: PCR amplicons visualized on a 2 % agarose gel. The first lane is the DNA ladder while the second lane and third lanes represent positive and negative controls respectively. Subsequent lanes contain DNA of the study samples**.**

Nucleotides are represented as colour coded peaks to aid identification. The sequences were trimmed and globally aligned with the reference sequence using the BioEdit 7.2.0 software (Hall, 1999) as shown in Figure 4.3.

Following the alignment, the sequences were trimmed with the analyzable sequences falling between nucleotide positions 16050 and 16460 with some bases falling outside the range of the rCRS sequence as analyzed by Anderson *et al.* (1981).

4.3 HVS-I SEQUENCES AND HAPLOGROUP CHARACTERIZATION

The sequence data for the HVS-I region from the four sampled ethnic groups are presented in Table 4.2. All samples were assigned to haplogroups by comparing to the rCRS motif (Anderson *et al.,* 1989) using the criteria developed Chen *etal.* (1995), Watson *et al.* (1997), Bandelt *et al.* (2001), Torroni *et al.* (2001) and Salas *et al.* (2002) with slight modifications. Forehand knowledge that the individuals drawn for the study are autochthonous to areas within Africa led to the decision to investigate the Continent specific SNPs of the macro-haplogroup L only, for all the samples. Distinction between the clades of macro-haplogroup L was based on restriction fragment length analysis using *HpaI, MboI* cutting sites andHVS-Imotifs. According to Chen *et al.* (2000) L0, L1 and L2 are defined by the presence of a restriction site for *HpaI* while the L3 sequences possess a restriction site for *MboI* as seen in Table 4.3.

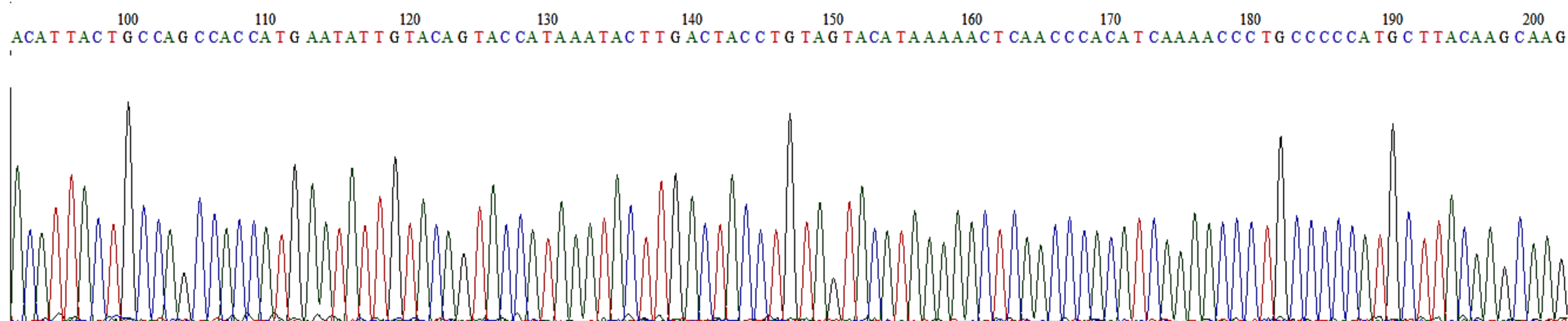


Figure 4.2: Representative mtDNA sequence chromatogram. Nucleotide peaks intensities and reading positions are colour coded to aid identification of sequenced nucleotide bases.

```
                    10        20        30        40        50        60        70        80        90        100
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
rCRS         TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a3_L15976    TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATCGTACGGTACCATAAATACTTGACCACCTGTAG
a5_L15976    TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a12_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a15_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAG
a18_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAG
a21_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATCGTACGGTACCATAAATACTTGACCACCTGTAG
a24_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a31_L15976   taccAccCaagtattGactCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAG
a36_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a53_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a64_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATCGTACGGTACCATAAATACTTGACCACCTGTAG
a65_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATGTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a75_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a83_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a84_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a85_L15976   taccAccCaagtattGactCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAG
a87_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a89_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCCATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a90_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATCGTACGGTACCATAAATACTTGACCACCTGTAG
a92_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATGTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a95_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a97_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a100_L15976  TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a47_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACCACCTGTAG
a49_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAG
a51_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATCGTACGGTACCATAAATACTTGACCACCTGTAG
a56_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
a58_L15976   taccAccCaagtattGactCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAG
a60_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b2_L15976    TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACTACCTGTAG
b8_L15976    TACCACCCAAGTATTGACTCACCCACCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b9_L15976    TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACTACCTGTAG
b22_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTAACCACCTGTAG
b23_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b25_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b29_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACTACCTGTAG
b30_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b33_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAG
b34_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b36_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b35_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATCGTACGGTACCATAAATACTTGACCACCTGTAG
b37_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b38_L15976   TaccacccaAgtATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAT
b10_L15976   taccacccaaGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b12_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTAGTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b15_L15976   TaccacccaaGtattGA-TCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAG
b41_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCATAAATACTTGACCACCTGTAG
b44_L15976   TACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCATAAATACTTGACTACCTGTAG
```

Figure 4.3: BioEdit window of aligned HVS-I sequences compared to the reference rCRS sequence (topmost sequence).

Table 4.2: Haplotypes and sequence segregating sites compared to the Cambridge reference sequence.

Haplotype	Segregating sites
	1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
	6666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666666
	0000001111111111111111111111111222222222222222222222222222222222222222222222223333333333333333333333333333333333334
	6778991122244455666778888889999000111122223333333444445555666666777788889999900000111111222333333444455666788889990
	6576231446956836378263567892357489136823580124579023582568013459036826790234513469013589027024679024756028314690192
rCRS	ATATTTCCTTGGACGGACCTCACCCCTCCTCGGTCGACCCCCATCCAACACTCCAGCACCTCAACGTCCCCACCACCCGTCAGTCTAGCACTCTGCCATCTCTCTTGTGTGGGAA
a3	........C..............................T...........................T....................................C..........
a5	....T...........................T....................................C..........
a12	....T................................T.................T........................
a15	..........A.............T.C............T..G...................C....................C...............................
a18	.........C.........C....T.C............T.....................T..T..T......G........C................G..............
a21	........C..............................T...
a24	..........................CT...........T...........................T.......T.....G.............................A...
a31	.........C.........C....T.C............T.....................T..T..T......G........C...............................
a36	...................C...................T...........................T...............C..G........................A...
a53	...C......................C..G......T..T.T...A.T..T..T...C......T..T......TT..TA.G.A.....T.A...AA.........A.A.AA...
a64	........C..............................T...........................T....................................C..........
a65	.....G.................................T..............................A...T..T.....C.................T..C........G.
a75	...................C...................T...........................T...A...
a83	......................................TT.............T....................T........C....................C........G.
b2	..........A..T....TC....TGC............T..G..C....T..........................
b8	.C.....................................T...........................T.............G.............................A...
b9	..........A..T....TC....TGC............T..G..C....T..........................
b22	...........A...........................T...........................T....................................C..........
b23	.................T.....................T...........................T.T.....T.....G.............................A...
b25	..........................C............T...........................T.......T................................A..A...
b29	..........A..T....TC....TGC............T..G..C....T..........................
b30	.................................C.....T..................................T........C...............................
b33	.........C.........C....T.C............T.....................T..T..T......G........C...............................
b34	....T...C....T.............C.C..........
b36	....T...C....T..........................
b35	........C..............................T...........................T....................................C..........
c73	[illegible]
c74	..........A..T....TC....TGC............T..G..C....T..........................
c75	......................................TT.............T....................T........C....................C..........
c76	..........A................T.........T.T................A..........................C....................C...C......
c78	........C..............................T...........................T......G........C....................C..........
c79	........C..............................T...........................T......G........C....................C..........
c80	.CC.......A......................C.....T..................................T........C...............................
c84	..........A..T....TC....TGC............T..G..C....T..........................
c86	........C..............................T...........................T.......T............................C..........
c87	.....C..C..............................T...
c90	.........C..............T.C............T.....................T..T..T......G........C...............................
c94	..........A..T....TC....TGC............T..G..C....T..........................
c96	....T..T..........................
c97	......T.C..............................T...
c98	........C..............................T...........................T.......T............................C..........
k11	.....C.................................T.....T..........T.................................T........................
k22	........C..............................T...........................T......................T.............C..........
k21	..........A.............T.C............T..................C....T.T.....T.......C...................T...........
k23	....C..................................T...........................T.T.....T...................................A...
k25	....T..T........................
k26	..........A................T.........T.T................A..........................C....................C..........
k27	......................................TT.............T....................T........C....................C..........

Following the assignment into the major L macro-haplogroups, the mtDNA haplogroup frequencies for the clades of the L macro-haplogroup of the study populations are presented in Table 4.4. The overall haplotype frequencies for the Atyap, Bajju, Chawai and Kagoro revealed the L3 haplogroup was the most frequent (62 %) haplogroup for the entire study population while the deep rooting L0 and L1 had lower frequencies. A few (five) haplotypes presented SNPs that could not be placed within the Sub-Saharan macro-haplogroup L, they were identified as Non-L and found among the Bajju, Chawai and Kagoro ethnic groups but not in the Atyap.

Figure 4.4 shows further breakdown of these haplogroups into their cladistic groupings across the ethnic groups. The only lineage present for haplogroup L0 is L0a1 present in the Bajju, Chawai and Kagoro (excepting the Atyap) constituting 10 % of the general population. The relative frequencies for L0a1 in the Bajju, Chawai and Kagoro were 14.3 %, 13.3 % and 9.4 % respectively. The L1 haplogroup with a frequency of 6.3 % across the groups was the least frequent among the four major classes of macro-haplogroup L. The L1 (6.3%) clusters identified in this study were L1a, L1b and L1c clades. L1a was found only in the Kagoro with a frequency of 3%. L1b had its highest frequency among the Bajju (6.7%) and Atyap (3.4%) but not present in the Chawai and Kagoro populations. While lacking the L1b haplogroup, Chawai and Kagoro both had the presence of L1c at frequencies of 6.7 % and 3.1 % respectively. Haplogroup L2 was more frequent (18.3 %) in the pooled data (Table 4.4) than the preceding haplogroups (L0 and L1).

Table 4.3: Sequence Polymorphisms Defining mtDNA Haplotypes

mtDNA HVS-1 Haplogroup	HVS-1 polymorphism motif	Atyap	Bajju	Chawai	Kagoro	3592 *HpaI*	2349 *MboI*	10084 *TaqI*	−8616 *DpnII*
L0a1	129A 148T 172C 187C C188G 189C 223T 230G 311C 320T		*3*	*1*	*1*	+	-	-	-
L0a1	129A 148T 168T 172C 187T T188G 189C 223T 230G 289G 311C 320T		*1*			+	-	-	-
L0a1	129A 148T 168T 172C 187T C188G 189C 223T 230G 311C 320T			3		+	-	-	-
L0a1	111T 124C 223T			1		+	-	-	-
L0a1	129A 189C 223T 278T 294T 311C 360T 384A G450C				1	+	-	-	-
L1a	93C 111T 126C 187T 189C 223T 239T 270T 278T 293G 311C				1	+	-	-	-
L1b	126C 172C 187T 189C 223T 264T 270T 278T 293G 311C	1				+	-	-	-
L1b	126C 187T 189C 223T 264T 270T 278T 311C		1			+	-	-	-
L1c	93C 129A 148T 168T 172C 187T C188G 189C 223T 230G 311C 320T			1		+	-	-	-
L1c	93C 129A 187T 189C 223T A265C 278T C286G 294T 311C 360T			1		+	-	-	-
L1c	146G 209C 223T 292T 295T 311C			1		+	-	-	-
L1c	129A 187T 189C 223T A265C 278T C286G 294T 311C 360T				2	+	-	-	-
L2a	189C 192T 223T 278T 294T 309G 390A	1	1			+	-	-	-
L2a	189C 195G 208T 223T 278T 294T 309G T368G 384A 390A G391C 434A G436T		1			+	-	-	-
L2a	223T 278T 286T 294T 309G 390A		1			+	-	-	-
L2a	189C 192T 208A 223T 273A T276A 278T 294T 310A 336A 384A 390A		1			+	-	-	-
L2a	189C 223T 230G 243C 311C 436A		1			+	-	-	-
L2a	189C 223T 278T 294T 309G 384A 390A 434A		1			+	-	-	-
L2a	223T 260T 286T 294T 309G 390A			5		+	-	-	-
L2a	223T 278T 294T 309G G384T 390A C444A			1		+	-	-	-
L2a	92C 111T 223T 311C 327T			1		+	-	-	-
L2a	66G 193T 213A 223T 239T 278T 294T 390A				1	+	-	-	-
L2a	189C 192T T195G 223T 276G 278T 294T 390A				2	+	-	-	-
L2a	92C 223T 278T 286T 294T 390A				3	+	-	-	-
L2a	193C 213A 223T 239C 278T 294T 390A				1	+	-	-	-
L2c	172C 223T 278T 311C 318G 390A		1			+	-	-	-
L2c	209C 223T 292T 311C			1		+	-	-	-
L2c	189C 213A 223T 225T C228A T231A A276A 278T 294T 311C 390A			1		+	-	-	-
L2c	223T 264T 278T 390A			1		+	-	-	-
L3b	086C 189C 223T 228T 234A 237T 242T 248T 258C 270T 278T A293T 294T G303T T304A 309G 384A 389A 390A	1				-	-	+	-
L3b	124C 223T 234A 278T 362C 384A		1			-	-	+	-
L3b	124C 223T 278T 362C	2	1			-	-	+	-
L3b	223T 278T 362C	1		1	1	-	-	+	-
L3b	124C 153A 223T 261T 278T		1			-	-	+	-
L3b	93C 223T 278T 362C			1		-	-	+	-
L3b	124C 223T 278T 294T 362T			2		-	-	+	-
L3b	124C 223T 278T 327T 362C				1	-	-	+	-
L3b	124C 223T 278T 311C 362C				1	-	-	+	-
L3b	86INDEL 223T 311C 320T 399G				1	-	-	+	-
L3d	124C 223T	1				-	-	-	+
L3d	93C 124C 223T			1		-	-	-	+

Table 4.3 continued

mtDNA HVS-1 Haplogroup	HVS-1 polymorphism motif	Atyap	Bajju	Chawai	Kagoro	3592 *HpaI*	2349 *MboI*	2349 DpnII
L3e	223T 290T 327T	2				-	+	+
L3e	126C 172C 187T 189C 223T 264T 270T 278T 293G 311C T347G	2				-	+	+
L3e	172C 223T 278T 390A	1				-	+	+
L3e	167T 223T 278T 286T 294T 309G 390A		1			-	+	+
L3e	223T 311C 320T 356C 362C		1			-	+	+
L3e	223T 311C 320T		1			-	+	+
L3e	126C G156T 187T 189C 223T 264T 270T 278T 293G 311C		1			-	+	+
L3e	223T 320T 399G		3	2		-	+	+
L3e	T92G 223T 320T 399G		2			-	+	+
L3e	223T 320T 384A 399G 436A		1			-	+	+
L3e	172C A183C 186T 189C 223T 320T		1			-	+	+
L3e	185T 223T 327T		1		1	-	+	+
L3e	172C A183C 189C 192C 223T 320T			1		-	+	+
L3e	172C A183C 189C C211A 223T 228T A258C A265C A269C T276A C313A T315A 320A			2		-	+	+
L3e	124C 223T 278T 293G 311C 362C			3		-	+	+
L3e	126C 187T 189C 223T 264T 270T 278T 293G 311C			1		-	+	+
L3e	223T 327T				2	-	+	+
L3e	93C 223T 234T 256T 327T				3	-	+	+
L3e	223T 234T 256T 311C 327T				1	-	+	+
L3e	93C 185T 223T 327T				1	-	+	+
L3e	176T 223T 327T				1	-	+	+
L3f	75C 223T 278T 309G 390A		1			-	-	-
L3f	93C 209C 223T 292T 295T 311C		1			-	-	-
L3f	75C A77C 129A 209C 223T 292T 311C				1	-	-	-
L3f	209C 223T 235G 292T 311C				1	-	-	-
L3f	129A 209C 223T 292T 295T 311C				2	-	-	-
L3f	209C 223T 311C				1	-	-	-
L3f	209C 218T 223T 292T 311C				1	-	-	-
L3*	129A 187T 189C 223T 230G A265C 311C	1				-	-	-
L3*	T93G 223T C287A A293T 301T 311C 355T 362C 399G	1				-	-	-
L3*	222T 223T 248T A293T 311C 362C 399G	1		1	1	-	-	-
L3*	145A 223T 278T 362C		1			-	-	-
L3*	223T 362C			2		-	-	-
L3*	129G 192T 218T 223T 256A 311C 362C			1	1	-	-	-
L3*	129A 192T 218T 223T C256A 311C 362C G384C 399G			1		-	-	-
L3*	C114A 129A 213A 223T 278T 342C 390A				1	-	-	-
Non L	189C 223T 278T 294T 384A 390A 434A		1			-	-	-
Non L	188C 192T 223T 239T C245G A252C G273A 278T 294T G310A G336A G384A T336A G390A G391A G434A G436A		1			-	-	-
Non L	163G 187T 189C 223T C228G 240G 278T 293G 294T 311C 434A 436A		1			-	-	-

Transitions are indicated by the nucleotide position minus 16,000 followed by a nucleotide and transversions are indicated by a nucleotide prefix and suffix. Some positions fall outside the common HVS-I segment.A plus (+) indicates a site gain, and a minus (-) indicates a loss of restriction site for the restriction enzymes.

Table 4.4: Major African mtDNA L clades frequencies across the Atyap, Bajju, Chawai and Kagoro ethnic groups

Population	n	Major L clades (Haplogroups)				
		L0	L1	L2	L3	Non-L
Atyap	29	-	1	3	25	-
Bajju	35	5	3	11	13	3
Chawai	30	4	2	4	19	1
Kagoro	32	3	2	5	21	1
Total	126	12(9.5 %)	8(6.3 %)	23 (18.3 %)	78(62 %)	5(4 %)

The L2 sub-clade L2a, was present in the four ethnic groups with the Bajju having the highest frequency of 23 % while the Atyap, Chawai and Kagoro had 7 %, 13 % and 16 % respectively. The other clade L2c, was present only in the Atyap (3.4 %) and Bajju (8.6 %).

The most common haplogroup, L3 (62 %), had five daughter clades viz L3*, L3b, L3d, L3e and L3f. This group lacking the *HpaI* cut site and representing the most recent clade in origins had L3e as the most frequent haplogroup across the groups with frequencies as high as 28 % in the Atyap, 17 % in the Bajju, 23 % in the Chawai and 19 % in the Kagoro. L3b also recorded the next most common frequency being present in all the ethnic groups. Its frequency reached 34 %, 9 %, 10 % and 13 % for the Atyap, Bajju, Chawai and Kagoro respectively. L3d and L3f were the least common having frequencies of 7 % and 8 % respectively. While the L3d was absent in the Bajju, L3f was also detected in all the groups except the Atyap. The L3* haplogroup connoted by an asterisk refers to differentiated lineages but not specifically identified clades. It constitutes 21 % of the total L3 lineages. It is present in all the study populations with frequencies 21 %, 8 %, 17 % and 9 % in the Atyap, Bajju, Chawai and Kagoro respectively.

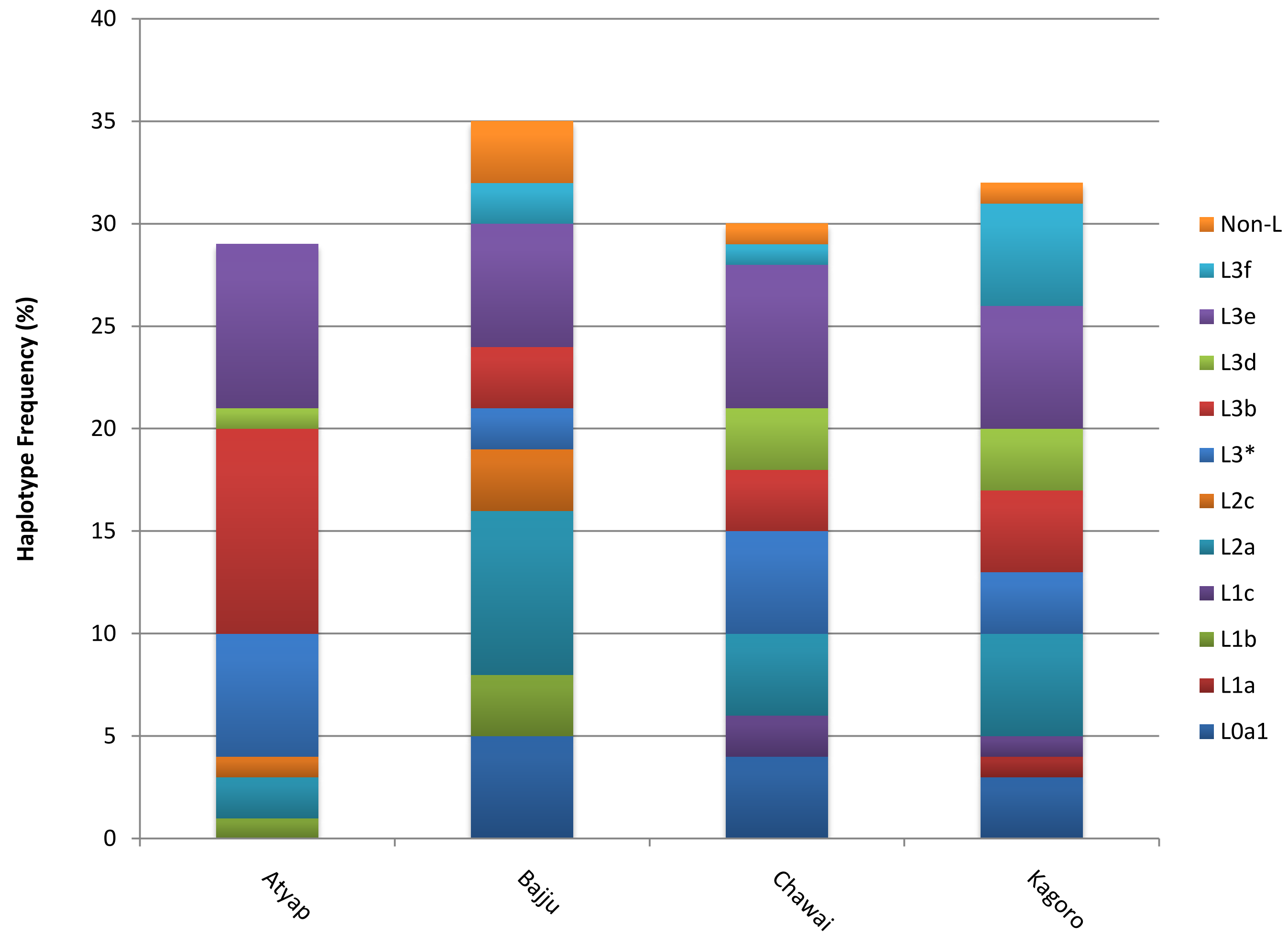


Figure 4.4: Haplogroup assignment of individuals within the four ethnic groups

4.4 DIVERSITY INDICES

Table 4.5 presents a general view of the diversity variables present within the mtDNA HVS-I sequence. They include the number of sites investigated, the total number of monomorphic and polymorphic sites observed among the four study populations. Other parameters included are the total number of singleton sites, non-informative sites, parsimony informative sites and polymorphic sites that are present at least twice.

For the four populations, there were 107 polymorphic sites (S) in all within the selected 16050-16460 region. The breakdown of diversity indices according to their ethnicities (Table 4.6) revealed 40, 68, 61 and 45 segregating sites for the Atyap, Bajju, Chawai and Kagoro respectively. 91 haplotypes were identified with a mean haplotype diversity (HD) of 0.991±0.003. The ethnic group with the least haplotype diversity was the Atyap (0.960) while the Bajju (0.992) had the most diverse haplotypes. The Nucleotide diversities (π) were 0.019, 0.026, 0.025 and 0.020 for the Atyap, Bajju, Chawai and Kagoro, showing a similar pattern with the haplotype diversity, the Atyap exhibited the least nucleotide diversity while the Bajju presented the highest. Further analysis according to the major sub-haplogroups L0, L1, L2 and L3 (Table 4.7) produced 4, 6, 16 and 48 haplotypes respectively and Nucleotide diversities of 0.5833, 0.8929, 0.977 and 0.9814 respectively.

The Jukes and Cantor model was applied to correct for the rate of nucleotide substitution. This is to make sure that the substitution rate for the four nucleotides (A, T, C and G) is the same. Thus, the model produces the maximum likelihood estimate of the number of substitutions (whether transitional or transversional) between the sequences.

Table 4.5: Descriptive data of the HVS-I polymorphisms among the four ethnic groups

Population	**Atyap**	**Bajju**	**Chawai**	**Kagoro**
Selected region	16050-16460	16050-16460	16050-16460	16050-16460
Number of sites	410	410	410	410
Total number of sites (excluding gaps or missing data	407	407	410	407
Sites with alignment gaps or missing data	3	3	0	3
Invariable (monomorphic) sites	366	338	349	362
Variable (polymorphic) sites	41	69	61	45
Total number of mutations relative to CRS	43	72	66	48
Singleton variable sites	17	36	30	28
Parsimony informative sites	24	33	31	16
Singleton variable sites (two variants)	17	34	29	26
Parsimony informative sites (two variants)	22	32	29	1
Singleton variable sites (three variants)	0	2	1	244
Parsimony informative sites (three variants)	2	1	1	2

Table 4.6: HVS-I diversity indices for the Atyap, Bajju, Chawai and Kagoro ethnic groups

Population	n	Segregating sites (S)	Haplotypes (h)	Haplotype diversity (HD)	Sd (HD)	Ave pairwise differences (K)	Nucleotide diversity(π)	Sd(π)	Nucleotide diversity (Jukes & Cantor) (πJC)
Atyap	32	40	23	0.960	0.024	7.58	0.019	0.003	0.019
Bajju	35	68	31	0.992	0.010	10.47	0.026	0.002	0.027
Chawai	30	61	22	0.977	0.014	10.12	0.025	0.003	0.026
Kagoro	29	45	25	0.990	0.012	8.13	0.020	0.001	0.021

Table 4.7: HVS-I diversity indices for the major L sub-haplogroups

Population	n	Segregating sites (S)	Haplotypes (h)	Haplotype diversity (HD)	Ave pairwise differences (K)	Nucleotide diversity(π)	Nucleotide diversity (Jukes & Cantor) (πJC)
L0	9	17	4	0.5833	4.944	0.0122	0.0125
L1	8	23	6	0.8929	8.679	0.0215	0.0219
L2	19	42	16	0.977	8.333	0.0207	0.0211
L3	70	76	48	0.9814	7.6017	0.0189	0.0192

Table 4.8: Diversity indices of comparative populations

Population	N	S	HD	K	D	mt-DNA source
West Africa						
Atyap	32	40	0.960	7.58	-1.03	This study
Bajju	35	68	0.992	10.47	-1.45	This study
Chawai	30	61	0.977	10.12	-1.47	This study
Kagoro	29	45	0.990	8.13	-1.25	This study
Hausa	20	30	0.995	5.77	-1.25	Watson *et al.,* 1996
Kanuri	14	32	0.989	6.90	-1.35	Watson *et al.,* 1996
Fulbe	60	43	0.972	6.82	-0.98	Watson *et al.,* 1996
Songhai	10	28	0.978	8.49	-0.68	Watson *et al.,* 1996
Tuareg	23	39	0.992	6.75	-1.4	Watson et al., 1996
Yoruba	32	43	0.995	7.25	-1.18	Watson *et al.,* 1996;1997
East Africa						
Turkana	37	54	0.991	9.52	-1.05	Watson *et al.,* 1996;1997
Somali	27	41	0.991	6.90	-1.32	Watson *et al.,* 1996
Kikuyu	24	45	0.993	8.17	-1.3	Watson *et al.,* 1996

a) N sample size

b) S number of segregating sites

c) HD haplotype (sequence) diversity

d) K average number of pairwise differences

e) Tajima's D statistic

Table 4.8 compares this study's diversity values with those from Watson *et al.* (1996; 1997), Torroni *et al.* (2001). Their high haplotype diversities closely matched those from other West and East African populations. This study's populations had lower haplotype diversities than the Hausa (0.995) in the West and Kikuyu (0.993) in the East. The Atyap had the least haplotype diversity within this study's population and the comparative population. The Tajima's D statistic showed negative values for all the populations. Compared to the West African groups, the East African populations (Turkana and Kikuyu) had higher number of sequence segregating sites. The Bajju and the Chawai had the highest pairwise differences with the Turkana of East Africa also following the pattern. Generally, the populations from the present study and the comparative populations appear to share similar characteristics.

4.5 ANALYSIS OF MOLECULAR VARIANCE (AMOVA)

AMOVA describes the partitioning of genetic variation (the average distance between randomly chosen haplotypes or alleles) into within and among population components. The results of the analysis are given in Table 4.9. The major language group model was adopted and the analysis reveals little variation among the populations (3.66 %) and among-populations within groups (1.72 %). The greatest amount of the variation (94.6 %) was found within the populations. The low fixation indices (F_{SC}: 0.01789, F_{ST}: 0.05381 and F_{CT}: 0.0365) further demonstrates this little variation. The fixation indices and the variance values here are statistically significant ($P<0.001$) based on the probability of observing the same or lower measure for each statistic.

Table 4.9: Results of the AMOVA analysis with populations grouped according to major language groups.

Source of variation	D.F	Sum of squares	Components of variation	Percentage of variation
Among groups	2	63.149	0.23168 *	3.66
Among populations within groups	10	88.445	0.10917	1.72
Within populations	337	2019.74	5.99330*	94.62
Total	349	2171.34	6.33415	

*Statistically significant at $P<0.001$

Fixation Indices

F_{SC} : 0.01789

F_{ST} : 0.05381

F_{CT} : 0.03658

FSC – the variance among subpopulations within groups
FST – the variance among subpopulations relative to the total variance
FCT – the variance among groups relative to the total variance

Table 4.10: Genetic distance matrix using Kimura-2p parameters.

	Fulbe	Kikuyu	Yoruba	Atyap	Bajju	Chawai	Kagoro	Somali	Tuareg	Hausa	Kanuri	Songhai	Turkana	European
Fulbe	0.000													
Kikuyu	0.051	0.000												
Yoruba	0.018	0.005	0.000											
Atyap	-0.005	0.011	0.017	0.000										
Bajju	0.016	0.004	-0.013	0.003	0.000									
Chawai	0.036	0.012	0.019	0.008	0.012	0.000								
Kagoro	0.048	0.028	0.035	0.028	0.034	0.013	0.000							
Somali	0.043	0.023	0.028	0.040	0.035	0.036	0.040	0.000						
Tuareg	0.007	0.024	0.001	0.017	0.006	0.035	0.024	0.011	0.000					
Hausa	0.006	0.038	-0.009	0.030	0.010	0.052	0.052	0.030	-0.005	0.000				
Kanuri	0.043	0.011	0.011	-0.005	0.017	0.034	0.034	0.025	0.010	0.005	0.000			
Songhai	0.032	0.022	0.006	0.021	0.011	0.047	0.087	0.056	0.016	0.016	0.007	0.000		
Turkana	0.112	0.027	0.053	0.033	0.059	0.060	0.070	0.067	0.070	0.063	-0.005	0.039	0.000	
European	-0.352	-0.516	-0.277	-0.288	-0.394	-0.353	-0.358	-0.431	-0.389	-0.240	-0.782	-0.379	-0.630	0.000

Table 4.11: Neutrality test scores for mtDNA HVS-I sequence data

	G+C content	Θ (per sequence) from Eta	Θ (per site) from Eta	Tajima's D	Fu's and Li's F	Fu's F
Atyap	0.475	10.68	0.026	-1.037*	-1.316	-9.133*
Bajju	0.473	17.48	0.043	-1.457	-2.348	-18.041
Chawai	0.473	16.66	0.041	-1.479*	-2.099	-6.181*
Kagoro	0.474	12.22	0.030	-1.253**	-1.203	-14.366*

*p<0.05, **p<0.01

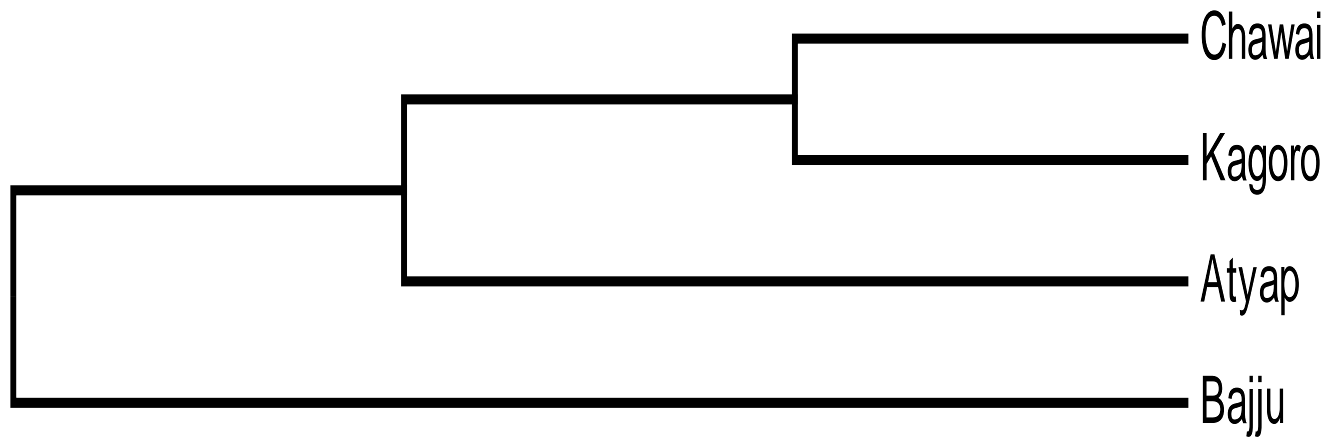


Figure 4.5: Neighbour-Joining tree (NJ) estimating the hierarchical relationship based on mtDNA data between the Atyap, Bajju, Chawai and Kagoro ethnic groups.

4.6 NEIGHBOUR-JOINING (NJ) TREE

To illustrate phylogenetic relationships among groups of related species or sequences, phylogenetic trees were constructed. The Neighbour-Joining (NJ) tree (Figure 4.5) was constructed for the study populations based on allele frequencies converted into Kimura-2p distances. These distances as represented in Table 4.10 provides the metric used in the NJ tree. This method uses the stepwise addition and star decomposition model rather than the cluster analysis to produce the nodes on a tree as against the taxa. The extracted tree is un-rooted and it consists of four major branches indicative of existing genetic variation. The clusters conform to the geographical pattern of an existing genetic differentiation.

The hierarchical topology of the tree in Figure 4.5 suggests the absence of an evolutionary root and branch lengths are less informative. The Chawai and Kagoro forming a monophyletic clade, appear more closely related. They both exhibit a shorter genetic distance to the Atyap than the more distant Bajju node.

4.7 MISMATCH ANALYSIS

The results of the pairwise differences as represented by histograms of the mismatch distribution for the major haplogroups (L0, L1, L2 and L3) displayed in Figures 4.6 – 4.9) and for the four ethnic groups as shown in figures 4.10 – 4.13. The histograms denote the frequency of pairwise differences among individuals within these haplogroups and ethnic groups. Accompanying the frequencies are the raggedness indices (r), used in the determination of the demographic history of populations. The frequency distribution for the combined ethnic groups sorted by the major West-African haplogroups showed a multi-modal distribution as represented (by multiple peaks) on the histograms except for L3.

The raggedness index for L0, L1 and L2 were 0.1489, 0.1492 and 0.0452, respectively while L3 was characterized by a smooth unimodal distribution and a raggedness index of 0.0056. Unimodal distributions were observed for the Atyap, Chawai and Kagoro yielding raggedness indices of 0.0151, 0.0135 and 0.0125 respectively while the Bajju showed a multimodal distribution with a raggedness index of 0.0324. The Tajima's D as a neutrality test determines the demographic process within populations. It measures how much deviation exists from neutrality. Even though the mtDNA is system presumed to be genetically neutral (evolving randomly), it does also reveal if the population is at equilibrium (neutral) or if expansion events have occurred or are occurring within a population. The Tajima's D relates the proportion of average number of pairwise differences to the total number of nucleotide differences. Table 4.11 shows significant negative values for both Tajima's D and Fu's Fs for the Atyap, Chawai and Kagoro except the Bajju that had non-significant negative values.

The G+C content for the light chain was almost uniform across the groups with the Bajju and Atyap showing slight deviations. Theta (θ) per sequence represents the amount of variation expected at each nucleotide site under neutral evolution. The value of θ per sequence was highest for the Bajju followed by Chawai and least for the Atyap, this pattern follows also for θ per site.

4.8 PRINCIPAL COMPONENT ANALYSIS

The Principal Component Analysis is aimed at producing a smaller number of artificial variables obtained from measures on a number of observed variables. The artificial variables are referred as 'principal components' that will account for most of the variance

in the observed variables. To place the Atyap, Bajju, Chawai and Kagoro ethnic groups within the perspective of other African mtDNA, the haplogroup frequencies of some West and East African ethnic groups were compared with those of the original study population as shown in figure 4.14. The West African populations of Fulbe, Yoruba, Tuareg, Hausa, Kanuri, Songhai and the eastern African groups were Kikuyu, Somali and Turkana were again used for the PCI. An additional European haplogroup was included to provide better resolution of the distribution. Based on linguistic and geographical distribution, the African populations show a characteristic clustering away from the single European sample. Within the African populations, the Atyap, Bajju, Kagoro and Chawai are found close to other Western-African populations. The variances are explained by 1st principal component (18.2 %) and the 2nd principal component (7.37 %).

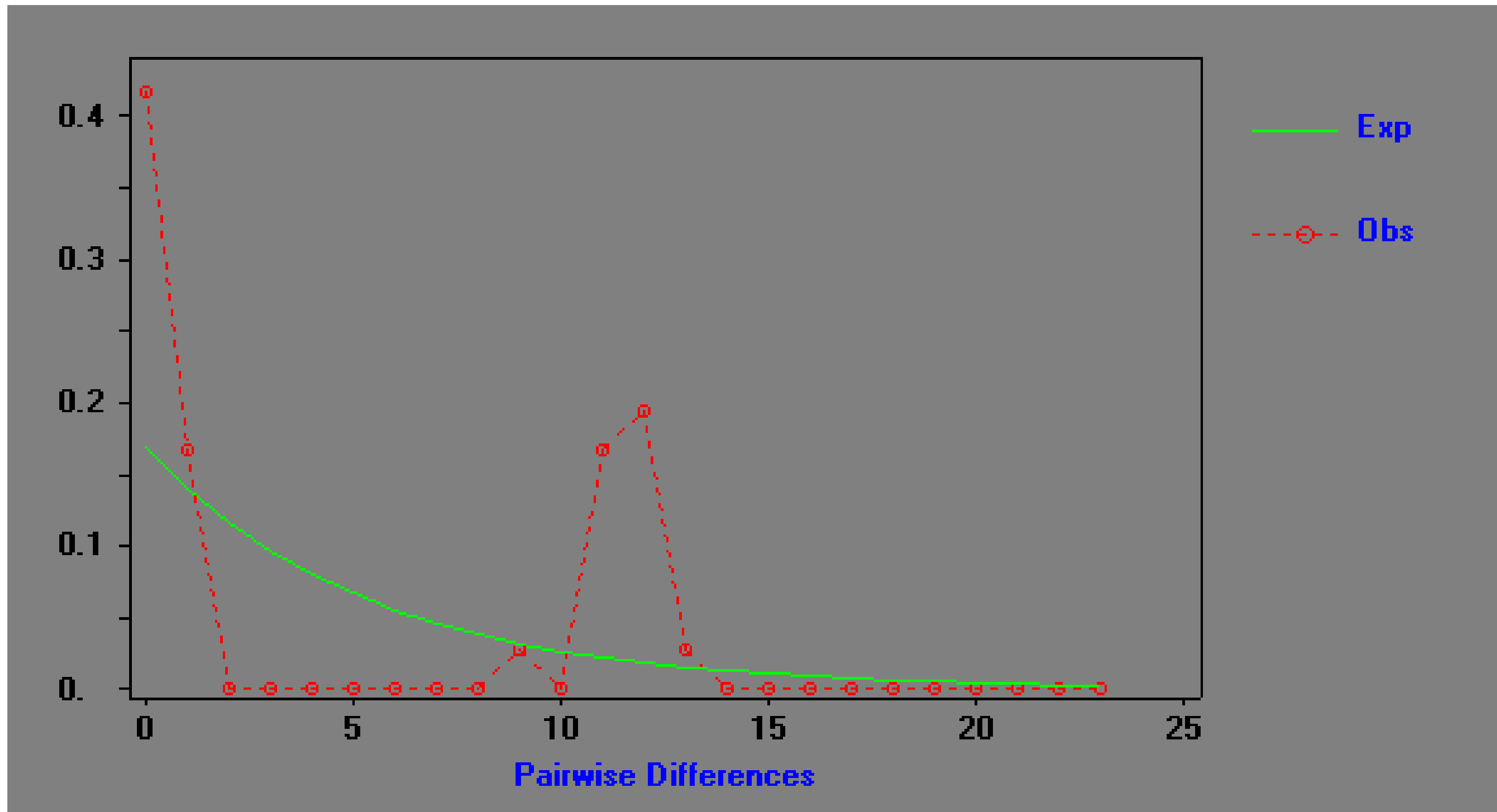


Figure 4.6: L0 haplotype mismatch distribution with multimodal peaks (Raggedness index, r : 0.1489)

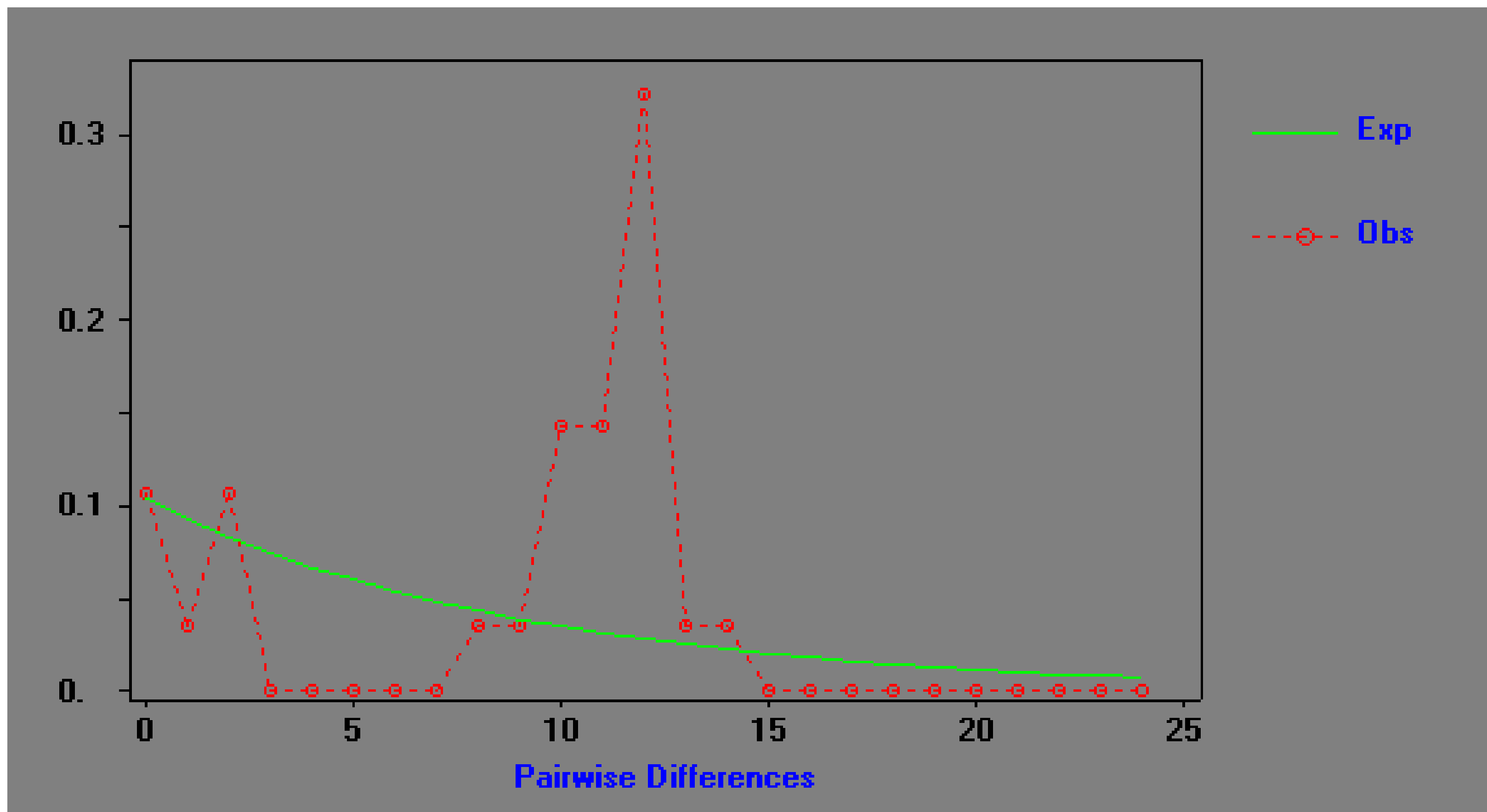


Figure 4.7: L1 haplotype mismatch distribution with multimodal peaks. (Raggedness index, r : 0.1492)

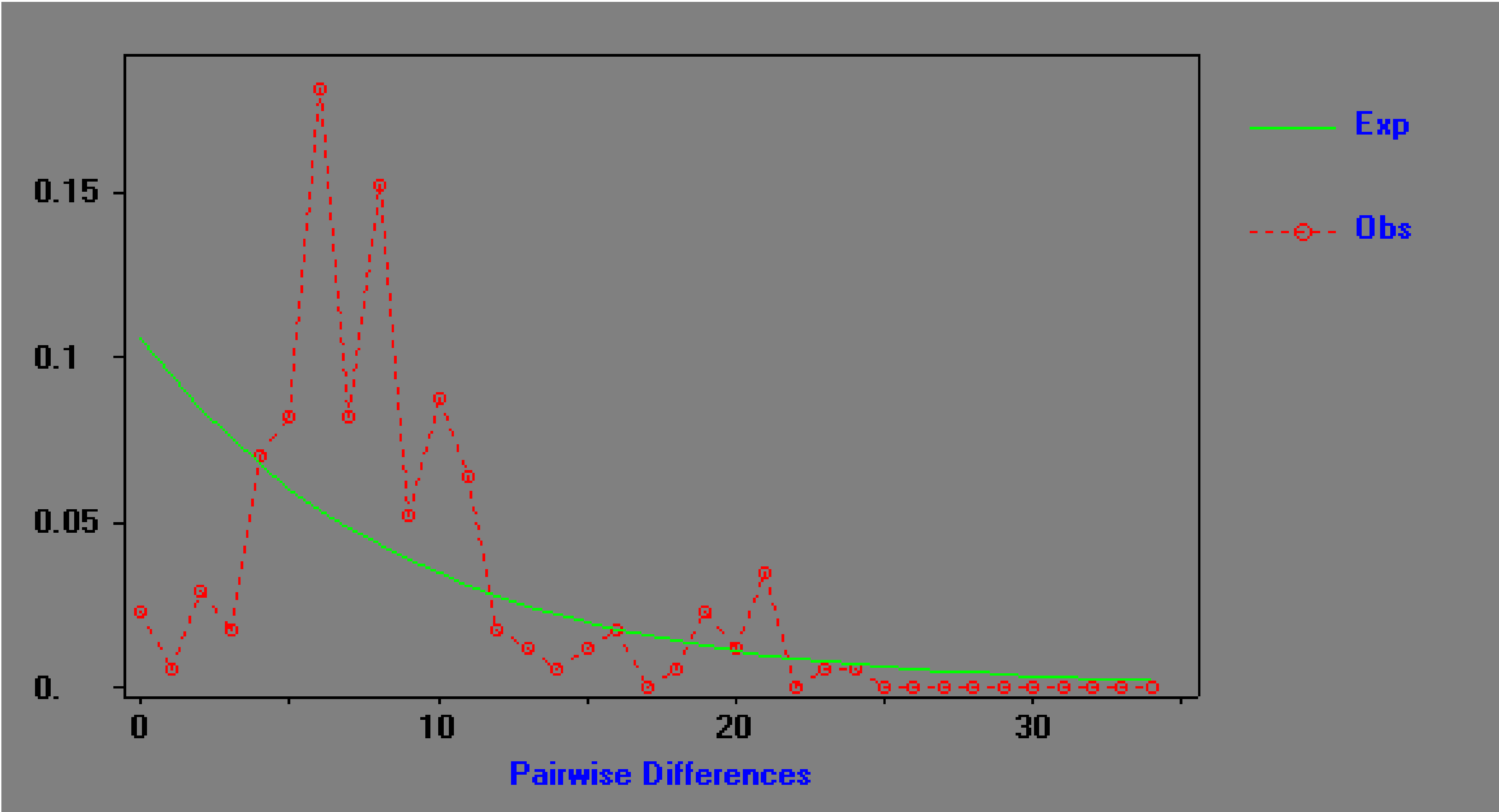


Figure 4.8: L2 haplotype mismatch distribution with multimodal peaks. (Raggedness index, r: 0.0452)

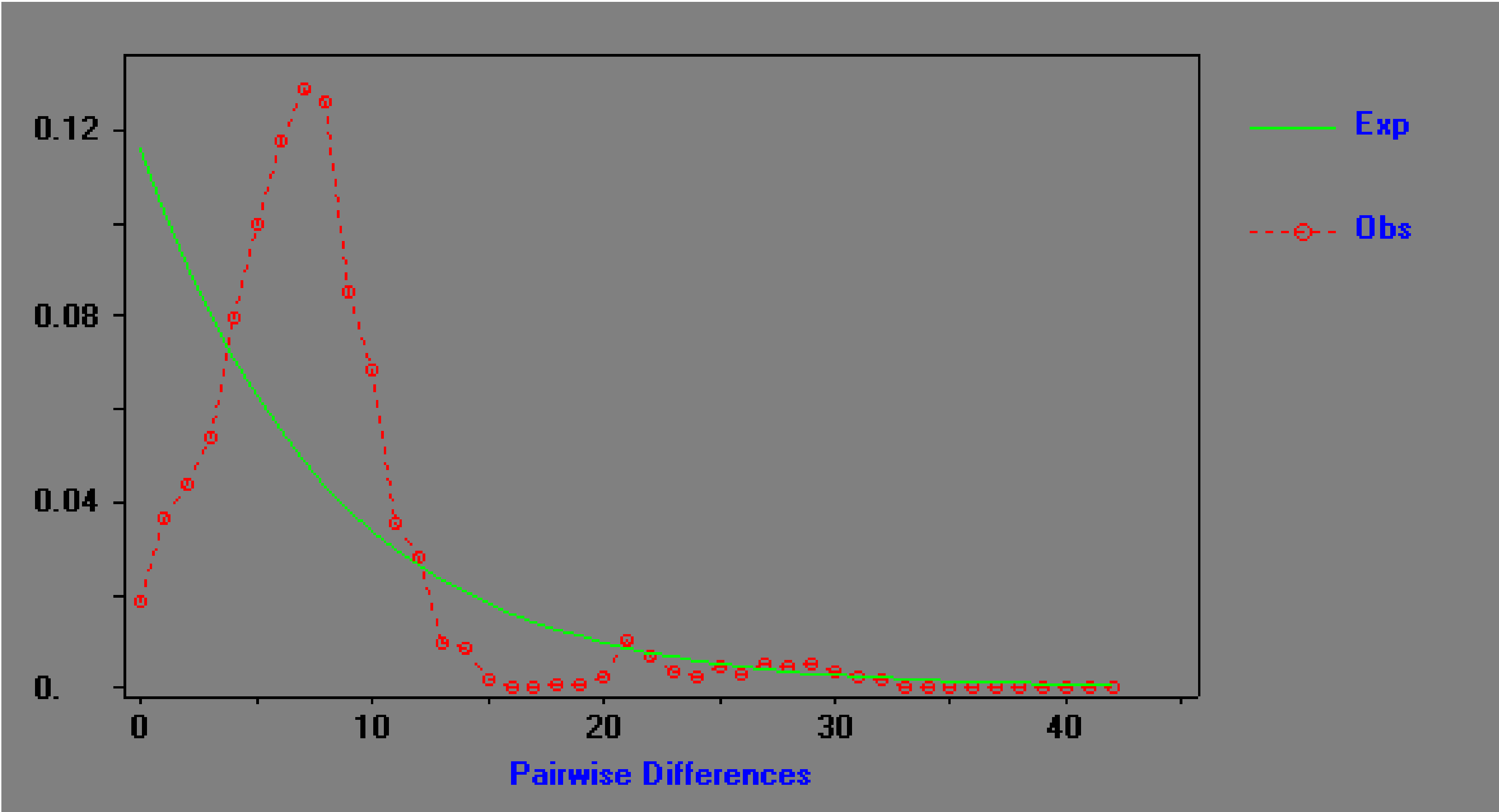


Figure 4.9:L3 haplotype mismatch distribution with a unimodal peak. (Raggedness index, r: 0.0056) y-axis = frequency

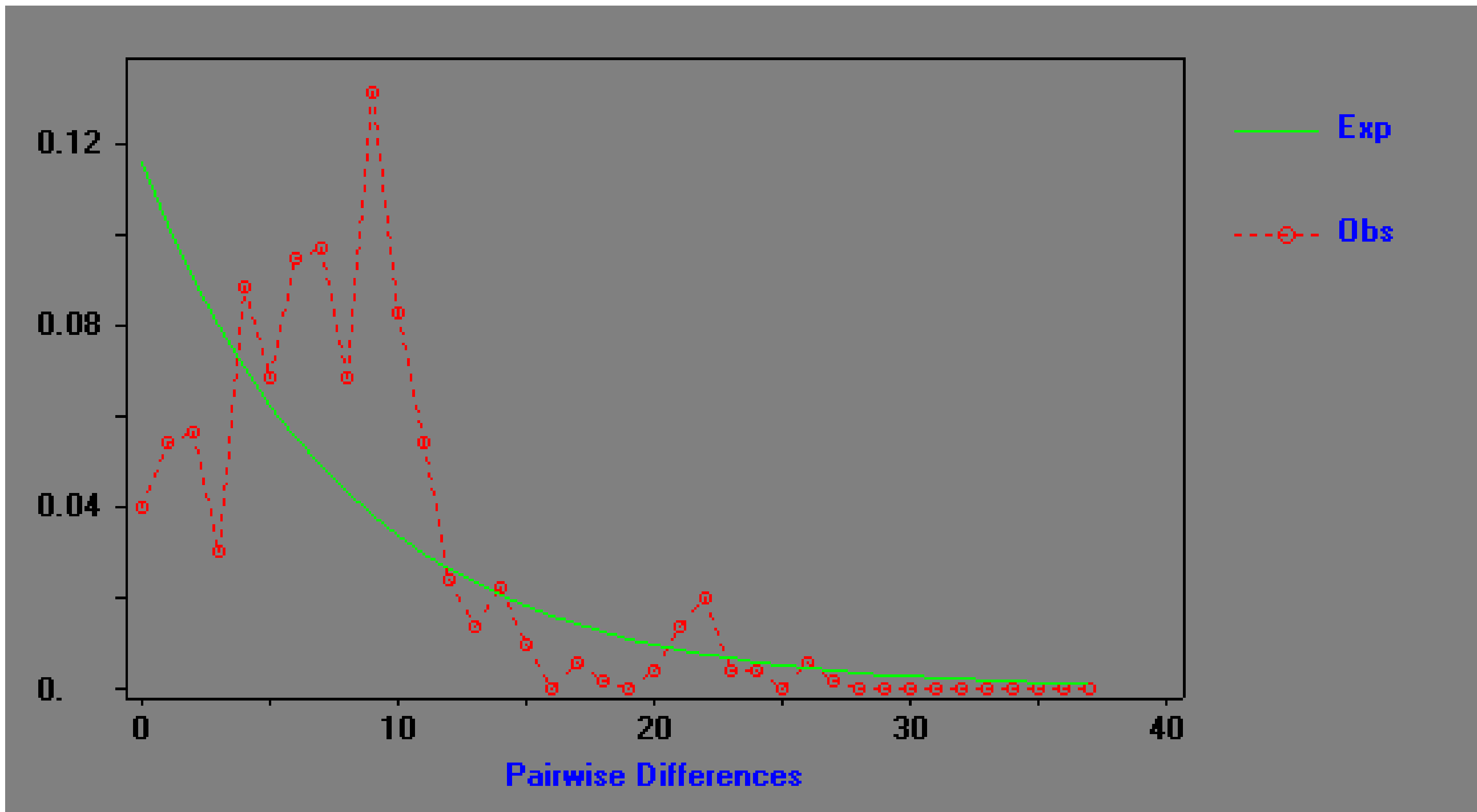


Figure 4.10: Atyap population mismatch distribution with a unimodal peak(Raggedness index, r =0.0151) (*y* axis = frequency)

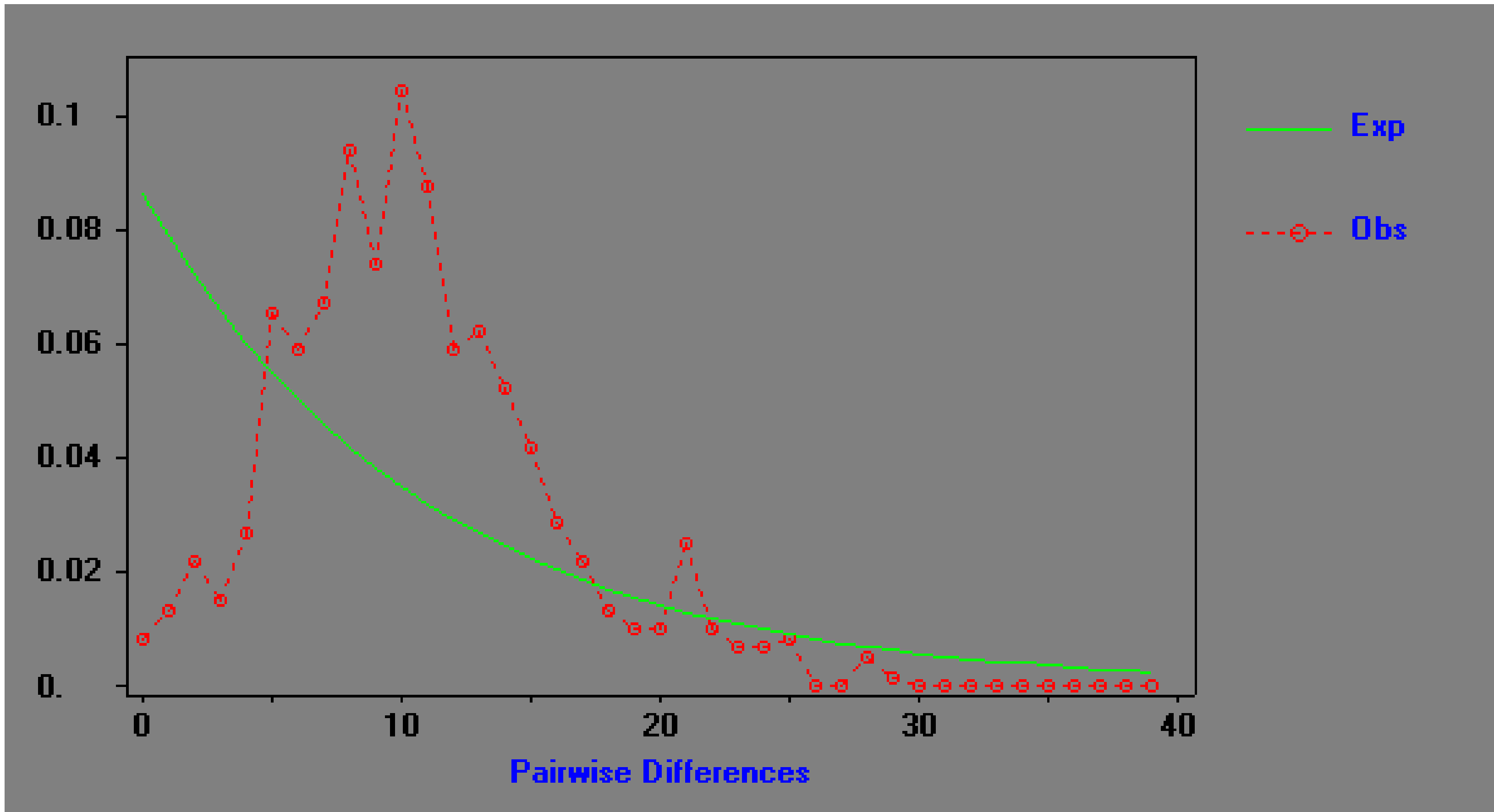


Figure 4.11: Bajju population mismatch distribution with multimodal peaks (Raggedness index, r: 0.0324) (*y* axis = frequency)

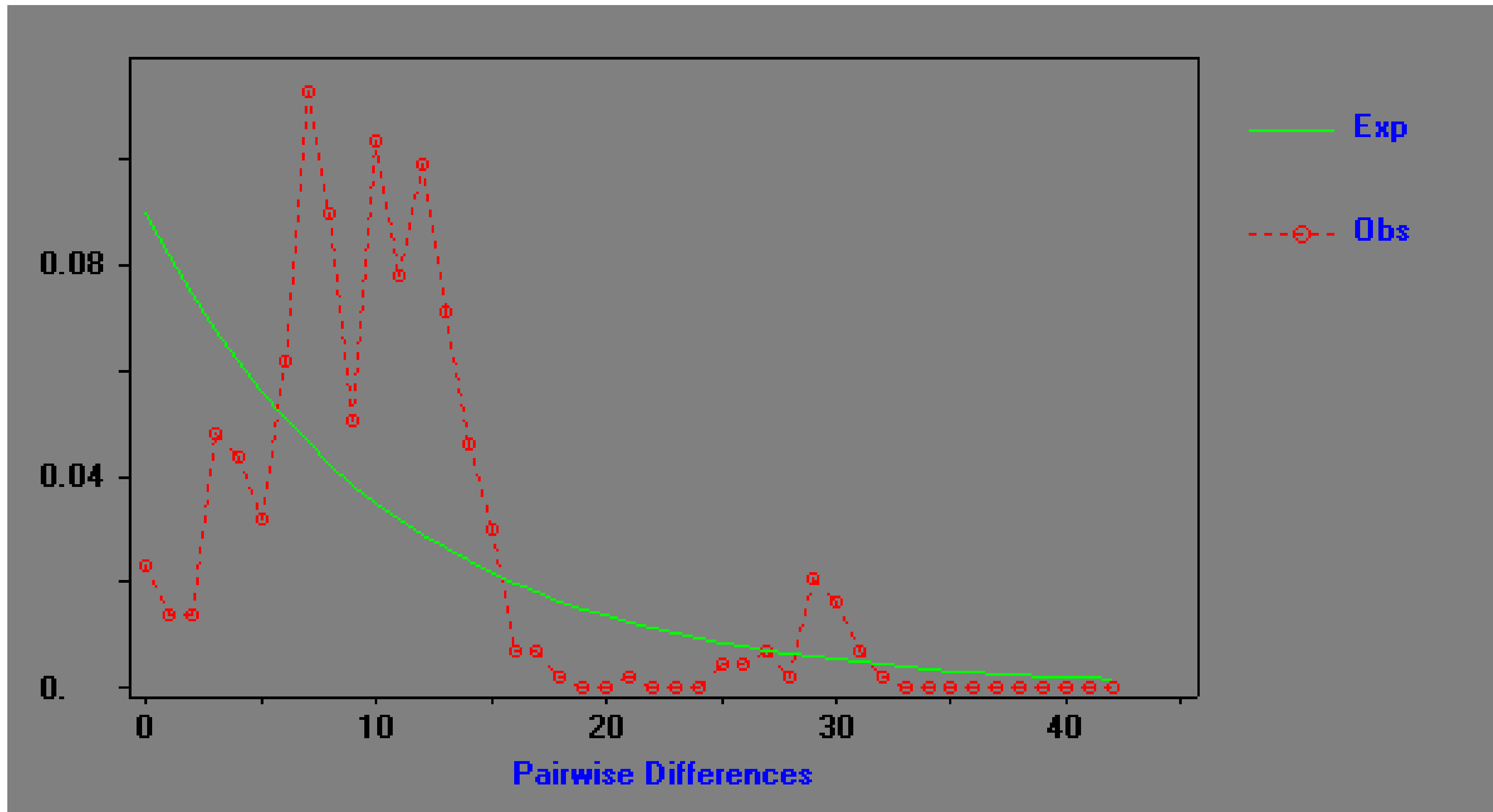


Figure 4.12: Chawai population mismatch distribution with a unimodal peak (Raggedness index r: 0.0135). (y-axis= frequency)

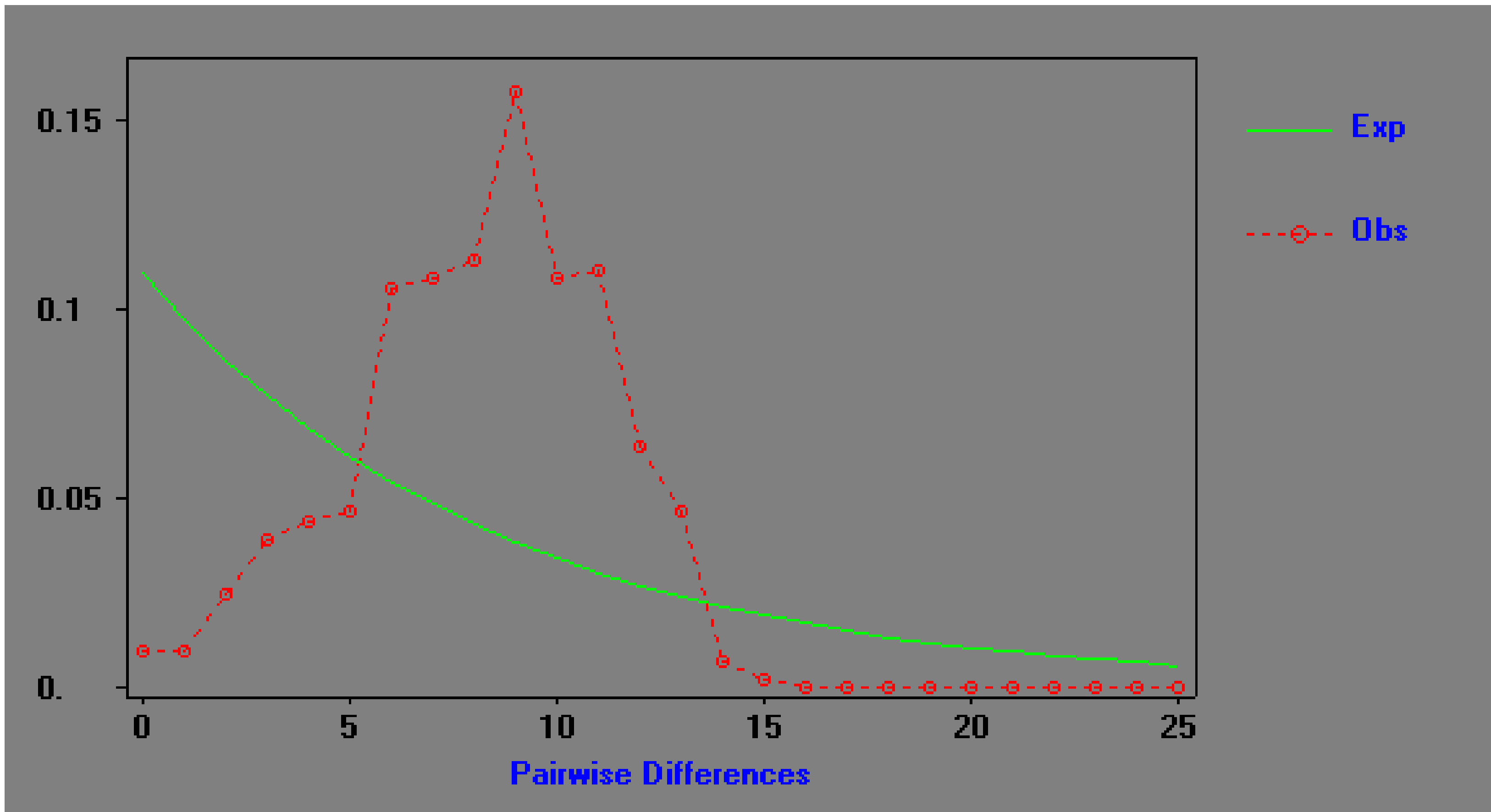


Figure 4.13: Kagoro population mismatch distribution with a unimodal peak (Raggedness index r: 0.0125) y-axis = frequency

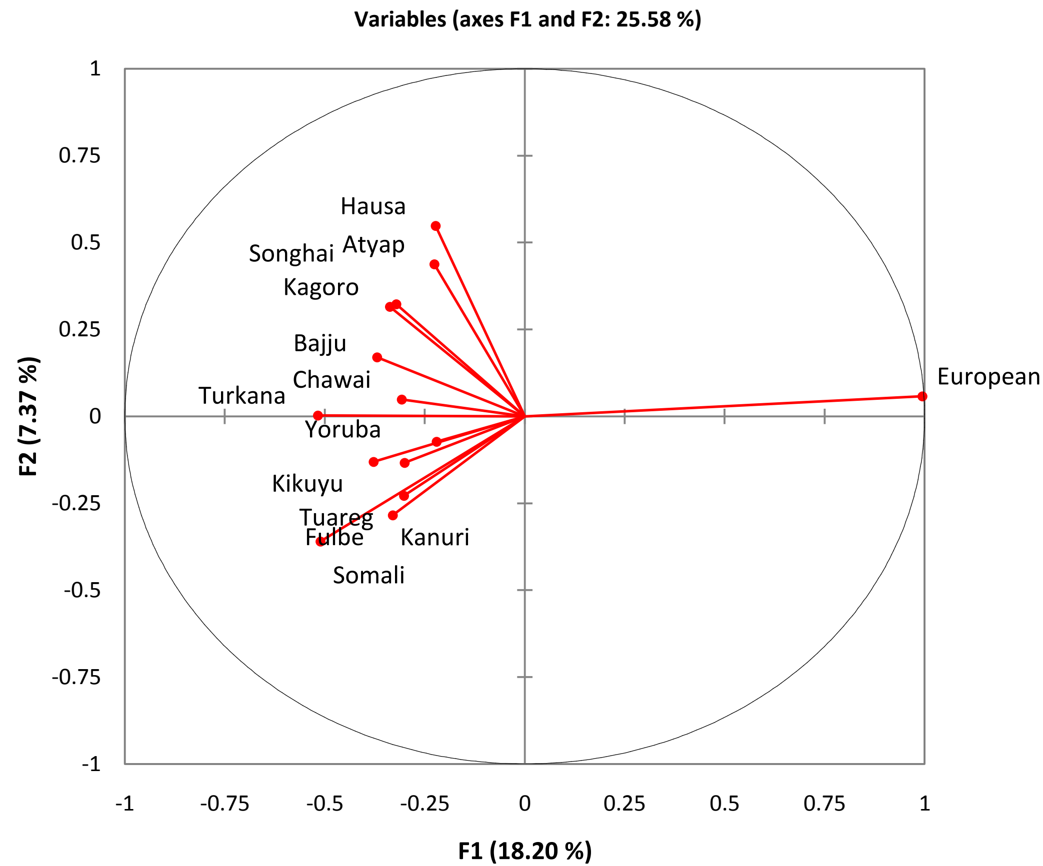
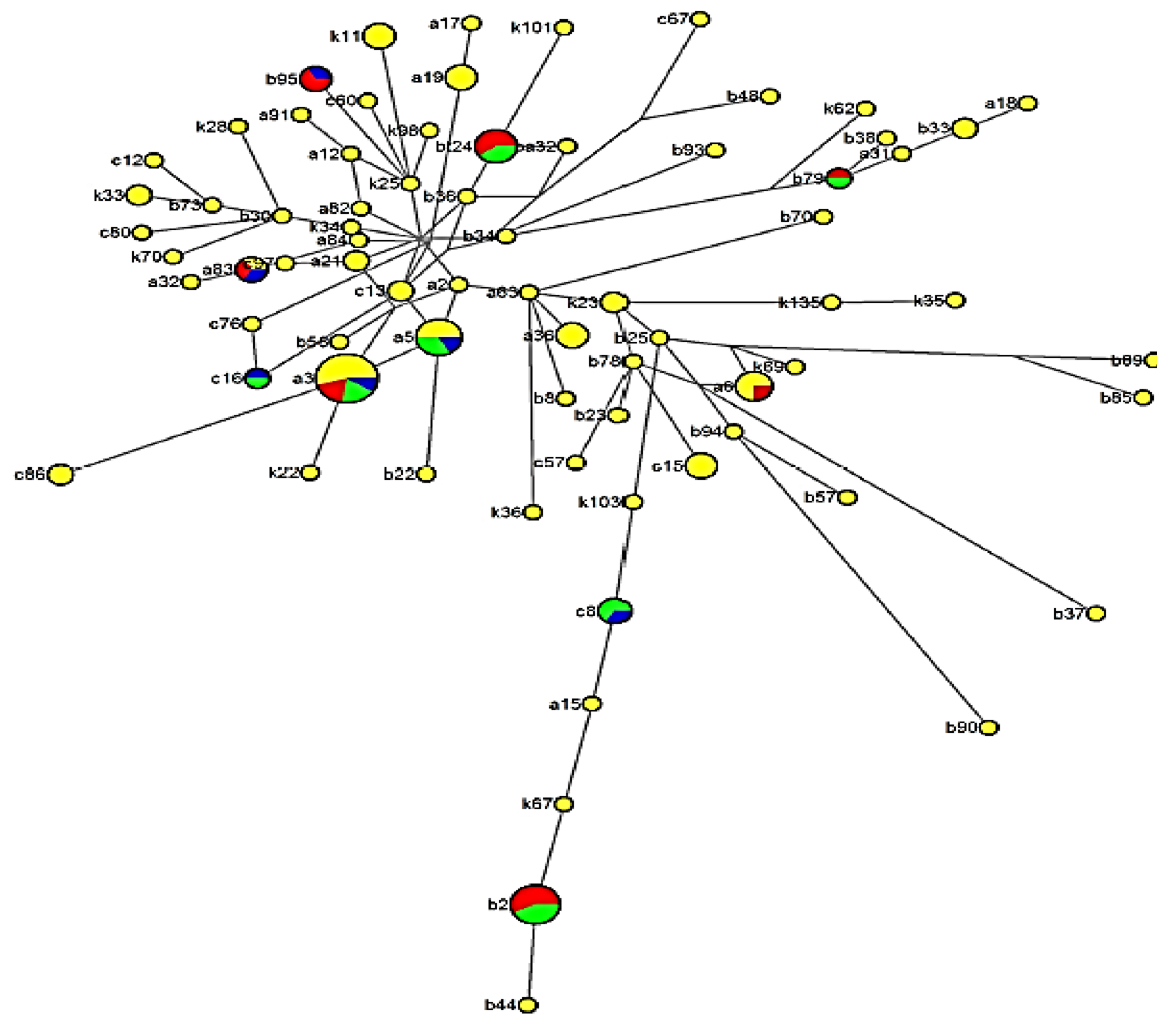


Figure 4.14: Principal Component Analysis. Derived from mtDNA sequences from populations in this study and other populations from published studies. The 1st PC and 2nd PC retain 18.2% and 7.3% of the variance respectively.

4.9 MEDIAN-JOINING NETWORK

From Figure 4.15 the mtDNA HVS-I network pattern illustrated a close type of relationship between the four study populations. The larger circles indicate clusters of the total sample. The smaller circles represent single, multiple, or star-contracted sequence. The circles (nodes) represent all haplotypes found within the populations with surface area proportional to their frequency. Identical haplotypes are shown to branch off beyond nodal points. All nodes are separated by mutational steps (not shown in the diagram for clarity) represented by lines joining the nodes. Overall, there were many shared haplotypes among the four populations with the general topology being star-like which offers a window into the processes of demographics in these populations.



Pie (only) Legend

Atyap

Bajju

Chawai

Kagoro

Figure 4.15: Median-Network of haplogroups based on HVS-I sequences. Alphanumerics represent taxa and mutational steps are not shown.

CHAPTER FIVE

5.0 DISCUSSION

The discussion focuses on the results of analyses from the previous chapter on the major observations of the study which centres on the hypotheses as well as the research objectives. The findings from the analysis of the study have been restricted to investigations on the following objectives: the major haplogroup types, the presence of a genetic structure, statistical relationship between the ethnic groups, language family, geography and genetics and also gene flow among the Atyap, Bajju, Chawai and Kagoro ethnic groups.

The study population's demographics have presented a patrilocal setting, where patrilocality is the phenomenon in which females (usually within a given ethnic background) take up residence in a man's village (Burton, 1996). It has been estimated that 70 % of modern societies are patrilocal (Jobling *et al.,* 2004). It is therefore implicit that the rate of mtDNA movement between villages is higher than the corresponding Y-chromosome in each generation therefore having a greater role in shaping genetic diversity (Jobling *et al.,* 2004). The use of the maternally inherited mtDNA as a marker of choice in biogeographic and phylogenetic studies is also based on the premise that within most populations, both the male and females share the same history (Hebert *et al.,* 2003; Hurst and Jiggins, 2005), this allows for the mtDNA locus from either sex within a particular population to sufficiently represent the entire group's genetic diversity.

The haplogroup profile for the Atyap, Bajju, Chawai and Kagoro showed the prevalence of haplogroups L0, L1, L2 and L3 with their sub- clades namely L0a1, L1a, L1b, L2a, L2c, $L3^*$,

L3b, L3d, L3e, L3f and L3h sub- haplogroups that have been well documented within the sub-Saharan Africa. This presents a consistent fit with findings from studies such as Salas *et al.* (2002) and Rosa *et al.* (2004), conducted on other ethnic groups through-out the African region. Few samples, however failed to cluster with the majority as they lacked SNPs belonging to the region and were merely identified as Non-L haplogroups.

Haplogroup L0 had only one branch of its five branches (L0a, L0b, L0d, L0f, L0k) which was L0a1 found in these ethnic groups except the Atyap. This clade was found in a little more than marginal frequencies (12 %) in the Bajju and the Chawai ethnic groups, suggesting a substantial gene flow from East to West. Salas *et al.* (2002) put forward a coalescent time of about 33 KYA in East Africa while Rosa*et al.* (2004) advocated two possible scenarios as to the presence of L0 in West Africa. The first is an East-West migration during the Holocene expansion that brought in Eastern founders to the region with subsequent migration known as the Bantu expansion towards the South, and the second was that of an already existing L0 pool with a later expansion within the region (Rosa *et al.,* 2004).

Haplogroup L1 is one of the oldest branches of L with a time depth of about 150 KYA (Ingman *et al.,* 2000) which was designated as the root of the mtDNA tree (the maternal most recent common ancestor). It was accounted for by only 8 % of the population, but its various clades were represented in all the ethnic groups. This frequency was less than the L1 lineages displayed for Ghana (16 %) as observed by Veeramah *et al.* (2010). The L1 subclades are given by L1a, L1b, L1c, L1d, L1e, L1f and L1k, but this nomenclature has been truncated by van Oven and Kayser (2008), who recognised only L1b and L1c. L1a is found in the Central, and South Eastern Africa regions but absent in West, North and Southern Africa (Salas *et al.,* 2002; Gonder *et al.,* 2007). But contrary to these findings, the present study reported the presence of L1a in

small frequencies within the Kagoro ethnic group of West Africa. L1b clade which has a West African origin (Watson *et al.,* 1997; Rosa *et al.,* 2004) even though a Central African origin has been proposed by Salas *et al.* (2002), was detected in low frequencies among the Bajju and Atyap. It has a coalescent time of about 30 KYA (Salas *et al.,* 2002). The presence of another L1 subclade, L1c was also confirmed among the Chawai and Kagoro of West Africa. The frequency of L1c in West Africa averages 5-8 % (Brehm *et al.,* 2002; Rosa *et al.,* 2004; Jackson *et al.,* 2005), as observed also in this present study. L1c is also known to have a high frequency in Central African populations but the highest frequencies of mtDNA haplogroup L1c are among Cameroon, Central African Republic, Angola and Republic of Congo (Destro-Bisol *et al.,* 2004; Plaza *et al.,* 2004; Beleza *et al.,* 2005).

Haplogroup L2 and its clades (L2a, L2b, L2c L2d and L2e) are widely found in sub-Saharan Africa but the present study only detected two clades (L2a and L2c) having a greater frequency (18 %) than L0 and L1 for the studied groups. The origin of L2 ascribed to West and Central Africa (Salas *et al.,* 2002) with a coalescent time of about 90 KYA (Pedro *et al.,* 2009). It was identified by HVS-I motif 16223-16278-16390 and restriction sites *HpaI* +3592, *DdeI* +10394, *MnlI* −10871, *HinfI* +16390/*AvaII* −16390 (Alves-Silva *et al.,* 2000; Pereira *et al.,* 2001a). The most common clade within L2 for the present study was L2a which was found in each of the original study populations, this agrees with the findings of Torroni *et al.* (2001) and Salas *et al.,* (2004) that highlighted the spread of L2a as the most frequent clade of L2 throughout Africa. Its origins from the Cameroonian Plateau have been suggested by Veeramah *et al.* (2010). L2c was only found in the Atyap and Bajju, both in moderate frequencies when compared to its occurrence in other studied West African populations such as those from Cape Verde, Guinea-Bissau and Senegal (Rosa *et al.,* 2004). The haplogroup with the greatest

frequency (62 %) among the four ethnic groups was L3. This finding is expected and unexpected as Salas *et al.* (2002) had ascribed both an East-African origin for L3 and the highest frequencies in the African continent. The presence of L3 within the West-African region can be attributed to series of migrations within the continent. It considered a more recent haplogroup compared to L0, L1 and L2 dated at about 57–87 KYA (Soares *et al.,* 2009), with the subsequent non-African haplogroups (M and N) taking direct origin from it. Thus, it was more related to Eurasian haplogroups when compared to the other African haplogroups (Maca-Meyer *et al.,* 2001). Out of the L3 clusters, sub-clade L3e was more frequent in all the ethnic groups. Its origins can be traced to Central Africa about 45 KYA (Bandelt *et al.,* 2001) and some of its clades can be traced back to East Africa (Rosa et al., 2004). Similar observation was made for the high frequency of L3e in the Cross Rivers region of Nigeria and neighbouring Ghanaian lineages (Veeramah *et al.,* 2010) but Salas *et al.* (2004), observed that it is a rare lineage in West Africa. L3b immediately follows L3e in terms of frequency and was also present within the four populations signifying putative gene flow amongst these populations. The L3b clade is predominant in western Africans who mainly speak languages of the Niger-Kordofanian family, of which Bantu is a member (Watson *et al.,* 1997).

Both L3d and L3f were only present in some of the study groups (low penetrance) and their low frequency when they occurred is indicative of a recent gene flow into the region. The estimated age of L3d was estimated by Salas *et al.* (2002) to be about 30 KYA while Behar *et al.* (2008) put it at 38 KYA. Rosa *et al.* (2004) detected L3d lineages in low frequencies in Senegalese and in parts of Niger and Nigeria. L3f is hypothesised to have arisen in Eastern Africa giving its high occurrence and diversity in that region with a coalescent age of about 40 KYA (Salas *et al.,* 2002). Three non-L clades where found in the Bajju but their origin could not be ascertained for

lack of definitive SNPs. From the ongoing, haplogroup lineages were consistent with existing literature from this region, though few could not be accounted for by the set parameters of this study. Each population had varied haplotypes depicting a heterogeneous setting within each population. Where some haplogroups or sub-clades were absent for certain ethnic groups in the study it may be as a result of genetic drift or attributable to sampling error. The presence of shared haplogroups (among the ethnic groups) present in this study may have resulted due to common ancestry or owing to high levels of genetic exchange as in an admixture through gene flow. This suggests that the Atyap, Bajju, Chawai and Kagoro cannot be placed in specific groups that will readily identify them as individual clades using the HVS-I informative sites. This similarity could be as a consequence of complex long range and short range migrations and invasions within the region (Rosa *et al.,* 2004). The Bantu migrations and the Lake Chad dispersals are examples of such movements (Bellwood 2005).

Within the 410 nucleotide sites investigated for the four populations, there were 91 haplotypes identified with a total of 107 polymorphic sites. The Bajju and Chawai had more segregating sites indicative of greater intra-population diversity than the Atyap and Kagoro that had values below average. The combined haplotype diversity index averaged 0.991(0.003) which was lower than that found in other West African groups but at par with values from East Africa (Watson *et al.,* 1996, 1997). The Bajju had the highest haplotype diversity value while the least was found in the Atyap. The other diversity measure, Nucleotide diversity, π, values showed similar variation as the haplotype diversity with the Bajju having the highest values and Atyap, the least. This suggests a high level of diversity within the individual population and the groups with the highest nucleotide diversity may be informative regarding their ancestral state. Sub-clade analysis based on L0, L1, L2 and L3, both haplotype and nucleotide diversities, π,

exhibited greater diversity for L3 and L2 while L0 had the least diversity. This is corroborated by the average pairwise differences.

Mismatch distributions detect irregularities in the distribution of the average number of pairwise differences measured using the raggedness index (Harpending, 1994), the outcome may be represented by higher values for stable populations or lower values for expanding populations. The calculated nucleotide pairwise differences for both mtDNA haplogroups and ethnic groups showed an interval of mismatch differences between 4.94 - 8.67 and 7.58 – 10.47 respectively. These values are consistent with the findings of Watson *et al.* (1996) and Rosa *et al.* (2004) on the West African region. They are normally higher among the African lineages depicting the great diversity associated with the older age of African gene pool, but usually lower in European population (Pereira *et al.,* 2001b).

A raggedness index of greater than 0.03 with a multi modal distribution curve provides evidence of a constant population in which mutation equals drift whereas an index less than 0.03 with a unimodal distribution curve is indicative of an expanding population. The resulting curves for the sub-clades of macro-haplogroup L showed L0, L1 and L2 (to a lesser extent) to be stable or at equilibrium giving by multimodal curves and high raggedness indices values (greater than 0.03). L3 presented a smooth curve (unimodal distribution) or bell-shaped distribution signifying a recent population expansion and supported by a low raggedness index (0.0056). This is in conformity with the assertion of Rosa *et al.* (2004) that haplogroups with younger coalescence age showed unimodal distributions. Ingman*et al.* (2000) and other workers have also pointed out that most individuals of African origin showed a ragged distribution consistent with constant

population size. The mismatch distribution according to ethnic groups exhibited unimodal distribution for the Atyap, Chawai and Kagoro with raggedness indices less than 0.03 and indicating that these populations have experienced a recent expansion event which is an evidence of population growth (Schneider and Excoffier, 1999). The Bajju displayed a multimodal curve expected of a constant sized population (in equilibrium) with a raggedness index value of 0.0324.

Signatures of selection are readily identified using statistical techniques that can also be used to assess demographic events in populations (Nielsen, 2005) especially in neutral systems. For this purpose the Tajima's D neutrality test that seeks out the correlation between mean pairwise sequence difference and the segregating sites on sequences, is usually used. Populations that retain positive *D* values show a balancing type of selection meaning that it is consistent with the proposition that the population has been of roughly constant size while negative values indicate that the population has experienced a period of population growth. Here, Tajima's D was significantly slightly negative for the Atyap, Chawai and Kagoro whereas it failed to reach significance in the Bajju, though it had a negative value. This corresponds to the findings of (Ingman *et al.,* 2000; Pereira *et al.,* 2001a; Salas *et al.,* 2002; Garrigan and Hammer 2006) on other African populations except for the Pygmies and Khoisan (Pereira *et al.,* 2001a). The negative values imply the presence of demographic expansion which is a departure from the null model of populations barring any selection and evolving at constant size in mutation-drift equilibrium. Fu's and Li's F also showed non-significant negative values while the potent Fu's F statistic corroborated the findings with strong statistically significant negative values for the Atyap, Chawai and Kagoro but not significant for the Bajju. The constant sized Bajju population corroborates the oral traditions of being older than the other ethnic groups that probably derived their origins from the Bajju, hence, their recent expansions. The population expansions did not

occur in isolation, they were accompanied by substantial gene flow among these neighbouring populations, leading to the presence of shared haplotypes. These genetic characteristics are likely to have resulted from a founder effect.

Principal component analysis was used to present genetic distances as a two dimensional graphical plot or genetic distance map, it provides an alternative to cluster analysis but cannot be used independently (Rosa *et al.,* 2004). The closer two points are on the map the more related they are (Mielke *et al.,* 2011). To place the study samples in a regional context, they were subjected to principal component analysis together with a few published groups from the African continent. The choice of a few sample populations was made in order to provide a better resolution in the reduced dimension. The samples generally clustered together with the other African samples as expected because of the extensive gene flow within the African continent but formed tighter associations with the West African groups than with the Eastern groups. This clustering pattern was also observed by Salas *et al.,* (2004), Rosa *et al.* (2004), Ramachandran *et al.* (2005) and Cerezo *et al.* (2012). Even though the variances are explained by the 1st principal component (PC) which is 18.2 % and the 2nd PC which is 7.37 %, the linguistics and geography may not be enough to explain the subtle differences among these populations therefore some stochastic underlying mechanisms could be at work, which will need further probing.

An analyses of molecular variance (AMOVAs) using the obtained mtDNA HVS-I sequences was conducted to determine the possible presence of a population substructure among the study's four populations. Genetic structure refers to any pattern in the genetic makeup of individuals within a population and its analysis helps to define the genetic composition of the population and the pattern of existing variation in relation to observed genotype frequencies (Chakraborty, 1993).

The four ethnicities were placed into groups according to their linguistic families. The Atyap, Bajju and Kagoro were grouped together as the Platoid branch of the Benue-Congo whereas Chawai was placed in the Kainji branch. AMOVA was carried out so that any observed genetic variation among groups (F_{CT}), within populations among groups (F_{SC}) and within individual populations (F_{ST}) can be easily partitioned. The amount of variation observed among the groups is 3.66 % ($F_{CT} = 0.03658$ and $p < 0.001$) which is statistically significant. This signifies a marginal population differentiation which implies a sub-structure among the groups. The amount of variation within groups ($F_{SC} = 0.01789$ and $p > 0.05$) described 1.72 % of the variation. The variation within populations ($F_{ST}= 0.05381$ and $p < 0.0001$) described 94.62 % (statistically significant) of the observed variation within these populations. The moderate F_{ST} is indicative of some appreciable level of population structure within the individual studied groups. In the absence of genetic structure in a population very little can be deduced about the genetic makeup of an individual by studying other members of the population (Slatkin, 1985). Usually, populations that are structured possibly by language or other factors may be continuous indicating an absence of a geographical boundary between the sub-populations. This introduces some degree of population diversification as seen in this study's populations. The result here proposes that notwithstanding the existing similarities in their sociocultural settings (lingustics) the high heterogeneity observed within the individual populations is consistent with the findings of Salas *et al.* (2004) and Rosa *et al.* (2004) among other Niger-Congo speakers, this however does not rule out common ancestry. If there is a question here it will be, "can the origin of individuals and their differences be ascertained using the preceding information? The findings of Bowcock *et al.* (1994), Wilson *et al.* (2001), Bamshad*et al.* (2003) provide answers in the affirmative that since not all the variation found lies within the population, the remainder of the

variation that exists between populations is significantly large enough to provide information about similarity and by extension, geographical origin. Genetic distance matrices were used to construct a phylogenetic tree using the Neighbor- Joining (NJ) algorithm written by Saitou and Nei (1987). This method clusters the samples and does not assume an evolutionary clock which allows for the production of unrooted trees. It is accurate when applied to related populations. The consensus tree illustrates relationships between the different HVS-I haplotypes identified within the study populations.

The branches of the tree placed the Chawai and Kagoro together (monophyletic) showing some evolutionary proximity whereas the Atyap and Bajju formed separate clades. The Atyap had a shorter genetic distance to both Chawai and Kagoro while the Bajju showed greater length when compared to the three other tribes, this is indicative of an older ethnic group. The topology could be supported by linguistic evidence where the Atyap and Kagoro have stronger similarity than they share with the Bajju but may not account for the Chawai giving their linguistic affiliation which places them within Kainji language family of the Benue-Congo division. Since no root was assumed for the tree, there was no ancestral node representing the most recent common ancestor (TMRCA).

A median joining network was constructed to show the relationship between the haplotypes of the four populations. Each haplotype is represented by a node linked to other nodes through branches. Unlike the Neighbor-Joining tree, a median joining network represents many equally good trees from which a reliable tree can be viewed without being subjective. The resulting network for the 109 haplotypes depicted a star-like phylogeny with multiple taxons derived from a group of common taxa. Haplotypes were separated by mutational steps and this leads to the branching patterns. There was displacement of the central node however; there was

the presence of several satellite nodes along with their branches resulting into smaller star structures indicating the presence of deep lineages. Typically, the West African landscape has a higher frequency of L2 (Salas *et al.,* 2004) but the study populations presented a higher presence of the East African L3 haplogroups especially sub-clade L3e.

The shape of the network from this study and the lack of haplotype clustering suggest the occurrence of a demographic event, which in this case will be a recent population expansion. This expansion probably took place after a long period of low effective population size and may be extrapolated to a point during the Chad basin dispersals. This area lacked geographicbarriers to migration, which may have led the Niger-Kordofanian language family existing among others to move through the Sahel corridor described as a bidirectional corridor of human migrations towards the West(Bellwood 2005; Bereir *et al.,* 2007; Cerny *et al.,* 2007; Tishkoff *et al.,* 2009).

CHAPTER SIX

6.0 SUMMARY, CONCLUSION AND RECOMMENDATION

6.1 SUMMARY

This study has estimated some genetic and demographic parameters of the Atyap, Bajju, Chawai and Kagoro ethnic groups from the Southern area of Kaduna Stateusing HVS-I sequences. Given their shared geographical landscape and ethnocultural characteristics, the study is exigent on the basis of understanding the evolutionary processes at work and to shed light on the possible application of uniparental markers for measures of diversity within and between these ethnic groups.Hence, this study concentrated on ethnic and linguistic units for its analysis. The mtDNA lineages presented in this study were consistent with those from surrounding areas and is in agreement with the findings within the neighbouring regions as corroborated by studies such as Salas *et al.* (2002; 2004), Rosa *et al.* (2004), Veeramah *et al.* (2010). The profile, frequency and diversity of the haplogroups and subclades of L0, L1, L2 and L3 were comparable with those found in other African populations resulting in the formation of tight clusters when compared. The discovery of Eastern and Central African lineages in this study is indicative of gene flow that may have occurred between them. This was estimated to a time during their passage within the Sahelian corridor which served as a connection between East and West African populations. However, lack of a pronounced genetic structure, evidenced by shared haplotypes between the

study populations point to the fact that these populations have been in prolonged contact. This is in agreement with archaeological findings, even though the Chawai have maintained that they do not intermarry with the surrounding populations. This leads to the understanding that after their initial contact either from origin or migrations they might have long sustained and passed down their inherited SNPs to subsequent generations. The populations' linguistic diversity may have been caused by past environmental events including migration with fragmentation potentials and periods of isolation resulting in the formation of ethnic units.The evolutionary force of genetic drift operating at the mtDNA locus appears weak in decreasing the amount of genetic variability in the study populations which can be attributed to their large effective population sizes. If these groups are endpoints of a migration event it is not clear when the "within" population genetic differentiation occurred.

6.2 CONCLUSIONS

The results of this study has for the first time provided evidence of close genetic ties among the recently diverged Atyap, Bajju, Chawai and Kagoro ethnic groups. This was achieved by use of extracted and sequenced molecular (mtDNA) markers and the application of various multivariate analytical techniques. The observed shared genetic relationship among these ethnicities is reflected in linguistics and other cultural elements arising as a consequence of a probable common origin (founder effect), gene flow and geographic proximity.

6.3 RECOMMENDATIONS

Due to the fact that human migration within Africa has been complex, tracing and understanding the patterns of differentiation in smaller groups within its genetically diverse

landscape will provide more information on the already under-represented continent. Therefore, further investigations are required using Y-chromosomal STR markers, combined HVS-I and HVS-II SNPs or whole mtDNA markers. In addition, mathematical coalescent models should also be developed for these groups to probe the timing of ethnogenetic events within these local populations. More ethnic groups should be included so as to have abroad map of theNigerian genetic landscape in relation to Africa as a whole.

REFERENCES

Abu-Amero, KK., Gonzalez, A.M., Larruga, J.M., Bosley, T.M. and Cabrera, V.M. (2007). Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evolutionary Biology*, **7**:32.

Achilli, A, Rengo, C., Magri, C., Battaglia, V,, Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E. and Carelli, V. (2004). The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. *American Journal Human Genetics,***75** (5):910-918.

Achilli, A., Rengo, C., Battaglia, V., Pala, M,, Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N, Santachiara-Benerecetti, A.S., Bandelt H.J., Semino, O. and Torroni, A. (2005). Saami and Berbers - an unexpected mitochondrial DNA link**.***American Journal Human Genetics,***76:**883-886.

Adebisi, S.S. (2008). Medical impacts of anthropometric records.*Annals of African Medicine.***7**(1):42-47

Agre, P.C., and Cartron J.P., eds. (1992).*Protein Blood Group Antigens of the Red Cell: Sturucture, Function, and Clinical Significance.* Baltimore: Johns Hopkins University Press.

Alexandra, R., Brehm, A., Kivisild, T., Metspalu, E., and Villems, R. (2004). MtDNA Profile of West Africa Guineans: Towards a Better Understanding of the Senegambia Region*Annals of Human Genetics,***68**:340–352 doi: 10.1046/j.1529-8817.2004.00100.x

Anderson, S., Bankier, S., Barrell, B., De Brujin, M., Coulson, A., Drouin, J., Eperon, I., Nierlich, D., Roe, B., Sanger, F., Schreier, P., Smith, A., Staden, R. and Young, I. (1981) Sequence and organization of the human mitochondrial genome.*Nature,* **290:**457-465.

Andersson, S.G., Karlberg, O., Canback, B., and Kurland, C.G, (2003). On the origin of mitochondria: a genomics perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **358**(1429):165-177.

Andrews, R., Kubacka, I., Chinney, P., Lightowlers, R., Turnbull, D., Howell, N. (1999). Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA.*Nature Genetics,***23**:147

Antonarakis SE, Krawczak M, Cooper DN (2001). The nature and mechanisms of human gene mutation. In: Scriver CR, Beaudet AL, Valle D (eds). *The metabolic and molecular bases of inherited disease.* McGraw-Hill, New York, pp 343–377

Atkinson, Q.D., Gray, R.D. and Drummond, A.J. (2009).Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa *Proceedings of the Royal Society of Biological Sciences,***276:**367–373 DOI:10.1098/rspb.2008.0785

Aumassip, G., Ferhat, N., Heddouche, A. and Vernet, R. (1994): Le milieu saharien aux temps préhistoriques**.** In Milieux, hommes et techniques du Sahara préhistorique. Problèmes actuels. Edited by Aumassip G *et al*. Paris: L'Harmattan:9-29.

Avise**,**J.C., Shapira, J. F., Daniel, S.W., Aquadro, C.F. and Lansman, R.A. (1983). Mitochondrial DNA differentiation during the speciation process in *Peromyscus***.***Molecular Biology and Evolution,***1**:38-56.

Baldauf, S.L. (2003). Phylogeny for the faint of heart: a tutorial *Trends in Genetics,***19**(6):345-351

Bamshad, M.J., Wooding, S., Watkins, W.S., Ostler, C., Batzer, M.A. and Jorde, L.B. (2003).Human population structure and the inference of group membership.*American Journal of Human Genetics,* **72**:578-589

Bandelt, H.J., Alves-Silva, J., Guimaraes, P., Santos, M., Brehm, M., Pereira, L., Coppa, A., Larruga, J.M., Rengo, C., Scozzari, R., Torroni, A., Prata, M.J,, Amorim, A, Prado, V.F,, Pena, S.D,J. (2001). Phylogeography of the human mitochondrial L3e: a snapshot of African prehistory and Atlantic slave trade. *American Journal of Human Genetics,***65**:549–563

Bandelt, H.J., Forster P., Sykes B.C., and Richards, M.B. (1995).Mitochondrial portraits of human populations using median networks.*Genetics*, **141**:743-753.

Bandelt H.J. and Parson W. (2008). Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *International Journal of Legal Medicine,* **122**(1):11-21.

Barbujani, G. and Bertorelle, G. (2001). Were Cro-Magnons too like us for DNA to tell? *Natur,*.**424:**127

Bartlett, J., Stirling, D. (2003).A Short History of the Polymerase Chain Reaction.*Methods in Molecular Biology,* **226**:3-6.

Behar, D.M., Villems, R., Soodyall H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan H., Tzur, S., Comas, D., Bertranpetit, J., Quintana- Murci, L., Tyler-Smith, C., Wells, R.S. and Rosset, S. (2008). The dawn of human matrilineal diversity.*American Journal Human Genetics,* **82**: 1130-1140.

Beleza, S., Gusmão, L., Amorim, A., Carracedo, A., Salas, A., (2005).Thegenetic legacy of western Bantu migrations.*Human Genetics,***117**, 366–375.

Bellwood, P. (2001) Early agriculturalists population diasporas? Farming, languages and genes.*Annual Review in Anthropology,***30**:181-207.

Bellwood, P. (2005). First farmers: the origins of agricultural societies. Blackwell Publishing: Oxford.

Bendor-Samuel, J.T. (ed.) (1989).*Niger-Congo*. Boston: University Press of America.

Bendor-Samuel, J. T. (1971). Niger-Congo, Gur in Linguistics in sub-Saharan Africa (The Hague: Mouton)

Bensasson, D., Feldman, M.W. and Petrov, DA. (2003). Rates of DNA duplication and mtDNA insertion in the human genome.*Journal of Molecular Evolution,***57**: 343-354.

Bereir, R.E., Hassan, H.Y., Salih, N.A., Underhill, P.A., Cavalli-Sforza, L.L. (2007).Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel.*European Journal of Human Genetics,***15**: 1183–1185.

Bermisheva, M.A., Viktorova, T.V., and Khusnutdinova, E.K. (2003).Polymor-phism of human mitochondrial DNA.*Russian Journal of Genetics,***39**(8):849-859.

Bisol, G., Donati. F., Coia, V., Boschi, I., Verginelli, F., Caglia, A., Tofanelli, S., Spedini, G. and Capelli, C. (2004). Variation of female and male lineages in Sub-Saharan populations: the importance of sociocultural factors. *Molecular Biology and Evolution,***21**(9): 1673– 1682.

Blench, R.M. (1998). The status of the languages of Central Nigeria. In: Brenzinger, M. ed. *Endangered languages in Africa*. 187-206.

Blench, R.M. (2003). Why reconstructing comparative Ron is so problematic. In: *Topics in Chadic Linguistics*. Papers from the 1st Biennial International Colloquium on the Chadic Language Family (Leipzig, July 5-8, 2001). Ekkehard W H. (ed.) 21-42.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, JR. and Cavalli-Sforza, LL. (1994).High resolution of human evolutionary tees with polymorphic microsatellites.*Nature,* **368**: 455-457

Brehm, A., Pereira, L., Bandelt, H.J., Prata, M.J. and Amorim, A (2002). Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Annals of Human Genetics,***661**:49-60.

Brinkman, F.S.L. and Leipe, D.D. (2001). Phylogenetic analysis In: *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins,* Baxevanis, AD Ouellette B.F.F (Eds) *Second Edition*

Brion, M., Salas, A., Gonzalez-Neira, A., Lareu, M.V. and Carracedo, A. (2003).insights into Iberian population origins through the construction of highly informative I-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite. *American Journal of Physical Anthropology*, **122**:147-161.

Brown, W.M. (1980). Polymorphism in mitochondrial DNA of human as revealed by restriction endonuclease analysis.*Proceedings of the National Academy of Science*,*USA*.**77**:3605–3609.

Brown, K. and Ogilvie, S. (2009). *Concise encyclopedia of languages of the world*. Elsevier Ltd., The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

Butler, J. (2005). *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers.* 2nd Edition.London. Elsevier Academic Press.

Burton, M.L., Moore, C.C., Whiting, J.W.M., Romney, A.K. (1996). Regions based on social structure. *Current Anthropology,***37**:87-123.

Callegari-Jacques, S.M., Grattapaglia, D, Salzano, F.M., Salamoni, S.P., Crossetti, S.G., Ferreira, M.E. and Hutz, M.H. (2003). Historical genetics: Spatiotemporal analysis of the formation of the Brazilian population. *American Journal of Human Biology,* **15**:824–834.

Campbell, M.C., and Tishkoff, S.A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics,***9**:403–433. doi:10.1146/annurev.genom.9.081307.164258

Cann, R.L., Stoneking, M, and Wilson, A.C. (1987) Mitochondrial DNA and human evolution.*Nature,***325**(6099):31-36.

Cattaneo, C., Gelsthorpe, K., Sokol, R.J. (2000). DNA Extraction Methods in Forensic Analysis. In: Meyers, RA, editor. *Encyclopedia of Analytical Chemistry: Applications, Theory, and Instrumentation*.John Wiley and Sons, Ltd.

Cavalli-Sforza, L.L. and Cavalli-Sforza, F, eds. (1995). The great human diasporas: the history of diversity and evolution. Reading, Mass.: Addison-Wesley.

Cavalli-Sforza, L.L, Menozzi, P., and Piazza, A.(1994).*The History and Geography of Human Genes.* Princeton: Princeton University Press.

Cavalli-Sforza, L.L., Piazza, A., Menozzi, P, and Mountain, J. (1988) Reconstruction of human evolution: Bringing together genetic, archaeological and linguistic data. *Proceedings of the NationalAcademy of Sciences of the USA,* **85**:6002-6006.

Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gómez-Carballa, A., Brisighelli, F., Lancioni, H., *et al.*, (2012). Reconstructing ancient mitochondrial DNA links

between Africa and Europe *Genome Research*.**22**(5): 821–826. doi: 10.1101/gr.134452.111

Cerny, V., Fernandes, V., Costa, M.D., Hájek, M., Mulligan, C.J. and Pereira, L. (2009). Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evolutionary Biology*, **9**:63.

Cerny, V., Hájek, M., Cmejla R., Bruek, J. and Brdiéka, R. (2004).mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. *Annals of Human Biol*ogy, **31**:554-569.

Cerny, V., Salas, A., Ha´jek, M., Zaloudkova, M. and Brdiéka, R. (2007). A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. *Annals of Human Genetics,***71**: 433–452.

Chakraborty, R. (1993). "Analysis of Genetic Structure of Populations: Meaning, Methods, and Implications". In Majumder, Partha P. *Human Population Genetics: A Centennial Tribute to J.B.S.Hadane*. London: Plenum Press. p. 189.

Chaudhmi, A. (1977). Cholera and blood-group.*Lancet,* **2**:404-405

Chawai History Project Transcripts (1993). Department of history, Ahmadu Bello University, Zaria pp 31-81

Chen, J.T., Sokal, R.R. and Ruhlen, M. (1995).Worldwide analysis of genetic and linguistic relationships of human-populations.*Human Biology,* **67**: 595– 612.

Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, and Wallace DC.(2000). mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations.*American Journal of Human Genetics,***66**:1362–1383.

Chen, X.J. and Butow, R.A. (2005).The organization and inheritance of the mitochondrial genome.*National Review of Genetics,* **6:**81 5-825.

Chinnery, P.F., Johnson M.A., Wardell, T.M., Singh-Kler, R., Hayes, C., Brown, D.T., Taylor, R.W., Bindoff, L.A. and Turnbull, D.M. (2000).The epidemiology of pathogenic mitochondrial DNA mutations.*Annals of Neurology*, **48:**188-193**.**

Cockerham, C.C. (1973). Analysis of gene frequencies.*Genetics*, **74**: 679–700

Crabb, D.W. (1974). Niger-Congo languages. West African Language Monographs 4, Cambridge

Crawford, M.H. (1973). The use of genetic markers of the blood in the study of the evolution of human populations. In: *Methods and Theories of Anthropological Genetics,* Crawford MH and Workman P L. (Eds) Albuquerque: University of New Mexico Press, pp 19-38.

Crawford, M.H. (2007). *Anthropological genetics: Theory, Methods and Applications*. Ed Crawford M.H, pp 1-16. University press Cambridge

Dambo, A.B.P. (2011). *Bajju tradition*. Nigeria: Kenyi Prints.

Darby, J. (2004). Comparison of complete mitochondrial genome sequence between different ethnic groups from Southern Africa. Unpublished Thesis Dissertation submitted for the degree Magister Scientiae (M.Sc.) in Biochemistry at the North-West University, South-Africa.

Diez-Sanchez, C., Ruiz-Pesini, E., Montoya, J., Perez-Martos, A., Enriquez, J.A, and Lopez-Perez, M.J. (2003). Mitochondria from ejaculated human spermatozoa do not synthesize proteins. *Federation of European Biochemical Societies Letters,* **55:**205–208.

Ebner, S., Lang, R., Mueller, E., Eder, W., Oeller, M., Moser, A., Koller, J., Paulweber, B., Mayr, J., Sperl, W. and Kofler, B. (2011) Mitochondrial Haplogroups, Control Region Polymorphisms and Malignant Melanoma: A Study in Middle European Caucasians.*PLoS ONE* **6**:e27192.

Efron, B. (1982). The jacknife, the bootstrap and other resampling plans. Regional conference series in applied mathematics, Philadelphia

Ennafaa, H., Cabrera, V., Abu-Amero, K.A., González, A.M,, Amor, M.B., Bouhaha, R., Dzimiri, N., Elgaaïed, A.B. and Larruga, J.M. (2009). Mitochondrial DNA haplogroup H structure in North Africa *BMC Genetics*, **10**:8 doi:10.1186/1471-2156-10-8

Ethnologue (2013). (www.ethnologue.com).

Excoffier, L. and Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resource,***10**(3):564-567.

Excoffier L, Smouse P, and Quattro J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,***131**(2):479.

Excoffier, L., and Yang, Z. (1999).Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees.*Molecular Biology and Evolution*, **16**(10):1357-1368.

Falkenberg, M., Larsson, N.G. and Gustafsson, C.M. (2007) DNA replication and transcription in mammalian mitochondria.*Annual Review of Biochemistry* **76:**679-699.

Forster, P., Harding, R., Torroni, A., and Bandelt, H.J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *American Journal of Human Genet*ics, **59**:935-945.

Freeman, B., Powell, J., Ball, D., Hill, L., Craig, I., and Plomin, R. (1997). DNA by mail: An inexpensive and noninvasive method for collecting DNA samples from widely dispersed populations. *Behavioural Genetics,***27**:251–257.

Garrigan, D. and Hammer, M.F. (2006).Reconstructing human origins in the genomic era.*National Review of Genetics,* **7**:669–80

Gonder, M.K., Mortensen, H.M., Reed, F.A., de Sousa, A. and Tishkoff, S.A. (2007).Whole-mtDNA genome sequence analysis of ancient African lineages.*Molecular Biology and Evolution*, **24**, 757–768. DOI:10.1093/molbev/msl209

Gonzalez, A.M., Cabrera, V.M., Larruga, J.M., Tounkara, A., Noumsi, G., Thomas, B.N. and Moulds, J.M. (2006).Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations.*Annals of Human Genetics,***70**(5):631-657.

Goodman, M. (1963). Serological analyses of the systematics of recent hominoids. *Human Biology,***35**: 377-443

Greenberg, J.H. (1954). Studies in Afrian linguistic classification. VIII. Further remarks on method; revisions and corrections. *Southwestern Journal of Anthropology,* ***10****:405-415*

Greenberg, J.H. (1963). The languages of Africa. 3rd ed. Bloomington: Indiana University Press.

Gregersen, E.A. (1977). *Language in Africa: An introductory survey*. Newyork: London and Breach

Guthrie, M. (1948).*The classification of the Bantu languages.* London: Oxford University Press for International African Institute.

Habu, B.D. (1993). *The Bajju people: an outline in social anthropology*. Barnawa-Kaduna, Nigeria

Hall, B.G. (2004). *Phylogenetic Trees Made Easy: A How-To Manual for Molecular Biologists.* (second edition), Sinauer Associates.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*.**41**:95-98.

Harpending, R.C. (1994) Signature of Ancient Population Growth in a Low-Resolution Mitochondrial DNA Mismatch Distribution. *Human Biology*, **66**:591-600.

Harpending, H. (1994). Gene frequencies, DNA frequencies, and human origins. *Perspectives in Biology and Medicine*,**37**:384-394.

Harpending, R.C., Sherry, S.T., Rogers, A.R., and Stoneking, M. (1993).The genetic structure of ancient human populations.*Current Anthropology*, **34**: 483-496

Harrison, C.J. and Langdale, J.A. (2006).A step by step guide to phylogeny reconstruction.*The Plant Journal*, **45**:561–572 DOI: 10.1111/j.1365-313X.2005.02611.x

Hazkani-Covo, E. and Graur, D. (2007).A comparative analysis of numt evolution in human and chimpanzee.*Molecular Biology and Evolution*,**24**(1): 13–18. DOI: 10.1016/j.fgb.2010.03.008

Hebert, P. D. N., Cywinska, A., Ball, S. L. and DeWaard, J. R. (2003).Biological identifications through DNA barcodes.*Proceedings of the Royal Society of Biological Sciences,* **270**: 313–321. (doi:10.1098/rspb.2002.2218.)

Helgason, A., Nicholson, G., Stefánsson, K., and Donnelly, P. (2003). "A Reassessment of Genetic Diversity in Icelanders: Strong Evidence from Multiple Loci for Relative Homogeneity Caused by Genetic Drift." *Annals of Human Genetics*,**67**: 281-297.

Hoelzel, A.R., Hancock, J.M. and Dover, G.A. (1991).Evolution of the cetacean mitochondrial D-loop region.*Molecular Biology and Evolution***8**: 475-493.

Hoisington, D., Khairallah, M. and Gonzalez-de-Leon, D. (1994).*Laboratory Protocols: CIMMYT Applied Biotechnology Center*. Second Edition, Mexico, D.F.: CIMMYT.

Holland, H.D. (1994). Early Proterozoic atmospheric change. In: Bengtson S. (ed) *Early life on Earth*. Columbia University Press: New York, pp 237-244.

Horai, S., Kondo, R., Hattori, Y.K., Hayashi, S., Sonoda, S. and Tajima, K. (1993). "Peopling of the Americas, Founded by Four Major Lineages of Mitochondrial DNA" *Molecular Biology of Evolution,* **10**(l):23-47

http://kagoro.online.cm/kaka/Lil%20historia.html accessed in 2/9/2013

http://www.atyap.net/atyap/AtyapNationality.aspx retrieved 2 3 2013

http://www.fluxus-engineering.com accessed in 2/12/2012

http://www.galter.northwestern.edu/Guides-and-Tutorials/Phylogenetics-Resources-for-the-NonPhylogeneticist retrieved 5/4/2013

http://www.uvm.edu/~biology/Classes/296D/10_Mitochondria.pdf

Hurles, M.E., Nicholson, J, Bosch, E., Renfrew, C., Sykes, B.C. and Jobling, M.A. (2002). Y chromosomal evidence for the origins of oceanic-speaking peoples.*Genetics,***160**:289-303.

Hurst, G.D.D. and Jiggins, F.M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of Biological Sciences,***272**:1525–1534 DOI:10.1098/rspb.2005.3056

Ingman, M., Kaessmann, H., Paabo, S. and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans.*Nature,***408:**08–713. (doi:10.1038/ 35047064)

Jackson, B.A., Wilson, J.L., Kirbah, S., Sidney, S.S., Rosenberger, J., Bassie, L., Alie, J.A., McLean D.C, Garvey, W.T. and Ely, B. (2005) Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *American Journal of Physical Anthropology,***128** (1):156-163.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech Z.A., Degnan, J.H., Wang, K. and Guerreiro, R. (2008).Genotype, haplotype and copy-number variation in worldwide human populations.*Nature,***451**:998–1003.

Jobling, M.A., Hurles, M.E., and Tyler-Smith, C. (2004).*Human Evolutionary Genetics - Origins, Peoples and Disease.*Garland Publishing: New York

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

Joshua, E. (2013). Kagoro tribe and tradition of origin Retrieved from http://kagorohistory.blogspot.com/2013/10/kagoro-heroes-from-1905-1945.html

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution,***16**(2):111-120.

Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E,, Parik, J., Geberhiwot, T., Usanga, E. and Villems, R. (2004). Ethiopian mitochondrialDNA

heritage: tracking gene flow across and around thegate of tears.*American Journal of Human Genetics*, **75**(5):752-770

Krause, J., Orlando, L., Serre, D., Viola, B., Pruffer, K, Richards, M.P., Hublin, J.J., Hanni, C., Derevianko, A.P. and Paabo, S. (2007). Neanderthatls in central Asia and Siberia.*Nature*, **449**: 902-90

Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J. and Khrapko, K. (2004)Recombination of Human Mitochondrial DNA.*Science, **304***:981.

Lander, E.S., Linton L., Birren B., Nusbaum, C, Zody, M., Baldwin J., Devon K., et al., (2001).Initial sequencing and analysis of the human genome.*Nature*, **409**:860-921

Lansing, I., Justesen, J. and Loeschcke.V. (2000). Variation in the expression of Hsp70, the major heat-shock protein, and thermotolerance in larval and adult selection lines of *Drosophila melanogaster*. *Journal of Thermal Biology***25**:443–450.

Li, J.Z., Absher D.M, Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M, Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science,***319**:1100–1104.

Librado, P. and Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,***25**:1451-1452

Lightowlers, R., Chinnery, P., Turnbull, D. and Howell, N. (1997). Mammalian Mitochondrial Genetics: Heredity, Heteroplasmy and Disease. *Trends in Genetics,**13***:450-455.

Lopez, J., Yuhki, N., Masuda, R., Modi, W. and O'Brien, S.J. (1994).Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat.*Journal of Molecular Evolution,***39**:174-190.

Lum, J.K., Cann, R.L., Martinson, J.J. and Jorde, L.B. (1998). Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations *American Journal of Human Genetics*, **63**:613-624.

Macaulay, V. and Richards, M. (2001). The mitochondrial gene tree comes of age *American Journal of Human Genetics*, **68**(6): 1315–1320.

Mann, M. and Dalby, D. (1987). *A Thesaurus of African Languages*

Mardis, E.R. (2008).Next-Generation DNA Sequencing Methods *Annual Review Genomics Human Genetics*.**9**:387–402

Margulis, L. (1970a). *Origin of eukaryotic cells*.Yale University Press, New Haven, CT.

Margulis, L. (1970b). Recombination of non-chromosomal genes in chlamydomonas: assortment of mitochondria and chloroplasts? *Journal of Theoretical Biology,* **26** (2):337-342.

Margulis, L. (1975). Symbiotic theory of the origin of eukaryotic organelles; criteria for proof.*Symposia of the Society of Experimental Biology,***29**:21-38.

Margulis, L. (1981). *Symbiosis in cell evolution: Life and its environment on the early earth.* WH Freeman and Co., NewYork.

Marks, J. (2012).The Origins of Anthropological Genetics.*Current Anthropology,* **53**(**S5**):S161-S172

Marrero, A.R, Bravi, C., Stuart, S., Long, J.C., Pereira, L., das Neves Leite, F., Kommers, T., Carvalho, C.M., Pena, S.D., Ruiz-Linares, A., Salzano, F.M. and Bortolini, M.C. (2007). Pre- and post-Columbian gene and cultural continuity: The case of the Gaucho from southern Brazil. *Human Heredity***64**:160–171.

Mazières, S., Sevin, A., Bonnet, F., Crubezy, E., Salzano, F.M. and Larrouy, G. (2007). Genetic studies in French Guiana populations: Synthesis. *American Journal of Physical Anthropol*ogy, **132**:292–300.

Mellars, P. (2006) Going East: New Genetic and Archaeological Perspectives on the Modern Human Colonization of Eurasia. *Science,* **313**:796-800

Meyer, N., Gonzalez, A.M., Pestano, J., Flores, C., Larruga JM., Cabrera VC. (2001). Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genetics*, **4**:15.

Meyer, S., Weiss, G., and von Haeseler A. (1999). Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics,***152**(3):1103-1110.

Mielke, J.H., Konigsberg, L.W., and Relethford, J. (2011). *Human biological variation.* (2nd ed.). New York: Oxford University Press.

Mishmar, D., Ruiz-Pesini, E., Brandon, M. and Wallace, D.C. (2004). Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. *Human Mutation,***23** (2):125-133.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M, Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A. and Wallace, D.C., (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Science USA,***100**:171–176.

MITOMAP, (2013).A human mitochondrial genome database.Center for Molecular Medicine. Emory University, Atlanta, GA, USA.,Retrieved from http://www .mitomap.org

Molnar S. (1998) Human Variation: races, types and ethnic groups. Prentice Hall Upper Saddle River, NJ 07458 pp34

Moritz, C., Dowling, T.E. and Brown, W.M. (1987).Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology, Evolution and Systematics*, **18:**269-292.

Mullis, K.B. and Faloona, F.A., (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain-reaction. *Methods in Enzymol*ogy.**155**: 335-350

Mullis, K.B, Faloona, F.A., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1992). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Biotechnology,***24**:17-27.

National Population Commission (2006). Report on 2006 Nigerian Census pp 206-210

Nei, M. (1987).*Molecular evolutionary genetics.* Columbia University Press, New York

Nei, M. (1995).Genetic Support for the out-of-Africa theory of human evolution.*Proceedings of the National Academy of Science*, *USA.***92**:6720-6722

Nengel, J., Adamu, A., Mohammed, S. and Gwadabe, M. (2002).*History of the Chawai of North-central Nigeria to 1932* pp13-14. Ahmadu Bello University press Ltd.

Neto, D., Montiel, R., Bettencourt, C., Santos, C., Prata, M.J., Lima, M. (2007). The African contribution to the present-day population of the Azores Islands (Portugal): analysis of the Y chromosome haplogroup E. *American Journal of Human Biology,* **19**:854–860.

Network 4.1.1.2 - Fluxus Technology Ltd**.**(2004). [www.fluxusengineering.com].

Nielsen, R. (2005). Molecular signatures of natural selection.*Annual Review of Genetics.***39**:197–218).

North, K.E, Martin, J.L, and Crawford, M.H (2000).The origins of the Irish travellers and the genetic structure of Ireland.*Annals of Human Biology,* **27**:453-465.

Omkar, D., Serafim, B., Marcus, W.F., and Cavalli-Sforza, L.L. (2009).A serial founder effect model for human settlement out of Africa.*Proceedings: Biological Sciences,* **276**(1655):291-300.

Oswald, N. (2007). “Determining DNA Concentration and Purity”.BitesizeBio. Available from http://bitesizebio.comdna-concentration-purity/

Oxygen Free Radicals and Mitochondrial Mutation (2013). Retrieved January, 2013, from http://www.geneticorigins.org/mito/theory4.html

Page, R.D.M. and Holmes, E.C. (1998).*Molecular Evolution: a Phylogenetic Approach*, Blackwell Science.

Pakendorf, B. and Stoneking, M. (2005).Mitochondrial DNA and human evolution *Annual Review of Genomics and Human Genetics.***6**:165-183 DOI: 10.1146/annrev.genom. 6.080604.162249

Pedro, S., Ermini, S., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V. and Richards, M.B. (2009) "Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock". *The American Society of Human Genetics*.**84**(6):82–93. doi:10.1016/j.ajhg.2009.05.001.

Pereira, L., Dupanloup, I., Rosser, Z.H., Jobling M.A. and Barbujani, G. (2001b) Y-chromosome mismatch distributions in Europe.*Molecular Biology and Evol*ution.**18**(7):1259-1271.

Pereira, L., Macaulay V., Torroni A., Scozzari, R., Prata, M.J., Amorim, A. (2001a). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Annals of Human Genetics*, **65**(5):439-458.

Pereira, L., Richards, A., Alonso, A and Albarrán C, (2005).High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refuge.*Genome Research,***15**:19–24.

Piganeau, G., Eyre-Walker, A. (2004) A Reanalysis of the Indirect Evidence for Recombination in Human Mitochondrial DNA.*Heredity,***92**:282-288.

Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, A. and Comas, D., (2004). Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human Genetics,***115**: 439–447.

Pogo, A.O. and Chaudhuri, A. (1997).The Duffy blood group system and its extension in nonhuman primates. In: *Molecular Biology and Evolution of Blood Group and MHC Antigens in Primates,* Eds. Blancher A, Klein J, and Socha W W. Berlin: Springer-Verlag, pp 219-235.

Price, R. and Price, S. (2004). Les Marrons.Culture en Guyane.2 e`me e´dition. Cayenne: Vents d’ailleurs. *Proceedings of the National Academy of Sciences of the USA,* **92**:6720-6722.

Quintans, B., Beleza, S., Brion, M., Sanchez-Diz, P., Lareu, M. and Carracedo, A., (2003) Population data of Galicia (NW Spain) on the new Y-STRs DYS437, DYS438, DYS439, GATA A10, GATA A7.1, GATA A7.2, GATA C4 and GATA H4, *Forensic Science International.***131**:220–224.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the USA,* **102**(44):15942-15947.

Rando, J.C., Cabrera, V.M., Larruga, J.M., Herna´ndez, M., Gonza´ les, A.M., Pinto, F. and Bandelt, H.J. (2000). Phylogeographic patterns of mtDNA reflecting the colonisation of the Canary Islands. *Annals of Human Genetics,***63**:413–428

Rando, J.C., Pinto, F., Gonza´ les, A.M, Herna´ndez M, Larruga, J.M., Cabrera, V.M. and Bandelt, H.J (1998). Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Annals of Human Genetics,***62**:531–550

Reed, F.A., and Tishkoff, S.A. (2006).African human diversity, origins and migrations.*Current Opinion Genetic Development,* **16**: 597–605.

Reid, M.E. and Lomas-Francis, C. (1997).*The Blood group antigen facts book.*San Diego: Academic Press.

Relethford, J.H. (2008). Genetic evidence and the modern human origins debate.*Heredity,***100**(6):555-563.

Reynier, P., May-Panloup, P., Chretien, M-F., Morgan, C.J., Jean, M., Savagner, F., Barriere, P., and Malthiery Y. (2001). Mitochondrial DNA content affects the fertilizability of human oocytes. *Molecular Human Reproduction,***7**(5):425-429.

Riccheti, M., Tekaia, F., and Dujon, B. (2004).Continued colonization of the human genome by mitochondrial DNA.*PLoS Biology,* **2**(9):1313-1324.

Richards, M., Rengo, C., Cruciani, F., Gratrix, F., Wilson, J.F., Scozzari, R., Macaulay, V., Torroni, A. (2003).Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations.*American Journal of Human Genetics*, **72**:1058-1064.

Richards M., and Macaulay, V. (2001).TheMitochondrialGeneTreeComesofAge *American Journal of Human Genetics,***68**:1315-1320

Rogers, A.R. and Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*.**9**(3):552-569.

Rogers, A.R. and Harpending, H.C. (1983).Population structure and quantitative characters.*Genetics*, **105:**985-1002

Rosa, A., Brehm, A., Kivisild, T., Metspalu, E. and Villems, R. (2004). MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Annals of Human Genetics*, **68**:340-352.

Rosenberg, N.A, Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. and Feldman, M.W. (2002).Genetic structure of human populations.*Science,***298**:2381–2385.

Rubicz, R.C., Melton, E.M. and Crawford, M.H. (2007). In: *Anthropological genetics: Theory Methods and applications.* Ed Crawford, M.H. Cambridge UK, Cambridge University Press 141-186

Rubinoff, D. and Holland, B.S. (2005). Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Systemic Biology,* **54**(6): 952-961. DOI: 10.1080/10635150500234674

Ruiz-Pesini, E., Lapena, A.C., Diez-Sanchez, C., Perez-Martos, A., Montoya, J., Alvarez, E.,Diaz, M., Urries, A., Montoro, L., Lopez-Perez, M.J., Enriquez, J.A., (2000). Human mtDNA haplogroups associated with high or reduced spermatozoa motility. *American Journal of Human Genetics,***67**:682–696.

Saccone, C., Pesole, G., Sbisà, E. and Preparata, G. (1992). Time and biosequences: a contribution to the origin of modern man *Human Evolution,* **7**(2): 37-46

Saccone, S., De Sario, A., Della Valle G., Bernardi, G. (1992). The highest gene concentra-tions in the human genome are in T-bandsof metaphase chromosomes. *Proceedings of the National Academy of Sciences of the USA*.**89**:4913-7

Saiki, R.K., Scharf, S., Faloona. F, Mullis, K.B., Horn G.T., Erlich, H.A. and Arnheim, N. (1985).Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anemia.*Science,***230**:1350-1354

Saitou, N. and Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution,***4**(4):406-25.

Salas, A., Richards, M., Lareu, M.V., Scozzari R, Coppa A, Torroni A, Macaulay V. and Carracedo, A. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade *American Journal of Human* Genetics, **74**:454-465.

Salas, A., Richards, M., De la Fe, T., Lareu, M. V., Sobrino, B., Sanchez-Diz, P., Macaulay, V. and Carracedo, A. (2002). The making of the African mtDNA landscape.*American Journal of Human Genetics,* **71:**1082–1111. (doi:10.1086/344348)

Samarin, W.J. (1971). 'Adamawa-Eastern.' In: Sebeok, T.A. (ed.) *Current trends in linguistics*, Vol. 7: Linguistics in Sub-Saharan Africa. The Hague: Mouton.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977).DNA sequencing with chain-terminating inhibitors.*Proceedings of the National Academy of Sciences of the USA,* **74**:5463-5467

Santos, S.E.B., Rodrigues, J.D., Ribeiro-dos-Santos, A.K., and Zago, M.A. (1999).Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and YDNA.*American Journal of Physical Anthropol*ogy, **109**:175–180.

Santos, T.A., El Shourbagy, S. and St John, J.C. (2006). Mitochondrial content reflects oocyte variability and fertilization outcome. *Fertility and Sterility*, **85**:584-591.

Sarich, V.M. and Wilson, A.C. (1967). Immunological time scale for hominid evolution *Science,***158**:1200-1203

Schneider, S. and Excoffier, L. (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics,***152** (3):1079-1089.

Schneider, S., Roessli, D. and Excoffier, L. (2000). Arlequin version 3.5: A software for population genetics data analysis. http://anthropologie.unige.ch/arlequin

Schurr, T.G. and Sherry, S.T. (2004). "Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence" *American Journal of Human Biology,* **16**(4):420-39.

Schwartz, M. and Vissing, J. (2002).Paternal inheritance of mitochondrial DNA.*New England Journal of Medicine,***347**:576-580

Shoubridge, E.A. (2001). Nuclear genetic defects of oxidative phosphorylation.*Human Molecular Genet*ics, **10**:2277-2284.

Skaletsky, H., Kuroda-Kawaguchi, T, Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikovak, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., …………. and Page, D.C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature, **423*** (6942):825-37 PMID: 12815422

Slatkin, M. (1985).Rare alleles as indicators of gene flow.*Evolution,* **39**:53

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M.B., (2009).Correcting for purifying selection: an improved human mitochondrial molecular clock.*American Journal of Human Genetics,***84**:740–759.

StatSoft, Inc. 2004.STATISTICA (data analysis software system), version 7. www.statsoft.com.

Stone, A.C. and Stoneking M. (1998). MtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *American Journal of Human Genetics,***62**:1153-1170.

Stoneking, M. (2008). Human origins: The molecular perspective. *EMBO reports* 9, S1, S46-S50 doi: 10.1038/embor.2008.64

Strachan, T. and Read, A.W. (2011).*Human Molecular Genetics* 4ed. Garland Science/Taylor and Francis Group,

Stringer, C. (2002).Modern Human Origins: Progress and Prospects. *Philosophical Transactions, Biological Sciences,* **357**:563-579.

Sykes, B., Leiboff, A., Low-Beer, J., Tetzner, S., Richards, M. (1995).The origins of the Polynesians: An interpretation from mitochondrial lineage analysis. *American Journal of Human Genetics*, **57**:1463-1475.

Tamura, K., and M. Nei.(1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.*Molecular Biology and Evolution*. **10**:512–526.

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A., Singh, L. (2005).Reconstructing the origin of Andaman Islanders.*Science*, **308**(5724):996.

Tishkoff, S.A., Dietzsch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K., Bonne-Tamir, B., Santachiara-Benerecetti, A.S., Moral, P., and Krings, M. (1996). Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science,* **271***:* 1380-1387

Tishkoff, S.A. and Gonder M.K. (2007). In: *Anthropological genetics: Theory Methods and applications.* Ed Crawford MH Cambridge UK, Cambridge University Press 141-186

Tishkoff, SA., Gonder, M.K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., . . . Mountain, J. L. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution*, **24**(10): 2180-2195. doi:10.1093/molbev/msm155

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C. and Ranciaro, A., (2009). The genetic structure and history of Africans and African Americans.*Science,***324**:1035–1044.

Tishkoff , S.A, Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A. ……… and Williams, S.M. (2009). The genetic structure and history of Africans and African Americans.*Science,***324**:1035–1044.

Tishkoff, S.A. and Verrelli B.C. (2003). Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. *Current Opinion in Geneticsand Development*, **13**:569-575

Tishkoff, S.A. and Williams, S.M (2002). Genetic analysis of African population: human evolution and complex disease. *Nature Reviews Genetics,* **3**: 611-21

Torroni, A., Achilli, A., Macaulay, V., Richards, M. and Bandelt, H.J. (2006).Harvesting the fruit of the human mtDNA tree.*Trends in Genetics,***22**(6): 339-345

Torroni, A., Huoponen, K., Francalacci, P. (1996). Classification of European mtDNAs from and analysis of three European populations.*Genetics,***144**:1835–1850.

Torroni, A., Neel, J.V., Barrantes, R., Schurr, T.G. and Wallace, D.C. (1994).Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America.*Proceedings of the National Academy of Sciences of the USA.*,**91**:1158-1162.

Torroni, A, Richards, M., Macaulay, V., Forster, P., Villems, R., Nørby, S. and Savontaus, M.L., (2000).mtDNA haplogroups and frequency patterns in Europe. *American Journal Human Genetics,***66**: 000–000

Torroni, A, Schurr, T.G, Cabell, M.F, Brown, M.D, Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M. and Wallace, D.C. (1993).Asian affinities and continental radiation of the four founding Native American mtDNAs.*American Journal of Human Genetics,***53**: 563-590

Torroni, A, Schurr T.G, Yang, C.C., Szathmary, E.J., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M. (1992). "Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations" *Genetics,***130**(1): 153-62

Tranah, G.J., Manini, T.M., Lohman, K.K., Nalls, M.N., Kritchevsky, S., Newman, A.B., Harris, T.B., Miljkovic, I., Biff,i A., Cummings, S.R. and Liu, Y., (2011). Mitochondrial DNA variation in human metabolic rate and energy expenditure, *Mitochondrion,*DOI:10.1016/j.mito.2011.04.00

Tripathy, V. and Reddy, B.M. (2007). Present status of understanding on the G6PD and natural selection. *Journal of Postgraduate Medicine*.**53**:193-202

Underhill, P.A, Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, R.W., Feldman M.W., Cavalli-Sforza, L.L. and Oefner, P.J. (2000) Y chromosome sequence variation and the history of human populations. *Nature Genetics*, **26**:358–361

van Oven, M. and Kayser, M. (2008) Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation *Human Mutation* Mutation in Brief #1039, **29**:E386-E394, Online DOI: 10.1002/humu.20921

Vansina, J. (1995). New Linguistic Evidence and 'the Bantu Expansion'.*The Journal of African History*, **36**(2):173-195

Veeramah, K.R., Connell, B.A., Pour, N.A., Powell, A., Plaster, C.A., Zeitlyn, D., Mendell, N.C., Weale, M.E., Bradman, N. and Thomas, M.G. (2010). Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evolutionary Biology*, **10**:92

Vellai, T., Takács, K. and Vida, G. (1998).A new aspect to the origin and evolution of Eukaryotes.*Journal of Molecular Evolution,***46**, 499–507

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., (2001).The sequence of the human genome.*Science,***291**:1304-1351

Wallace, D.C. (2007). Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine.*Annual Review of Biochemistry,* **76:**781-821.

Wallace, D.C. and Lott, M.T. (2004)."MITOMAP: A Human Mitochondrial Genome Database" Retrieved from http://www.mitomap.org.

Wallace, D.C., Ruiz-Pesini, E. and Mishmar, D. (2003).mtDNA variation, climatic adaptation, degenerative diseases, and longevity. *Cold Spring Harbour Symposium on Quantitative Biology*, **68**: 479–486.

Watkins, W.S., Rogers, A.R., Ostler, C.T., Wooding, S., Bamshad, M.J., Brassington, A.M.E., Carroll, M.L,, Nguyen, S.V., Walker, J.A. and Prasad, B. (2003). Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. *Genome Research,***13**:1607–1618.

Watson, E., Forster, P., Richards, M. and Bandelt, H. J. (1997). Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics,* **61**:691–704. (doi:10.1086/515503)

Welmers, W. E. (1956). Review of studies in African linguistic classification, by Joseph H. Greenberg, *Language,***32**: 56-63

Williamson, K. and Blench, R.M. (2000). Niger-Congo. In: *African languages: an introduction*. Heine, B. and Nurse, D. eds. 11-42. Cambridge: Cambridge University Press.

Wilson, J.F., Weale M.E. and Smith, A.C. (2001).Population genetic structure of variable drug response.*Nature Genet*ics, **29**:265-269

Wolpoff, M.H., Hawks, J. and Caspari, R. (2000), Multiregional, not multiple origins. *American Journal of Physical Anthropology*.**112**: 129–136. DOI: 10.1002/ (SICI)1096-8644(200005)112:1<129::AID-AJPA11>3.0.CO;2-K

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

Zitt localization project May 8, 2013.Retrieved from http://zitt.sourceforge.net/ zitt.php? Su=eng&ibe=4

Appendix I

Letter of Consent

This letter is to give you information in the hope that you will participate in a study for a Dissertation as part of my PhD anatomy program. This study seeks to find out variation in the genetic structure of your ethnic group. I hope to learn more about the topic area and develop my research skills.

Participation in this study is voluntary. It will involve the painless swabbing of your mouth and collection of buccal rinse for the sole purpose of DNA extraction from the cells.

If you request, any personal identifying information will not appear in the course of this study and after it.

The sample collected from this interview will only be used for the above objectives and scientific publications, it will not be used for any other purpose and will not be recorded in excess of what is required.

There are no known or anticipated risks to you as a participant in this study.

I assure you that this study has been reviewed and approved by my Supervisors and the University committee on ethics. Thank you for your assistance in this research.

Yours Sincerely,

Timbuak James Abrak
Department of Anatomy
Faculty of Medicine
Ahmadu Bello University,
Zaria

I have read the information presented in the information letter about the study being conducted by Timbuak James Abrak for his PhD thesis.

I have had the opportunity to ask any questions related to this study, and received satisfactory answers to my questions, and any additional details I wanted.

I was informed that I may withdraw my consent at any time by advising the student researcher.
With full knowledge of all foregoing, I agree to participate in this study.

Participant Name: _____________________________
Participant Signature: _____________________________
Interviewer Name: _____________________________
Interviewer Signature: _____________________________

Appendix II

Ethical committee clearance form

POSTGRADUATE OFFICE

FACULTY OF MEDICINE

AHMADU BELLO UNIVERSITY TEACHING HOSPITAL

SHIKA - ZARIA, NIGERIA

P.M.B. 06 Samaru Zaria
Phone: 069 -551399
Fax: 069 - 555001
E-mail: abuthpg@yahoo.com

Your Ref:

Our Ref: ABUTH/PGO/COMM/9

Date: 6th March, 2011

James Abrak Timbuak
Dept. of Anatomy
Faculty of Medicine
ABU
Zaria,

ETHICAL CERTIFICATE

Your application for ethical clearance on your proposal titled **"ANALYSIS OF MTDNA AND Y-CHROMOSOME VARIABILITY: A STUDY OF 4 ETHNIC GROUPS FROM THE SOUTHERN PART OF KADUNA STATE, NIGERIA"** refers

This is to convey ethical approval for you to commence the study. You would be required to update the committee of progress on your work.

Thank you.

DR AISHA I MAMMAN
Asst Dean Postgraduate Studies/
Residency Training Co-ordinator

Appendix III

Human Subject committee Lawrence Campus ethical clearance

7/17/09

HSCL #18079

Michael Crawford
Anthropology
5 Lippincott Hall

The Human Subjects Committee Lawrence Campus (HSCL) has reviewed your research project application

18079 Crawford/ (ANTHRO) Investigation of the Genetic Structure of the in Nigeria

and approved this project under the expedited procedure provided in 45 CFR 46.110 (F) (3) Prospective collection of biological specimens for research purposes by noninvasive means. As described, the project complies with all the requirements and policies established by the University for protection of human subjects in research. Unless renewed, approval lapses one year after approval date.

The Office for Human Research Protections requires that your consent form must include the note of HSCL approval and approval date, which has been entered on the consent form sent back to you with this approval.

1. At designated intervals until the project is completed, a Project Status Report must be returned to the HSCL office.

2. Any significant change in the experimental procedure as described should be reviewed by this Committee prior to altering the project.

3. Notify HSCL about any new investigators not named in original application. Note that new investigators must take the online tutorial at http://www.rcr.ku.edu/hscl/hsp_tutorial/000.shtml.

4. Any injury to a subject because of the research procedure must be reported to the Committee immediately.

5. When signed consent documents are required, the primary investigator must retain the signed consent documents

for at least three years past completion of the research activity. If you use a signed consent form, provide a copy of the consent form to subjects at the time of consent.

6. If this is a funded project, keep a copy of this approval letter with your proposal/grant file.
Please inform HSCL when this project is terminated. You must also provide HSCL with an annual status report to maintain HSCL approval. Unless renewed, approval lapses one year after approval date. If your project receives funding which requests an annual update approval, you must request this from HSCL one month prior to the annual update. Thanks for your cooperation. If you have any questions, please contact me.

Sincerely,

Jan Butin
Associate Coordinator
Human Subjects Committee – Lawrence

Human Subjects Committee Lawrence
Youngberg Hall | 2385 Irving Hill Road | Lawrence, KS 66045-7563 | (785) 864-7429 | Fax (785) 864-5049 | www.rcr.ku.edu/hscl

Appendix IV

PCR Amplification (HVS1_mini DNA) protocol Date:_____________

Samples: Population:

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48

Reaction Mix Per Sample:
10.0 ul 5x buffer
8.0 ul MgCl2
2.0 ul BSA
0.4 ul GoTaq Flexi
1.0 ul dNTPs
23.6 ul dd H_2O
2 ul For Primer (__________)
2 ul Rev Primer (__________)

x 50 samples:
500.0 ul 5x buffer (250+250)
400.0 ul MgCl2 (200+200)
100.0 ul BSA (50+50)
20.0 ul Taq Pol (10+10)
50.0 ul dNTPs (25+25)
1180.0 ul ddH_2O (590+590)
100.0 ul For Primer (50+50)
100.0 ul Rev Primer (50+50)
2450 ul total PCR mix

50.0 ul total rxn volume

Aliquot 49 ul PCR mix into each tube: _____
Add 1.0 ul DNA dilution (1 ng/ul): _____

Thermal Profile: Annealing temperature for this primer pair: ___°

Date of Gel: _____________
Date of Digest: _____________

Appendix V

RFLP (Digestion) Gel protocol

Samples: ______________________ For Haplogroup: _________________

Electrophoresis will use a 3% agarose gel which will be poured at 50° C because this appears to produce sharper bands. Allowing the gel to cool overnight improves visualization of bands. The bands will be visualized using ethidium bromide. For a 3% 150 ml gel: 4.5g NuSieve 3:1 agarose in 150 mls of 1x TBE*. The gel will be run at 97V for about 1 ½ hours for maximum resolution. For the digestion gel, load 10µl digested sample (already contains loading buffer) into each well.

samples in the gel:

1. _________________
2. _________________
3. _________________
4. _________________
5. _________________
6. _________________
7. _________________
8. _________________
9. _________________
10. ________________
11. ________________
12. ________________
13. ________________
14. ________________
15. ________________
16. ________________
17. ________________
18. ________________
19. ________________
20. ________________
21. ________________
22. ________________
23. ________________
24. ________________
25. ____________________
26. ____________________
27. ____________________
28. ____________________
29. ____________________
30. ____________________
31. ____________________
32. ____________________
33. ____________________
34. ____________________
35. ____________________
36. ____________________
37. ____________________
38. ____________________
39. ____________________
40. ____________________
41. ____________________
42. ____________________
43. ____________________
44. ____________________
45. ____________________
46. ____________________
47. ____________________
48. ____________________

Gel Start Time: ________
Gel Start Volts: ________
Gel Start mAmps:_______

Gel End Time: ___________
Gel End Volts: ___________
Gel End mAmps: _________

DNA standard used:

____

Date of Gel: ___________________
Date of PCR: __________________

Specifically, the protocol for preparing the
is as follows

1) Tape and seal the casting tray for the
2) Using FMC NuSieve 3:1, weigh out 4. gms.
3) Put 150 mls of 1x TBE in a flask (or beaker) and add a stir bar.
4) With the stir bar moving on the magn stir plate, add the agarose.
5) Remove the stir bar and weigh the fla (and record the weight): __________
6) Place plastic wrap over the top of the flask and poke a hole in the plastic w
7) Heat the agarose in the microwave fo about 5 minutes or until the agarose completely dissolved, pausing every minute to swirl the contents.
8) Remove the plastic wrap and weigh th flask.
9) Add room temperature distilled wate the side of the flask to bring it to the original weight.
10) Add a stir bar and a thermometer to flask and allow the agarose to cool as stirs on the magnetic stir plate.
11) When the flask reaches 50°C remove stir bar and thermometer and add 5 µ of ethidium bromide.
12) Carefully pour the gel into the casting tray, trying to avoid bubbles.
13) Add the casting comb.
14) The gel will set in about 10-15 minute If you need to store the gel for severa hours (or overnight), it can be wrapp plastic wrap and stored in the refrigerator.

Appendix VI

Exosap purification protocol

Purification using Exo-sap (47 samples)

1. Dilute to a working stock of 20 % (30 ul of stock into 120ul of molecular grade water)
2. Use 2 ul of this working mix for every 25 ul of PCR sample.
    a. Amount of PCR product: 15 ul
    b. Amount of Exo-sap dilution to be added: 1.2ul
3. Once the Exo-sap is added to the PCR sample, place it on a PCR machine.

Thermocycle Profile:
37° C -30:00 (1 cycle)
80° C – 15.00 (1 cycle
4° C - ∞

1	12	25	37
2	13	26	38
3	14	27	39
4	15	28	40
5	16	29	41
6	17	30	42
7	18	31	43
8	19	32	44
9	20	33	45
10	21	34	46
11	22	35	47