POPULATION STRUCTURE AND DEMOGRAPHICS IN NIGERIAN POPULATIONS UTILIZING Y-CHROMOSOME MARKERS

Curtis Lanre Cole-Showers

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor Philosophiae in the Department of Biotechnology, University of the Western Cape

Supervisor: Prof. Maria Eugenia D'Amato

Co-Supervisor: Prof. Sean Davison

JUNE 2014

KEYWORDS

Nigerian populations

Male subjects

Linguistic affiliations

Geographical origins

Y-STRs

Y-SNPs

Summary statistics

Genetic structure

Demographic parameters

Forensic values

To**: The glory of the Lord God Almighty**

ABSTRACT

Population structure and demographics in Nigerian populations utilizing Y-chromosome markers

C. L. Cole-Showers

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

Nigeria is peopled by ethnically and linguistically diverse populations of which little were known until the last few millennial. The absence of major natural geographical barrier increases the possibility of the populations being affected by the same demographic events. The aim of this thesis was to ascertain the genetic variations and demographics in five major Nigerian populations using Y-markers. This was done by determining the genetic structures of the Afro-asiatic speaking Hausa (n=78) of Northern Nigeria and the Niger Congo speaking populations of Igbo (n=119), Yoruba (n=238), Bini (n=13) and Ijaw (n=15) of Southern Nigeria all spread over 22 geographical origins and four (North, South east, south west and South south) geographical regions. They were compared with more than 2000 individuals from 46 populations of 20 other African and Middle Eastern countries, in published literature. The Scientific Working Group on DNA Analysis Methods (SWGDAM) recommended Y-Short Tandem Repeats (STRs) and nine Y-Single Nucleotide Polymorphisms (SNPs) haplogroups were typed with multiplex Polymerase Chain Reaction (PCR), Restriction Fragment Length Polymorphisms (RFLP) and High Resolution Melting (HRM). Summary statistics and measures of diversity were determined. Population structure was assessed with Population Pairwise Differences, hierarchical Analysis of Molecular Variance, Multidimensional scaling and correspondence analysis plots. Mantel's test was used to assess the correlation of genetic distances with geographic distances. Demographic inferences were assessed with lineage based Network reconstruction, Spatial autocorrelation plots, effective migrants per population and both Inter and Intra-lineages Times to the Most Recent Common Ancestor (TMRCA). The patterns of diversity of the Y-markers showed a North-South gradient and a notable sub-structure among the Hausa populations. The Niger-Congo speakers displayed rare presence of haplogroups R and E1b1b but a preponderance of E1b1a7. Overall, the Y markers showed high diversities and significant genetic sub-structure within the Hausa populations of Nigeria with stronger linguistic than geographical bias. The demographic evaluations gave credence for genetic validation of both historical records and archeological findings among these Nigerian populations. These populations showed stronger affiliations with other sub-Saharan African populations rather than with North African or Middle Eastern populations, lacking evidence for the Middle Eastern origins of the male founders of these populations. Finally, the contribution of these Nigerian dataset would greatly enhance the Africa meta-population on the YHRD with more than 274 new haplotypes of forensic estimation significance.

June2014

DECLARATION

I declare that *Population structure and demographics in Nigerian populations utilizing Y-chromosome markers* is my own work, that it has not been submitted for any degree or examination in any other University, and that all the sources I have used or quoted have all been indicated and acknowledged by complete references.

Full Name: Curtis Lanre Cole-Showers Date: June, 2014

Signature: 

ACKNOWLEDGMENT

First and most importantly, I thank the Lord God Almighty for His providential sustenance throughout my stay in South Africa. I want to thank all the DNA donors for this work. I want to appreciate my institution in Nigeria, the Lagos State University, for granting me the leave to pursue this degree. I want to thank all my friends, especially Prof. Oluwafemi Oguntibeju of Cape Peninsular University of Technology (CPUT), Dr. Sunday Adewara and Dr. Olabisi Falowo (both of the University of Cape Town) for their spiritual and moral support. I thank my colleagues in the Forensic DNA Laboratory, FDL, of the University of the Western Cape, for making my stay in South Africa productive. I appreciate the mentorship and guidance of my Supervisors, Prof. Maria Eugenia D'Amato and Prof. Sean Davison. I want to also appreciate Prof. Ralf Henkel, Prof. David Fisher and Prof. Lorna Holtman, all of whom I never met physically but whose fastidious interest and interventions delivered me from cumbersome bureaucracy that could frustrate all the great effort that have gone into this thesis. I appreciate my family especially my wife, Ruth and my children William and Wesley, for their sacrifices throughout my absence from Nigeria to study in the Republic of South Africa.

Table of Content

List of tables

List of Figures

Abbreviations

μ	Mutation rate
AMOVA	Analysis of Molecular Variance
ASD	Average Square Distance
BSA	Bovine Serum Albumin
CEPH	Centre d'Etude du Polymorphisme Humain
CI	Confidence Interval
D	Gene Diversity
DC	Discrimination Capacity
DNA	DeoxyriboNucleic Acid
dNTP(s)	Deoxyribonucleotide Triphosphate
HD	Haplotype Diversity
HRM	High Resolution Melting
MCMC	Markov Chain Monte Carlos
MDS	Multidimensional Scaling
MgCl2	Magnesium Chloride
mtDNA	Mitochondrial DNA
NIST	National Institute for Standards and Technology
NRY	Non-recombining Region of the Y-Chromosome
PCR	Polymerase Chain Reaction
r	Correlation Coefficient
RFLP(s)	Restriction Fragment Length Polymorphism(s)
RMP	Random Match Probability
SNP(s)	Single Nucleotide Polymorphism(s)
SRM	Standard Reference Material
STR(s)	Short Tandem Repeat(s)
SWGDAM	Scientific Working Group on DNA Analysis Methods
Taq	Thermophilus aquaticus
TMRCA	Time to the Most Recent Common Ancestor
UEP(s)	Unique Event Polymorphism(s)
VNTR(s)	Variable Number Tandem Repeat(s)
YHRD	Y-Chromosome haplotype Reference Database

UNIVERSITY of the
WESTERN CAPE

1.0 CHAPTER ONE: INTRODUCTION

The non-recombining part of the human Y chromosome (NRY) was widely used in human population (Underhill and Kivisild, 2007) and forensic genetics (Kayser, 2007) because it showed a male inheritance and substantial structuring in human populations (Karafet, *et al.*, 2008). With its particular susceptibility to genetic drift caused by low effective population size (Jobling and Tyler-Smith, 2003) and the additional influence of patrilineal cultural practices (King and Jobling, 2009; Oota, *et al.*, 2001; Seielstad, *et al.*, 1998) the NRY provides the strongest genetic differentiation over geographic distance when compared with other parts of the genome (Hammer and Zegura, 2002; Kayser, *et al.*, 2000). This has made the NRY exceptionally valuable for the reconstruction of human population history (Kayser, *et al.*, 2000; Underhill, *et al.*, 2000) including estimation of demographic parameters (Shi, *et al.*, 2010) as well as for genealogical relationships (Kayser, 2007) and male lineage determination in forensic applications (Kayser, *et al.*, 2007; Dettlaff-Kakol and Pawlowski, 2002; Roewer, 2009).

van Oven and others (van Oven, *et al.*, 2014) also enumerated the uses of the largely non-recombining portion of the human Y chromosome, to include, as molecular markers in fields like population history and evolutionary anthropology (Chiaroni, *et al.*, 2009; Underhill and Kivisild, 2007), population structure (Larmuseau, *et al.*, 2013; Ottoni, *et al.*, 2011), genetic genealogy (King and Jobling, 2009; Larmuseau, *et al.*, 2012), Forensic

genetics (Kayser, 2007); sex chromosome structure and evolution (Hallast, *et al.*, 2013; Trombetta, *et al.*, 2010) and medical genetics (Charchar, *et al.*, 2012).

The skeleton of a detailed phylogenetic tree of Y-chromosome was primarily based on binary polymorphisms (single nucleotides polymorphism (Y-SNP) and insertion/deletion polymorphisms) and specific branches are assigned to haplogroups following a hierarchical pattern (**Figure 1.0** below). This was because apart from being bi-allelic, they are relatively stable (Xue, *et al.*, 2009) and easy to genotype (van Oven, *et al.*, 2011; 2013).

The Y chromosome tree consists of at least 20 major clades containing more than 311 distinct haplogroups defined by more than 600 mutational events (Karafet, *et al.*, 2008, Jobling, 2012; van Oven, *et al.*, 2014). Furthermore, by typing Y-chromosome Short Tandem Repeats (Y-STRs), haplotypes are generated, which are then used for finer resolution within the haplogroups (Underhill and Kivisild, 2007). Prior to recent times, the two primary splits in the Y-chromosome tree leads to two branches, Haplogroups A and BT. The core Haplogroups A and B are associated with the distribution of ancient hunter-gatherer tribes before the expansions of pastoralists (Underhill, *et al.*, 2001; Underhill and Kivisild, 2007; Tishkoff, *et al.*, 2007; Pickrell, *et al.*, 2013). The rest of the Y-chromosome tree was defined by the M168 mutation, which represented the most common African lineages (Haplogroup E) as well as all the non-African clades. Haplogroup A was defined by the M91 and P97 mutations and contained more than 12

branches determined by not less than 45 (internal) mutations (Karafet, *et al.,* 2008; Batini, *et al.,* 2011). A strict regional distribution was particularly pronounced for haplogroup A.

Within Haplogroup A, A1 was found in Mali and Morocco (Underhill, *et al.,* 2000; Scozzari, *et al.,* 2001), A3b2 was found in East Africa (Sudan, Ethiopia, Tanzania, Kenya) and in lower frequencies in northern Cameroon (Scozzari, *et al.,* 1999; Underhill, *et al.,* 2000; Cruciani, *et al.,* 2002; Semino, *et al.,* 2002; Knight, *et al.,* 2003), while A3b1 and A2 are found exclusively among the Khoe-San (Scozzari, *et al.,* 1999; Underhill, *et al.,* 2000).

Haplogroup B was defined by four mutations (M60, M181, P85, and P90) and contained 17 branches with 28 internal markers (Karafet, *et al.,* 2008, Batini, *et al.,* 2011). Haplogroup B occurred throughout Africa but have higher frequencies among Pygmies, Khoe-San and Hadza, with some lineages being restricted to them (Underhill, *et al.,* 2000; Cruciani, *et al.,* 2002; Semino, *et al.,* 2002; YCC, 2002; Knight, *et al.,* 2003). There was a clear-cut difference between the B haplogroups associated with the Pygmies, Khoe-San and Hadza versus all the other African populations. Pygmies, Khoe-San and Hadza populations have mainly Haplogroup B haplotypes defined by the M112 mutation, while other populations have the M150 mutation. Within haplogroup B-M112, haplogroups B2b2, B2b3 and B2b4b were restricted to the Pygmy populations while B2b1 (P6) and B2b4a (P8) were restricted to Khoe-San groups. The B2b* ancestral

haplotype occurred in both Pygmy and Khoe-San groups (Underhill, *et al.*, 2000; Cruciani, *et al.*, 2002; Semino, *et al.*, 2002; YCC, 2002; Knight, *et al.*, 2003).

Eighteen mutations currently define Haplogroup E. It was the most mutationally-diverse of all the major Y-chromosome clades and contained 83 polymorphisms that define 56 distinct haplogroups (Karafet, *et al.*, 2008). The E haplogroups are found at very high frequencies in Africa, moderate frequencies in the Middle East and southern Europe but scarce in Central and South Asia. Although Haplogroup E groups are widespread all over Africa, the distributions of the numerous distinctive haplogroups are not homogeneous across the continent (Hammer and Horai, 1995; Hammer, *et al.*, 1998; Qamar, *et al.*, 1999; Bosch, *et al.*, 2001; Hammer, *et al.*, 2001a; Underhill et al., 2001; Cruciani, *et al.*, 2002; Cruciani, *et al.*, 2004). Haplotypes carrying the mutations M75 (E2) and M33 (E1a) are present at low frequencies across Africa but with different individual distributions. Haplogroups E1b1a and E1b1b are the most frequent and widespread of the E haplogroups (Hammer, *et al.*, 2001a; Underhill, *et al.*, 2001; Cruciani, *et al.*, 2002; Cruciani, *et al.*, 2004; Semino, *et al.*, 2004). Haplogroup E1b1a, defined by M2 and seven other mutations was mainly limited to sub-Saharan populations and was associated with the expansion of Bantu-speaking populations (Hammer, *et al.*, 1998; Passarino, *et al.*, 1998; Scozzari, *et al.*, 1999). The E1b1a subgroups have different distributions and frequencies. The M191 mutation defines the most frequent E-M2 subgroup and was an evidence of a founder effect that resulted from the Bantu-expansions (Hammer, *et al.*,

2001a; Underhill, *et al.,* 2001; Cruciani, *et al.,* 2002; Cruciani, *et al.,* 2004; Semino, *et al.,* 2004).

The non-African distribution of haplogroup E was associated with haplogroup E1b1b characterized by the M35 and M215 mutations (Hammer, *et al.,* 1998; Semino, *et al.,* 2000; Underhill, *et al.,* 2001; Semino, *et al.,* 2004). This haplogroup, however, also have a widespread African representation (Hammer, *et al.,* 2001a; Underhill, *et al.,* 2001; Cruciani, *et al.,* 2002; Cruciani, *et al.,* 2004; Semino, *et al.,* 2004). Compared to other E haplogroups, M35 occur at very low frequencies within Bantu speakers but is widely, though not uniformly, dispersed throughout Africa. Among the different lineages carrying the M35 mutation, haplotypes defined by M78 occurs in East Africa, North Africa, the Middle East and Europe. It is the E-M35 subgroup with the highest frequency and the widest distribution outside Africa. This marker has a northeastern African origin and multiple exodus routes out of Africa have been demonstrated (Cruciani, *et al.,* 2007).

Karafet and others (Karafet, *et al.,* 2008) identified eight mutations: M207, M306, P224, P227, P229, P232, P280, and P285 which characterized haplogroup R. A total of 42 mutations identified the 28 subclades nested in it. It has also been found in significant proportion in Central/West Africa (van Oven, *et al.,* 2014), as R1b2-V88, a version of R1b (SNP V88 = rs 180946844), despite being preponderant among Europeans.

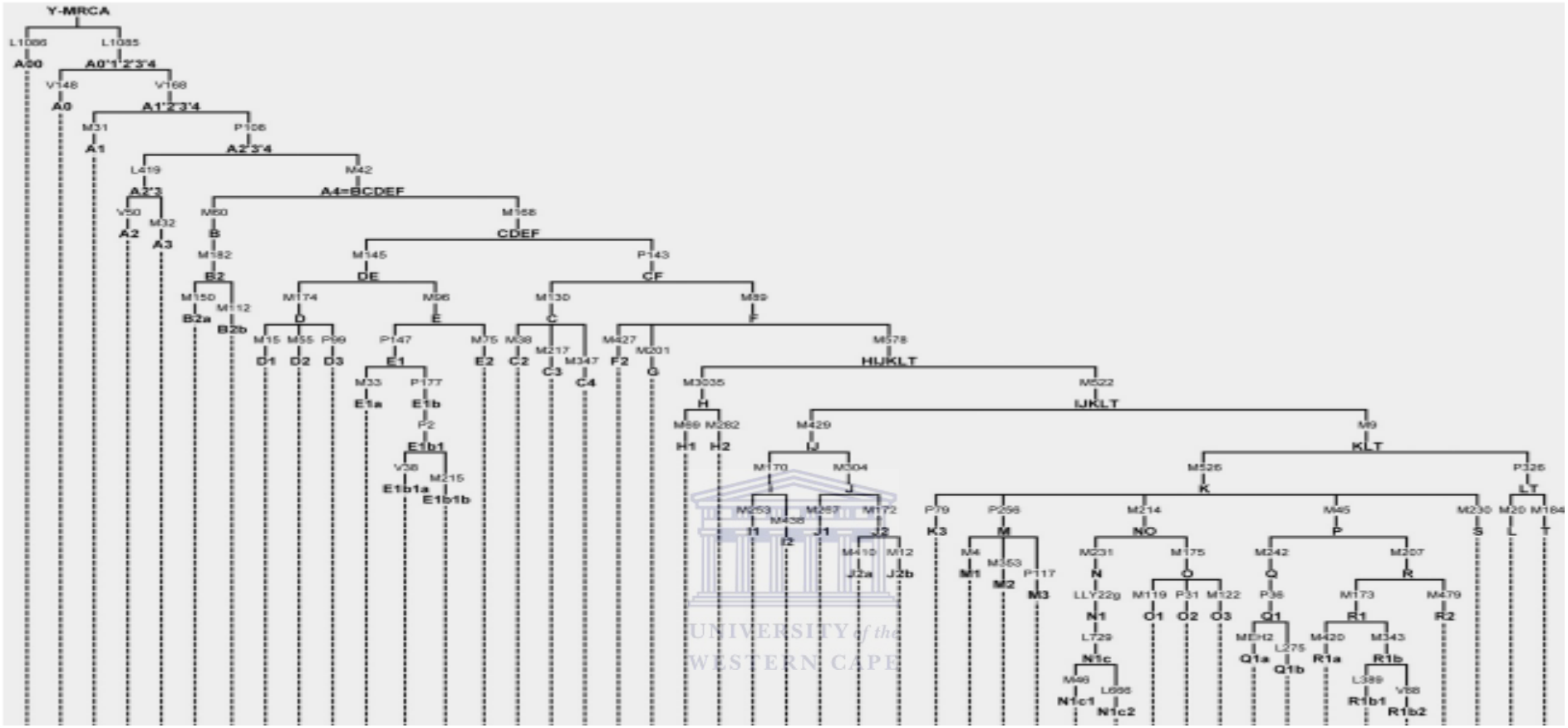


Figure 1.1: Skeleton of the human Y-chromosome phylogeny with basal haplogroup nomenclature (in bold) and defining Y-SNP markers indicated on the branches. The most recent common patrilineal ancestor of all modern humans (Y-MRCA) is indicated at the top. The deep-rooting A and B clades (on the left) are specific to Africans thus supporting the African origin of modern humans (van Oven, *et al.*, 2014)

The number of Y chromosome short tandem repeat (Y-STR) loci available for use in human identity testing has increased considerably since the turn of the century. In the 1990s only few Y STR markers were characterized and available for use. The first STR locus to be identified on the Y chromosome was DYS19 (Roewer, *et al.*, 1992). A series of highly polymorphic Y-specific microsatellites have been identified and tested on different population samples. These markers show high levels of heterogeneity within and between populations and thus very useful for population genetic, evolutionary and forensic applications (de Knijff, *et al.*, 1997).

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

In 2004, the Scientific Working Group on DNA Analysis Methods (SWGDAM) voted to adopt 11 Y-STR loci for forensic casework analysis. The decision was based on availability to the scientific community and the large amount of published performance and database information for most to these loci. The committee encourages further study of additional loci as to their suitability for forensic use. The first nine loci comprise the

minimal haplotype complement of markers, plus two other additional markers DYS 438 and DYS 439 (Ayub, *et al.*, 2000; Daniels, *et al.*, 2004).

The limitation of Y-STRs compared with autosomal STRs was a reduced power of discrimination due to a lack of recombination throughout most of the Y-chromosome (Mulero, *et al.*, 2006). However the advantages of Y-STR analysis over autosomal STRs include: a) male profile can be obtained in the presence of large amounts of female DNA; b) differential extraction of sperm and non-sperm fraction is not necessary; c) analysis of azoospermic semen samples from vasectomized males is feasible; d) the number of male contributors often can be determined in multiple rape cases because of the haploid nature of the Y-STRs; e) rapid exclusion of suspects can occur; f) interpretation is simplified due to single allele per locus profile; g) in deficient paternities and h) multigenerational male lineage studies can be performed (Roewer, 2009).

Due to the duplicated, palindromic regions of the Y chromosome, some Y-STR loci occur more than once and when amplified with a locus specific set of primers produce more than one PCR product. This fact can lead to confusion in terms of counting the number of loci present in a haplotype. A single set of primers can produce two amplicons, which may be thought of as 'two loci' for a Y chromosome haplotype. For example the Y-STR locus DYS385 is present in two regions along the long arm of the Y chromosome. These duplicated regions are located about 40 kbp apart and can generate

two different alleles when amplified with a single set of primers. The two alleles are typically labeled ‘a’ and ‘b’ with ‘a’ designation going to the smaller sized allele. It is also possible to have both ‘a’ and ‘b’ alleles be the same size in which case only a single peak would appear in an electropherogram. Due to the presence of two alleles, this duplicated locus is usually referred to as DYS 385a/b. Two PCR products can also be generated at the locus DYS389I using a single set of primers resulting in DYS389II which is a subset of DYS389I (Butler, 2005). In some cases duplications or even triplications of Y-STR locus have been reported, particularly for DYS 19 (Butler, 2005). It was important to keep this fact in mind so that two peaks at the DYS 19 locus are not automatically interpreted as coming from a mixture of two males.

Both of these issues, mutations and duplications of loci, impact analysis and therefore confusing mixture interpretation, suggest that analysis of additional Y-STR loci can be helpful in these circumstances (Butler, 2005).

As of May 2014, the Y Chromosome Haplotype Reference Database (YHRD) contains more than 126,000 haplotypes with information from various markers (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635, and YGATAH4) as well as Y-SNPs from 710 populations sampled worldwide and can accessed on the website http://www.yhrd.org. The website database has become an important tool for use in

comparing Y-STRs from different populations across the world. The website standardizes Y-STR nomenclature by ensuring the successful completion of a quality assurance exercise. All Y-STR population data accepted are done with the understanding that a population is defined as a group of more than 50 individuals living in the same area. When using Y-STRs, it is advisable to understand the data obtained with a look towards the origin of the population (Willuweit and Roewer, 2007). The Y-STRs are thus becoming routine and necessitating the development of international networks and databases for crime investigation. Their usefulness was attributable to their high discrimination capacity and haplotype diversities. However, not until recently have many undertaken to bridge the wide gap on populations data between Africa and the rest of the world (Alves, *et al.*, 2003; Arroyo-Pardo, *et al.*, 2004; Melo, *et al.*, 2010; Rosa, *et al.*, 2007; Omran, *et al.*, 2008; de-Filippo, *et al.*, 2011; Coelho, *et al.*, 2009; Berniel-Lee, *et al.*, 2009; Brandt-Casadevall, *et al.*, 2003; Gomes, *et al.*, 2010 and Barbieri, *et al.*, 2012).

The purpose of the thesis was to investigate in more detail, the combined Y-chromosomal variation of bi-allelic and microsatellite markers in Nigeria to gain insights into pre-historic population movements in this part of West Africa. It was also to analyze the demographics of the Nigerian populations Y-chromosome differentiation and to evaluate the correlation of the regional differentiation.

The thesis would explore the genetic differentiations among five Nigerian populations as revealed by Y-Chromosome analysis by asking the following questions: was there any structure in the genetic variation on the non-recombining region of the Y-Chromosome (NRY) of the major Nigerian populations? Based on the answer to the above question, what factors drove this? Are there correlations between genetic distances and geographical distances in the Nigerian populations? Was inter-haplogroup STR profile sharing a common or rare occurrence in this Nigerian dataset? It would also discuss the forensic significance of this Nigerian data set and its potential enhancement of the Africa meta-population on the Y-chromosome haplotype reference database (YHRD). It would then explore how genetic analysis seems to validate the demographic history of some Nigerian populations by providing insights on how genetic evidences corroborate the historical and archeological pre-existence of the Nigerian populations in their current geographical locations and the nature of the associations among the STR haplotypes in the Nigerian lineages.

The chapters (two to five) that follow would then present the sequential order of the experiments and emphasize the principles surrounding the use of the different analytical tools employed in the thesis. This would be followed by a chapter (three) summarizing the key results of these experiments and then a chapter (four) discussing their overall importance to the set aims and objectives of the thesis. Finally, this would be followed by

a chapter (five) highlighting the key contributions of the thesis to knowledge and some recommendations for future investigations arising from the limited scope of the thesis.

CHAPTER TWO: MATERIALS AND METHODS

2.1 Nigeria populations sampling and comparative data from the literature

Buccal swabs from a total of 463 adult males, unrelated at the grandfather level, were collected between December 2008 and February 2009 and also between December 2009 and January 2010 in Lagos, southwest Nigeria with informed consent. Their self-described ancestries represent five (5) distinct ethnic groups (Hausa, Igbo, Yoruba, Bini and Ijaw) and four (4) geographical regions (North, South East, South West and South South) with twenty-two geographical origins (States of the Nigerian Federation) and two linguistic groups (Afro-asiatic and Niger-Congo). Some samples were merged because they are very few (i.e. less than five) and mostly from the same geographical region. These include the Hausa populations from Niger (7) with Benue (2), Plateau (2), Kogi (1), Bauchi (1) and Adamawa (1); The Hausa populations from Kebbi (5) with Zamfara (2); The Hausa populations from Kaduna (3) with Katsina (6); The Yoruba populations between Ondo (21) with Edo (3); The Igbo populations from Delta (7) with Rivers (1) and finally the Ijaw populations from Bayelsa (7) with Delta (4), Ondo (3) and Rivers (1). The summary of the sampling are on Table 2.1 and Figure 2.1 below. The swabs were immediately stored on Ice and transported to a -20^{0}C freezer for long term storage.

The comparative data from the literature were such that have the full complements of the Y Chromosome Short Tandem Repeats (STR) loci of the Scientific Working Group on DNA methods (SWGDAM) or more. Some of the sample sizes were very small (See

Namibian SAN and South African Bantu populations in Table 2.2) but it was discovered that these were still able to be substantially differentiated from other larger populations. Throughout this report, Bonferroni correction was implemented by dividing the statistical significant P- value 0.05 by the number of pairwise comparisons. Also, four of the five major linguistic groups found in Africa were represented namely, Khoisan, Afro-Asiatic, Niger Congo and Nilo-Saharan languages. In cases were the language of the populations were not defined in the literature, the linguistic characterization was left vacant. A non-African population (Arabs from Saudi Arabia) was included to confirm its proximity to the North African Populations and utilized to resolve some questions. Overall, Five Geographical regions of Africa were represented namely: North Africa, West Africa, Central Africa, East Africa and Southern Africa by more than 2200 male subjects from 21 Countries.

Table 2.1: The Nigerian populations studied

S/No	Region	Geographical origin	Abbreviation	Geographical Latitude , longitude	Linguistic group	Ethnicity	Total sampled
No1	North	Niger	NG	10.2155388, 5.3939551	Afro-asiatic	Hausa	14
No2	North	Jigawa	JG	12.4460001, 9.7232673	Afro-asiatic	Hausa	6
No3	North	Kaduna	KD	10.1589593, 8.133855	Afro-asiatic	Hausa	9
No4	North	Kano	KN	11.7574188, 7.6114217	Afro-asiatic	Hausa	10
No5	North	Kebbi	KB	11.6781241, 4.0695454	Afro-asiatic	Hausa	7
No6	North	Sokoto	SK	13.1177202, 5.3939551	Afro-asiatic	Hausa	32
No7	South east	Abia	AB	5.4308, 7.5247	Niger-congo	Igbo	19
No8	South east	Anambra	AN	6.2757656, 7.0068393	Niger-congo	Igbo	23
No9	South south	Delta	DT	5.5324624, 5.8987139	Niger-congo	Igbo	8
No10	South east	Ebonyi	EB	6.177973, 7.9592863	Niger-congo	Igbo	5
No11	South east	Enugu	EN	6.6095187, 7.351658	Niger-congo	Igbo	12
No12	South east	Imo	IM	5.5214533, 6.9209135	Niger-congo	Igbo	52
No13	North	Kogi	KG	7.561891, 6.5783387	Niger-congo	Yoruba	7
No14	North	Kwara	KW	8.9847995, 4.5624426	Niger-congo	Yoruba	22
No15	South west	Ekiti	EK	7.6655813, 5.3102505	Niger-congo	Yoruba	24
No16	South west	Lagos	LA	6.5232765, 3.5407909	Niger-congo	Yoruba	48
No17	South west	Ogun	OG	6.9098333, 3.2583626	Niger-congo	Yoruba	38
No18	South west	Ondo	ON	6.8959293, 4.8935627	Niger-congo	Yoruba	24
No19	South west	Osun	OS	7.5875843, 4.5624426	Niger-congo	Yoruba	35
No20	South west	Oyo	OY	8.119567, 3.4195527	Niger-congo	Yoruba	40
No21	South south	Edo	ED	6.5438101, 5.8987139	Niger-congo	Bini	13
No22	South south	Bayelsa	BY	4.86777767, 5.8987139	Niger-congo	Ijaw	15

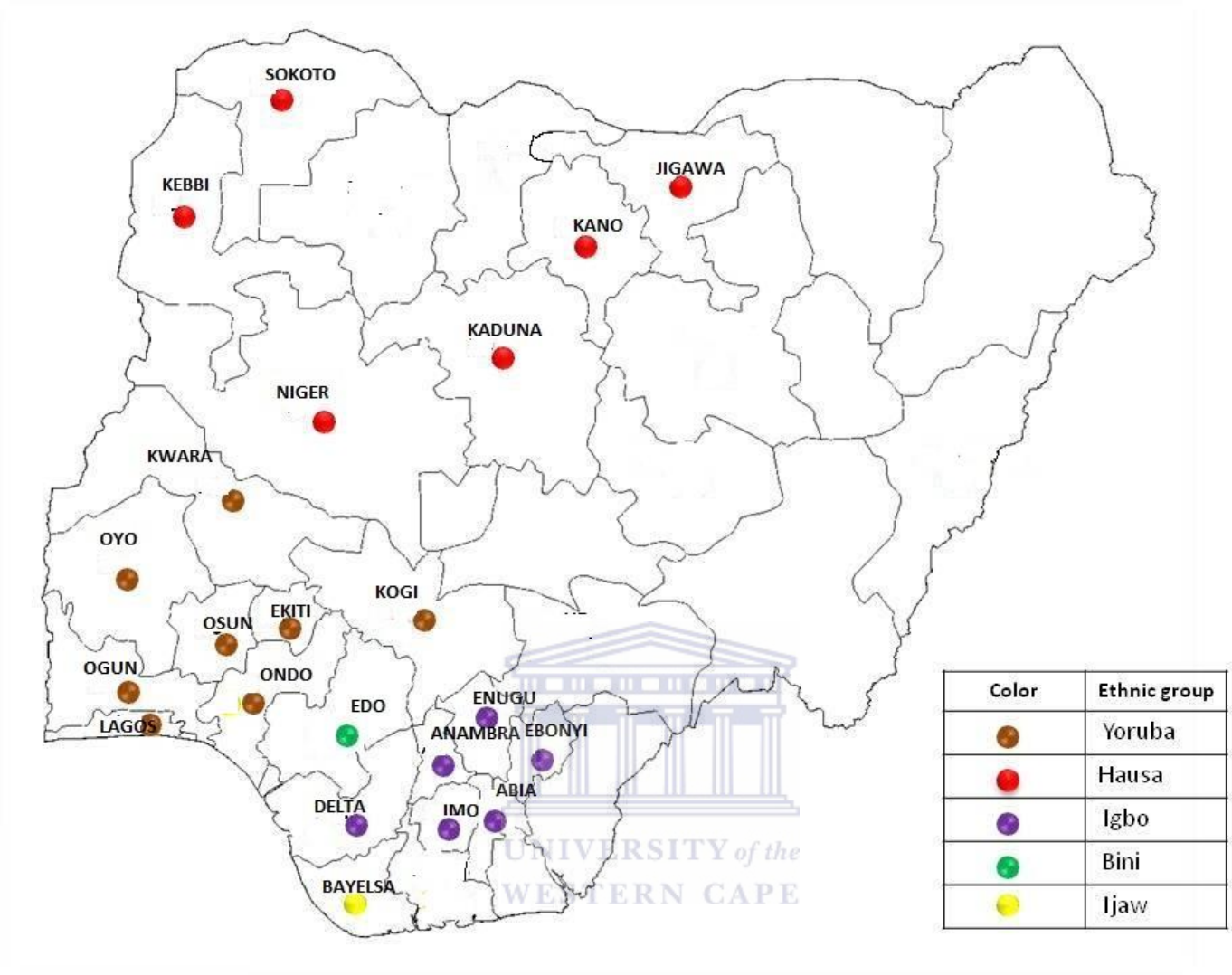


Figure 2.1: Map showing the 22 geographical origins of sampled Nigerian populations

Table 2.2: Data from published literature used for comparative analysis

Serial No.	Country	Population(Size)	Population Code	Linguistic Affiliation	Continental region	References
1	Nigeria	Hausa (78)	NIG-H	Afro-Asiatic	West Africa	This Thesis
		Igbo (119)	NIG-IG	Niger-Congo	West Africa	This Thesis
		Yoruba (214)	NIG-Y	Niger-Congo	West Africa	This Thesis
		Bini(13)	NIG-B	Niger-Congo	West Africa	This Thesis
		Ijaw(15)	NIG-IJ	Niger-Congo	West Africa	This Thesis
2	Burkina Faso	Kassena (33)	BFKSN	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Marka(33)	BFMRK	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Lyela(38)	BFLYL	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Mossi(36)	BFMSS	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Nuna(29)	BFNN	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Pana(24)	BFPN	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Samo-North(34)	BFSN	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Samo-South(41)	BFSS	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
		Samoya(14)	BFSMY	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
3	Guinea Bissau	Mixed(161)	GBIS	Niger-Congo	West Africa	Rosa, *et al.*, 2006
4	Senegal	Mandeka(15)	SENMDK	Niger-Congo	West Africa	De Filippo, *et al.*, 2011
5	Saudi Arabia	Arab(48)	S,ARBI	Arabic	Middle east	Abu-Amero, *et al.*, 2009
6	Morocco	Bedoin(166)	MRCO	Afro-Asiatic	North Africa	Laouina, *et al.*, 2011
7	Tunisia	Bedoin(100)	TUNs	Afro-Asiatic	North Africa	Brandt-Casadevall, *et al.*, 2003
8	Algeria	Mozabite(20)	ALGMZ	Afro-Asiatic	North Africa	De Filippo, *et al.*, 2011
		Afroasiatic(99)	ALGAA	Afro-Asiatic	North Africa	
9	Equatorial Guinea	Mixed(194)	EGUI	Niger-Congo	Central Africa	Arroyo-Pardo et al., 2004
10	Cameroon	Bakola(25)	CAMBK	Niger-Congo	Central Africa	Berniel-Lee, *et al.*, 2009
		Ngumba(24)	CAMNGMB	Niger-Congo	Central Africa	Berniel-Lee, *et al.*, 2009

11	Central African Republic(CAR)	Biaka(24)	CARBIK	Niger-Congo	Central Africa	De Filippo, *et al.*, 2011
12	Democratic Republic of Congo (DRC)	Mbala(10)	DRCMBL	Niger-Congo	Central Africa	De Filippo, *et al.*, 2011
		Mbuti(10)	DRCMBT	Niger-Congo	Central Africa	De Filippo, *et al.*, 2011
		Mbun(9)	DRCMBN	Niger-Congo	Central Africa	De Filippo, *et al.*, 2011
		Pende(10)	DRCPD	Niger-Congo	Central Africa	De Filippo, *et al.*, 2011
		Yansi(15)	DRCYNS	Niger-Congo	Central Africa	De Filippo, *et al.*, 2011
13	Gabon	Akele(50)	GBAKL	Niger-Congo	Central Africa	Berniel-Lee, *et al.*, 2009
		Bekwil(32)	GBBK	Niger-Congo	Central Africa	Berniel-Lee, *et al.*, 2009
		Benga(47)	GBBG	Niger-Congo	Central Africa	Berniel-Lee, *et al.*, 2009
		Duma(45)	GBDM	Niger-Congo	Central Africa	Berniel-Lee, *et al.*, 2009
14	Angola	Bantu(27)	ANGBNT	Niger-Congo	Southern Africa	Coel, *et al.*, 2009
		Kuvale(25)	ANGKVL	Niger-Congo	Southern Africa	Coel, *et al.*, 2009
		Nyaneka Nkhumbi(74)	ANGNYK	Niger-Congo	Southern Africa	Coel, *et al.*, 2009
		Umbundu(93)	ANGUBD	Niger-Congo	Southern Africa	Coel, *et al.*, 2009
15	Namibia	San(4)	NAMSAN	Khoisan	Southern Africa	De Filippo, *et al.*, 2011
16	South Africa	Bantu south(8)	SABAS	Niger-Congo	Southern Africa	De Filippo, *et al.*, 2011
17	Botswana	Kalanga(20)	BOTKLG	Niger-Congo	Southern Africa	De Filippo, *et al.*, 2011
		Tswana(19)	BOTTSW	Niger-Congo	Southern Africa	De Filippo, *et al.*, 2011
18	Zambia	Luyana(60)	ZAMLYN	Niger-Congo	Southern Africa	De Filippo, *et al.*, 2011
		Kwamashi(26)	ZAMKWS	Niger-Congo	Southern Africa	De Filippo, *et al.*, 2011
		Lozi(94)	ZAMLZ	Niger-Congo	Southern Africa	De Filippo, *et al.*, 2011
		Bisa(34)	ZAMBS	Niger-Congo	Southern Africa	De Filippo, *et al.*, 2011
19	Ethiopia	Omo Valley(65)	ETHOV	Nilo-Saharan	East Africa	De Filippo, *et al.*, 2011
20	Kenya	Bantu(10)	KENBNT	Niger-Congo	East Africa	De Filippo, *et al.*, 2011
		Masai(44)	KENMSI	Afro-Asiatic	East Africa	De Filippo, *et al.*, 2011
21	Uganda	Nilo-saharan(118)	UGDNS	Nilo-Saharan	East Africa	De Filippo, *et al.*, 2011

2.2 DNA extraction and quantification

Genomic DNA was extracted using a modification of the salt extraction protocols of Medrano *et al*., 1990. The principle was essentially to recover nuclei material after disrupting cell membrane using detergent and then salting out proteins and finally precipitating out genomic DNA. Specifically, the swabs were placed in separate Eppendorf tubes and digested in a lysis buffer containing 10mM NaCl, 10 mM Tris HCl, pH 8.0, 0.5% SDS and 0.1mg/ml Proteinase K and incubated at 56°C overnight. The digested content was further treated with 0.3volume 4.5M NaCl and shaken vigorously for 15 seconds. The content was spun at 5,000rpm and the supernatant was precipitated with ice cold Isopropanol at -80°C for 30 minutes. The content was spun at 15,000rpm for 30 minutes. The precipitate was washed with 70% ethanol and then dissolved in 30µl double distilled water. The DNA solution obtained was quantified on a Nanodrop ND100 Spectrophotometer (Applied Biosystem) and had an average of 80ng/µl from which a serial dilution of 5ng/µl was prepared for Y chromosome marker typing. The Nanodrop gave readings for both the concentration of DNA and the ratio of its purity given as the ratio of Absorbance reading at 260nm against absorbance reading at 280nm. The acceptable value for further utilization was 1.80, as recommended in Molecular Cloning – A laboratory manual (Sambrook, Fritsch and Maniatis, 1989).

2.3 Primers design

Most of the primers utilized in this report have been reported in the literature (Butler, *et al., 2002*). However, the primer pairs were re-designed for the Y Short Tandem Repeat (STR) DYS390. This was done with *Oligo* v1.4 (Rychlick and Rhoads, 1989) and melting temperature *Tm* was determined according to the nearest neighbour method. The possibility of these primers matching the human genome elsewhere but the Y-chromosome was tested with BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and BLAST (http://blast.ncbi.nlm.nih.gov) with BLASTN 2.2.18 (Altschul, *et al.,* 1997). Self and heterodimers, potential hairpins and secondary structures were checked using *Oligo* Analyzer software (Rychlick and Rhoads, 1989).

2.4 Y-STR Typing

The 11 loci STR recommended by the Scientific Working Group on DNA Analysis Methods SWGDAM, were typed using some modification of Butler and others method (Butler, *et a*l., 2002) which was optimized for Capillary Electrophoresis (CE). The 20-plex PCR protocol of Butler and others (Butler, *et a*l., 2002) was substantially modified as follows:

1.) The DYS390 primers were redesigned with the Forward primer: PET-Labeled TGA CAG TAA AAT GAA AAC ATT GC and the reverse primer: CCC GGG TTT TTA CAC ATT TTA TA.

2.) The primer concentrations were reduced with none exceeding 1.0μM.

3.) The annealing temperature was increased to 58.0°C.

4.) The cycle number was increased to 32.

5.) Only 9 pairs of primers were multiplexed.

6.) 5ng genomic DNA was used for amplification.

7.) Electrophoresis was run on Slab-based electrophoresis.

All remaining components in the PCR amplification cocktail were as described by Butler and others (Butler, *et a*l, 2002).

Table 2.3: The 11-loci SWGDAM primers used for Y-STR typing.

Locus	Dye	Sequence	[Primer]	Length
DYS19	NED	F 5'ACT ACT GAG TTT CTG TTA TAG TGT TTT T 3'	1.0µM	28
		R 5' GTC AAT CTC TGC ACC TGG AAA T 3'	1.0µM	22
DYS389	FAM	F 5' CCA ACT CTC ATC TGT ATT ATC TAT G 3'	0.4µM	25
		R 5' GTT ATC CCT GAG TAG TAG AAG AAT G 3'	0.4µM	25
DYS390	PET	F 5' CAA TGT GTA TAC TCA GAA ACA AGG 3'	0.8µM	24
		R 5' CAC ATA TAT TTT ACA CAT TTT TGG G 3'	0.8µM	25
DYS391	FAM	F 5' TTC AAT CAT ACA CCC ATA TCT GTC 3'	0.2µM	24
		R 5' GAT AGA GGG ATA GGT AGG CAG GC 3'	0.2µM	23
DYS392	NED	F 5' TAG AGG CAG TCA TCG CAG TG 3'	0.6µM	20
		R 5' GAC CTA CCA ATC CCA TTC CTT 3'	0.6µM	21
DYS393	VIC	F 5' GTG GTC TTC TAC TTG TGT CAA TAC 3'	0.25µM	24
		R 5' GAA CTC AAG TCC AAA AAA TGA GG 3'	0.25µM	23
DYS438	FAM	F 5' CCA AAA TTA GTG GGG AAT AGT TG 3'	0.25µM	23
		R 5' GAT CAC CCA GGG TCT GGA GTT 3'	0.25µM	21
DYS439	FAM	F 5' TCG AGT TGT TAT GGT TTT AGG TCT 3'	0.18µM	24
		R 5' GTG GCT TGG AAT TCT TTT ACC C 3'	0.18µM	22
DYS385	VIC	F 5' AGC ATG GGT GAC AGA GCT A 3'	0.2µM	19
		R 5' GCC AAT TAC ATA GTC CTC CTT TC 3'	0.2µM	23

2.5 Multiplex PCR cycling conditions for Y- STR amplification

All the 11 loci STRs were amplified in a single multiplex reaction of 10µl reaction volume. The reaction mixture comprised of 0.5µL SuperTherm Gold brand *Thermophilus aquaticus (Taq)* (5units/µL) (MEDOX (PTY) Ltd), 5ng genomic DNA, 1.5mM Magnesium Chloride buffer, 200µM of each dNTP (Roche), sterile water, 1.6mg/ml BSA, and 0.5% glycerol. Amplification was performed in a GeneAmp 2720 Thermal cycler (Applied Biosystems), under the following cyclic conditions: 1 cycle of enzyme activation at 95°C for 10 minutes; 32 cycles of denaturation at 94°C for 30 seconds, primer annealing at 58°C for 1 minute and elongation at 72°C for 1minute, and final extension of 1cycle at 68°C for 75 minute then a hold at 15°C until the samples were removed from the thermal cycler and stored in the fridge at 4°C.

2.6 Detection of PCR products and scoring of alleles

Analysis of amplified products was performed by mixing 1µL of amplified product with 1µL of loading mix which consisted of de–ionized formamide; Dextran Blue dye (*Applied Biosystems*) and LIZ500 size standard (*Applied Biosystems*) in a ratio of 5:2:1 (µl). The mixture was then denatured using GeneAmp 2720 Thermal cycler PCR System set at 95 °C for 5 minutes. Immediately after denaturation, the samples were placed sharply on ice and loaded on 4% Poly Acrylamide Gel slab and ran for two and half hours on the ABI 377 Sequencer (Applied Biosystems).

The protocols for the preparation of this gel and associated solutions are included in the appendix. The outputs of the run were analyzed with the associated software Gel processor (Applied Biosystems), Genescan 3.0.0 (Applied Biosystems) and Genotyper 3.7 (Applied Biosystems). The Genotyper macros for allele calling were validated with the National Institute for Standards and Technology reference sample SRM 2395. The STR profiles derived from the above protocols for Nigerian populations were contrasted with published results for Tunisia (Brandt-Casadevall, *et al., 2003*), Angola (Coelho*, et al., 2009*), Uganda (Gomes *et al., 2010*), Cameroon (Berniel-Lee, *et al., 2009*) and Burkina Faso (De-Filippo, *et al., 2011*). A more comprehensive analysis with all the comparative dataset in Table 2.2 is shown in the R_{st} distance and P value matrix in Supplementary Table 3 of this report.

2.7 Y- SNPs Typing

The Y-SNP typed were M60 for Haplogroup B, Y chromosome Alu polymorphism YAP for Haplogroup DE, SRY_{4064} for Haplogroup E, M2 for E1b1a, M9 for haplogroup KR, M207 for Haplogroup R, U175 for Haplogroup E1b1a8, U186 for Haplogroup E1b1a7 and M215 for E1b1b. The first 6 were done with Restriction Fragment Length Polymorphism (RFLP) analysis while the last

three were carried out with High Resolution Melting (HRM) analysis. The phylogenetic relationship among the SNPs tested was shown in Figure 2.2 below.

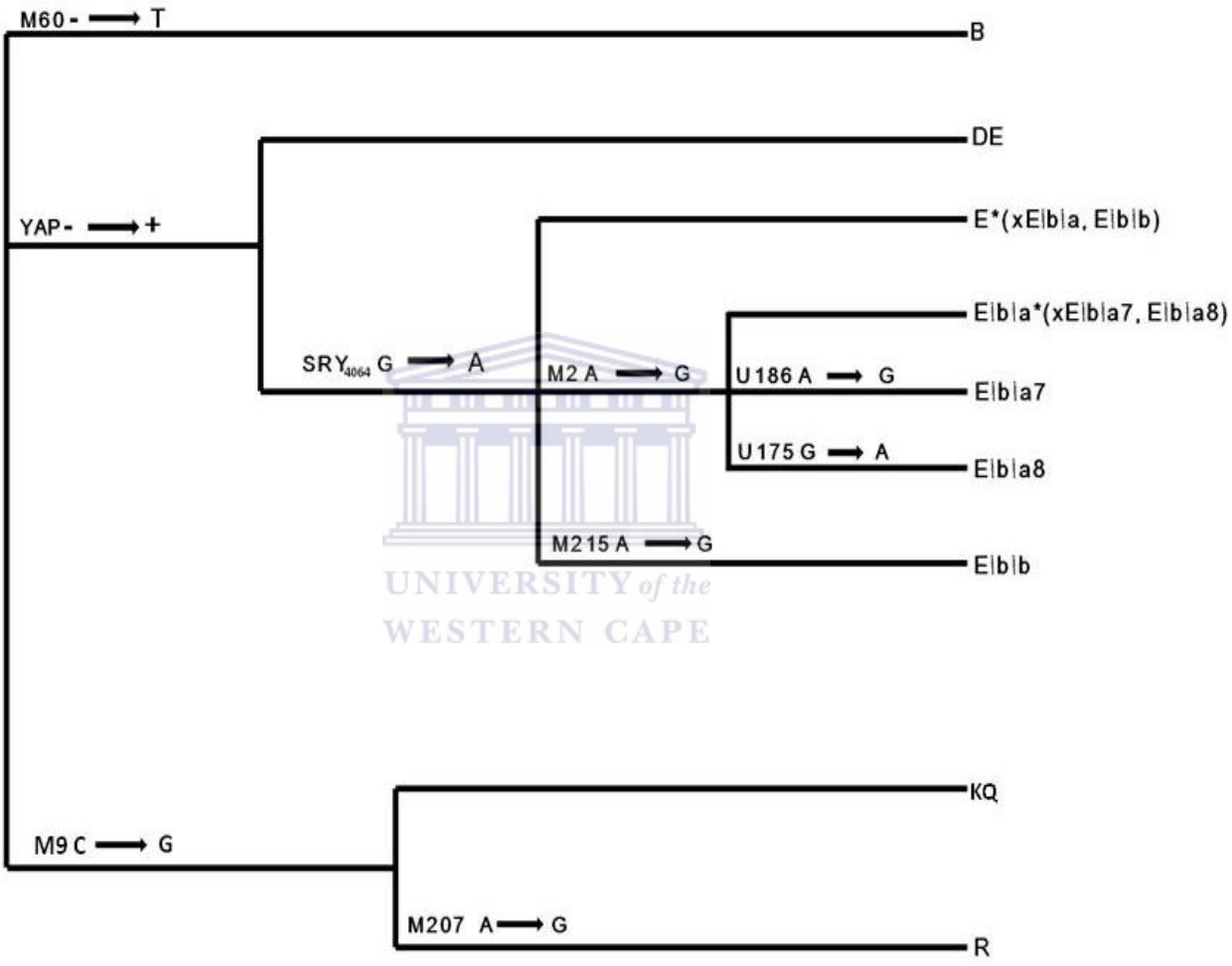


Figure 2.2: The Tree of Y-SNPs evaluated in this thesis.

2.8 Restriction Fragment Length Polymorphisms (RFLPs)

In principle, RFLP was considered to be the simplest and earliest method to detect SNPs. SNP-RFLP made use of the many different restriction endonucleases and their high affinity to unique and specific restriction sites. By performing a digestion on a genomic sample and determining fragment lengths through a gel assay it was possible to ascertain whether or not the enzymes cut the expected restriction sites. A failure to cut the genomic sample results in an identifiably larger than expected fragment implying that there was a mutation at the point of the restriction site which rendered it protected from nuclease activity. Unfortunately, the combined factors of the high complexity of most eukaryotic genomes, the requirement for specific endonucleases, the fact that the exact mutation cannot be necessarily resolved in a single experiment and the slow nature of gel assays make RFLP a poor choice for high throughput analysis. However, the Y SNPs evaluated in this thesis had Restriction Enzymes that specifically or indirectly resolved them as shown in the **Table 2.4** below as modified from Thomas and others method (Thomas, *et al., 1999*). Three primer pairs were retained from the methodology namely M2, SRY_{4064} and YAP. The remaining SNP primers were designed as described in the primer design section above (**Table 2.3**).

Specifically, two experiments were carried out namely a multiplex PCR with fluorescently labeled primers and secondly, both a tetraplex and a singleplex restriction enzyme digest.

The 10.0µl multiplex PCR reactions consisted of the stated concentrations of both forward and reverse SNP primers (New England Biolabs), in **Table 2.4** in addition to 1.5mM Magnesium Chloride, 200µl dNTPs, 1.6µg of acetylated Bovine Serum Albumin (BSA), 1Unit *Taq* polymerase (MEDOX (PTY) Ltd) and 5ng of genomic DNA and finally with distilled water to make the volume. This was subjected to the following cyclic conditions: 1 cycle at 95^0C for 10 minutes, followed by 33 cycles of 94^0C for 1 minute, 60^0C for 1 minute and 72^0C for 1 minute then a third and final cycle at 68^0C for 75 minutes and a hold at 15^0C for ∞.

2.9 Restriction Enzyme Digest of PCR products

Digestions were performed in a 10.0µl PCR tube in a final volume of 8 µl. Each reaction contained 2 µl of PCR product, NEB buffer 4 (New England Biolabs) to 1x concentration, 0.16 mg/ml acetylated BSA, 0.3 U *Bsr*BI, 0.5 U *Bsm I*, 0.3 U *Nla*III and 0.5 U *Mbo I*. This was incubated at 37°C for 10 hours and later re-incubated at 65^0C for another 10 hours as *Bsm I* is activated at the higher temperature which also inactivated the other Restriction Enzymes (RE) in the tube. The digestion was however terminated by incubating again at 85^0C for 20 minutes

to inactivate the *Bsm I.* Predicted sizes and associated polymorphic status for each dye-labeled PCR product was as given in Table 2.4. The samples that were 'derived' for SNP M9, were subjected to a single RE digest with 0.3 U *Dra I* to determine if they belong to Haplogroup R in a reaction only differing in the number of RE included

2.10 SNP resolution by High Resolution Melting

In principle (Liew, *et al.*, 2004; Zuccarelli, *et al.*, 2011), the same thermodynamic properties that allowed for the ingeniously crafted gel techniques to work apply here, and in real-time. A fluorimeter monitors the post-PCR denaturation of the entire double stranded DNA amplicon. Primers were designed around the specific site of the SNP of interest to produce amplicon of not more than 150 base pairs. The amplicon integrates with a double-strand specific dye, included in the PCR mix, in the process making the entire amplicon a probe. The melting temperature (Tm) of the entire amplicon was determined and the SNPs had sufficiently different melting temperature (T_m) to genotype.

Specifically, Type-it ®HRM® kit was purchased from QIAGEN and utilized as recommended by the manufacturer. The HRM cyclic conditions were implemented on the Rotor Gene Q Real Time PCR machine with in-built HRM capabilities. In order to obtain broader divergence between melting profiles and increase peak

resolution, Amplicon's length was less than 125 base pairs in all cases. The HRM primers are listed in Table 2.5. All the three SNPs evaluated are Class 1 SNP, that is, Guanine/Adenine base change, guaranteeing sufficiently different T_m http://www.appliedbiosystems.com/etc/medialib/appliedbio-media-library/documents/application-and-technology/real-timepcr/hrm.par.73223.file.pdf. While E1b1a7 and E1b1a8 were optimized for a duplex, E1b1b was done as a single PCR.

Samples used as controls for Y-SNPs were confirmed by sequencing analysis of Y-chromosome loci rs169808588 (U175), rs16980370 (U186) and rs2032654 (M215). Sequences were manually aligned and edited using BioEdit vs. 7.0.5.2. Three hundred and ninety-six (396) samples earlier determined to be positive for SNP M2 mutation by RFLP were further resolved with HRM.

Table 2.4: Restriction Fragment Length Polymorphism (RFLP) primers

SNP	Hg	F or R	Primer sequence 5'-3'	µM	DRE	NDRE	PDT	Ancestral	Derived
M2	E1b1a	F-VIC	ATG GGA GAA GAA CGG AAG GA	0.05	*Nla III*		142	105	142
		R	TGG AAA ATA CAG CTC CCC CTT	0.05					
SRY4064	E	F	CCA CGC CCA GCT AAT TTT TTG T	0.15					
		R-PET	CAT TTC AGT AAA TGC CAC ACA AG	0.15	*Bsr BI*	*MboI*	180	90	104
M9	KR	F-PET	GTG CGG CGT CTT TGA TCT C	0.05	*Bsm I*	*Nla III*	324	241	268
		R	GAA GTA AGC GCT ACC TTA CTT AC	0.05					
YAP	DE	F-FAM	CAG GGC CAA CTC CAA CCA AG	0.05					
		R	GGA CTA GCA ATA GCA GGG GAA G	0.05			88/413	88	413
M60	B	F	CCA ACA CTG AGC CCT GAT G	0.05					
		R-FAM	GAG AAG GTG GGT GGT CAA GA	0.05	*Mbo I*	*Nla III*	216	100	108
M207	R	F-NED	CAC ATC TCT ATT TAG TCT AAA TTC	0.3	*Dra I*		233	141	233
		R	GAA GGA AAA GTG GAG TCT GAC	0.3					

F or R is Forward or Reverse primer

M.wt is Molecular weight

DRE is Discriminating Restriction enzyme

NDRE is Non-Discriminating restriction Enzyme

(A), (G), (C) and (-) are nucleotides at the site cut by Restriction Enzyme. The dash "-" indicates an indel.

VIC, PET, FAM and NED are fluorescent dyes attached at the 5' ends of primers.

Table 2.5: Primers for the High Resolution Melting (HRM) Experiments

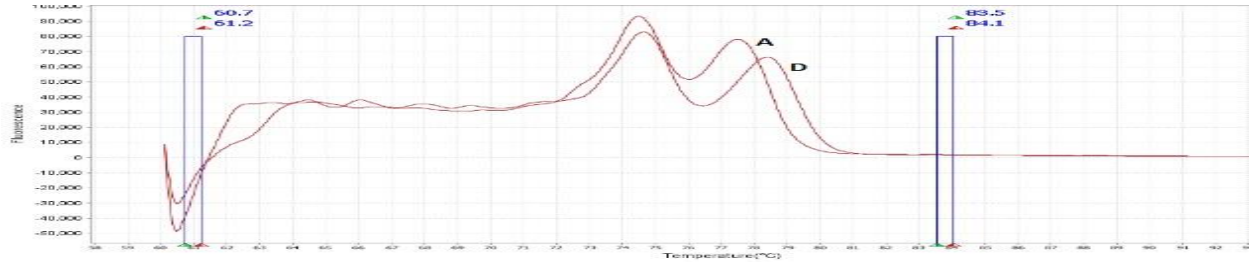
Primers for HRM analysis of Y SNPs

SNP	UEP	Forward Primer 5'-3'	Reverse Primers 5'-3'	PCR Product	Mutation
E1b1b	M215	AAA GAA ATA TTC TCA AAC TGT TGG	TCC AGC ACA GAA GCA TCA G	108	A>G
E1b1a8	U175	GCT TAT ACT GGT CAC ACT AAG GC	TCT AAT GAC CAG GAG AAG TCA AG	106	G>A
E1b1a7	U186	CCT TCT CGT AAG GGG CTG	CTG GAT AAG AGT CCT TGG AG	73	A>G

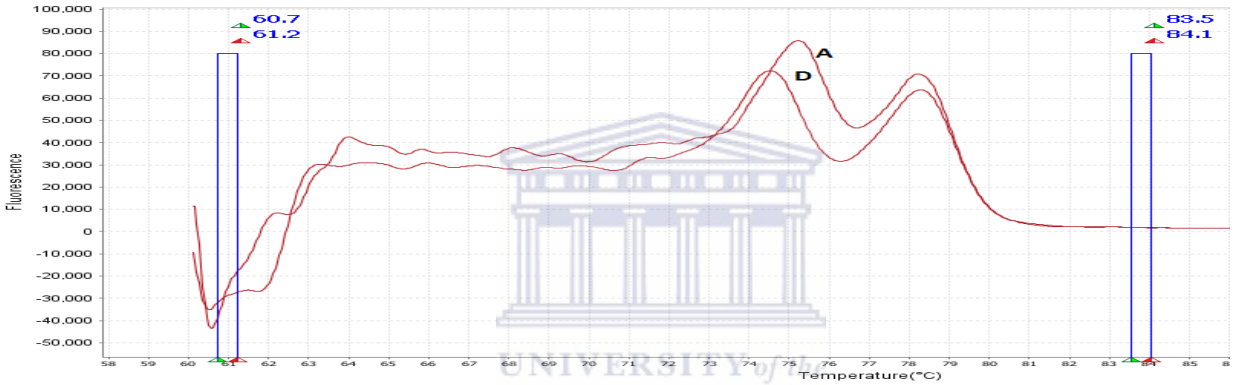
Primers for Sequencing of Y SNPs

SNP	UEP	Forward Primer 5'-3'	Reverse Primer 5'-3'	PCR Product	Mutation
E1b1b	M215	TCC CAT GAA ATA TAC ACA GAA AC	CGT TCA TTA GGA ATC ACT GTC T	401	A>G
E1b1a7	U186	ACA GAT GTT GCT GGA TGA AAA GTG	ATC CCT GGG CTT GTG GTT ATA TC	430	G>A
E1b1ba8	U175	CAC ATT CCA TAA CCT TTA ACA CAC	AGG AAT CAG TGG TTT GTT TGA G	411	A>G

(a)



(b)



(c)

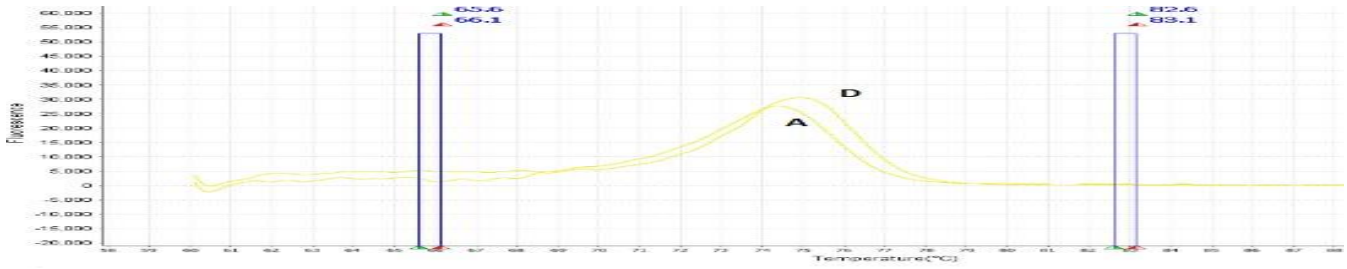


Figure 2.3: Derivative HRM plots of (a) E1b1a7, (b) E1b1a8 and (c) E1b1b YSNP haplogroups. "A" and "D" represented the Ancestral and Derived versions of the haplogroups

2.11 Statistical analysis

Allelic frequencies (AF), Gene Diversity (GD), Haplotype frequencies (including both Unique and shared haplotypes), Genetic diversity measures (i) number of haplotypes and haplogroups with (ii) haplotypes and haplogroup diversity and (iii) Mean Pairwise Differences were calculated using the software package Arlequin 3.11 (Excoffier, *et al.*, *2005*). Apart from the frequencies and relative frequencies which were calculated by mere counting and dividing individual count against overall in populations, the diversity parameters were calculated using the equation $D= (n/n-1)(1-\Sigma P_i^2)$ where Diversity is D, n is sample size and P_i is the relative frequency of the *ith* allele of the haplotype or haplogroup respectively. Mann-Whitney U test was used to compare the differences in the mean values of the three major ethnic populations (Hausa, Igbo and Yoruba) to ascertain their closeness or distinctions.

Inter population distances and their associated P values were calculated by generating (a) R_{st} (Slatkin, 1995) distance matrices for STR haplotypes, and (b) F_{st} distance matrices for SNP haplogroups calculated in Arlequin v3.11 (Excoffier, *et al.*, 2005). R_{st} was based on allele sizes and compared to the corresponding F-statistics estimated following Weir and Cockerham (1984), (Michalakis and Excoffier, 1996). Bonferroni correction was given as $P=\alpha/n$ where α is 0.05 and n is the number of non-independent tests.

At the level of the 51 Africa populations, phylogenetic tree was constructed using genetic distance matrices based on Goldstein and others (Goldstein, *et al*., 1995a) δμ2 pairwise distance measure and 1,000 bootstrap datasets were created for internal node confidence value using the Neighbor-Joining (NJ) algorithm. This clustering approach was used because it does not assume an evolutionary clock and produces more accurate results when closely related populations, such as human groups, are analyzed (Saitou and Nei 1987). The basic principle of the NJ method was to minimize the total evolutionary distance in the tree (Hartl and Clark 1997). The NJ phylogeny was constructed using PAST v.1.54 (Hammer *et al.*, 2001b).

The Multidimensional scaling (MDS) distance coordinates were plotted using PAST v.1.54 (Hammer *et al.*, 2001b) based on (a) the microsatellite pairwise $\mathbf{R_{st}}$ and (b) pairwise $\mathbf{F_{st}}$ of SNP haplogroups to access the spatial differentiation of the populations. These two plots were reported for (i) the different Nigerian populations across the 22 geographical locations, (ii) five ethnic pooled populations and finally (iii) with some comparative data from the literature for some representative African Countries based comparison. This multivariate method defines for each population coordinates so that the distances among them are as close as possible to the original genetic distances. The stress was a measure of "goodness of fit" that indicates how similar was the distance matrix based on the new coordinates to the original genetic distance matrix and it was actually smaller for better fits.

A correspondence analysis was performed with the frequencies of the Y-SNP haplogroups by means of the PAST software. This multivariate method plots in the same graphical representation both columns and rows of a contingence table (in this case, populations and haplogroups based on Y-SNPs). By plotting both populations and haplogroups in the same graphical representation opens the possibility to assess, which haplogroups are contributing to the distribution and differentiation of the populations in the plot.

Population genetic structure was estimated using hierarchical Analysis of Molecular Variance (AMOVA) (Excoffier, *et. al.*, 1992) based on a particular mutation model to generate a single Fixation Index statistic, F_{st}, when a simple structure of populations within a single group was defined. Also, three other Fixation Indices, F_{st} (the within-population Fixation Index), F_{sc} (the among-populations within-group Fixation Index) and F_{ct} (the among-group Fixation Index), could be defined for a more complex structure of populations within multiple groups. Significances of Fixation Indices are assessed by randomly permuting individuals (given that only haploid systems are considered) among populations or groups of populations, depending on the Fixation Index being tested and after every round of permutations, of which 10,000 were performed, Fixation Indices are recalculated to create a null distribution. Population pairwise genetic distances were estimated from Analysis of Molecular Variance. Hierarchical analysis was assessed based

on the (a) five ethnic, (b) four geographical regions and (c) 2 linguistic classifications shown in **Table 2.1**, to determine the major driving force of the population structure.

The degree of genetic differentiation between populations was quantified by means of the analysis of molecular variance (AMOVA) using the Arlequin 3.11 package (Excoffier, *et al. 2005*). This method allows us to define the percentage of the genetic variation that is explained by (1) among groups of population defined a priori, (2) between the populations of the same group, and (3) within the populations. Making Bonferroni correction, significant p was < 0.005, <0.0083 and 0.05 for 5, 4 and 2 groupings respectively.

The geographical location of putative genetic barriers was analysed by means of the barrier version 2.2 program (Manni, *et al.*, 2004). This program computes Monmonier's algorithm to detect a spatial abrupt rate of change in terms of the genetic differentiation between geographically neighbouring populations. This was evaluated separately for both haplotypes and haplogroups data of the sampled Nigerian populations.

Mantel test were performed between R_{st} or F_{st} genetic distances and geographic distances matrices using the software SPAGeDi v. 1.3 (Hardy and Vekemans, 2002). Significance was assessed by permuting the rows and columns of the matrices 1,000 times and determined to be <0.0002. Geographic distances were Great Circle distances estimated from latitude and longitude data. The Mantel test was a method for testing the

significance of the correlation between two or more matrices, and was used in this study to investigate the relationship between genetic and geographic distances of the Nigerian populations Y-STR haplotype and Y-SNP haplogroup data sets.

A general limitation of all relevant analyses dealing with patterns of genetic marker frequencies should be noted here: frequencies of different genetic markers are not independent from each other in the way that a high frequency of one marker in a population consequently leads to a lower frequency of one (or more) different marker(s) in that same population.

Genetic relationships between haplotypes inside of specific haplogroups were analyzed using Network 4.5.0.0 software. Networks of STR haplotypes on the background of Y SNP haplogroup lineages B, E(xE1b1a), E1b1a(xE1b1a7, xE1b1a8), E1b1a7, E1b1a8 and R with the respective mutations M60, SRY_{4064}, M2, U186, U175 and M207 with the Y STR loci DYS19, DYS389I, DYS389c, DYS390, DYS391, DYS392, DYS393, DYS385, DYS438 and DYS439 were constructed using the Median Joining (MJ) algorithm (Bandelt, *et al.*, 1999) of Network v4.5.0.0 (Fluxus-engineering, 2008). DYS389c was obtained by subtracting the value of DYS389I from DYS389II. DYS385ab were added together to obtain a new DYS385. The advantage of this type of cluster analysis was it allows for cycles or reticulations within evolutionary pathways in order to accommodate the elevated mutation rates and corresponding homoplasy of particular genetic systems such as Y-STR loci (Kayser, *et al*., 2000). The MJ networks presented in

this study were subjected to maximum parsimony post-analysis using the Steiner maximum parsimony (MP) algorithm (Polzin and Daneshmand, 2003) within Network 4.5.0.0. All the STRs used were weighted equally.

The spatial distribution of Y-SNP haplogroups were analyzed by means of spatial autocorrelation analysis (Sokal and Oden 1978) using the PAST program. The spatial autocorrelation analysis computes the level of autocorrelation between pairs of points that are within a certain geographic distance. The plot of the level of autocorrelation in relation to increasing geographic distance classes gives information about the spatial pattern of the data. In the case of a clinal pattern of the data, it is expected that the shape of the autocorrelogram will decrease from positive autocorrelation values for the closest geographical distances to negative values for the longest geographic distance classes (Barbujani, 2000). This was constructed for the three most common haplogroups among the Nigeria populations namely Haplogroup B, Haplogroup E1b1a7 and Haplogroup E1b1a8, to assess the presence of any serial founder effect.

The N*p* parameter (denoting effective migrants per population) incorporates effective population size, migration rate and mutation rate and was calculated by application of the formula N*p*= (1/Fst)-1, according to the island model of migration for haploid systems (Cavalli-Sforza and Bodmer, 1971; Destro-Bisol, *et al.*, 2004; Hassan, *et al.*, 2008).

Where N is the effective population size and *p* is the sum of migration (m) and mutation rate (μ); more precisely *p*=m+μ-mμ (Cavalli-Sforza and Bodmer, 1971). The contribution

of mutation rate to the N*p* parameter in the genetic systems used in this thesis (the SWGDAM Y-STR loci) may be considered constant (Destro-Bisol, *et al.*, 2004) and that the fluctuations of N*p* values can be effectively assumed to be the results of differences in migration rate and effective population size among populations (Wjisman, 1984; Seielstad, Minch and Cavalli-Sforza, 1998).

Average Squared Distance (ASD) (Goldstein, *et al.*, 1995b; Goldstein and Pollock, 1997), as employed in Populations v.1.2.30 (Langella, 2002), were determined among specific lineages for the estimation of time to the most recent common ancestors (TMRCA). This was estimated as T =ASD/2μ (Batini, *et al.*, 2011) for inter-haplogroup and T=ASD/μ for intra-haplogroups (Batini, *et al.*, 2011; de Filippo, *et al.*, 2011). Where μ is the STR mutation rates averaged among the 11 STRs and T is "Generations ago" since the common ancestor. It was converted to "years ago", by multiplying by the generation time. For all inter-haplogroup estimates, a generation time of 31 years was used (Helgason, *et al.*, 2003; Batini, *et al.*, 2011). The mutation rate used for the dating estimates was calculated to be 0.0024 (0.0016-0.0031) mutations per locus per generation based on the 11 set of STR markers used from the mutation rates reported in the Y-chromosome Haplotype Reference Database (http://www.yhrd.org). This value was not substantially different from other estimates based on pedigree data, and was approximately 3 times faster than the more general and non-locus specific 'evolutionary' rate (0.00069 ± 0.00057) mutations per locus per generation (Ravid-Amir and Rosset,

2010; Zhivotovsky, *et al.*, 2004). The TMRCA of a clade was estimated by calculating the ASD between all chromosomes in a lineage, and the founder haplotype, which was reconstructed by combining the modal alleles at every single STR locus in the haplotypes in the lineage (Thomas, *et al.*, 1998, Batini, *et al.*, 2011). This was computed from the Y TMRCA calculator using pedigree mutation rates that excluded DYS385 value from the website *http://ehelix.pythonanywhere.com/init/default/instructions*.

3.0 CHAPTER THREE: RESULTS

3.1 Summary of Y-SNP and Y-STR diversities in Nigerian Populations

Table 3.1**:** Y-Chromosome Y-STR and Y-SNP diversity in the Nigerian Populations studied

Population	Geographical location/ region	# of subjects/ Haplotypes	Haplotype diversity	Mean Pairwise Differences (Haplotypes)	#. of Haplogroups (HGP)	HGP Diversities
Hausa	Niger	14(14)	1.000000	7.560440±3.734862	4	0.6709
Hausa	Jigawa	6(6)	1,000000	7.133333±3.844590	4	0.9331
Hausa	Kano	9(9)	1.000000	6.800000±3.473038	4	0.8062
Hausa	Kaduna	10(10)	1.000000	7.194940±3.692061	5	0.8000
Hausa	Kebbi	7(7)	1.000000	7.333333±3.864546	5	0.8565
Hausa	Sokoto	32(30)	0.993952	7.935484±3.779179	5	0.8344
Hausa overall	North	78(75)	0.998009	7.680986±3.615490	7	0.8298
Igbo	Abia	19(19)	0.999999	5.526316±2.768115	5	0.4442
Igbo	Anambra	23(22)	0.996048	4.636363±2.351883	5	0.3864
Igbo	Delta	8(8)	1.000000	5.500000±2.929858	4	1.0000
Igbo	Ebonyi	5(5)	1.000000	4.300000±2.619363	3	0.7000
Igbo	Enugu	12(12)	1.000000	5.609642±2.875806	4	0.4393
Igbo	Imo	52(48)	0.996983	5.291101±2.593475	6	0.3382
Igbo overall	South East	119(103)	0.998150	5.208375±2.536377	6	0.3598
Yoruba	Ekiti	24(24)	1.000000	6.190476±3.311041	3	0.8088
Yoruba	Kogi	7(7)	0.999999	6.056277±2.987899	6	0.5021
Yoruba	Kwara	22(22)	0.999999	5.028986±2.523037	7	0.4347
Yoruba	Lagos	48(43)	0.996299	5.370028±2.632036	3	0.4663
Yoruba	Ogun	38(37)	0.998577	5.914651±2.880960	3	0.5905
Yoruba	Ondo	24(24)	1.000000	5.797721±2.853696	3	0.4716
Yoruba	Osun	35(34)	0.998320	5.492437±2.700743	5	0.5896
Yoruba	Oyo	40(40)	1.000000	5.719231±2.792509	6	0.4713
Yoruba overall	South West	238(214)	0.99883	5.622490±2.706762	8	0.5209
Bini	Edo	13(13)	0.999999	5.692308±2.842048	4	0.5256
Ijaw	Bayelsa	15(14)	0.99048	5.619048±2.842048	3	0.3620

The three major populations (Hausa, Yoruba and Igbo) spread across 20 geographical origins were contrasted to observe the most diverse. The Mann-Whitney U test did not reveal any significant difference when the six Hausa populations' haplotype diversities were compared with those of the six Igbo or the eight Yoruba populations i.e. z= -0.6368, p= 0.5243 and z= -0.8187, p= 0.413 respectively. The six Igbo and eight Yoruba populations haplotype diversities were also not significantly different, z= -0.06542, p= 0.9478. However, at the level of comparison of the Mean STR Pairwise Differences (MPD) within each population, there were significant differences among the populations with the Hausa most diverse, followed by the Yoruba. For comparison between Hausa (range 6.800000- 7.935484, pooled 7.680986) and Igbo (range 4.300000-5.609642, pooled 5.208375) was z= -3.067, p= 0.002165. The Hausa with the Yoruba (range 5.028968- 6.190476, pooled 5.622490) was the same as for the Igbo. However, the Mann-Whitney U test between the Igbo and Yoruba MPD was z= -2.044 and p= 0.04091. The Bini and Ijaw were represented by single population each unlike the Hausa, Igbo and Yoruba, hence were not contrasted.

On the background of Haplogroup diversities also, the Mann-Whitney U comparison of the three major populations of Hausa, Yoruba and Igbo showed that while the six Hausa populations (HGD range 0.6709- 0.9331, pooled 0.8298) could be distinguished from the six Igbo populations (HGD range 0.3382-1.000, pooled 0.3598) z = -2.044 and p = 0.04091, and the eight Yoruba populations (HGD range 0.4347- 0.8088, pooled 0.5209) z

=-2.683and p =0.00729 both the six Igbo and eight Yoruba populations are not distinguishable (z = -1.15, p = 0.2502). This result seems to suggest that the distinction among the major populations was language dependent on the basis that the Hausa populations being Afro-Asiatic language speakers while both Igbo and Yoruba are both non-Bantu Niger-Congo language speakers. The inference from both STR and SNP data set was that there were some distinctions between recent events and pre-historic events affecting the genetic diversities in the Nigerian populations especially among the Niger Congo language speakers. This inference was undetectable with haplotype diversities but fairly discernable with both MPD and haplogroup diversities among the populations.

Table 3.2: The Nigeria ethnic populations' relative haplogroup frequencies across their geographical origins. The * is not a geographical location. The Hausa populations are represented with Red, The Igbo with blue, The Yoruba with yellow, The Bini with Cyan and the Ijaw with Purple colour. The Haplogroup "Others" were not resolved.

Geographical Location	B	E (xE1b1a xE1b1b)	E1b1a (xE1b1a7, xE1b1a8)	E1b1a7	E1b1a8	E1b1b	R	Others
Niger	0.071	0.143	0.000	0.571	0.071	0.000	0.143	0.000
Jigawa	0.000	0.167	0.167	0.333	0.000	0.167	0.167	0.000
Kaduna	0.222	0.333	0.111	0.000	0.000	0.000	0.333	0.000
Kano	0.000	0.000	0.000	0.200	0.200	0.200	0.400	0.000
Kebbi	0.143	0.143	0.143	0.143	0.429	0.000	0.000	0.000
Sokoto	0.094	0.281	0.063	0.219	0.188	0.031	0.125	0.000
All Hausa*	0.090	0.210	0.060	0.260	0.150	0.050	0.180	0.000
Abia	0.053	0.000	0.000	0.737	0.158	0.000	0.000	0.053
Anambra	0.000	0.045	0.000	0.772	0.182	0.000	0.000	0.000
Delta	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
Ebonyi	0.000	0.000	0.200	0.600	0.200	0.000	0.000	0.000
Enugu	0.000	0.083	0.000	0.750	0.167	0.000	0.000	0.000
Imo	0.038	0.019	0.019	0.808	0.115	0.000	0.000	0.000
All Igbo*	0.030	0.030	0.020	0.790	0.130	0.000	0.000	0.008
Kogi	0.000	0.143	0.143	0.429	0.286	0.000	0.000	0.000
Kwara	0.045	0.000	0.045	0.682	0.227	0.000	0.000	0.000
Ekiti	0.042	0.000	0.000	0.750	0.125	0.042	0.042	0.000
Lagos	0.021	0.042	0.042	0.710	0.188	0.000	0.000	0.000
Ogun	0.105	0.026	0.026	0.605	0.211	0.000	0.000	0.026
Ondo	0.042	0.042	0.000	0.708	0.208	0.000	0.000	0.000
Osun	0.086	0.000	0.086	0.600	0.229	0.000	0.000	0.028
Oyo	0.050	0.050	0.025	0.650	0.200	0.000	0.000	0.025
All Yoruba*	0.050	0.004	0.004	0.660	0.200	0.000	0.000	0.013
Edo	0.000	0.077	0.000	0.692	0.154	0.077	0.000	0.000
Bayelsa	0.000	0.000	0.133	0.800	0.067	0.000	0.000	0.000
All Nigeria*	0.050	0.060	0.040	0.630	0.190	0.010	0.030	0.009

The five ethnic populations were resolved into eight haplogroups (Table 3.2) based on Karafet and others nomenclature (Karafet, *et al.,* 2008). The Hausa populations had the highest pooled diversity (0.8298) and the greatest number of haplogroups (7). Among the six Hausa populations, Jigawa Hausa (JG) was the most diverse while Niger Hausa was the least. While Haplogroup E1b1a7 and E1b1a8 were prevalent in the entire Country, they were not noticed in the Kaduna Hausa (KD) populations. Even though Haplogroup E1b1b has been suggested as the signature Afro-Asiatic language haplogroup in the broader African context (De Filippo, *et al.,* 2011), it was only found in 5% of the entire Afro-Asiatic speaking Hausa of Nigeria. There was however a significant proportion of haplogroup R (18% found among these Hausa populations) suggested to be mainly found in Central Africa (Hassan, *et al.,* 2008). This was the third most abundant haplogroup among the entire Hausa Populations and its preponderance was between 12.5% in Sokoto Hausa (SO) to 40% in Kano Hausa (KN) but was not observed in the Westernmost Kebbi Hausa (KB). Haplogroup R was almost exclusively found among the Hausas where it represented more than 90% of the entire Haplogroup in the Country (i.e. 15 in 16 subjects). On the background of the M2 mutation (i.e. pooling E1b1a with all E1b1a7 and E1b1a8), this was found in less than 50% of the Hausa populations of Nigeria despite its been represented not less than 90% in the four different Southern Nigeria populations (Igbo – 94%, Yoruba –90%, Bini -92% and Ijaw 100%). These were not substantially different from the reported observations in some recent publications (Montano, *et al.,*

2011; Barbieri, *et al.*, 2012). However when those not found to be E1b1a7 and E1b1a8 were excluded, as new haplogroups, E1b1a (xE1b1a7, xE1b1a8) was generally low among all the Nigerian populations. Another significant observation was the preponderance of Haplogroup E, not including Haplogroups E1b1a and E1b1b. This very broad Haplogroup was observed in more than 20% of the Hausa population and ranged from 14% to 33% except in Kano Hausa were it was not observed at all. These two haplogroups (E and R) could easily , with their unresolved sub-clades, have contributed to the high Haplogroup Diversities of the Hausa populations as they were both found in less than 10% of each of the four different Southern Nigerian populations (Igbo – 3% and 0%, Yoruba -4% and 0.4%, Bini - 8% and 0% and Ijaw – 0% and 0% respectively). Haplogroup B had its highest frequency among the Hausa in contrast with the four Southern Nigerian populations.

In the overall Nigerian pooled population, on the background of M2 mutation, Nigeria had 84% (E1b1a (xE1b1a7, E1b1a8) =4%, E1b1a7 =63% and E1b1a8 =19%) of this mutation which was comparable to earlier report of Veeramah *et al.* (Veeramah, *et al.*, 2010) which reported more than 90% for Cross Rivers Nigerian populations. Haplogroups E1b1b and R were found in about 1% and 3% respectively while both haplogroup B and E (xE1b1a, E1b1b) were found in about 5% and 6% respectively. Only four individuals (<1%) could not be resolved and are grouped as others.

Table 3.3a: Matrix of R_{st} distances based on Y-STR data among the Five Nigerian ethnic populations. R_{st} values are below diagonal while P values are above diagonal. Significant P<0.005 are represented in blue coloured print.

	Hausa	Igbo	Yoruba	Bini	Ijaw
Hausa		0.000	0.000	0.000	0.000
Igbo	0.076		0.009	0.054	0.162
Yoruba	0.060	0.006		0.018	0.459
Bini	0.053	0.021	0.026		0.261
Ijaw	0.053	0.011	0.000	0.007	

Table 3.3b: Matrix of F_{st} distances based on Y-SNP data among the Five Nigerian ethnic populations. F_{st} values are below diagonal while P values are above diagonal. Significant P<0.005 are represented in blue coloured print.

	Hausa	Igbo	Yoruba	Bini	Ijaw
Hausa		0.000	0.000	0.000	0.000
Igbo	0.241		0.036	0.793	0.405
Yoruba	0.154	0.018		0.405	0.279
Bini	0.153	-0.032	-0.016		0.676
Ijaw	0.180	-0.010	0.015	-0.031	

Tables 3.3a and 3.3b above, revealed the distinctions of the Northern Nigeria Afro-Asiatic Hausa populations from the four Niger Congo language speaking populations of Igbo, Yoruba, Bini and Ijaw in Southern Nigeria. Both the R_{st} and F_{st} distances showed similar pattern. The farthest population from the Hausa population was the Igbo population in both cases. There were no statistically significant differences among all the southern Nigeria populations.

This pattern was found to be consistent even when the Nigerian populations were resolved along the 22 different geographical origins (Supplementary Tables 1 and 2 which represented R_{st} and F_{st} distances of STR and SNP variations respectively). Most STR haplotypes manifest Continental and sometimes Country based sub-structure but not sub-structuring within populations.

All the major populations (Hausa, Yoruba and Igbo) were homogenous based on R_{st} distances but F_{st} distances showed that the Hausa populations were not homogenous (supplementary Table 2).

Table 3.4a: Matrix of Pairwise R_{st} distances of Y-STR data and P values of Nigerian pooled population and regionally representative African Countries. Significant P<0.003 is presented in blue colour print. R_{st} values are below diagonal while P values are above diagonal. The regions represented are North Africa (Tunisia), Southern Africa (Angola), West Africa (Burkina Faso), East Africa (Uganda) and Central Africa (Cameroon). Nigeria is also from West Africa

	Tunisia	Angola	Burkina Faso	Uganda	Cameroon	Nigeria
Tunisia		0.000	0.000	0.000	0.000	0.000
Angola	0.338		0.000	0.000	0.000	0.000
Burkina Faso	0.283	0.043		0.000	0.000	0.000
Uganda	0.174	0.145	0.123		0.000	0.000
Cameroon	0.265	0.089	0.057	0.099		0.000
Nigeria	0.286	0.046	0.019	0.139	0.031	

Table 3.4b: Matrix of Pairwise F_{st} distances of Y-SNP data and P values of Nigerian pooled population and regionally representative African Countries. Significant P<0.003 is presented in blue colour print. F_{st} values are below diagonal while P values are above diagonal. The regions represented are North Africa (Algeria), Southern Africa (Zambia), West Africa (Burkina Faso), East Africa (Uganda) and Central Africa (Democratic Republic of Congo). Nigeria is also from West Africa

	Nigeria	Algeria	Burkina Faso	Democratic Republic of Congo	Ethiopia	Zambia
Nigeria		0.000	0.000	0.009	0.000	0.000
Algeria	0.4763		0.000	0.000	0.000	0.000
Burkina Faso	0.1888	0.346		0.000	0.000	0.000
Democratic Republic of Congo	0.109	0.410	0.094		0.000	0.000
Ethiopia	0.361	0.186	0.228	0.267		0.000
Zambia	0.211	0.420	0.109	0.028	0.324	

Tables 3.4a and 3.4b revealed the genetic distinction of Nigeria in West Africa with other populations in West Africa (Burkina Faso) and Central Africa (Cameroon and Democratic Republic of Congo) followed by Southern Africa (Angola and Zambia). The farthest distinctions are those of North Africa (Tunisia and Algeria) and East Africa (Uganda and Ethiopia). Considering the time depth measured by two genetic markers (recent and pre-historic events by STR and SNP respectively) the inference seemed to suggest that structure since pre-historical times have persisted till present or that recent events has obliterated pre-historical structure.

While all the pooled populations are significantly distinct ($P<0.003$), however, some Nigerian populations seem not to be distinguishable from some from the Democratic Republic of Congo (Table 3.4b and Supplementary Table 3). The influence of the historic Bantu expansion can be detected from some of the relatively shorter R_{st} and F_{st} distances in the regional representative population from West Africa through Central Africa to Southern Africa especially, which was more geographically distant than both East Africa and North Africa from West Africa. However, the F_{st} distances showed more distinctiveness than R_{st} distances apparently because of their different mutation rates with the SNP rates slower than STR rates.

Supplementary Table 3 showed the comprehensive pairwise R_{st} distances ($p<0.000043$) among the comparative data of African and Middle Eastern populations from published literature. A break down showed that 12 West African populations from Guinea Bissau

(1), Senegal (1), and Burkina Faso (10) were included. 13 Central African populations from Equatorial Guinea (1), Cameroon (2), Central African Republic CAR (1), Democratic Republic of Congo DRC (5) and Gabon (4) were included. Five North Africa / Middle East populations include Algeria (2), Tunisia (1), Morocco (1) and Saudi Arabia (1) which were all Afro-asiatic language speakers. Four East African populations including Ethiopia (1), Uganda (1) and Kenya (2) were all Nilo-Saharan language speakers except a Kenyan Bantu population. Finally the 12 Southern African populations included Angola (4), South Africa (1), Namibia (1), Botswana (2) and Zambia (4).

Of these, the Ijaw population of Nigeria was not distinguishable from 9 West African, 12 Central African, 1 East African and 6 Southern African populations (Supplementary Table 3). The Bini population was not distinguishable from 3 West African, 9 central African, one east African and 4 southern African populations. The Hausas are not distinguishable from 2 West African, 5 Central African, one east African and 3 southern African populations. The Igbo population was not distinguishable from 4 Central African, one east African and four southern African populations. The Yoruba population was not distinguishable from 1 West African, 4 central African, one east African and four southern African populations.

It is noteworthy that based on this experiment; there was not one Nigerian population that was NOT distinguishable from the five North African / Middle Eastern populations (Supplementary Table 3).

3.2 Significant Forensic Parameters

Tables 3.5a, 3.5b and 3.5c summarized the allelic frequencies and genetic diversities of the 11-loci SWGDAM recommended Y-STRs in the major Nigerian populations of Hausa, Igbo and Yoruba respectively. The results for both the Bini **(**Table 3.5d) and the Ijaw (Table 3.5e) where included even though their sample sizes were much smaller than the other three which were in a ratio of 1:1.5:3. This was to have a rough idea of the heterogeneity of the overall Nigerian population and most importantly, to enrich the Nigerian dataset. This is very useful when considering that these Bini and Ijaw populations could be useful in explaining the sub-structuring along geographical locations and linguistics and might be useful in assessing the impact of population admixture in the two large southern Nigerian populations of Igbo and Yoruba respectively. Taking all these five tables together, it was observed that DYS19 revealed three patterns. The Yoruba and Igbo populations had Alleles 15 and 17 as the two most common in that order. The Bini and Ijaw had the reverse that is, 17 and 15, to that of the Yoruba and Igbo. The Hausa had the third pattern which showed a descent from 15 through 16 to 17. On the DYS389I and DYS389II loci, the most common in all the five populations are alleles 13 and 30 respectively. Considering that the DYS389II already incorporated DYS389I values, subtracting DYS389I values from those of DYS389II, the most common allele difference was Allele 17. The most common alleles among DYS390, DYS391 and DYS392 were alleles 21, 10 and 11 respectively, among all the five

populations. Different patterns were noticed with respect to locus DYS393 also among the five populations. The Hausa, Igbo and Yoruba had predominantly allele 13 (54%, 39% and 48% respectively) while the Bini and Ijaw had allele 15 at a prevalence of 47% and 54% respectively. All the five populations had allele 11 as the most common for locus DYS438 and 12 for locus DYS439. The Yoruba population showed an intermediate level of gene diversities across most of the 11 loci STRs except DYS19 and DYS439 where it was the least with values of 68% and 58% respectively. On the other hand, the Hausa generally showed the most diversities at most loci ranging 44% (DYS391) to 94% (DYS385ab) suggesting their very high heterogeneity among their population. This observation about the Hausa populations were confirmed by the Mann-Whitney U test which revealed a significant difference between the gene diversities of the Hausa and Igbo populations with a z and p values of -2.23 and 0.02575 respectively. However in contrast with other southern populations of Yoruba, Bini or Ijaw, these were not significant as the z and p values showed -1.701, 0.08897; -1.399, 0.1618 and -0.8696, 0.3845 for the Yoruba, Bini and Ijaw populations respectively. All the comparisons of individual southern Nigerian populations with each other were not significantly different.

Table 3.5a: Nigerian Hausa Population Allelic Frequencies And Gene Diversities (N=78)

Alleles	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS385ab Haplotypes	DYS385ab freqs
6										11,12	0.02564
7										12,12	0.01282
8					0.01282	0.01282		0.03846		11,13	0.02564
9					0.03846			0.01282		12,13	0.02564
10					0.74359	0.08974		0.15385	0.02564	13,13	0.07692
11		0.02564			0.08974	0.53846	0.01282	0.57692	0.32051	12,14	0.01282
12	0.01282	0.16667			0.0641	0.10256	0.0641	0.14103	0.39744	13,14	0.02564
13	0.07692	0.44872			0.05128	0.11538	0.53846	0.07692	0.16667	14,14	0.02564
14	0.0641	0.30769				0.12821	0.23077		0.08974	13,15	0.0641
15	0.42308	0.01282				0.01282	0.10256			14,15	0.01282
16	0.21795	0.02564					0.05128			12,16	0.01282
17	0.20513	0.01282								14,16	0.01282
18				0.02564						15,16	0.03846
19										16,16	0.03846
20				0.02564						15,17	0.05128
21				0.44872						16,17	0.08974
22				0.12821						14,18	0.01282
23				0.05128						16,18	0.17949
24				0.29487						17,18	0.08974
25				0.01282						17,19	0.02564
26				0.01282						19,19	0.01282
27			0.05128							15,20	0.03846
28			0.01282							18,20	0.02564
29			0.15385							20,21	0.03846
30			0.38462							22,22	0.01282
31			0.14103							22,23	0.01282
32			0.21795								
33			0.03846								
GD	0.7342	0.68754	0.76972	0.70598	0.44363	0.6744	0.6524	0.62903	0.71581		0.94079

Table 3.5b: Nigerian Igbo Population Allelic Frequencies And Gene Diversities (N=119)

ALLELES	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS385ab Haplotypes	DYS385ab Freqs
8								0.0084		11,11	0.0084
9					0.03361			0.01681		11,12	0.01681
10					0.85714			0.01681	0.0251	13,15	0.0084
11		0.0084			0.10084	0.93277	0.0084	0.92437	0.30252	15,15	0.0084
12		0.11765			0.0084	0.05042	0.05042	0.02521	0.5042	13,16	0.0084
13		0.7563				0.0084	0.0084	0.0084	0.15966	14,16	0.0084
14	0.05042	0.11765				0.0084	0.0084		0.0084	15,16	0.02521
15	0.36134									16,16	0.02521
16	0.27731									14,17	0.0084
17	0.31092									15,17	0.06722
18										16,17	0.11765
19										17,17	0.10924
20				0.01681						14,18	0.0084
21				0.89076						15,18	0.0084
22				0.06723						16,18	0.09247
23										17,18	0.2353
24				0.01681						18,18	0.07563
25				0.0084						16,19	0.0084
26										17,19	0.10924
27			0.0084							18,19	0.0084
28			0.01681							15,20	0.0084
29			0.09244							17,20	0.01681
30			0.5042							15,21	0.0084
31			0.28571							16,21	0.0084
32			0.08403								
33			0.0084								
34											
GD	0.7018	0.40873	0.6566	0.20987	0.26242	0.13573	0.5151	0.1527	0.6365		0.89939

Table 3.5c: Nigerian Yoruba Populations Allelic Frequencies and Gene Diversities (N=238)

ALLELE	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS385ab Haplotype	Freq.	DYS385ab Haplotype	Freq.
8								0.0126		11,11	0.0084	16,20	0.0084
9					0.0462			0.0252		12,12	0.0126	17,20	0.0168
10					0.8487	0.0378	0.0042	0.0714	0.0126	11,13	0.0042	18,20	0.0042
11		0.0042			0.0882	0.8824		0.8361	0.2395	13,13	0.0042	19,20	0.0042
12		0.1387			0.0126	0.0588	0.0462	0.0336	0.5966	11,14	0.0042	17,21	0.0084
13	0.0168	0.6933			0.0042	0.0168	0.3908	0.021	0.1092	13,14	0.0042	19,21	0.0042
14	0.042	0.1513				0.0042	0.3403		0.042	14,14	0.0126	17,22	0.0042
15	0.4496	0.0042					0.1765			13,15	0.0042	20,22	0.0084
16	0.2101	0.0084					0.042			15,15	0.0084	21,22	0.0042
17	0.2731									13,16	0.0084		
18	0.0084									14,16	0.0084		
19				0.0042						15,16	0.03357		
20				0.0084						16,16	0.04205		
21				0.8697						14,17	0.021		
22				0.0714						15,17	0.0042		
23										16,17	0.18642		
24				0.0294						17,17	0.10084		
25				0.0042						14,18	0.0126		
26				0.0084						15,18	0.0168		
27			0.0084	0.0042						16,18	0.0126		
28			0.021							17,18	0.15546		
29			0.0882							18,18	0.05042		
30			0.4202							15,19	0.0042		
31			0.2647							16,19	0.0252		
32			0.1639							17,19	0.02016		
33			0.021							18,19	0.0084		
34			0.0126							19,19	0.0042		
GD	0.68	0.47917	0.72067	0.2384	0.271	0.2172	0.6994	0.2947	0.5752				0.92653

Table 3.5d: Nigerian Bini Population Relative Allelic frequencies and Gene Diversities (N=13)

ALLELES	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS385ab Haplotypes	DYS385ab Freqs
9					0.07692					16,17	0.15385
10					0.76923			0.07692		16,18	0.07692
11					0.07692	0.76923		0.76923		17,18	0.23077
12		0.07692				0.15385		0.07692	0.76923	18,18	0.07692
13		0.84615			0.07692	0.07692	0.23077	0.07692	0.15385	17,19	0.07692
14	0.0769	0.07692					0.15385		0.07692	18,19	0.23077
15	0.2308						0.53846			19,20	0.07692
16	0.3077									17,21	0.07692
17	0.3846										
18											
19											
20											
21				0.92308							
22											
23				0.07692							
24											
25											
26											
27											
28											
29			0.07692								
30			0.61538								
31			0.23077								
32											
33			0.07692								
34											
GD	0.782	0.35553	0.63955	0.2253	0.4739	0.462	0.7165	0.4739	0.6188		0.92357

Table 3.5e: Nigerian Ijaw Population Allelic Frequencies and Gene Diversities (N=15)

ALLELE	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS385ab haplotypes	DYS385ab freqs.
9								0.06667		14,15	0.06667
10					0.93333			0.06667		14,16	0.06667
11					0.06667	0.93333		0.8	0.33333	15,16	0.06667
12		0.13333						0.06667	0.46667	16,17	0.33333
13	0.06667	0.73333				0.06667	0.2		0.2	17,17	0.13333
14	0.06667	0.13333					0.2667			17,18	0.13333
15	0.33333						0.4667			16,21	0.06667
16	0.13333						0.06667			20,22	0.06667
17	0.4									21,22	0.06667
18											
19											
20											
21				0.93333							
22				0.06667							
23											
24											
25											
26											
27											
28											
29			0.13333								
30			0.53333								
31			0.2								
32											
33			0.13333								
34											
GD	0.77365	0.4981	0.71143	0.1959	0.1959	0.19588	0.7762	0.3714	0.6762		0.8981

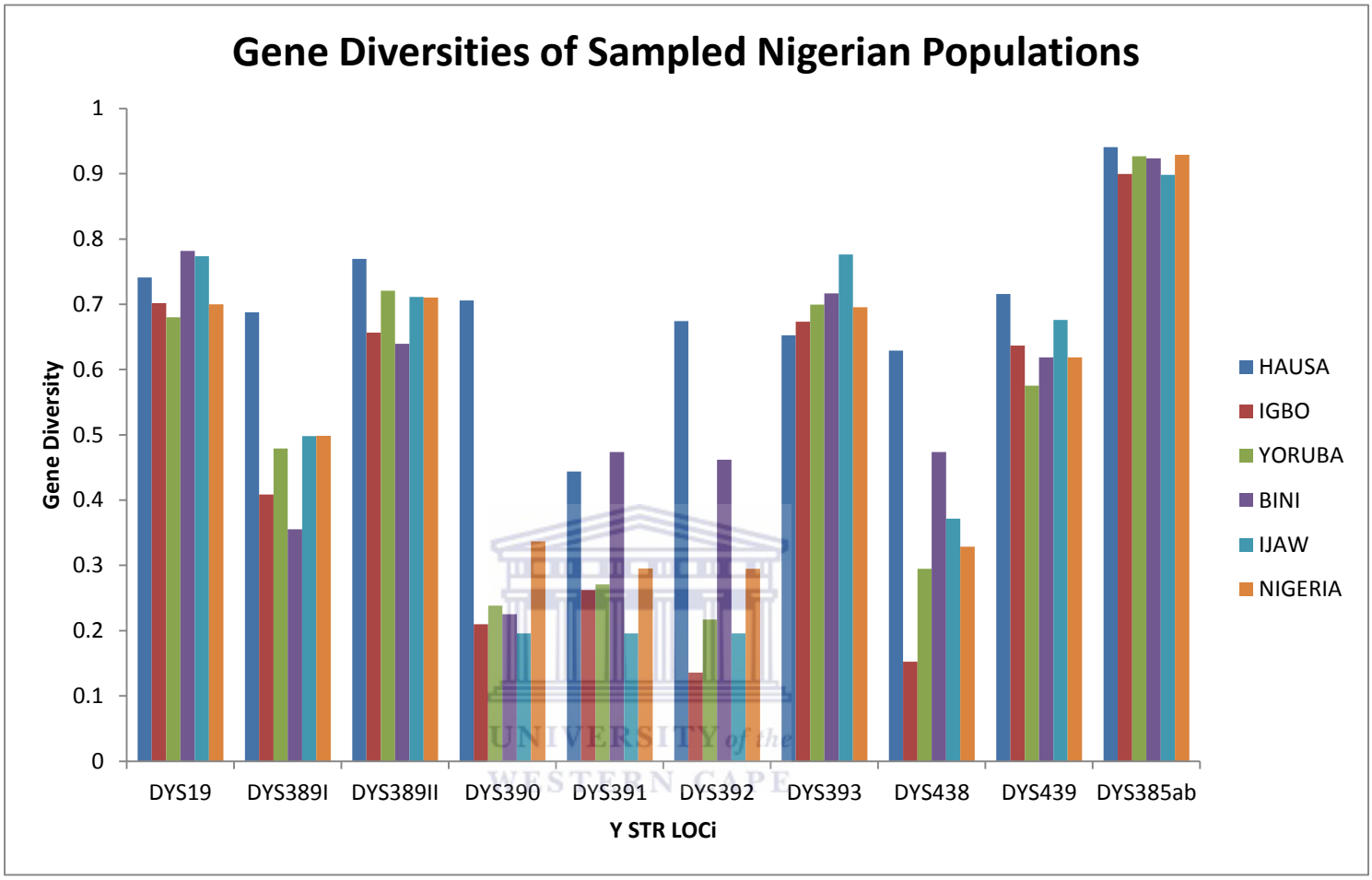


Figure 3.1**:** Comparative gene diversities of the five major Nigerian populations with the overall Nigerian population.

The Figure 3.1 above, showed a graphical contrast of the gene diversities of the major Nigerian populations both among themselves and also with the overall Nigerian pooled population. This plot also confirmed the outstanding heterogeneity of the Y-STRs among

the Hausa, even when the southern Nigerian populations showed relatively lower diversities for DYS 390 and DYS 392.

Table 3.6a: Forensically significant data from the five Nigerian populations based on the 11-loci SWGDAM recommended Y-STRs. N is sample size, K is the number of Haplotypes, HD is Haplotype diversity and RMP is the Random Match Probability and UH is Unique Haplotypes

Population	N	K	UH	DC	HD	RMP
Hausa	78	75	74	0.962	0.999167	0.000833
Igbo	119	108	99	0.908	0.99815	0.00185
Yoruba	238	214	193	0.899	0.998834	0.001166
Bini	13	13	13	1	1	0.000000
Ijaw	15	14	13	0.933	0.990471	0.009529

Table 3.6b: Forensic parameters based on SWGDAM recommended Y STRs of representative African Countries. N is sample size, K is the number of Haplotypes, HD is Haplotype diversity and RMP is the Random Match Probability and UH is Unique Haplotypes

Country	N	K	DC	HD	RMP
Tunisia	100	71	0.71	0.9764	0.0236
Angola	230	165	0.7174	0.9929	0.0071
Burkina Faso	322	242	0.7516	0.9957	0.0043
Uganda	118	73	0.6186	0.9929	0.0071
Cameroon	53	44	0.8302	0.992	0.008
Nigeria	463	394	0.851	0.999	0.001

Table 3.7: Discrimination Capacity (DC) of the 11 loci Y-STR extended haplotype compared with more African countries. The table compares several African countries Y-STRs. Most were modified to represent 10 STRs or 11 loci while others 11 STRs or 12 loci[a] and 16 STRs or 17 loci[b] since the DYS 385 is a duplicated STR.

CONTINENTAL REGION	COUNTRY	NUMBER OF INDIVIDUALS SAMPLED	NUMBER OF HAPLOTYPES OBSERVED	DISCRIMINATION CAPACITY	REFERENCE
West Africa	Nigeria	463	394	0.850972	This report
West Africa	Burkina Faso	334	253	0.757485	De Filippo et al. (2011)
West Africa	Guinea Bissau[a]	161	154	0.956522	Rosa et al. (2006)
Central Africa	Equatorial Guinea[b]	101	94	0.930693	Arroyo-Prado et al. (2005)
Central Africa	Democratic Republic of Congo	62	54	0.870968	De Filippo et al. (2010)
Central Africa	Gabon	828	552	0.666667	Berniel-Lee et al. (2009)
Central Africa	Cameroon	55	46	0.836363	Berniel-Lee et al. (2009)
East Africa	Ethiopia	69	48	0.695652	De Filippo et al. (2011)
East Africa	Uganda	118	73	0.618644	Gomes et al. (2009)
East Africa	Kenya	61	52	0.852459	De Filippo et al. (2011)
East Africa	Somalia[a]	201	96	0.477612	Hallenberg et al. 2005
Southern Africa	Zambia	546	380	0.695971	De Filippo et al. (2011)
Southern Africa	Mozambique[b]	112	101	0.901786	Alves et. al (2003)
Southern Africa	Angola [a]	166	138	0.831325	Melo(2010),Coelho et al. (2009)
North Africa	Egypt[b]	208	204	0.980769	Omran et al. (2008)
North Africa	Morocco[b]	166	164	0.987952	Laouina et al. (2011)
North Africa	Tunisia[a]	100	71	0.71	Brandt-Casadevall et al. (2003)
North Africa	Algeria	20	16	0.8	De Filippo et al. (2011)

Tables 3.6a and **3.6b** summarized the forensically significant parameters based on the SWGDAM Y-STRs for the Nigerian major populations and overall on one hand and a comparison with representative African Countries on the other hand. The countries were selected to represent North Africa (Tunisia), Southern Africa (Angola), West Africa (Burkina Faso), East Africa (Uganda) and Central Africa (Cameroon) specifically.

Even though the Bini and the Ijaw populations were apparently under represented among the Nigerian populations, their capability to enrich the overall Nigerian dataset cannot be overlooked especially as when one observes that these two populations showed apparently greater discrimination capacities than their two major southern populations of Igbo and Yoruba. Even the relatively fewer Hausa populations revealed a greater discrimination capacity than the Igbo and Yoruba populations. These apparently lower discrimination capacities among the Yoruba and Igbo were only observed when the two different populations were pooled across several geographical locations but not within each sampling location. This suggested a substantial admixture or haplotype sharing among the peoples of these two populations which was not observed among the Hausa populations of Northern Nigeria (see **Table 3.6a**). From Table 3.6a, there was apparent haplotype sharing which could be mere coincidence due to apparently high mutation rates of STRs generally or slight hint of admixture among the populations. A look at Y SNP might assist to resolve this conclusively. The later suggestion with respect to the haplotype sharing might be very likely among the people that speak the same language

but for those that speak different languages it might be necessary to analyze the structure behind this observation and the forces driving it. However, it was not absolutely out of place for there to be admixture considering the relatively recent history of the current Nigerian populations of slavery and slave integration and free human migration and relocation for economic reasons allowing people to learn the languages of neighbouring tribes and thus facilitate the possibilities of inter-ethnic marriages. This might affect the ethnic distinctions over time. In actual fact, the YHRD anticipated this when it introduced the concept of meta-population groups like the African metapopulation. It was left to be proven if there was absolutely justified for the African populations where the SWGDAM loci showed substantially lower discrimination capacities (<0.85) within the respective countries (Table 3.6b). Obviously, increasing the number of loci analyzed will definitely increase the discrimination capacity as shown among data from some African Countries (Table 3.7). But it must be stressed that the SWGDAM loci has relatively low haplotype diversities and discrimination capacities among the countries of Africa examined in this thesis.

One of the weaknesses of Y-STRs is their generally high Random Match Probabilities (RMPs) when compared with autosomal STRs which have been observed to have RMP in the range of 10^{-18}. These high RMPs were observed for individual populations as well as at country level (Tables 3.6a and 3.6b) respectively. It must be noted that comparatively speaking, Nigeria has relatively greater DC, HD and lower RMP than the

other African countries in this thesis (Table 3.6b). Also, it must be noted than even though the Bini Population had relatively small sample size, it had notably significant DC, HD and RMP in the Nigerian population. The Hausa population, among the three major populations of Nigeria, has the most heterogeneity, DC and RMP, followed by the Yoruba before the Igbo. Overall, the Nigerian population has close to 75% unique haplotypes, a testament to its heterogeneity.

Searches of the 394 STR haplotypes of the Nigerian populations on the YHRD revealed that 120 haplotypes appeared at least once while 274 have never been reported. Among the Nigerian populations alone, 49 haplotypes were shared among the five populations while 345 were unique. On respective populations, 74, 99, 193, 13 and 13 haplotypes were unique to the Hausa, Igbo, Yoruba, Bini and Ijaw populations respectively. The two most common haplotypes found in the overall population were shared by 6 individuals each (Table 3.8 below), cutting across at least two different populations.

The inclusion of the Nigerian data set has the potential of enriching the African meta-population of the YHRD with more than 270 new SWGDAM haplotypes and increasing the sampled populations by five, and total haplotypes by 393. Until date (May 2014), the YHRD had 126, 931 haplotypes with Africa contributing a little above 5% despite the relative recent increased contributions of African data

Table 3.8: The most common SWGDAM-STR based haplotypes shared among the Nigerian populations and their matches numbers on the YHRD

Haplotypes	DYS19	DYS 389I	DYS 389II	DYS390	DYS391	DYS392	DYS393	DYS 385a	DYS 385b	DYS438	DYS439	Frequency	YHRD MATCHES
N001	17	13	30	21	10	11	14	17	18	11	12	6	18
N002	15	13	31	21	10	11	13	16	17	11	12	6	13
N003	17	13	30	21	10	11	13	17	18	11	12	4	9
N004	15	13	31	21	10	11	14	16	16	11	12	4	3
N005	15	13	32	21	10	10	13	16	18	11	11	4	0
N006	15	13	30	21	10	11	13	16	18	11	12	4	4
N007	17	13	30	21	10	11	14	18	18	11	12	3	9
N008	16	13	31	21	10	11	13	16	17	11	12	3	2
N009	15	13	30	21	10	11	14	17	18	11	12	3	3
N010	17	13	30	21	10	11	14	17	17	11	12	3	18
N011	16	13	30	21	10	11	13	16	18	11	11	3	0
N012	15	13	30	21	10	11	13	17	18	11	12	3	1
N013	16	13	30	21	10	11	13	17	18	11	12	3	8
N014	17	13	30	21	10	11	13	18	18	11	12	3	8
N015	15	13	30	21	10	11	15	17	18	11	12	3	5
N016	17	13	30	21	10	11	14	16	17	11	12	3	6
N017	15	13	32	21	10	11	13	16	17	11	11	3	3

3.3 Further assessment of Genetic Structure in the Nigerian Populations

Figure 3.2 below was the Neighbour Joining tree showing how the five ethnic populations clustered with other African populations. While the Yoruba, Igbo and Ijaw populations clustered with only populations from Central and West African populations, The Hausa population clustered with Central, West and Southern African populations. Again, not a single Nigerian population clustered with populations from North or East Africa! All the North African populations clustered together. The East African population of Ethiopia and other East African populations like Uganda and Kenya also clustered differently despite their geographical proximity.

This tree, even though limited to genetic variations, does not seem to reveal any linguistic affinity among clustering populations except among the North African populations. This however, suggested the possibility of significant gene flow across linguistic groups (Bantu and Non Bantu populations) and regional (West, Central and Southern African) populations

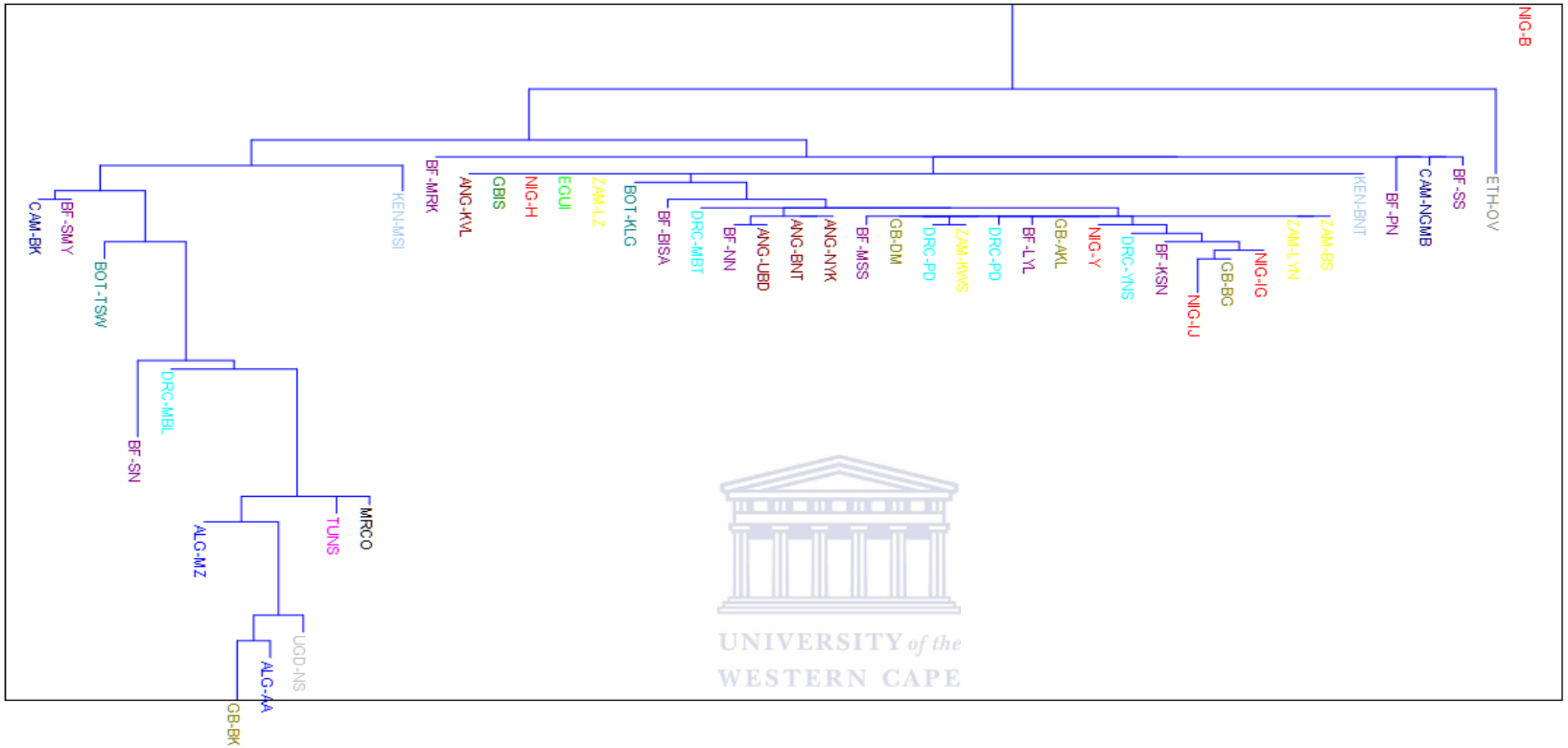


Figure 3.2: Neighbour joining tree of 46 African populations (five populations with samples sizes less than 10 from Tables 2.2 were excluded). The Nigerian populations were represented in red and other country-based populations were represented in unique country colours. The population codes are as described in Table 2.2.

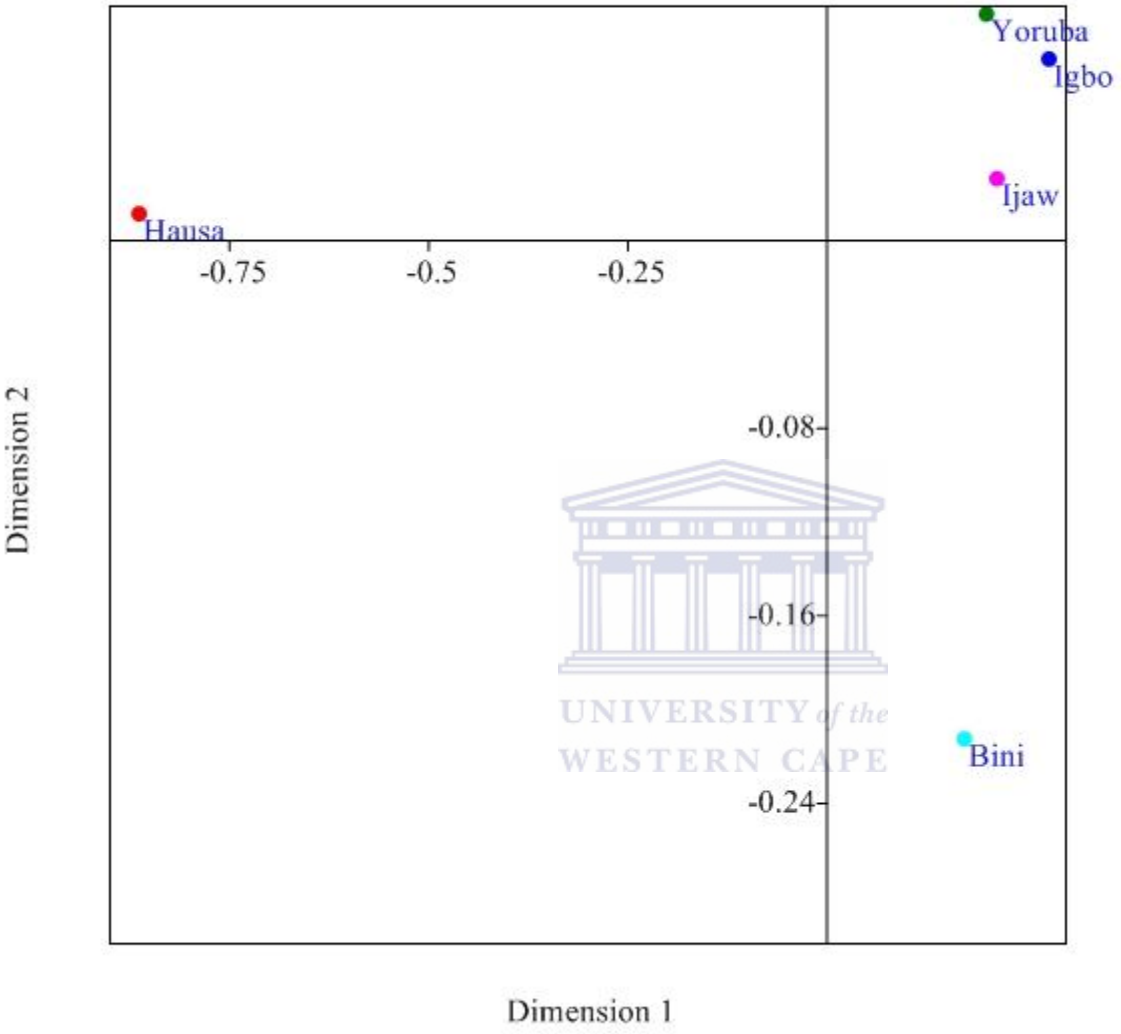


Figure 3.3a**:** The multidimensional scaling (MDS) plot of R_{st} distances based on Y-STR data of the five Nigerian ethnic populations (stress 0). The Hausa, Igbo, Yoruba, Bini and Ijaw populations were represented with red, blue, green, cyan and pink dots respectively.

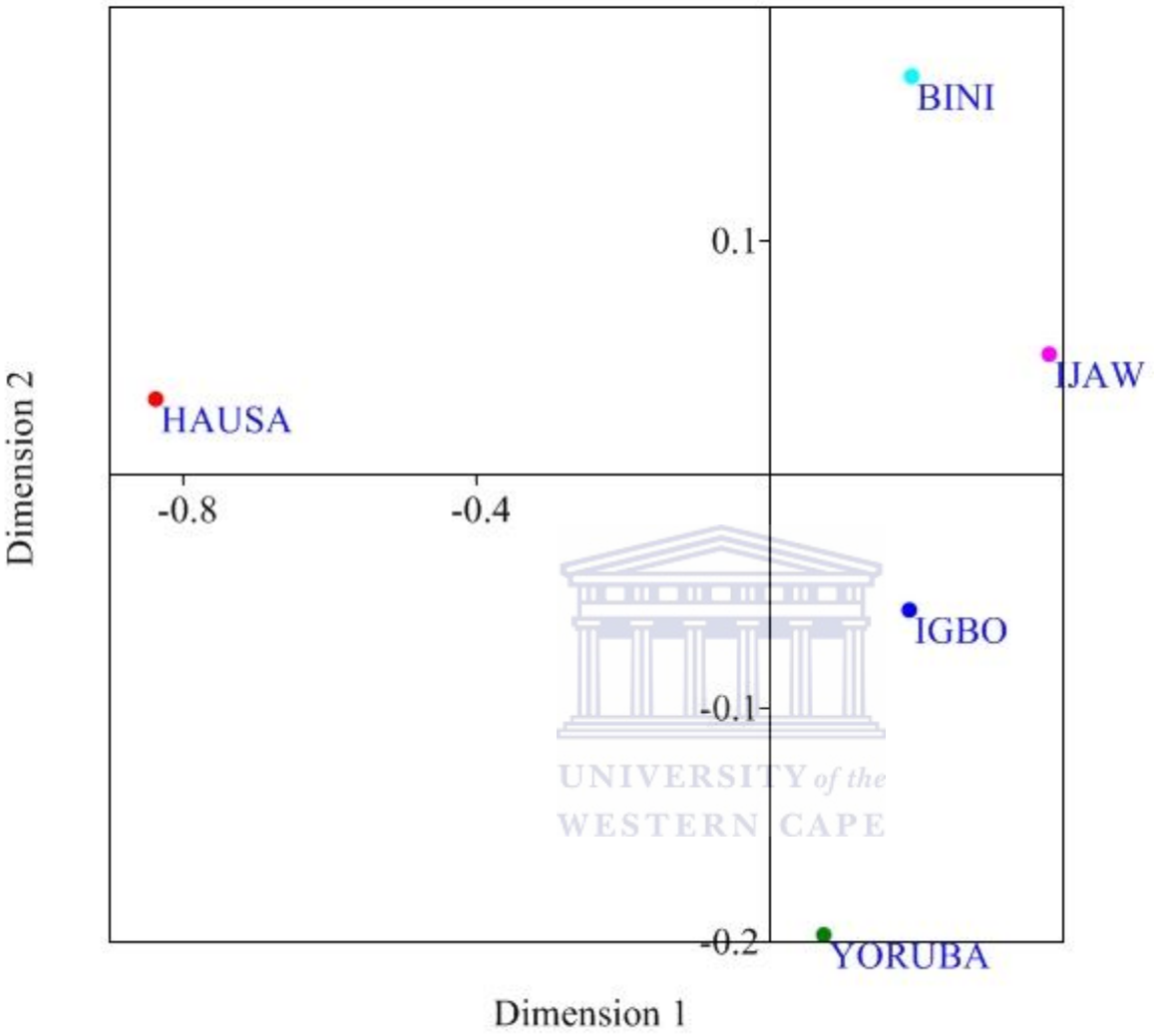


Figure 3.3b: The Multidimensional Scaling (MDS) plot based on F_{st} distances of Y-SNP data of the five Nigerian ethnic populations. The Hausa, Igbo, Yoruba, Bini and Ijaw populations were represented with red, blue, green, cyan and pink dots respectively.

Although Figures 3.3a and 3.3b presented genetic distances measured at different time depth (R_{st} (for STR) for recent events and F_{st} (for SNP) for pre-historical events), they shared some similarities that the Afro-asiatic language speaking Hausa population of

Northern Nigeria are spatially distinct from the four Niger Congo speaking Igbo, Yoruba, Bini and Ijaw populations of Southern Nigeria. The apparent distinction of the Bini population in Figure 3.3a apparently was a result of some stochastic effects of the sample peculiarities. However, very clearly revealed here was that both the Igbo and Yoruba populations are spatially close as they clustered together in both plots.

When these samples of ethnic populations were sub-divided into their respective geographical origins (Figures 3.4a and 3.4b respectively as shown below), The Igbo, Yoruba, Bini and Ijaw clustered together. The Hausa populations were observed to have three distinct groupings: (i) The Niger Hausa clustered with the Southern Nigeria populations, (ii) The Kaduna Hausa was an outlier and (iii) The remaining four Hausa populations of Jigawa, Kano, Kebbi and Sokoto formed a third and intermediate cluster between the Niger and Kaduna populations.

As these four MDS plots (Figures 3.3a, 3.3b, 3.4a and 3.4b) only showed the spatial representation of the populations in different contexts, they do not reveal the level of significance of the populations' differences as clearly as Tables 3.3a, 3.3b, 3.4a and 3.4b summarized earlier above. However, these figures confirmed that (i) Most Southern Nigerian populations are not distinguishable, (ii) The Hausa populations are distinct from the Southern Nigerian populations and (iii) The Hausa populations are not homogenous.

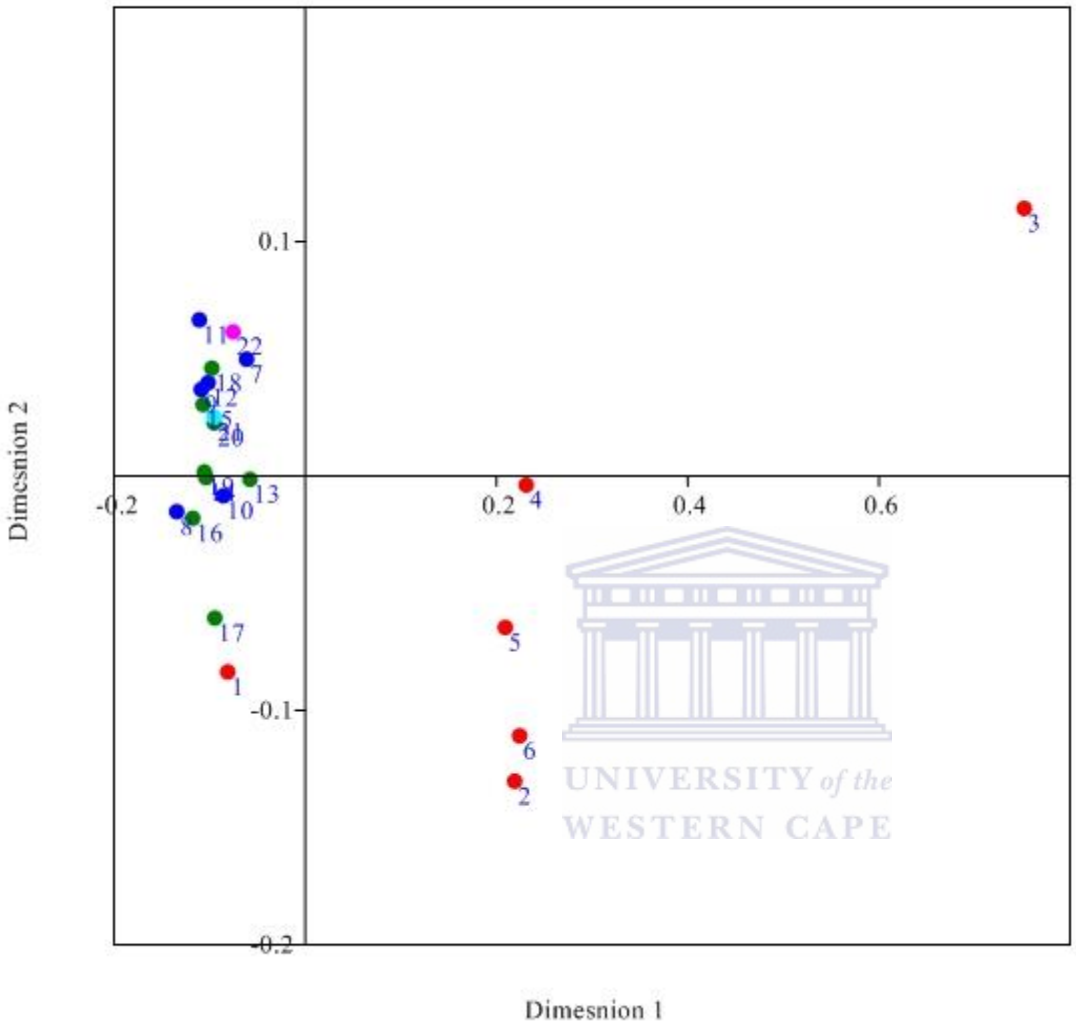


Figure 3.4a: The multidimensional scaling (MDS) plot of R_{st} distances of Y-STR data of the Nigerian populations across the 22 geographical origins (stress 0.0902). The populations were represented with numbers where 1-6 are the Hausa populations (red dot) from Niger, Jigawa, Kaduna, Kano, Kebbi and Sokoto respectively; 7-12 are the Igbo populations (blue dot) from Abia, Anambra, Delta, Ebonyi, Enugu and Imo respectively, 13-20 are the Yoruba populations (green dots) from Ekiti, Kogi, Kwara, Lagos, Ogun, Ondo, Osun and Oyo respectively and 21 and 22 are the Bini (cyan dot) and Ijaw (pink dot) populations from Edo and Bayelsa respectively.

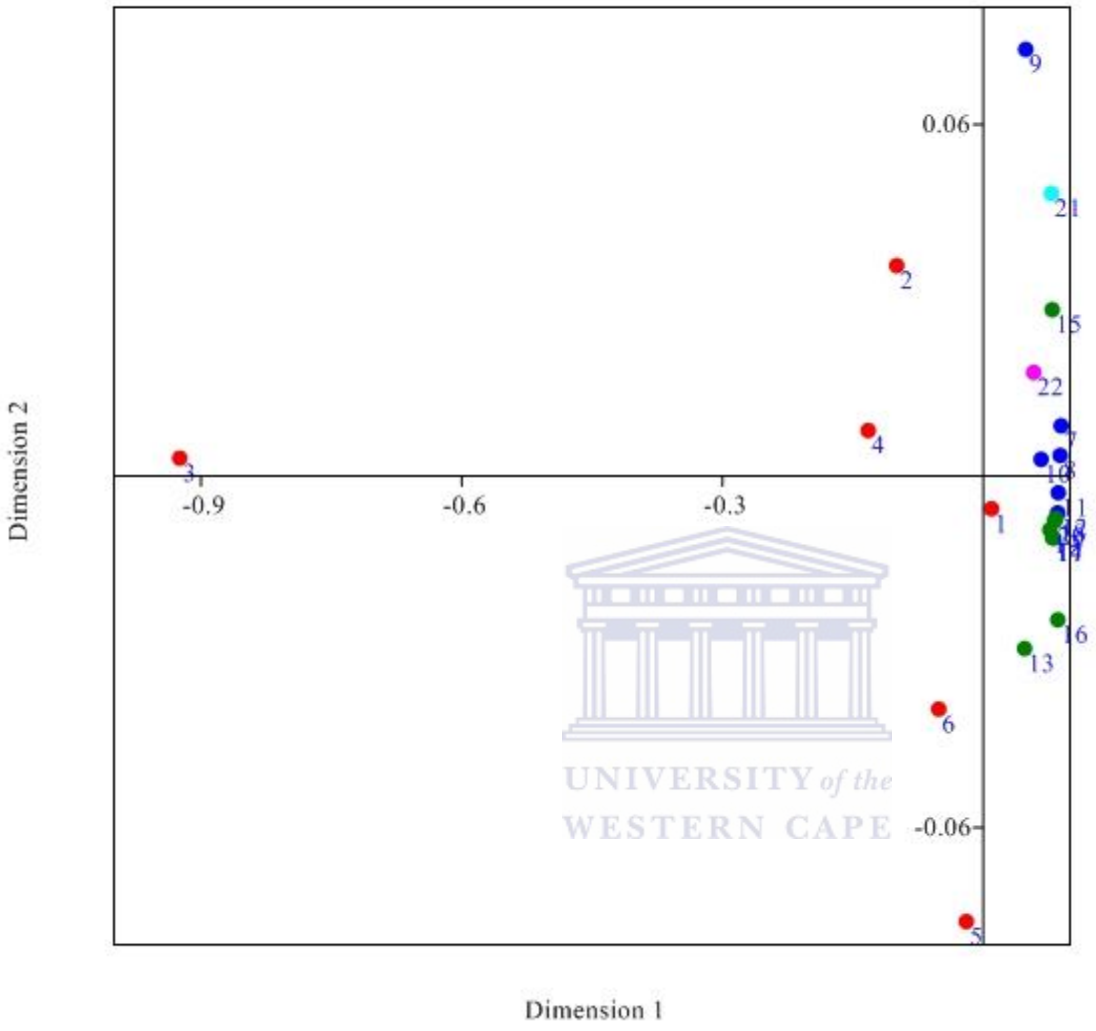


Figure 3.4b: The multidimensional scaling (MDS) plot of F_{st} distances of the Y-SNP data of Nigerian populations along 22 geographical locations (Stress 0.032). The populations were represented with numbers where 1-6 are the Hausa populations (red dot) from Niger, Jigawa, Kaduna, Kano, Kebbi and Sokoto respectively; 7-12 are the Igbo populations (blue dot) from Abia, Anambra, Delta, Ebonyi, Enugu and Imo respectively, 13-20 are the Yoruba populations (green dots) from Ekiti, Kogi, Kwara, Lagos, Ogun, Ondo, Osun and Oyo respectively and 21 and 22 are the Bini (cyan dot) and Ijaw (pink dot) populations from Edo and Bayelsa respectively.

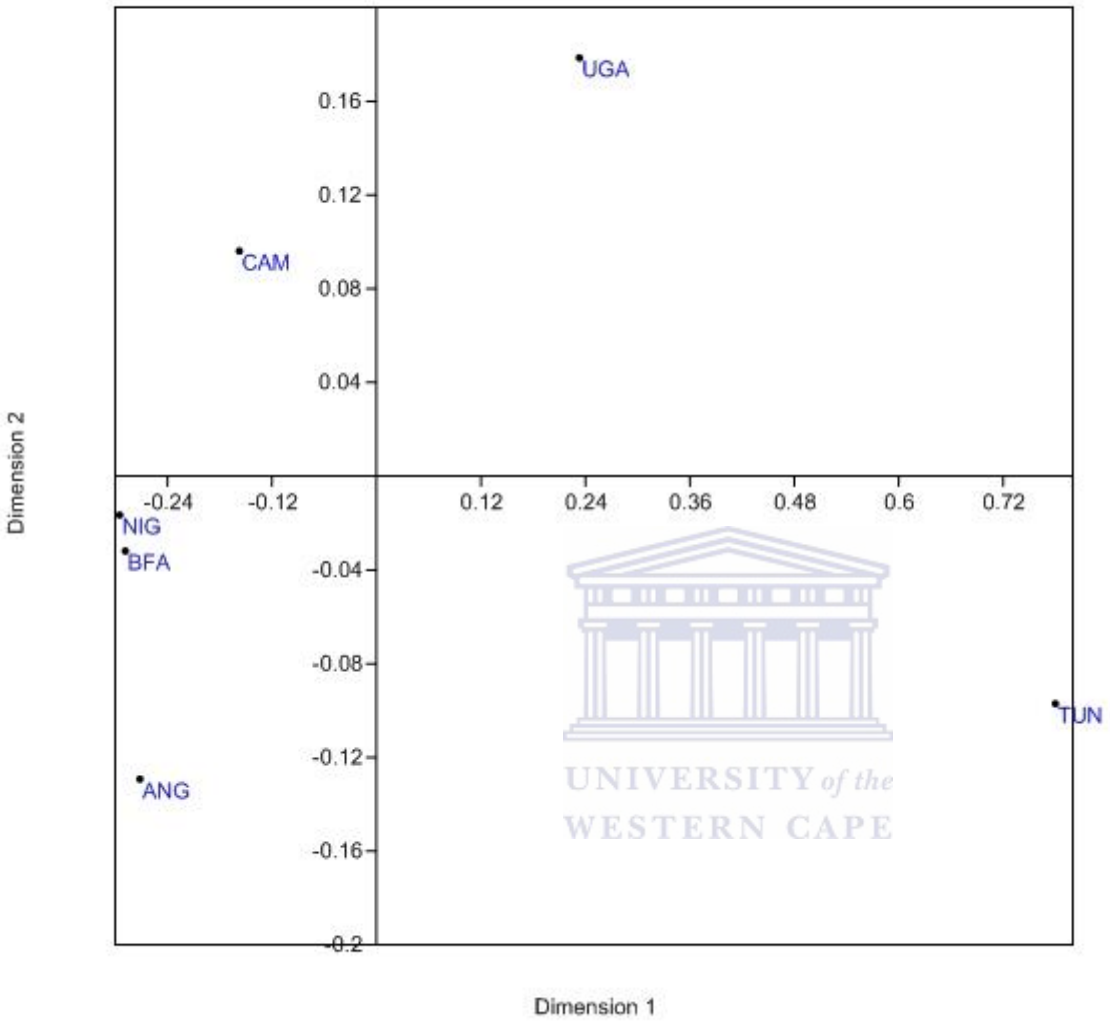


Figure 3.5a: The multidimensional scaling (MDS) plot of R_{st} distances of Y-STR data of the pooled Nigerian and five regionally representative African Countries (stress 0). The North Africa, West Africa, Central Africa, East Africa and Southern Africa regions were represented by Tunisia (TUN), Burkina Faso (BFA), Cameroon (CAM), Uganda (UGD) and Angola (ANG) respectively.

Figure 3.5a above confirmed some distinctively geographical region and broad linguistic differentiations among the contrasted African populations R_{st} distances. North African Afro-asiatic language speaking Tunisian population was distinct from the other sub-Saharan Africa populations as shown from left to right of the plot. The sub-Saharan countries are also distinguished along linguistic groupings with Nilo-Saharan language speaking Uganda of East Africa (upper part of the plot) and the majorly Niger Congo language speaking Burkina Faso (West Africa) and Angola (Southern Africa) with Cameroon (Central Africa) intermediate of these, to the middle of the plot. The Nigerian population clustered with the West African and Southern African Countries.

The MDS plot (Figure 3.5b**)** below comparing the Nigerian population F_{st} distances with some mostly new set of African populations (whose SNP haplogroups were accurately, not speculatively reported) also showed similar pattern to that of the R_{st} distances among the other African countries above. The differentiation of Algeria (ALG) and Ethiopia (ETH) from the others on the left and right side of the plot along the linguistic and geographic divide of Afro-Asiatic/Nilo-Saharan and Niger Congo languages with North/East Africa and West/Central/Southern Africa countries respectively was evident. The Niger Congo bantu speaking populations of Central (Democratic Republic of Congo, DRC) and Southern Africa (Zambia, ZAM) clustered together with the non Bantu Burkina Faso (BFA) population but spatially distinct from the Nigeria (NIG) population, her non-bantu Western African neighbour with which they shared much geographical proximity.

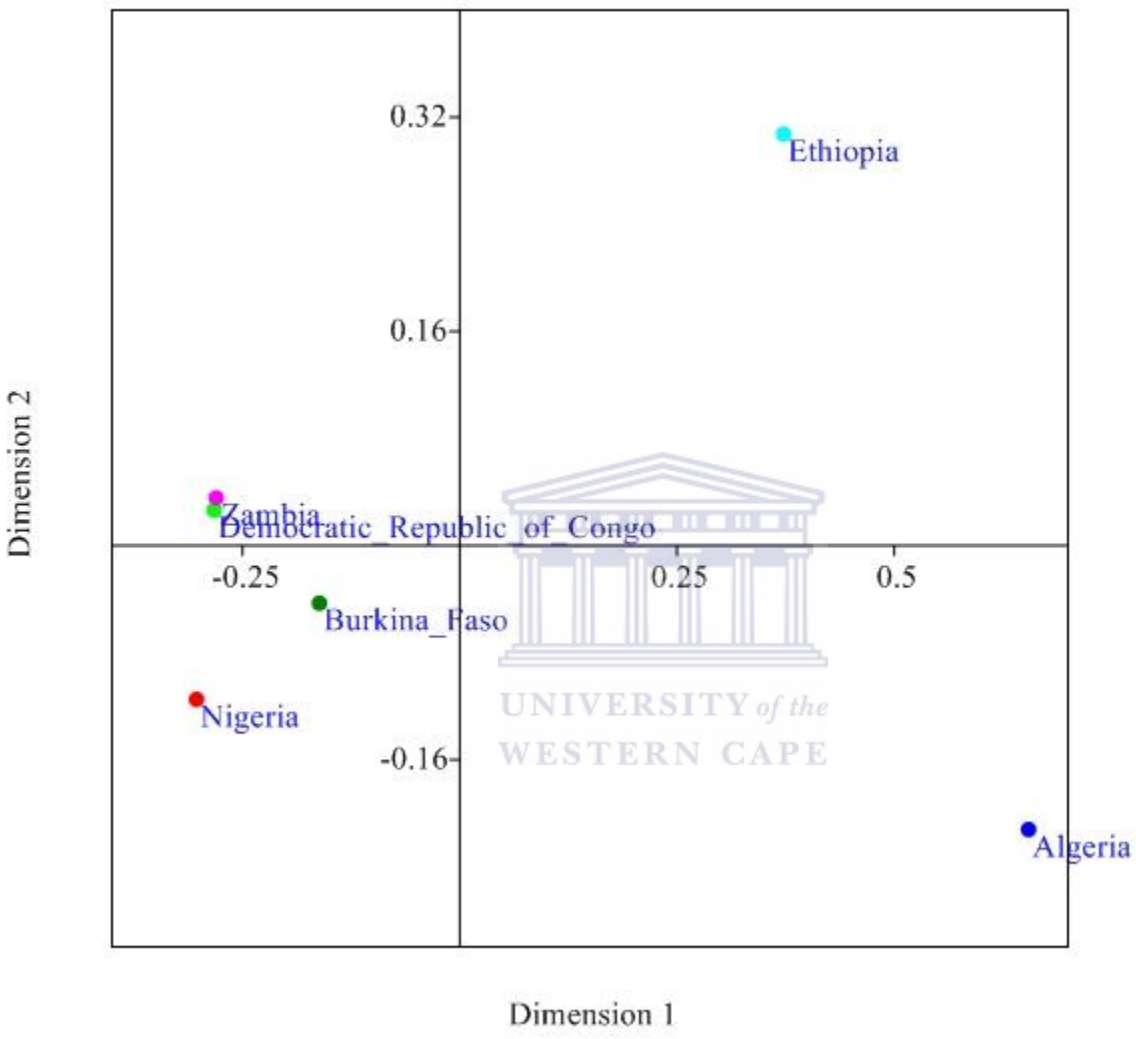


Figure 3.5b**:** The multidimensional scaling (MDS) plot of F_{st} distances Y-SNP data of the pooled Nigerian and five regionally representative African Countries (stress 0). The North Africa, West Africa, Central Africa, East Africa and Southern Africa regions were represented by Algeria (Algeria), Burkina Faso (BFA), Democratic Republic of Congo (DRC), Ethiopia (Algeria) and Zambia (ZAM) respectively.

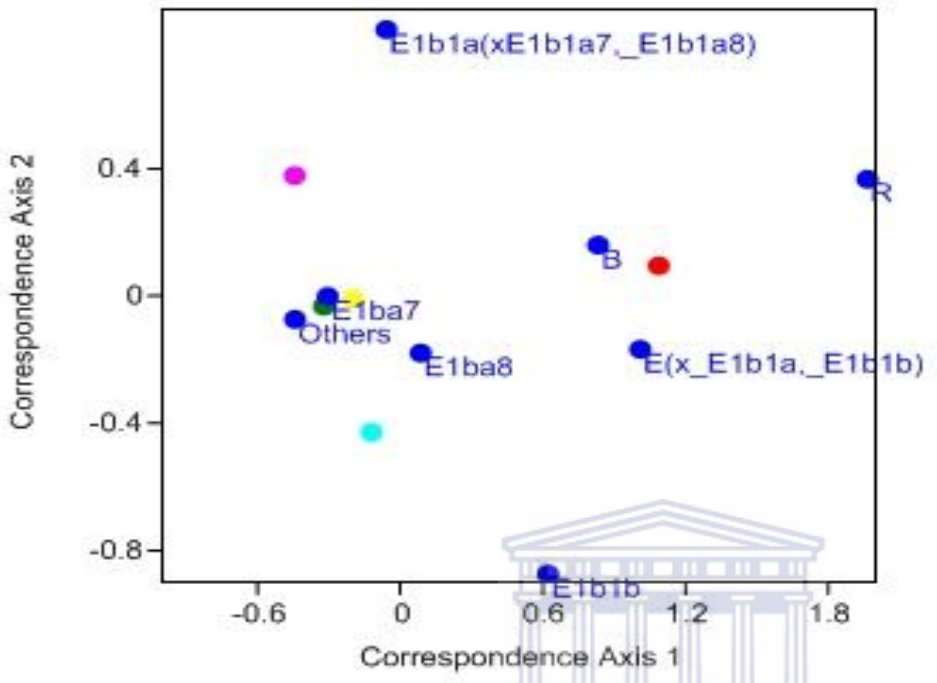


Figure 3.6a: The Correspondence plot of relative haplogroup frequencies and the five Nigerian ethnic populations. The haplogroups were represented with blue dots and named. The Hausa, Igbo, Yoruba, Bini and Ijaw populations were represented with Red, Green, Yellow, Cyan and Pink dots respectively.

Figure 3.6a above revealed the significant influence of haplogroup E1b1a7 on both the Yoruba (Yellow dot) population and the Igbo (Green dot) populations. The Hausa (Red dot) population was influenced majorly by three other haplogroups namely B, R and E(xE1b1a and E1b1b). The genetic proximity of the Igbo population was also obvious from their overlapping with one another. The Ijaw (Pink dot) and the Bini (Cyan dot)

populations were influenced by E1b1a7 and to a lesser extent, E1b1a (xE1b1a7, E1b1a8) and E1b1b respectively. The most influential haplogroups in the Nigerian populations generally speaking are in descending order E1b1a7, E1b1a8 and B. It must be noted that "Others" in this plot and Figure 2.3.5b below, referred to the non-specific haplogroups of the individuals that were not resolved into the seven specific haplogroups observed in this report.

When the sampled populations were resolved based on their different geographical locations (#1-22) in Figure 3.6b below, the influences of haplogroups E1b1a7 and E1b1a8 on the Southern Nigerian populations, especially, became more obvious (see clustering around the equator in the plot). Also very significant from this Figure was the obvious influences of different haplogroups on the different Hausa populations (#1-6) confirming the genetic structure within this population. Specifically the Kano (#4) Hausa population was influenced by haplogroup E1b1b, Jigawa (#2) Hausa population by haplogroup R, Niger (#1) Hausa population by E1b1a7 and probably E1b1a8, Kaduna (#3) Hausa population by haplogroup E (xE1b1a, E1b1b), Sokoto (#6) Hausa population by haplogroup E1b1a (xE1b1a7, xE1b1a8) and Kebbi (#5) Hausa population by haplogroup B.

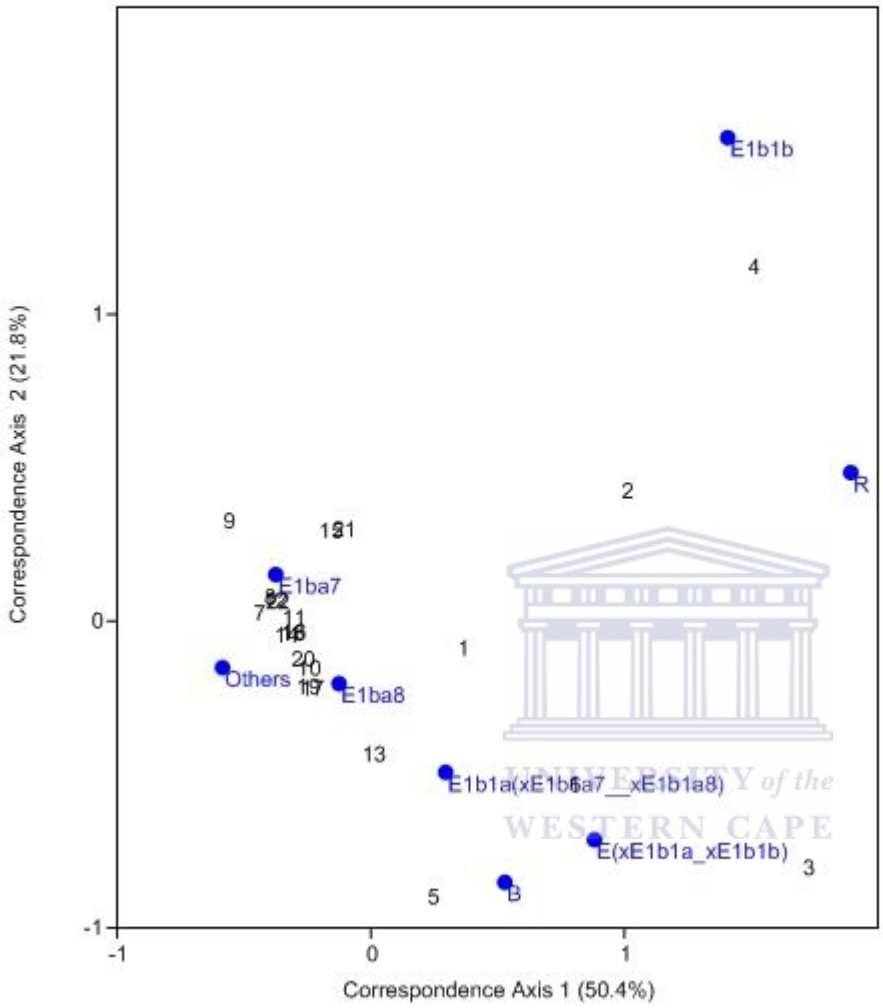


Figure 3.6b: The Correspondence plot relative haplogroup frequencies and Nigeria populations across the 22 different geographical origins. The haplogroups in blue dots were named while the geographical location specific populations represented with the numbers (#). The numbers1-6 were the Hausa populations from Niger, Jigawa, Kaduna, Kano, Kebbi and Sokoto respectively; 7-12 were the Igbo populations from Abia, Anambra, Delta, Ebonyi, Enugu and Imo respectively, 13-20 were the Yoruba populations from Ekiti, Kogi, Kwara, Lagos, Ogun, Ondo, Osun and Oyo respectively and 21 and 22 were the Bini and Ijaw populations from Edo and Bayelsa respectively.

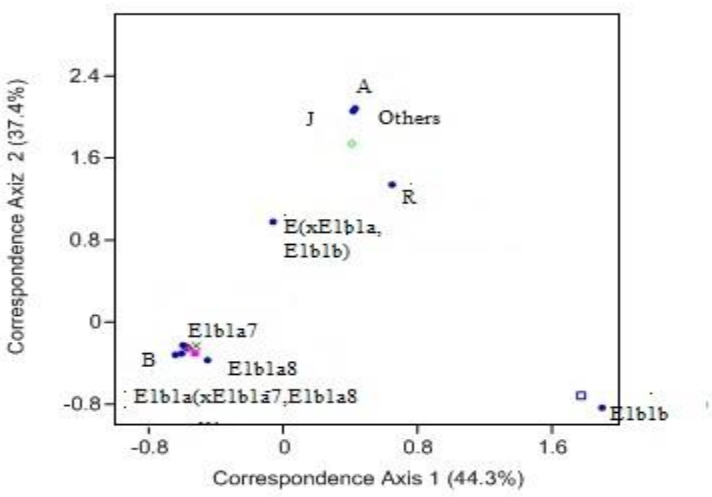


Figure 3.6c: The Correspondence plot of pooled Nigerian populations with five regionally representative African Countries and their relative haplogroup frequencies. Blue dots are the haplogroups labeled in bold. The Countries are Algeria (Blue Square), Ethiopia (Lemon Diamond), Nigeria (Red Cross), Burkina Faso (Green Cross), Zambia (Pink Square) and the Democratic Republic of Congo (Purple Circle)

When the pooled Nigerian populations were assessed with their African neighbours in **Figure 3.6c** above, there was a very clear clustering of the Niger Congo Bantu and Non Bantu speaking populations in the center influenced by haplogroup E1b1a and its sub-haplogroups E1b1a7 and E1b1a8. The Afro-Asiatic Ethiopia population was strongly influenced by the presence of both haplogroups A and J especially confirming that haplogroup J is "Easternly" distributed in Africa. Also, the North Africa population of Algeria was mostly influenced by the suggested Afro-Asiatic signature haplogroup E1b1b.

Table 3.9: The AMOVA F-statistics with significant groupings

	Y-STR (R_{st})		Y-SNP (F_{st})	
Ethnic grouping(5)				
F_{st}	0.03947	P<0.000001	0.07801	P<0.000001
F_{sc}	0.00727	P=0.07918	0.00906	P=0.18573
F_{ct}	0.03244	P<0.000001	0.06957	P<0.000001
Geographic region(4)				
F_{st}	0.03665	P<0.000001	0.08279	P<0.000001
F_{sc}	0.00659	P=0.12317	0.01421	P=0.13587
F_{ct}	0.03026	P<0.000001	0.06957	P<0.000001
Linguistic grouping(2)				
F_{st}	0.073	P<0.00001	0.15772	P<0.000001
F_{sc}	0.00965	P=0.01369	0.01433	P=0.08798
F_{ct}	0.06347	P<0.00001	0.14548	P<0.000001

Significant P value was calculated to be <0.005, <0.0083 and <0.05 for the three groupings respectively. The F_{st}, F_{sc}, F_{ct} are the fixation indices for within population, among populations within group, and among groups respectively.

Table 3.9 and Table 3.10 were results of AMOVA utilizing two different parameters. The first used F-Statistics and the later relied on the percentage of total genetic variations in experiments. The significant revelations from both are that: (i) Ethnicity, Geography and language all contributed significantly to genetic variations among the Nigerian populations; (ii) Language was the strongest factor among these three, affecting these genetic variations; and thirdly both ethnicity and geography have similar effects on genetic variation suggesting that the populations of Nigeria are ethnically structured along specific geographical regions.

Based on Wright's (1978) suggestion on interpreting Fst values, Y-SNP based Fst distance measure ranged from moderate to very great genetic variations while based on Y-STR based Rst distance measure, the Fst values ranged from little to moderate genetic variations. In both cases however, all the Fst values are significant ($P<0.000001$) with the values for linguistic grouping being double that of either ethnic or geographical groupings.

Based on the percentage of total variation in an hierarchical experiment also, even though the percentages of total variation were relatively small (between 3.03% and 6.4 % for Rst measure and between 6.96% and 14.6% for Fst distance measure) among groups for both marker (Y-STR and Y-SNP) systems, language grouping doubled those of ethnic and geographical groupings- a pattern observed for F-Statistics.

Another inference from these Tables is that based on the strong correlation of the two genetic markers (r=0.9623; p=0.00016534), it was either the present genetic structure has been retained from pre-historical times till present or that some more recent demographic events had completely obliterated the signatures from pre-historical times in the regions of modern Nigeria.

Table 3.10: The AMOVA percentage of variation with significant groupings

Source of variation	Percentage of variation	
	Y-STRs(R_{st})	Y-SNPs(F_{st})
Ethnic grouping (n=5)		
Among Groups	3.24	6.96
Among populations within groups	0.07	0.84
Within populations	96.05	92.2
Geographical grouping (n=4)		
Among Groups	3.03	6.96
Among populations within groups	0.64	1.32
Within populations	96.33	91.72
Linguistic grouping (n=2)		
Among Groups	6.4	14.55
Among populations within groups	0.9	1.22
Within populations	92.7	84.23

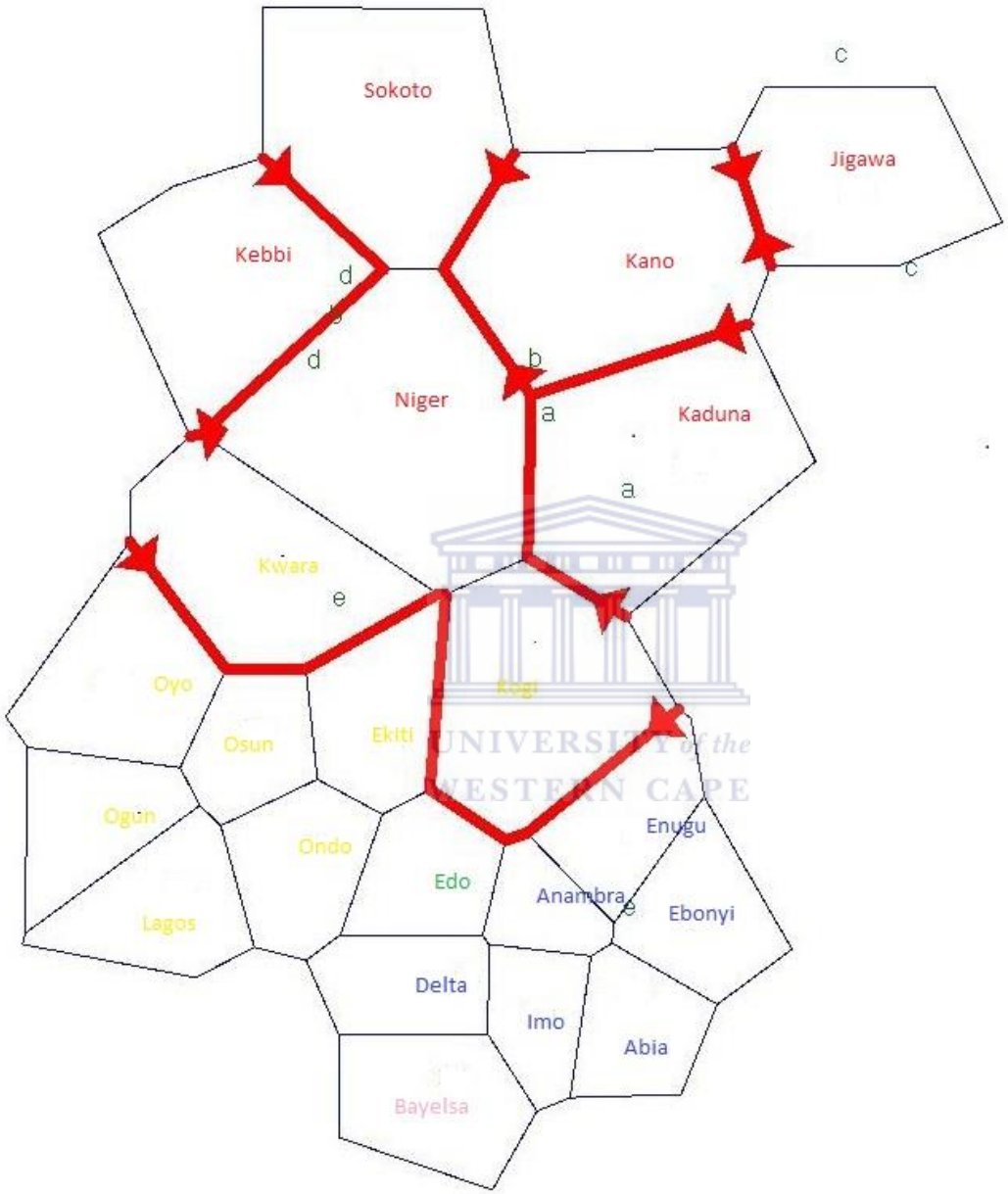


Figure 3.7a Plot of putative genetic barrier across the geographical locations of Nigerian population based on the Rst distances Y-STR data of the Nigerian populations. K was set at K=5 which revealed a Northern/Southern Nigerian populations divide. Hausa, Igbo, Yoruba, Bini and Ijaw populations were labeled with red, blue, yellow, green and pink colours respectively. Italics letters are the location of the genetic barrier while the bold red arrows are the barrier path.

Figure 3.7a above showed the putative spatial genetic barrier constructed to further test for geographical population sub-structure. This was done be performing Monmonier's analysis for detecting presence of genetic barrier given the spatial distribution of the populations using Y-STR-based Rst value. Based on this data, the fifth barrier revealed a clear north–south divide of the Nigerian populations with the Northern group incorporating the two Yoruba populations of Kogi and Kwara that are the transition s between the Northern Hausa and the Southern Nigerian populations analyzed in this report. The first 4 genetic barriers were found among the Hausa populations of northern Nigeria, corroborating earlier suggestions that the Hausa populations are not homogenous.

Figure 3.7b below showed the first seven putative genetic barriers in the Nigerian population on the background of the Y SNP haplogroups. These are (a) barrier between Jigawa Hausa (JG) and other Hausas in Northern Nigeria, (b) barrier between Anambra Igbo (AN) and Other Igbos in South eastern Nigeria, (c) barrier between Sokoto Hausa (SK), Kano Hausa (KN) and Jigawa Hausa (JG) and other Hausas population in northern Nigeria, (d) barrier between Kogi Yoruba (KO) in North Central Nigeria and their Northern, Western and eastern Neighbours, (e) barrier within the third barrier distinguishing Kano Hausa (KN) from both Sokoto Hausa (SK) and Jigawa Hausa (JG) in Northern Nigeria, (f) the barrier between the Oyo Yoruba (OY) and other Yoruba in South West Nigeria, and (g) the barrier distinguishing Ebonyi Igbo (EB) from other Igbo

populations in South East Nigeria. This suggested that there were differences within the three major ethnic populations of Nigeria- Hausa, Yoruba and Igbo that was deep rooted in their history. That these signals have persisted into the present despite the absence of major geographical and linguistic barrier was quite revealing.

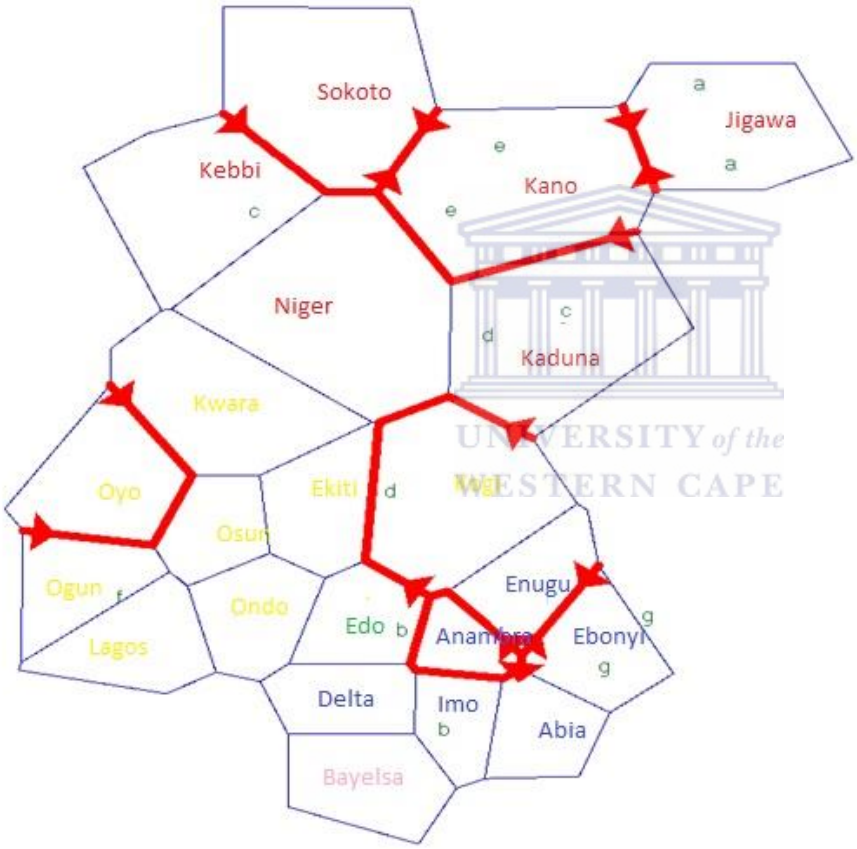


Figure 3.7b: Plot of putative genetic barrier across the geographical locations of Nigerian population based on the Fst distances Y-SNP data of the Nigerian populations. K was explored from 2 to 7. Hausa, Igbo, Yoruba, Bini and Ijaw populations were labeled with red, blue, yellow, green and pink colours respectively. Italics letters are the location of the genetic barrier while the bold red arrows are the barrier path.

However, the Barrier software, revealing the most likely areas to explore for genetic structure is rather speculative. This was because the level of significance of the presumed structure was not determined unlike the other software employed in this thesis. Its value is therefore useful in the preliminary evaluation of genetic and geographic data. More definitive assertions are made with more advanced software like Allele-In-Space (AIS), not used in this thesis.

Table 3.11: Mantel's test of correlation of genetic and geographic distances in the Nigerian populations (significant $p<0.0002$).

Marker	Measure of Distance	r	p value
11 STR	Rst	0.2034	0.005
SNP	Fst	0.4626	0.0001

Mantel's test was carried out to compare the genetic matrices from population pairwise Y-STR-based Rst analysis and population pairwise Y-SNP based Fst distance populations in order to test for correlation with the geographical distance. Significantly positive correlation ($r=0.4626$, $P=0.0001$) of the Y-SNP systems was observed (Table 3.11).

3.4 Network reconstructions

Network reconstruction was made for the six lineages (B, E (xE1b1a), E1b1a (xE1b1a7, E1b1a8), E1b1a7, E1b1a8 and R) represented by more than 10 individual per lineage (Figures 3.8a-f). The haplogroup E1b1b has been incorporated with haplogroup E (xE1b1a). Y-DNA network analysis of Y-STR haplotypes showing a non-star cluster was understood to indicate Y-STR variability due to multiple founding individuals, while that yielding a star cluster was regarded as a population descending from a single ancestor.

Haplogroup B (Figure 3.8a) showed three (3) branches, two (2) of which have only four (4) individuals each while the most extensive branch has more than 8 individuals. The predominant Haplogroup B sub-clade in central and western Africa has been reported to be Haplogroup B2b (Batini, *et al*., 2011). This was not evaluated in this thesis.

Haplogroup E (xE1b1a) network (Figure 3.8b) showed two distinct branches which were also observed to have two branches each. As the E1b1b populations were incorporated in this network, their contribution could be seen towards the center of the network.

Haplogroup E1b1a (xE1b1a7, xE1b1a8) network (Figure 3.8c). This haplogroup was known to possess at least nine branches but, having excluded both E1b1a7 and E1b1a8, seven were expected. The network however revealed two distinct branches apart from the reticulation. The inference from this network was that apart from the E1b1a7 and E1b1a8 haplogroups, Nigerian populations has at least two more sub-clades of E1b1a that could be investigated in future work.

Haplogroup E1b1a7 network (Figure 3.8d) showed high levels of reticulations with no clear structure. This was the most common haplogroup in the Nigerian populations. The network was consistent with what would be expected from lineages with relatively short evolutionary history, associated with 'recent' demographic expansions. Veeramah and others (Veeramah, *et al.,* 2010) observed a similar pattern among populations of Cross rivers region of Nigeria on the background of this lineage. Seven subclades have been detected in this haplogroup (Karafet, *et al.,* 2008) but the most commonly found among the Yoruba reported was E1b1a7a (Barbieri, *et al.,* 2012). Montano and other workers (Montano, *et al.,* 2011) working with some Tiv, Igala and Idoma populations from central Nigeria also revealed some E1b1a7a1, E1b1a7a2 and E1b1a7a3 in those populations.

Haplogroup E1b1a8 network (Figure 3.8e) also showed many reticulations but no clear structure. It has been reported to have at least six sub-clades (Karafet, *et al.,* 2008) which could not be inferred from the network. Montano and other workers (Montano, *et al.,* 2011) reportedly observed E1b1a8a and E1b1a8a1 among some Tiv, Igala and Idoma populations from Central Nigeria.

Haplogroup R network (Figure 3.8f) revealed two branches. The network was made up of mostly singleton nodes from the Hausa population. The most common R haplogroup reportedly found in West/Central Africa was the R1b2 variant on the V88 SNP (van Oven, *et al.,* 2014; Hassan, *et al.,* 2008) which was not evaluated in this thesis.

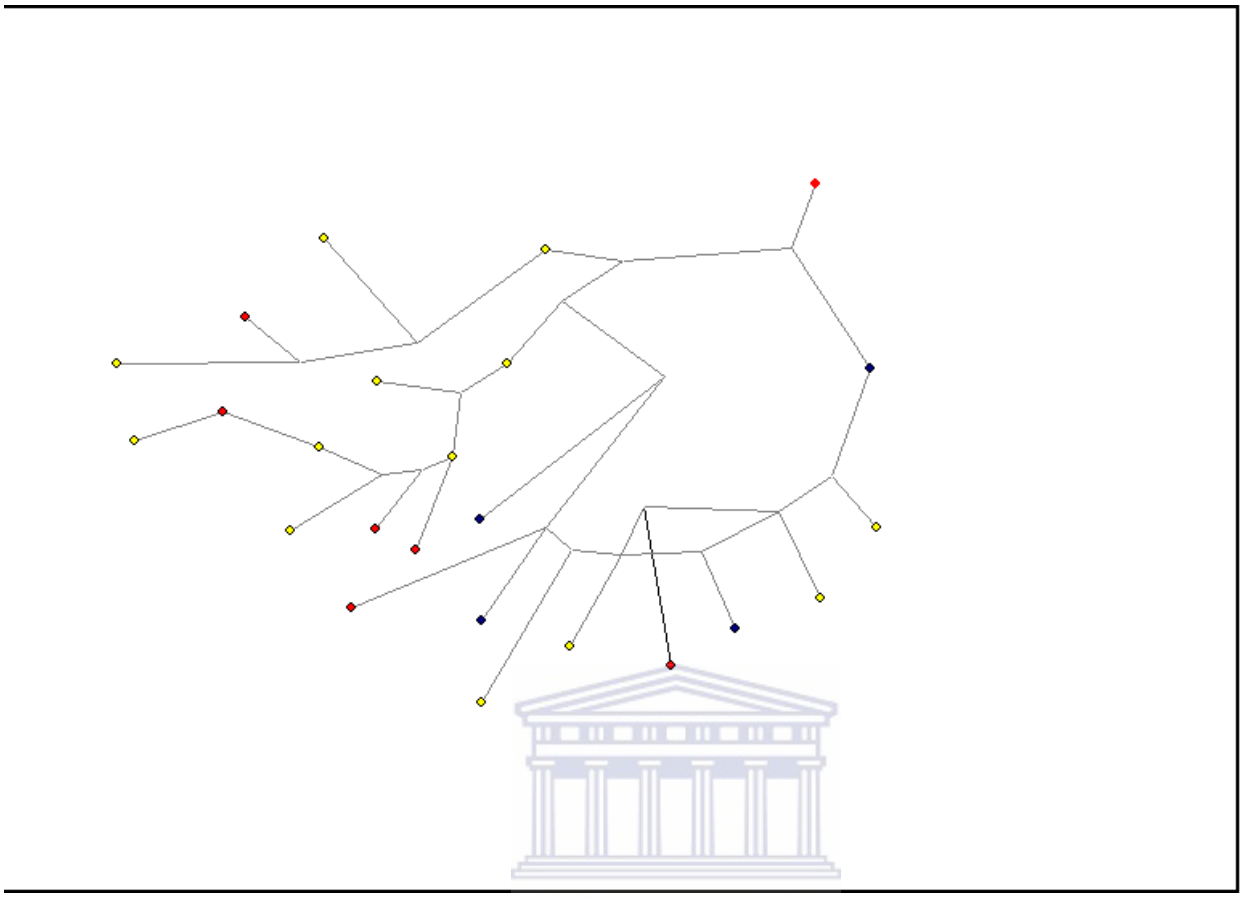


Figure 3.8a: Network reconstruction of the M60 mutation for haplogroup B in the Nigerian populations based on the SWGDAM Y-STR loci. The Hausa populations were represented with red, Igbos with blue and Yoruba with Yellow. The circle size is equivalent to the number of haplotypes.

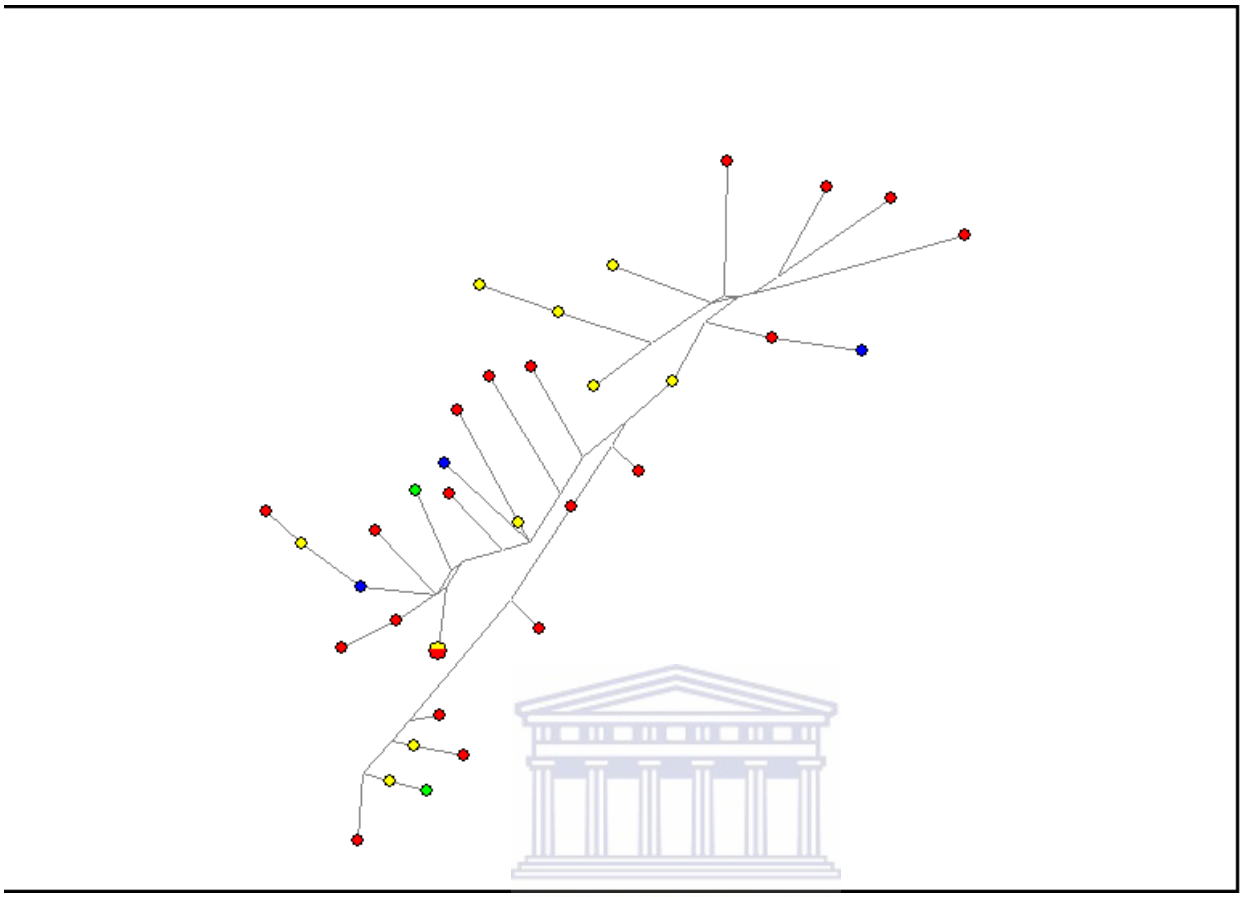


Figure 3.8b: Network reconstruction of the SRY_{4064} mutation for haplogroup E (xE1b1a) in the Nigerian populations based on the SWGDAM Y-STR loci. The Hausa populations were represented with red, Igbo with blue, Yoruba with yellow and Bini with lemon. The circle size is equivalent to the number of haplotypes

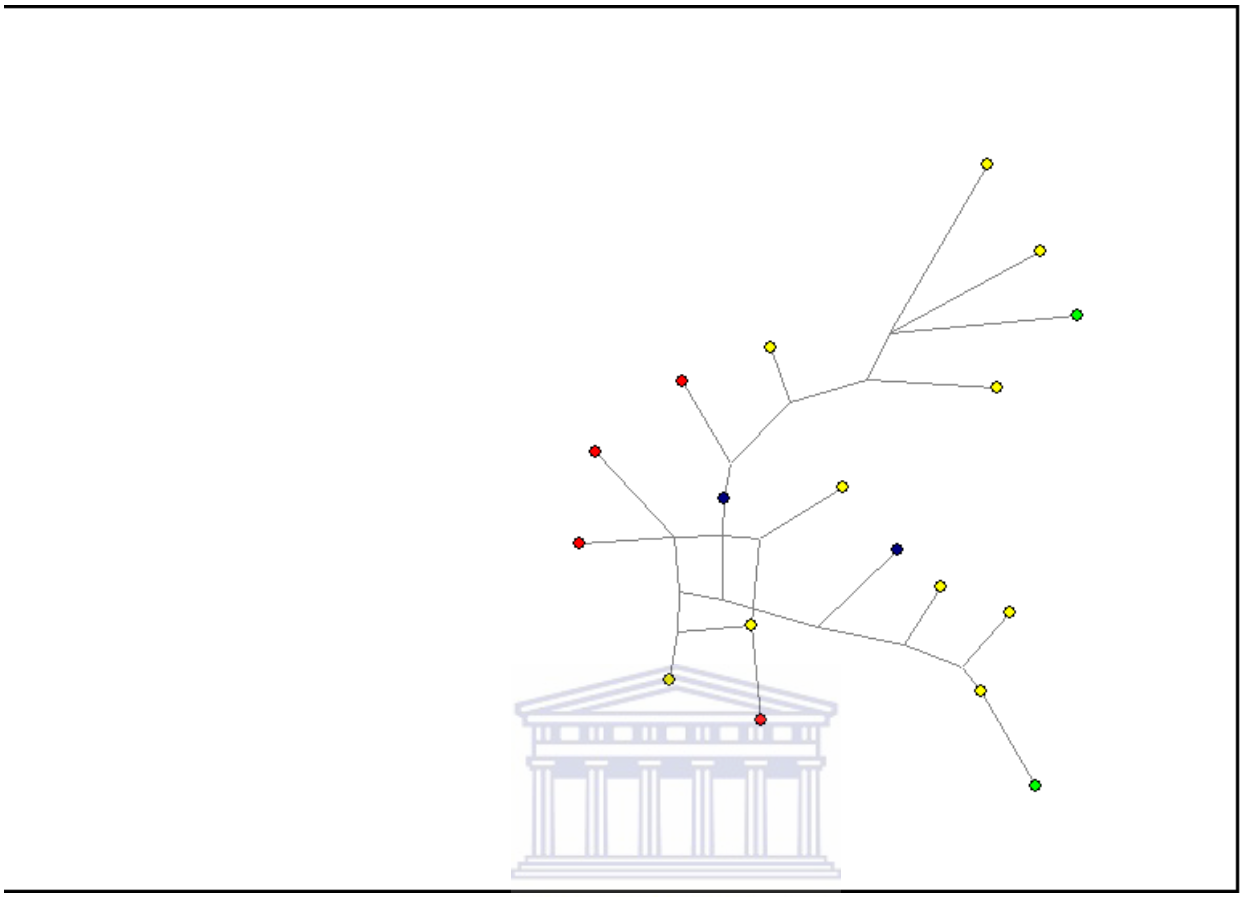


Figure 3.8c: Network reconstruction of the M2 mutation for haplogroup E1b1a (xE1b1a7, xE1b1a8) in the Nigerian populations based on the SWGDAM Y-STR loci. The Hausa were represented with red, Igbos with blue, Yoruba with Yellow and Bini with Lemon. The circle size is equivalent to the number of haplotypes.

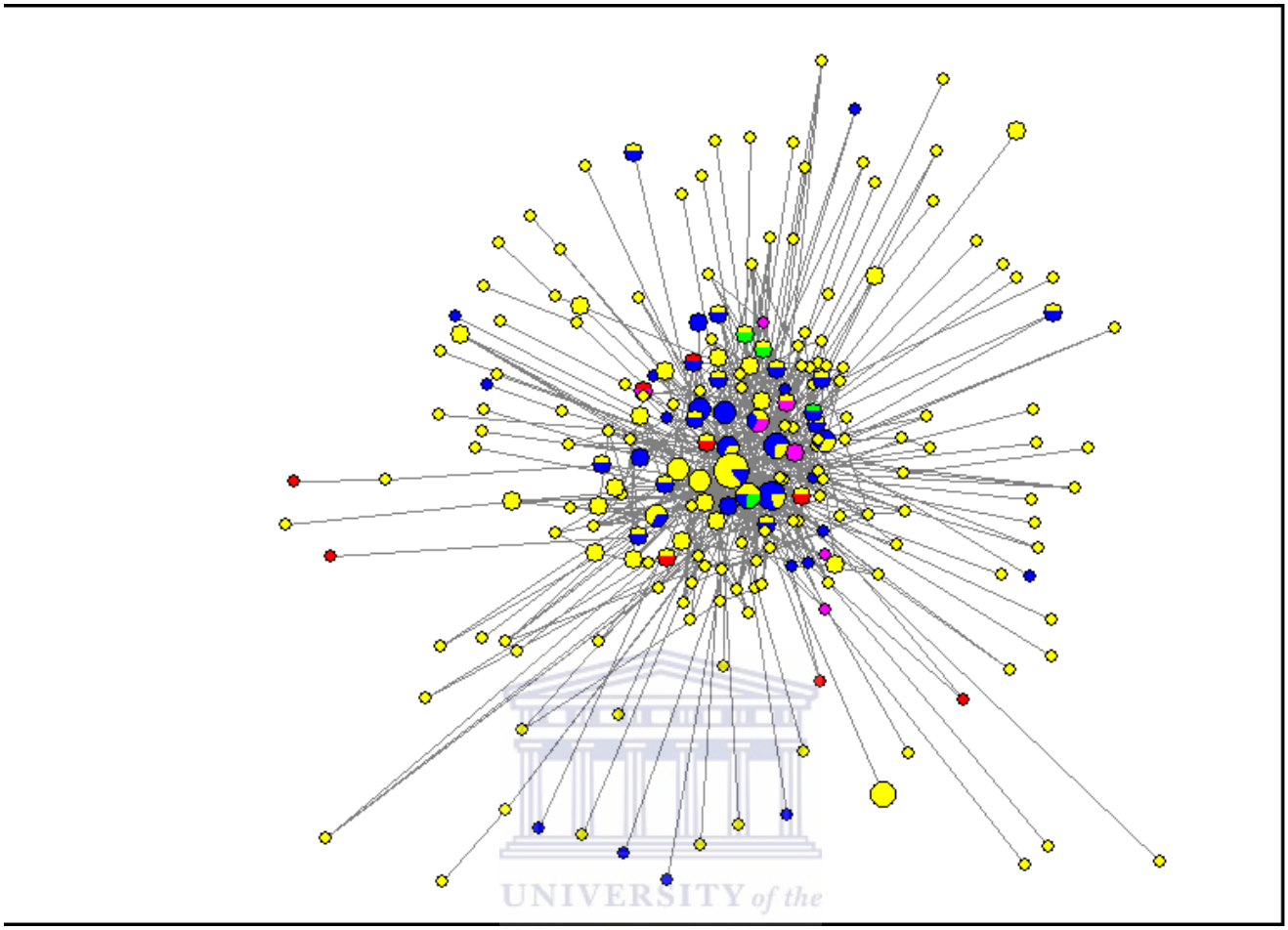


Figure 3.8d: Network reconstruction of the U186 mutation for haplogroups E1b1a7 in the Nigerian populations based on the SWGDAM Y-STR loci. The Hausa populations were represented with red, Igbo with blue, Yoruba with Yellow, Bini with Lemon and Ijaw with purple. The circle size is equivalent to the number of haplotypes.

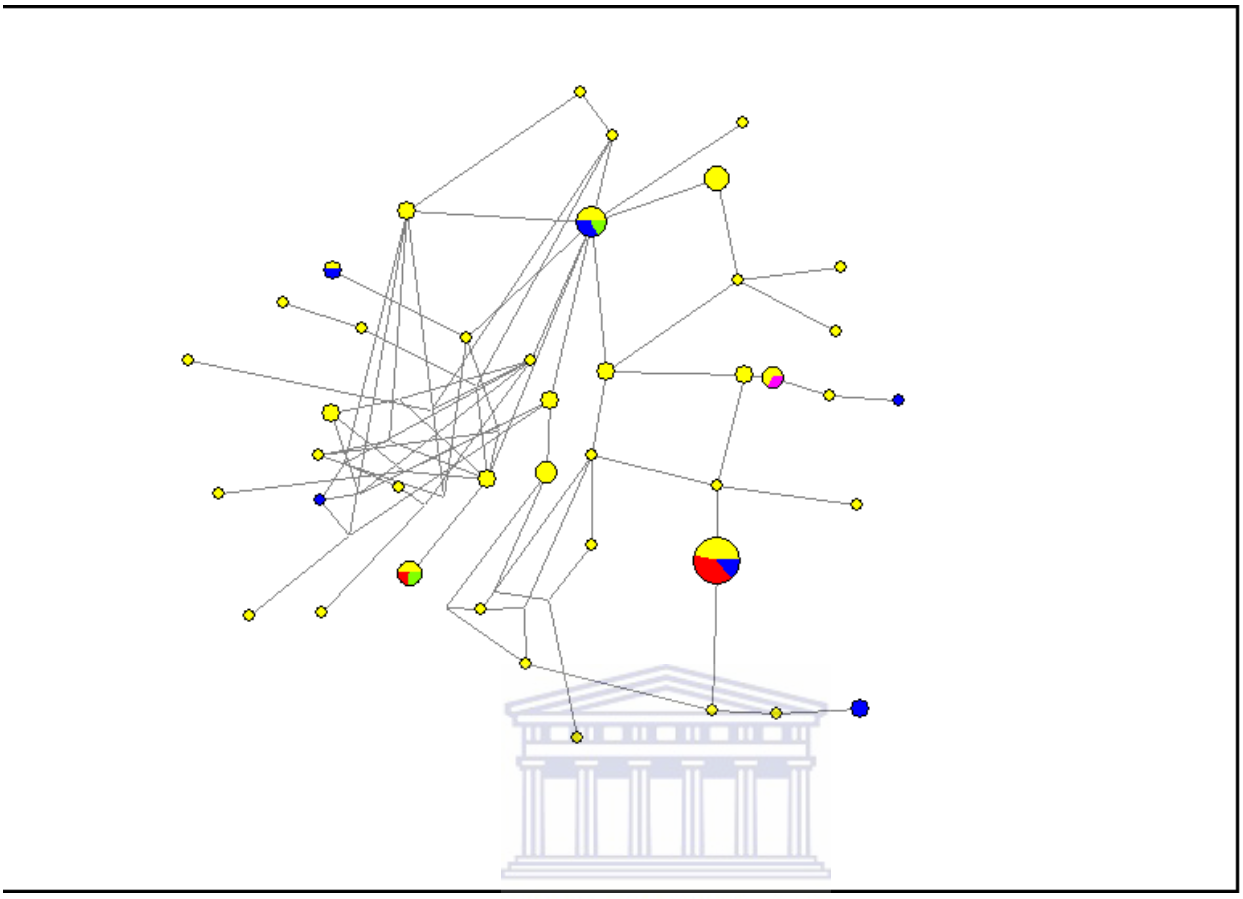


Figure 3.8e: Network reconstruction of the U175 mutation for haplogroup E1b1a8 in the Nigerian populations based on the SWGDAM Y-STR loci. The Hausa populations were represented with red, Igbo with blue, Yoruba with Yellow, Bini with lemon and Ijaw with purple. The circle size is equivalent to the number of haplotypes

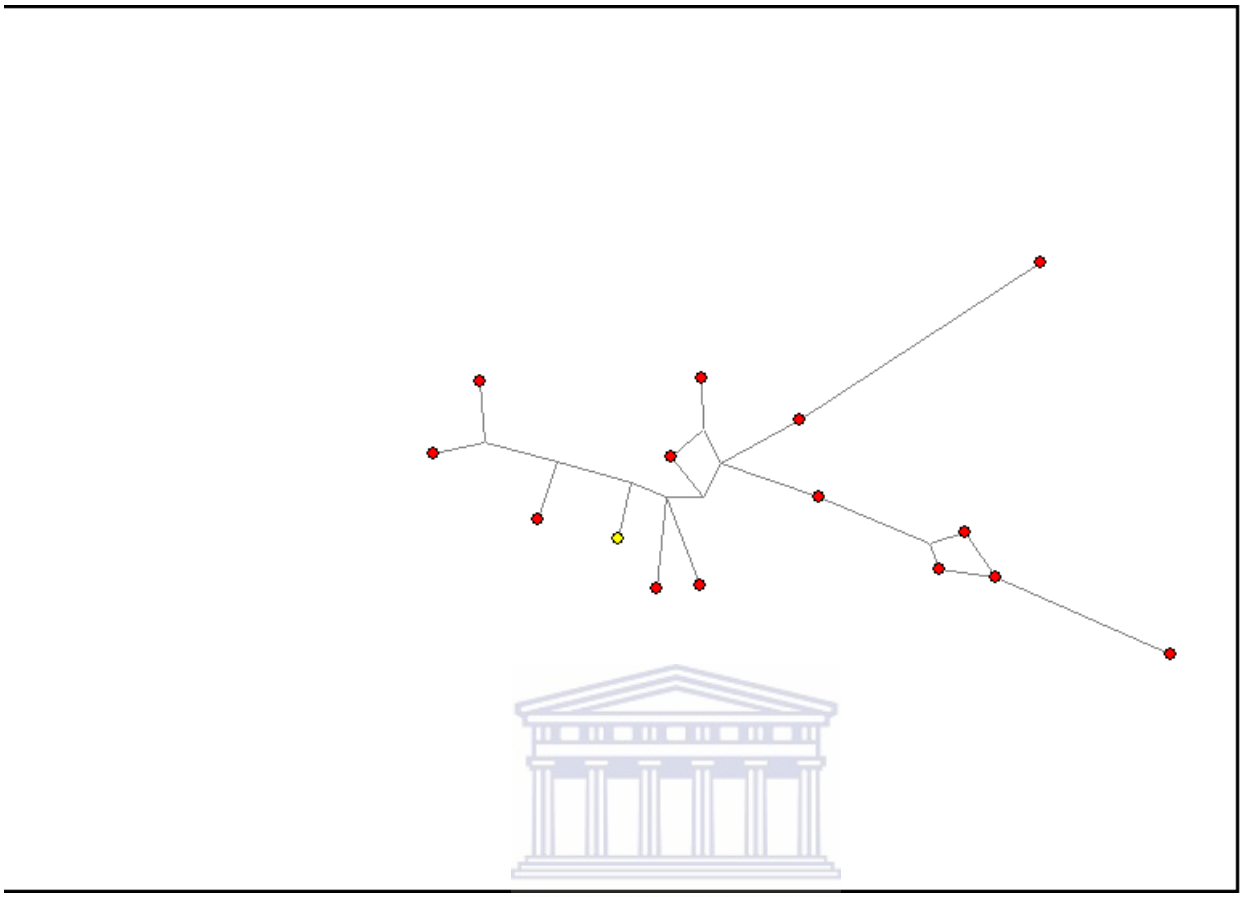


Figure 3.8f: Network reconstruction of the M207 mutation for haplogroup R in the Nigerian populations based on the SWGDAM Y-STR loci. The Hausa populations were represented with red and Yoruba with Yellow. The circle size is equivalent to the number of haplotypes.

3.5 Spatial Autocorrelation Plots

The three most common haplogroups found among the Nigerian populations were B, E1b1a8 and E1b1a7 in that increasing order. These were explored to determine whether any or all of these had clinal distribution in the Nigerian populations. The results shown in Figures 3.9a, 3.9b and 3.9c for Haplogroups B, E1b1a7 and E1b1a8 respectively showed that the spatial patterns deviated from pattern expected from haplogroup that showed clinal distribution from the literature (Kayser, *et al.*, 2005; Barbujani, 2000).

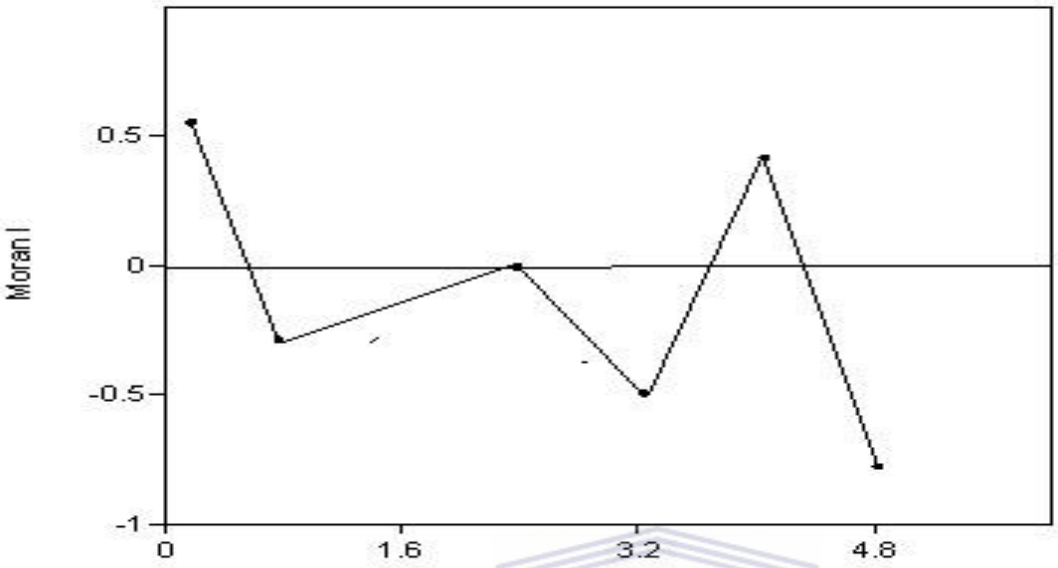


Figure 3.9a: The spatial autocorrelations for Haplogroup B relative frequencies among the Nigerian populations. The Y axis was the Moran Index, I while the X axis was the distance classes computed from the coordinates of the longitude and latitude of the geographical sampling locations.

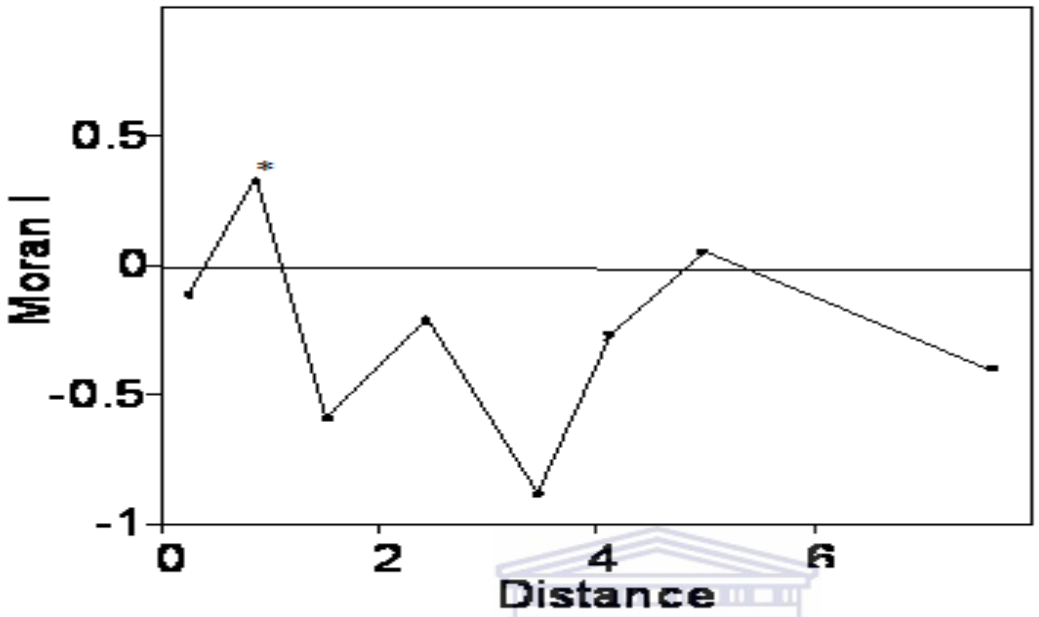


Figure 3.9b: The spatial autocorrelations for Haplogroup E1b1a7 relative frequencies among the Nigerian populations. The Y axis was the Moran Index, I while the X axis was the distance classes computed from the coordinates of the longitude and latitude of the geographical sampling locations.

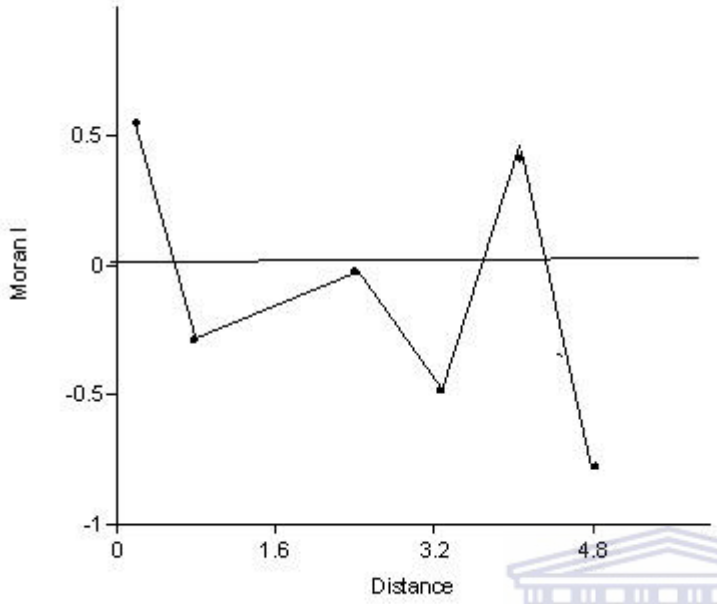


Figure 3.9c: The spatial autocorrelations for Haplogroup E1b1a8 relative frequencies among the Nigerian populations. The Y axis was the Moran Index, I while the X axis was the distance classes computed from the coordinates of the longitude and latitude of the geographical sampling locations.

Table 3.12 F_{ST} and N_V estimates of the five ethnic populations of Nigeria.

Population	F_{ST}	N_v
Hausa	0.02468	40
Igbo	0.03122	31
Yoruba	0.03001	32
Bini	0.03097	31
Ijaw	0.031	31

F_{ST} is a measure of inter-population variability, whereas N_v is the effective number of migrants.

Table 3.12 revealed that the Afro-asiatic Hausa population of Northern Nigeria has relatively more effective migrants than the four Niger-Congo speaking populations of southern Nigeria.

Table 3.13: Ancestral haplotypes of the Haplogroups in the Nigerian lineages*

Haplogroup	**Haplotype**
B	15, 13, 30, 21, 10, 11, 13, 11, 12, 10, 12
E(xE1b1a,E1b1b)	17, 13, 30, 22, 10, 11, 13, 14, 16, 11, 12
E1b1a(xE1b1a7, E1b1a8)	15, 13, 30, 21, 10, 11, 13, 15, 17, 11, 12
E1b1a7	17, 13, 30, 21, 10, 11, 14, 17, 18, 11, 12
E1b1a8	15, 13, 31, 21, 10, 11, 13, 16, 17, 11, 11
E1b1b	13, 14, 31, 24, 10, 11, 13, 17, 18, 11, 11
R	15, 14, 30, 24, 10, 13, 13, 13, 13, 11, 12

*Haplotypes are in the order of DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS 385a, DYS385b, DYS438 and DYS439 respectively.

Only haplogroup E1b1a7 still had the ancestral haplotype represented in two individuals, the rest of the haplogroups do not currently retain these haplotypes.

Table 3.14: Y-STR mutation rates as obtained from the YHRD fact sheets.

Y-STR	Mutation rates (95% confidence interval)
DYS19	0.0022 (0.0016-0.0032)
DYS389I	0.0029(0.0018-0.0035)
DYS389II	0.0041(0.0032-0.0053)
DYS390	0.0021(0.0013-0.0034)
DYS391	0.0025(0.0018-0.0036)
DYS392	0.0005(0.0002-0.0009)
DYS393	0.0011(0.0006-0.0018)
DYS385	0.0024(0.0017-0.0034)
DYS438	0.0004(0.0001-0.0009)
DYS439	0.0054(0.0039-0.0069)

Average mutation rate μ was calculated to be 0.0024 (0.0016-0.0031)

Table 3.14 revealed that DYS392 and DYS438 have very low mutation rates while DYS439 has fastest rate but the remaining are moderate according to the scale reported by Bird (Bird, 2012) that may have a significant effect on the average rate computed here and the overall application in TMRCA computation.

Table 3.15: Intra-lineage Time to the Most Recent Common Ancestors (TMRCA) TMRCA was calculated from (ASD/μ) multiplied by 31 (years ago).

Haplogroup	Generations ago	TMRCA* (Years ago)
All B	808	25,048
All E	469	14,539
All E1b1a	328	10,168
All E1b1b	287	8,897
All R	788	24,428

*Values represented Mean.

Table 3.16: Inter-Haplogroups E1b1a7 with E1b1a8. TMRCA was calculated from (ASD/2μ) multiplied by 31 (years ago). μ was computed to be 0.0024 from Table 3.14 above.

Inter-Haplogroups	Generations ago	TMRCA* (Years ago)
E1b1a7 versus E1b1a8	260	8,060

*Values represent Mean

3.6 Inter-lineages and intra-lineage time to the most recent common ancestor (TMRCA)

The general pattern of intra-lineage TMRCA (Table 3.15) above revealed that haplogroup B was older than haplogroup E which was also significantly older than its sub-lineages E1b1a and E1b1b in the Nigerian population. It also revealed that the TMRCA for haplogroup R was comparable with that from the literature (Karafet, *et al.*, 2008). It confirmed that E1b1a7 and E1b1a8 are relatively younger. However, considering the ages of E1b1a7 and E1b1a8, it gives room for the possibilities of them having sub-lineages of their own, which have been confirmed elsewhere. E1b1a7 was now known to have at least seven sub-lineages while E1b1a8 had at least six sub-lineages (Karafet, *et al.*, 2008; Montano, *et al.*, 2011 and Trombetta, *et al.*, 2011) none of which were further evaluated in this thesis.

Because the other lineages were not fully represented in all the Nigeria populations, inter lineage TMRCA splits were calculated for only E1b1a7 with E1b1a8 (Table 3.16) because they were the most common in the Nigerian populations.

On the basis of E1ba7 lineage, the network of E1b1a7 (Figure 3.8d) showed no structure thus providing support that haplogroup E1b1a7 was in rapid expansion, in no particular pattern.

Overall, this inter-haplogroups TMRCA might be suggesting that the pre-historical demography of these major populations were significantly influenced by the lineages B and E1b1a7 to a greater and perhaps E1b1a8 to a lesser extent.

4.0 CHAPTER FOUR: DISCUSSION

4.1 Was there any structure in the genetic variation on the non-recombining region of the Y-Chromosome (NRY) of the major Nigerian populations?

Based on the plethora of evidence presented earlier in this thesis, the unequivocal response in answer to this question was yes.

Looking at Tables 3.1, 3.2a, 3.2b with Supplementary Tables 1 and 2, which showed the mean pairwise differences of STR haplotype and haplogroup diversities, R_{st} and F_{st} distances at the level of the five ethnic populations with the R_{st} and F_{st} distances at the levels of the populations across their geographical locations respectively, the obvious structure between the Hausa populations and the other four populations was very clear. These tables were supported with the graphics shown through the MDS plots at the five ethnic groups' level, at the geographical origins level for both R_{st} and F_{st} distances (Figures 3.2a, 3.2b and 3.3a, 3.3b respectively. The correspondence plots in Figures 3.5a and 3.5b at the five ethnic and 22 geographical origins level respectively also confirmed this. Finally, the Genetic barriers for both R_{st} and F_{st} distances also confirmed that there were structures between the Hausa population and the other four southern Nigeria populations. Considering the distinct geographical locations and linguistic groups that are represented, this revealed both a North/South geographical and Afro-asiatic and Niger-Congo linguistic divide.

A closer look at these results seemed to corroborate the findings of Veeramah and others that there was very little genetic differentiation among people of the same or proximate geographical locations despite their language differences (Veeramah, *et al.*, 2010). Their inferred language differences perhaps was not across major linguistic groups as shown between Afro-asiatic and Niger Congo here but within the same broad classifications. This thesis too, could not differentiate among the four Niger Congo populations of Southern Nigeria namely Igbo, Yoruba, Bini and Ijaw at the very low level of significance threshold ($P<0.0005$). Even the seemingly within individual ethnic population obvious from the barrier plots and some of the MDS plots across geographical locations mostly disappear at this threshold. There was however significant within population genetic differentiation within only the Hausa populations based on F_{st} distances (Table 3.2b; Supplementary Table 2).

4.2 What factors drove the structure in the genetic variation on the non-recombining region of the Y-Chromosome (NRY) of the major Nigerian populations?

To answer this question, hierarchical AMOVA experiments were carried out along the five ethnic populations, 22 geographical origins and two linguistic differences suggested by the observations above. The AMOVA results revealed that most F_{st} differences were noticed among groups and that three factors contributed to the genetic differentiations among the Nigerian populations. A significant finding was that while both ethnicity and

geography had comparable impacts, the linguistic effect was more than double each of these (Table 3.9 and Table 3.10).

Campbell and Tishkoff had suggested that in addition to the three factors observed in this thesis, climatic and ecological factors also affect the genetic differentiations among sub-Saharan African populations (Campbell and Tishkoff, 2008). However, Veeramah and others (Veeramah, *et al.*, 2010) did not observe a similar trend to this thesis while working on populations from Cross Rivers region in Southern Nigeria. Apparently the different geographical span covered by our two different results could be the reasons for this difference in conclusions, as their contrasted populations were geographically closer than the ones evaluated in this thesis. Unfortunately these two reports, (theirs and this thesis) are the only reports to date on extensive genetic analysis of Nigerian populations Y-STRs and Y-SNPs. Montano and others (Montano, *et al.*, 2011) who also worked with some other Nigerian populations, did not utilize any of the populations in these other reports (Veeramah and this thesis) but focused on the Tiv, Igala and Idoma populations from Central Nigeria with the intent to resolve the guiding hypothesis of the Bantu expansion.

The fact that these factors (ethnicity, geography and language) had the same magnitude in the different genetic markers, Y-STR and Y-SNP which measured different time depth of mutations, also suggested the concordance of pre-historical genetic structure with modern time's observations. This, though consistent with other results in this thesis was however

surprising considering the massive impacts of past historical events and causes of human migrations within the Nigerian area in the last two millennial, especially the impact of the Trans-Sahara slave trade.

Although some of the myths of origins of the present Nigerian populations seem exaggerated (Yoruba from Mecca, Hausa from Baghdad, Igbo from Israel), strong links among peoples around the Niger-Benue confluence have been observed (Ehgosa, 2002). The links in origin among Borgu (in present day Republic of Benin and Togo), Oyo and Nupe; Yoruba and Edo (Bini was used in the discussion of this thesis to distinguish their geographical location from the people, as the two are interchangeable); Edo, Igala and Nupe; Edo, Onitsha, Igala and Nri; Jukun, Idoma and Igala are quite established. Modern day Nigeria region is a compact geographical zone. The close link among the peoples was apparently due to the basin of the lower Niger with its tributaries, compactness and most importantly, the absence of any major barrier within the geographical space, encouraging more migration and interaction within rather than with people outside (Falola and Heaton, 2008). The exact number of ethnic groups in the country was not known (Ehgosa, 2002; Blench, 2011). This was principally because different criteria have been used to estimate the number of groups. Classification based on languages can be problematic because most languages have dialects and sub-languages that may ambiguously be regarded as ethnic group. Among the Yoruba populations in South western Nigeria, they have dialects like *Ijebu, Ekiti, Ikale, Owo, Oyo* and so on while

among the Igbo populations of South eastern Nigeria, they have dialects like *Onitsha, Ika, Erei, Effium, Izi, Ikwo, Mtezi, Okpoto* and other dialects (Blench, 2011). Another problem was because many ethnic groups are spread across and beyond certain regional boundaries. The Yoruba populations spread across South-west, South-south and North-central Nigeria. The Igbo populations spread across South-east, South-south and North-central Nigeria too. The Ijaw populations are spread around the coastal swamps of the South-south and South-west. The Hausa populations spread over the vast expanse of the entire Northern zones. These four (4) ethnic populations are the largest homogenous ethnic groups although the population censors do not distinguish the populations along this line but simply by places of residence (Falola and Heaton, 2008). Another problem with classification was that some names by which some groups which are essentially the same are known changed from location to another, especially among clusters of minority groups in Benue, Kogi, Kwara and Nasarawa States like the *Eggon, Alago and Etulo* which claim their origins from *Igala and Ebirra* and speak dialects of these languages (Blench, 2011). Several ethnic groups are linked by myth of origin and shared common cultural practices. To resolve these conceptual problems, the use of combined criteria of language, name, core territory, culture and myth of origin rather than a single criterion like language was employed. Estimates of ethnic groups based mainly on languages spoken have ranged from 248 (Coleman, 1958) to 374 (Otite, 1990) and 550-610 (Wente-Lukas and Jones, 1985; Blench, 2011). The Hausa and Fulani are found all-over Northern

States (19 States of the entire 36 in Nigeria). Kanuri populations are found in Borno, Yobe, Kano, Niger, Nasarawa and Adamawa States of Nigeria; *Tiv* populations are found in Benue, Taraba and Nasarawa States; *Jukun* populations are in Benue, Nasarawa, Taraba and Bauchi States; *Bassa* populations are found in Niger, Benue, Kaduna, Kwara, Nasarawa States and the FTC Abuja; *Angas* populations are found in Plateau, Bauchi and Kano; *Gwari* populations are found in the FTC, Niger, Kaduna and Nasarawa States; *Uncinda* populations are in Niger, Kaduna and Sokoto States; *Kambari* populations are found in Niger, Kwara and Sokoto States; Yoruba in the six (6) South-west States and Kwara and Kogi in North-central; Igbo are spread across the five (5) South-east states, Delta, Cross rivers and Rivers States in the South-south and Benue State in the North-central region; Ijaw populations are in Bayelsa, Rivers, Edo and Delta in the South-south and Ondo in the South-west; *Mbembe* populations are found in Cross Rivers and Anambra States. The highest cluster of ethnic groups with the greatest heterogeneity was found in the Gongola cluster of Adamawa and Taraba states with 113 ethnic groups (Otite, 1990). Other states with enormous heterogeneity are Kaduna, Bauchi, Benue, Borno, Cross River, Kwara, Niger, Plateau and Nasarawa. By contrast, most homogenous states include the Igbo-speaking states of Anambra, Abia and Imo; Oyo and Osun which are Yoruba-speaking and Katsina state where the dominant groups are Hausa and Fulani. Most towns and villages are however ethnically mixed, especially Lagos, Kano, Port Harcourt, Onitsha, Abuja, Jos, Kaduna, Ibadan and Benin City (Ehgosa, 2002). It was

known that the major languages spoken in Nigeria represent three major families of African languages. The majority are Niger-Congo languages such as Yoruba, Igbo and Ijaw. The Hausa language is Afro-Asiatic and the Kanuri, spoken in the North-east, primarily Borno state, is a member of the Nilo-Saharan family (Blench, 2011).

The boundaries of present-day Nigeria were created by the British colonial administration in the late nineteenth and early twentieth centuries (Elochukwu, 1997: Hodgkins, 1984; Falola and Heaton, 2008). While political boundaries often coincide with physical boundaries, such as bodies of water or mountain ranges, or established by mutual agreement between societies over generations, the boundaries adopted to create modern Nigeria never had any geophysical boundary or physical significance to the indigenous peoples of the region. The only geophysical boundary of Nigeria is the Atlantic Ocean, which forms the southernmost border of the country. Her eastern, western and northern borders are all relatively arbitrary, having been negotiated at drafting tables in Europe rather than through local processes of societal development (Falola and Heaton, 2008). The Country hence was a conglomerate of hundreds of ethnic groups, many of which straddle these arbitrary borders. The Nigerian people of today thus have many different indigenous languages, historical memories, traditional lifestyles and social frameworks with roots reaching into the distant past.

4.3 Are there correlations between genetic distances and geographical distances in the Nigerian populations?

The Mantle's test (Table 3.11) was used to resolve this question conclusively. Based on the R_{st} distance matrix for Y-STR and F_{st} distance matrix for Y-SNP there were strong correlations of these with the geographical distances in the Nigerian populations. Even though the linguistic distances were not constructed for the Nigerian populations, the distinct manner that the Nigerian populations settled (Afro- Asiatic populations in the Northern part and all Niger-Congo populations in the Southern part and even surprisingly the Nilo-Saharans populations in the North East across the Lake Chad region, this would not have been surprising. The Afro-asiatic language-speaking Hausa populations provided all the distinctions by their geographical origin and language from the others to make notable difference.

Again, when this was contrasted with the report of Veeramah and other workers (Veeramah, *et al.*, 2010), they got no correlation at all the distances they contrasted except when including the wide geographically distant Ghanaian and Cameroonian populations that they also analyzed.

The most in-depth genetic studies on the populations of any part of Nigeria were the fine-scaled dense sampling of the peoples of the Cross rivers area (Veeramah, *et al.*, 2010). Using six Y-STRs and some Y-SNP and the Hyper Variable Region1 (HVRI) of the

mitochondrial DNA (mtDNA) of more than 1000 male subjects, they concluded that there was little genetic differentiation among the populations despite significant language difference. They also extoled the merits of fine scaled sampling over a region rather than the broader geographical distances utilized in most literature. Other studies included the 12 Yoruba males (Barbieri, *et al.*, 2012; Shi, *et al.*, 2010; de Filippo, *et al.*, 2011, Hausa (Tishkoff, *et al.*, 2009), Igbo (Adeyemo, *et al.*, 2005), Tiv (Montano, *et al.*, 2011), Idoma, Fulani and a few others that are not purely resident in Nigeria (Alloco, *et al.*, 2007; Disotell, 2000; Salas, *et al.*, 2002, Hassan, *et al.*, 2008; Zictkiewiicz, *et al.*, 1998, Deka, *et al.*, 1999; Shi, *et al.*, 2010; Woods, *et al.*, 2005; Montano, *et al.*, 2011). The literatures revealed a common denominator, that these genetic markers form a significant part of the heterogeneous Africa genetic landscape (Tishkoff, *et al.*, 2009). They established that the Fulani population exhibited low levels of European/ Middle Eastern ancestry, consistent with possible gene flow from those regions. Nilo-Saharan speaking populations from Nigeria (the Kanuri population) also shared a genetic cluster with peoples of Southern Sudan. A similar cluster with Bantu-speakers of West-Central Africa and East Africa was observed among individuals from Nigeria who speak non-bantu Niger-Kordofanian languages (the Yoruba, Igbo and Ijaw populations) (Tishkoff, *et al.*, 2009).

Generally however in a broader West African or sub-Saharan African context, there was a clear genetic separation of Nigerian and sub-Saharan African populations from North African populations. The latter more closely resembling Middle Eastern and Eurasian

populations in almost all mtDNA, NRY and autosomal studies (Poloni, *et al.*, (1997); Scozzari, *et al.*, (1999); Luis, *et al.*, (2004); Cruciani, *et al.*, (2002); Salas, *et al.*, (2002); Terreros, *et al.*, (2005)), demonstrating the major genetic barrier that the Sahara Desert has been through much of modern man's occupation of the African continent. However there was evidence of contact in both directions involving both male and female mediated gene flow in populations lying close to the boundaries of the Sahara: the Chad Basin, Guinea Bissau and Algeria (Salas, *et al.*, (2002), Coia, *et al.*, (2005), Rosa, *et al.*, (2004), Rosa, *et al.*, (2007), Cerny, *et al.*, (2007), Flores, *et al.*, (2001), Richards, *et al.*, (2003)), with the expansions of Berbers, migrations of Fulani and the Arab slave trade possibly being major influences.

Both NRY and mtDNA lineages appear to have been spread through much of sub-Saharan Africa as a result of Bantu expansion though the patterns observed for men and women are quite distinct. The majority of men in Bantu-speaking populations possess the NRY defined haplogroup E1b1a (Underhill, *et al.*, 2001; Cruciani, *et al.*, 2002) and a particular haplotype on this E1b1a background defined by six microsatellites (15-12-21-10-11-13 for DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393 respectively) was the modal type in numerous Bantu-speaking populations, stretching all the way from Cameroon and Nigeria to Southern Africa (Thomas, *et al.*, 2000; Pereira, *et al.*, 2002; Berniel-Lee, *et al.*, 2006). Given its predominant presence in Bantu-speaking populations, its relatively low within-haplogroup diversity (Scozzari, *et al.*, 1999) and an

estimated time for the most recent common ancestor for South African Bantu possessing E1b1a chromosomes of 3000-5000 years before present (Thomas, *et al.,* 2000) this distribution was best interpreted as a signature of Bantu-speaking males expanding across sub-Saharan Africa.

NRY data from Senegal, Guinea Bissau, Gambia and Ghana showed a high frequency of haplogroup E1b1a (Semino, *et al.,* 2002; Wood, *et al.,* 2005; Rosa, *et al.,* 2007), which was interesting given that the haplogroup was a putative signature of the Bantu expansion (Underhill, *et al.,* 2001), suggesting it has an older and geographically more widespread significance, possibly being a signature of the original proto Niger-Congo speakers. In addition or alternatively, the low NRY haplogroup diversity in West Africa may be a product of agricultural expansion throughout the region or another part of the same expansion that included that of the Bantu-speaking peoples.

The Chad Basin presented a very heterogeneous genetic profile that differed significantly between male- and female-specific lineages, consistent with the complex population movements the region had experienced. The NRY, mostly assessed by datasets from northern Cameroon, showed a substantial proportion of R1*-M173 types (Scozzari, *et al.*,1999; Cruciani, *et al.,* 2002; van Oven, *et al.,* 2014), a clade not usually found elsewhere in sub-Saharan Africa but present in Asia (Luis, *et al.,* 2004). This has been presented as evidence for a possible back migration from Asia to sub-Saharan African through the Levantine corridor. However, mtDNA showed no such signal, suggesting that

admixture of the immigrating group was primarily male-mediated; at least once they reached their destination (Coia, *et al.*, 2005). The mtDNA data are still very heterogeneous with many different types showing a mostly Central African connection but with possible gene flow from East Africa and from West Africa (Cerny, *et al.*, 2004; Cerny, *et al.*, 2007) as well as a small North African influence (Coia, *et al.*, 2005), demonstrating that the Sahel along which the Chad Basin lies, has been a major corridor for human migration in Africa.

Montano and co-workers explored previously unsampled populations in Central Nigeria with a view to reviewing some aspects of the Bantu expansion problems (Montano, *et al.*, 2011). They discovered distinct genetic substructures on the backgrounds of Haplogroups E1b1a7 and E1b1a8 when they explored the different sub-haplogroups of these lineages between Bantu and Non Bantu speakers and among different Bantu populations of Central Africa. These, with the published mtDNA data for the Yoruba populations by Barbieri and co- workers, studies on Burkina Faso populations (Barbieri, *et al.*, 2012) are good attempts to fill the dearth of genetic data on Nigeria and contribute to the on-going discussions on the resolution and significance of demographic events in sub-Saharan Africa.

4.4 Was inter-haplogroup STR profile sharing a common or rare occurrence in this Nigerian dataset?

This question arose from the observations from the Southern Nigerian dataset set of Veeramah and other workers (Veeramah, *et al.,* 2010) that there were significant inter haplogroup STR profile sharing among the populations of the Cross Rivers region of Nigeria. This perhaps may result in influencing their conclusions of little genetic differentiation among those populations. However, based on the SWGDAM STR profiles evaluated in this report, no inter-haplogroup STR sharing was observed in these Nigerian populations. Apparently, the number of loci evaluated might have resulted in this different outcome. The conclusions from results in this thesis overall, were based solely on the 11 SWGDAM STR profiles. It has however been shown that it was not just increasing the number of the STR but the particular qualities of the STR that improves their forensic usefulness (Hedman, *et al.,* 2011; D'Amato, *et al.,* 2011 and Ballantyne, *et al.,* 2010).

The incorporation of haplogroup profiles along with STR was recommended to enhance the value of the YHRD and when this was explored with the Nigerian dataset, there was no improvement in the discrimination capacity among this Nigerian dataset. This was apparently because, at the level of the SWGDAM STRs utilized in this report, no inter-lineage STR sharing was observed.

De Filippo and other workers (de Filippo, *et al.*, 2011), using a technique called Linear Discriminant Analysis (LDA), were able to achieve a more than 92% accuracy in predicting the haplogroups from Y-STR profiles of populations within and very close to the Bantu-language speaking populations of sub-Saharan Africa. As this technique was not applied to this dataset, it is hoped that future studies will be done to evaluate this process of LDA experiment with the Nigerian population. This is to take forward efforts at achieving accuracy as was observed by the Japanese and others in their populations (Mizano, *et al.*, 2010; Schlecht, *et al.*, 2008). However, towards this process, it must be borne in mind that, even de Filippo and others (de Filippo, *et al.*, 2011) recognized the constraint of extrapolating this process for populations beyond the Bantu language speakers where the result was much less accurate.

Based on the well-reported conclusion of Hammer and other workers (Hammer, *et al.*, 2000) that Jewish and Middle Eastern non-Jewish populations shared a common pool of Y-chromosome bi-allelic haplotypes, the expectation of finding Nigerian populations clustering with such populations might lend credence to the historical myth of origin of the male founders of many Nigerian populations coming from such exotic places and thus inspire a full investigation of this hypothesis. However the results in this thesis, which revealed North African populations (Algeria, Morocco and Tunisia) clustering with a Middle Eastern population (Saudi Arabia), was assessed with the Nigerian populations at the SWGDAM STR profile level (as this was synonymous with the SNP result in the

Nigerian populations), did not provide a prima facie support such conclusions on the myth of origin.

The Neighbour Joining Tree result in this thesis confirmed the relatively close association of the North African with the Middle Eastern populations but glaringly exposed the Nigerian populations as clustering with other sub-Saharan African populations. Most of the other results in this thesis showed a much closer relationship of sub-Saharan African populations among themselves rather than with populations afar off. This was also observed in the Veeramah and others report, establishing that the claims of some Southern Nigerian populations in the Cross Rivers region originating from the Middle East have no genetic evidence so far (Veeramah, *et al.*, 2010). The caveat in this thesis however was that, those subjects whose haplogroups were not resolved could perhaps be the carriers of the Middle Eastern signature. However, considering their number which is obviously less than 5% of the sampled subjects, this hypothetical situation could be explained as counseled by Jobling and Tyler-Smith (Jobling and Tyler-Smith, 2003), as more recent introgressions from the Middle East as a result of the advent of Islam (in the last 1000 years) with its demand of regular pilgrimages (and other social interactions) to Saudi Arabia and from many countries of the world including Nigeria.

4.5 Are there other important issues from these experiments?

The first issue addressed here was the ascertainment bias associated with the choice of Y-SNP markers evaluated in this thesis. The SNPs were resolved at the very roots of the major haplogroups found in Africa, to ensure that deep ancestral origins were captured. This approach was fairly representative considering that the haplogroup diversities confirmed that three major haplogroups were observed in significant proportion. The major Y-SNP haplogroups found in Africa are A, B, E, J and R (Underhill, *et al.*, 2001; Cruciani, *et al.*, 2002; Tishkoff, *et al.*, 2007; Batini, *et al.*, 2011). Both haplogroups J and R are geographically restricted to Eastern and Central Africa respectively. Haplogroup A is very rare in West Africa (Batini, *et al.*, 2011). A cautionary note with Y-SNP profile generally is that, the lack of evidence does not confirm an absence (Jobling and Tyler-Smith, 2003), as sample sizes in some cases might be extremely small that it could not have represented all possible haplogroups present in a population.

In this thesis, all the subjects, except four individuals, were resolved into three major haplogroups namely B, E and R. This immediately raises a bias issue and a need to assimilate the report with caution. However, a comparison of this result with that reported by Veeramah and others (Veeramah, *et al.*, 2010) on 1081 individuals from the Cross Rivers region of Southern Nigeria, which found only 5 and 1 individual respectively belonging to haplogroups A and J respectively (less than 0.5% and 0.1% respectively), erases such concern. In the same report (Veeramah, *et al.*, 2010), they observed 8.4%

haplogroup B and more than 91% with haplogroup E background which was very similar to the findings of this thesis (**Table 3.2**).

The impact of historical freed slaves' re-integration in these regions of modern Nigeria might also significantly undermine the generally self-ascribed ethnicities of most people found in these regions. Lovejoy (Lovejoy 1986; Lovejoy, 2000) estimated that the Trans-Sahara slave trade resulted in the demic-diffusion of more than five million peoples in Africa. Jobling and Tyler-Smith (Jobling and Tyler-Smith, 2003) advised that more recent explanations of genetic admixtures must be excluded before inferring pre-historic interpretation of population genetics data.

Geographic proximity apparently influenced substantially the gene flow among these Niger Congo speaking southern Nigerian populations rather than with the linguistically different Afro-Asiatic Hausa populations. This could explain the apparent non-differentiations of the Southern populations from their linguistically different neighbours. The relatively few gene exchange between Northern Nigeria Hausa populations and the four Southern Nigeria neighbours could be explained from the demographic influences of the last two millennial including, but not limited to, the trans-Sahara slave trade and other commercial contacts, inter-ethnic wars, and most importantly, the various cultural practices like polygamy, patrilocality, exogamy, and even, the re-integration of freed slaves from other parts of the continent (Falola and Heaton, 2008; Wood, *et al.,* 2005).

4.6 Forensic significance of these dataset

The forensic science infrastructures in Nigeria are being upgraded to cope with a rising security challenge by home-grown terrorist groups. Prior to this upsurge in violent crimes, rape and indecent assault cases ranked 4th on the scale of most serious crimes against persons reported in Nigeria (Alemika and Chukwuma, 2010). In some few cases when prosecutions were attempted, they were unsuccessful on the strength of weak evidence. The established autosomal marker based forensic analysis was unhelpful in the face of overwhelming female victims' cells over the comparatively small male signals. The absence of a reference population database of Nigerian populations too, might not allow for a statistical haplotype frequency estimates. However for this, the YHRD African meta-population database provides some temporary relief. The ultimate would be to have a database that fully represents all the different alleles in the Nigerian populations with high haplotype diversities within and among populations but without geographic structure (Tables 3.5a-f).

This thesis provides a complete profile of 463 males, representing five major Nigerian populations based on the SWGDAM recommended 11-loci STRs. Two notable features of these profiles are: (a) the high haplotype diversities among individual populations, and (b) the low discrimination capacities (DC) in 60% of the sampled populations (Igbo, Yoruba and Ijaw). The haplotypes obtained in this report could be incorporated and used

in a DNA reference population database that could be deployed in forensic match frequency estimation.

The issue of DNA reference database was a critical requirement for the use of DNA profiles from crime scenes. The Y-STR analyzed here would be very useful in resolving male specific forensic cases despite their non-specific identification and the population stratification associated with them at both national and continental levels.

It has been suggested that between 100 and 150 individuals per population could provide an adequate sampling for a genetic locus (Chakraborty, 1992). At this locus, allele frequencies below 1% would not be used in forensic calculations. Foreman and Evett also suggested that collecting information from more samples usually only improves the precision of a result rather than the accuracy of the allele count (Foreman and Evett, 2001). It was also recommended that each allele should be observed at least 5 times to be included in reliable statistical calculations. However, Hale and others (Hale, *et al.*, 2012) and Pruett and Winker (Pruett and Winker, 2008) both argued that, at least 20-30 individuals be sampled in microsatellite studies that assess genetic diversity when working in a population that has unknown level of diversity. They suggested that when large sample sizes are not possible or extremely difficult to obtain (as in the case of endangered species), research should include measures of genetic diversity if sample sizes are less than ideal. They tested sample sizes as low as 10 and even 5 individuals. This advice was followed in this thesis (**Table 2.1**) as seen in the low overall sample

sizes of the Bini and Ijaw populations. This was the reason the results of Bini and Ijaw populations forensic parameters were assimilated with caution in this thesis. An alternative solution to this sample size issue was rarefaction (-a statistical technique that compares observed taxon richness at a standardized sampling effort using confidence intervals) which provides a useful way to compare estimates of allelic diversity.

A major drawback of Y-STRs in forensics was the sometimes low diversities and geographical structures within specific populations (Hedman, *et al.,* 2011). These two issues could be resolved in two ways; (1) Increase the number of polymorphic markers and/or (2) explore or add new and fast mutating STRs (Hedman, *et al.,* 2011; D'Amato, *et al.,* 2011; Ballantyne, *et al.,* 2010). The first approach was the standard method and the rationale behind the 17 loci commercial kits. The second approach was applied successful by these cited publications to Finish population (Hedman, *et al,* 2011), South African populations (D'Amato, *et al.,* 2011) and European populations (Ballantyne, *et al.,* 2010). Ballantyne and others were able to demonstrate even a distinction between close and distant relatives of a male DNA profile with their new rapidly mutating 13 STRs system (Ballantyne, *et al.,* 2010; Ballantyne, *et al.,* 2012; Ballantyne, *et al.,* 2014). They overcame the associated problems that affect Y-STR mutation rates namely (a) the total repeat number, (b) repeat complexity, (c) length of base pairs repeat motifs and (d) the age of the father in a father son pair. They highlighted earlier problems associated with the approach of applying average mutation rates for Y-STRs and suggested the use of

only locus specific knowledge in subsequent applications in evolutionary and forensic studies, as utilized in this thesis, de Filippo and others (de Filippo, et al., 2011) and also Batini and others reports (Batini, *et. al.,* 2011). The profiles reported in this thesis allowed for comparison of different Nigerian populations' profiles with several others in the YHRD, especially within the African meta-population (a great potential to enrich the YHRD when added). There has not been any reported data of this nature from Nigeria with which to compare. In Veeramah and others report of more than 1000 Nigerian chromosomes (Veeramah, *et al.,* 2010), where they observed no genetic differentiation among the populations evaluated, they limited the scope of their investigation to basically evolutionary and ancestral studies, and thus used only six (6) Y-STRs, not even the minimal haplotypes as used in Forensics.

On some 51 African-wide populations comparison for the SWGDAM recommended loci also evaluated (Supplementary Table 3) however, geographic structuring was observed, even among those formerly assumed to be homogenous Bantu speaking populations (De Filippo, *et al.,* 2011; Berniel-Lee, *et al.,* 2009; Montano, *et al.,* 2011). Three different proposals to resolve this geographic structure are as follows: (1) to increase the number of STR loci to 17 as used in the Y-filer or other commercial kits (instead of the 11 used in this thesis) and to sample more previously unsampled Nigeria populations (from the more than 250 ethnic nationalities). (2) To explore the new 10-plex STRs developed in South Africa for their likely African populations' specificity (D'Amato, *et al.,* 2011). (3) To

explore the 13 rapidly mutating STRs validated for European populations for their world wide applications (Ballantyne, *et al.,* 2010; Ballantyne, *et al.,* 2014). All these to varying degrees will definitely change the current situation not only with the forensic application of Y-STRs in Nigeria but within the African continent and even world-wide. Some African population data (Table 3.8) for Y-STR have reported the 17- loci systems but with very limited number of populations within Countries, making it difficult to assess the impact among several populations within the same country (Omran, *et al.,* 2008; Laouina, *et al.,* 2011; Gomes, *et al.,* 2010; Arroyo-Prado, *et al.,* 2005; Alves, *et al.,* 2003).

4.7 Was there any genetic evidence to corroborate the historical and archeological pre-existence of the Nigerian populations in their current geographical locations?

The inter-lineages TMRCA, apart from confirming and describing a consistent age distinctions among the major haplogroups found in Nigeria also revealed the presumably shortest time of human existence in the area to be not less than 10,000 to 30,000 years ago. This corroborates the age ascribed to an archeological findings in the central Nigerian region dated to approximately 11,000 years ago (Falola and Heaton, 2008; Alabi, 2005). Apparently, the ethnicity might not be as well defined as today.

At the level of inter-lineages however, Shi and other workers (Shi, *et al.,* 2010) established that the Yoruba population has a minimum TMRCA of 17,000 years ago and

an expansion time of at least 12,000 years ago (based on population-wide estimation regardless of the specific haplogroups as estimated in this report). Apparently, their observations could have been majorly influenced by the comparatively older haplogroup B. It was also possible that about this time, the population was not known as Yoruba which now only retained the genetic make-up of their pre-historic ancestors.

The relatively lower TMRCA for Haplogroup E1b1a7 represented the Pre-bantu expansion presence of this haplogroup among some Nigerian populations (Hausa, Bini and Ijaw). As the date of the Bantu expansion was estimated to be between 3000 and 5000 years ago (Veeramah, *et al.*, 2010), the overall results in this thesis suggested that the E1b1a7 lineage (and other lineages) was in rapid expansion with no particular pattern in the Nigeria area. Indeed, the signature Bantu haplotype (15-21-10-11-13 for DYS19, DYS390, DYS391, DYS 392 and DYS393 respectively, as DYS388 was not assessed in this thesis) was present in significant proportion among the four southern Nigerian Non-Bantu speaking Niger Congo Nigerian populations of Bini, Ijaw, Igbo and Yoruba.

The distinction between the dates of haplogroups B and both haplogroups E1b1a7 and E1b1a8 seems to lend credence to the historical observance of different waves of settlements in the Nigeria area at significantly different times. The older settlers could be presumed to be haplogroup B while the later settlers were ancestors of haplogroup E1b1a7 and E1b1a8 (Falola and Heaton, 2008). The results in this thesis also hinted at the relatively equal migrants from both the Afro-asiatic speaking Hausa population (40)

of Northern Nigeria and the Niger-Congo language speaking Igbo, Yoruba, Bini and Ijaw populations (31) of Southern Nigeria respectively (Table 3.12).

The conclusion above seemed to be supported by other lines of evidence. For instance, given the rapid expansion, especially of Haplogroups E1b1a7 and E1b1a8 from the network plots in Figures 3.8a-f, and a diminished presence of these haplogroups in the Hausa populations, Tables 3.1 and 3.2, the Hausa population could not have been the source migrants of these haplogroups. Apparently, as shown from the results in this thesis, the modern Nigeria area has been a site of massive migration, without any strong evidence on the pattern of these migrations. This was quite possible with the absence of any physical or natural barrier to migration around the Niger-Benue River confluence, from which Nigeria got her name.

The insight from genetics in this thesis had significant corroboration from both archeology evidence and history of the peoples living in the area of modern day Nigeria.

The Late Stone Age (LSA), between roughly 12,000 and 4,000 years ago (Falola and Heaton, 2008; Alabi, 2005) was a period of major firsts for human development in the territories in and around modern-day Nigeria. The first known human remains were found in the *Iwo Eleru* rock shelter in what is now southwestern Nigeria, and have been dated to around 11,000 years ago. While humans must have lived in the area well before this time, the LSA was unique historically for several other reasons. It was widely

postulated that this was characterized by unprecedented levels of migration in the greater Nigeria area, particular as people moved south from the savanna into the forest zones to escape the rapid desiccation of the Sahara. Secondly, it was during this period that humans in the greater Nigeria began to use stone tools, called microliths. These tools led to the development of pottery around 5,000 years ago in most areas and ultimately to the development of agriculture between 6,000 and 3,000 years ago, depending on the specific area in focus. Development of agriculture allowed for the establishment of permanent settlements. This also meant a departure from hunting and gathering activities to the centralization of food resources, which allowed dwelling together on a lasting basis.

Much evidence has been uncovered in the various ecological region of modern-day Nigeria dating from the LSA period. These include *Apa I, Iwo Eleru, Ifetedo* and *Ita-Ogbolu* in the southern forested region to the *Itaakpa* in the middle belt region, *Mejiro* cave, *Afikpo* and *Rop* in the savanna zone, and *Daima* and other sites in the Chad basin area of the Sahel (Alabi, 2005). *Iwo Eleru I* in particular showed evidence of both the aceramic and ceramic periods of between 12,000 to 7,000 years ago and 7,000 to 3,500 years ago respectively. *Afikpo* in the southeastern part revealed three phases of development covering a few microlithic tools before 5,000 years ago, flaked axes between 5,000 and 2,500 years ago and different types of ceramics commencing around 2,100 years ago, different from those of the second phase. The transition from hunting to agricultural tools seemed to confirm this theory (Alabi, 2005). Lake Chad region revealed

evidence from *Daima* and *Kursakata*, the presence of domesticated animals like sheep, goats and cattle about 4,000 years ago. Archeologists believed that animal husbandry reached here across the Sahara from the east and north between 5,000 and 4,000 years ago (Breunig, *et al.*, 1996). Remains of horses from North Africa have been found in archeological sites in the savanna and Sahel dating to roughly 3,000 years ago. Unlike those in Europe, or the Near east, most west African societies transited directly from the use of stone tools to iron without an intervening period of using softer tools like copper or bronze. More evidence of iron-working and iron tools at archeological sites dates from about 9,000 years ago at *Taruga*, near Abuja, in the middle belt region. This *Taruga* was also known to be the Centre of the Nok culture, most famous in archeological circles for the large terracotta sculptures found within 500 km radius of Taruga. This technology which could have come from other region through trade contacts or migration indicated a local knowledge of iron production. Other areas in the region include *Tadun Wada*, *Kuchamfa, Jemaa Maitumbi*, *Kawu* and *Kagara*, all of which were smelting iron between 2,900 years ago and 1,800 years ago, the recognized dates for the duration of the Nok culture (Aremu, 2005). Eventually, this technology became rampant in the greater Nigeria areas like *Uffe Ijumu*, in the south west, 1860 years ago (Oyelaran, 1998*), Opi* in the southeast, as early as 7,000 years ago (Okafor, 1993), (Connah, 1968) and (Hartle, 1996).

The time-consuming nature of iron smelting, underscored the presence of professional Blacksmiths responsible for fashioning these tools in the places mentioned. The finished products of iron smelters in one region were evidently spreading to other regions through either migration or trade or both, indicating the extent to which distinct communities across long distances had contact with each other during this period (Falola and Heaton, 2008).

4.8 What were the genetic associations among the haplotypes in the Nigerian lineages?

Apart from the network of haplogroup E1b1a7 which revealed the pattern of a rapidly expanded lineage with no structure, the remaining lineages revealed clearly that there were some branches of haplotypes in the lineages. As most lineages evaluated had sub-clades that were not specifically evaluated, yet the networks revealed signatures of these. These branches, as summarized in the results section of this thesis, were not population specific, as could be seen from all the network plots (Figures 3.8a-f) and spatial auto correlation plots (Figures 3.9a-c). Clinal relationships among the populations were not observed within the three major lineages (B, E1b1a7 and E1b1a8) found among Nigerian populations.

The results also showed limited gene flow, particularly between geographical Northern and Southern populations as well as along linguistic lines of Afro-asiatic and Niger

Congo language speakers, the southern Nigerian populations were not distinguishable from each other at the levels of the two different genetic markers utilized, namely Y-STRs and Y-SNPs. This might thus account for the non-population specific branches in the networks and also suggested that the current genetic structure of the Nigerian population had persisted from ancient date and has not be significantly obliterated by relatively recent events of slave trading (especially trans-Saharan) in the last 1,500 years and other demographic or cultural practices.

5.0 CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

The Y-STRs are believed to be suitable for investigating more recent events, whereas Y-chromosome single-nucleotide polymorphisms (Y-SNPs) are suitable for more ancient events (de Knijff, 2000) in human history. This was because of the observed 10,000 times slower mutation rate of Y-SNPs compared with Y-STRs (Thomson, *et al.,* 2000). This thesis contributes to the growing number of systematic studies to compare the power of both marker systems in detecting the time depth of events in human population history and their potential forensic significance.

5.1 Conclusions

New Y-chromosome marker profiles of 463 unrelated Nigerian males representing five major populations namely Yoruba (238), Igbo (119), Hausa (78), Ijaw (15) and Bini (13) are reported. This significantly increases the amount of Nigerian genetic data available for comparison with other world populations and also serves to fill the scarcity of DNA data from Nigeria. When these are eventually contributed and incorporated into the existing Y haplotypes reference database (YHRD), they will be adding more than 270 new Scientific Working Group on DNA Analytical Methods (SWGDAM) recommended haplotypes and also increase the frequency of previously uploaded haplotypes, which are also found in this report, by 120 on the YHRD.

The SWGDAM recommended 11 loci Short Tandem Repeats (STRs) and nine (9) Unique Event Polymorphisms (UEP) have been used to resolve 463 Nigerian male chromosomes into 394 Y-haplotypes and 7 Single Nucleotide Polymorphisms (SNPs) Y-haplogroups. Only four of these were not assigned as they tested ancestral to all haplogroup tested.

The STR haplotypes revealed a very high diversity within and among populations but no significant geographical sub-structures.

The SNP haplogroups revealed that haplogroup E1b1b and haplogroup R are very rare in the four (4) Niger- Congo populations of Nigeria evaluated.

Even though haplogroup E1b1b was thought to be the signature haplogroup of the Afro-asiatic language speaking populations of Africa, the Hausa populations of Nigeria and Sudan seems to be exceptions to this rule demonstrating about 5% and 3% in these populations respectively. They also shared 18% and 13% respectively along the R lineage.

The combined haplotype and haplogroup dataset revealed that while the southern Nigerian populations were substantially undistinguishable, there was notable sub-structure among the Hausa populations hence revealing contrasting patterns among broad linguistic and geographical groupings. These patterns were different that while there was limited gene flow among the Afro-asiatic Hausa populations, there was substantial gene-

flow among the Niger Congo populations of southern Nigeria. Language was found to be a greater factor in accounting for genetic variance among these populations than geography and ethnicity which all have significant effects.

The most common haplogroups among the individual populations were E1b1a7, E1b1a8 and B with geographically varying frequencies.

The patterns of diversities of both haplotypes and haplogroups showed a North to South decreasing gradient. This pattern of diversities between STRs and SNPs also showed strong correlations to conclude that the structures observed in the Nigerian populations had an older rather than a more recent date.

The five Nigerian populations demonstrated a glaring lack of evidence that their male founders descended from the Middle East by clustering with other sub-Saharan African populations rather than with the populations from both North Africa and the Middle East.

There was strong correlation between both R_{st} and F_{st} distances and geographical distances among the Nigerian populations.

There seems to be no pattern in the spread of the E1b1a7 lineage between the Hausa populations of Northern Nigeria and the southern Nigerian populations of Bini, Ijaw, Yoruba and Igbo.

Also the distinctions between the oldest TMRCA (haplogroup B) and the youngest TMRCA (haplogroup E1b1a8) corresponded with historical waves of settlements found in the literature.

5.2 Recommendations

To resolve the potential issue of Y-STR haplotypes sub-structure within population for forensic applications, three proposals with increasing effectiveness are recommended:

Increase the number of STR loci in future studies from the 11-loci STR used in this thesis to 17 loci. This will definitely increase both the diversity and reduce potential sub-structure.

Based on the successfully developed 10-STR multiplex system for South African populations, which was able to increase both the discrimination capacity (DC) and haplotype diversity (HD) and most importantly, its greater efficiency than the available 17-loci commercial kits in those populations, apply this to the Nigerian populations to ascertain its African populations' specificity.

The newly developed rapidly mutating 13 STRs that were found to even resolve both close and distant relations of male haplotypes in European populations, apart from their superiority to the 17 loci commercial kits in terms of their increased DC, HD and absence of substructures among nations, should be applied to the Nigerian populations and even other African populations to ascertain their world-wide applicability for forensic and evolutionary studies.

To enhance the data on Nigerian population structure and demographics, a proposal to survey and analyze Nilo-Saharan populations of North eastern Nigeria and many

populations from the North central that were not reported in this thesis, which should specifically include in the minimum Idoma, Igala, Igbirra, Jukun, Kanuri, Kilba, Margi and Tiv populations, is recommended. This will give a fuller and more complete view of the Y profiles of the entire Nigerian landscape, having combined the earlier report on the Cross rivers region and the latest report in this thesis; it is proposed that the haplotype and haplogroup diversities of the North-Eastern Nilo-Saharan populations be evaluated. In addition to this, some sparsely sampled Hausa populations in some geographical locations could also be re-evaluated to confirm the complete none representation of the major haplogroups (E1b1a7 and E1b1a8) in them.

Haplogroups E1b1a7 and E1b1a8 have several sub-haplogroups each. These could be very informative in revealing patterns of migration and dispersal of the apparently homogenous Igbo, Ijaw and Bini populations. A proposal to explore their signatures is recommended.

Finally, despite the plethora of publications on the mitochondrial DNA (mtDNA) profiles of African populations which suggested a vastly homogenous maternal profiles among populations, it is still necessary to confirm this observation in the Nigerian populations too, to complete the pictures of uniparental markers signatures in Nigeria, especially for their evolutionary significance.

6.0 References

Adeyemo, A.A., Chen, G., Chen, Y., Rotimi, C. (2005). Genetic structure in four West African population groups. *BMC Genetics*. 6. 38-46.

Alabi, R. A. (2005). Late Stone Age technology and agricultural beginnings. In: Pre-colonial Nigeria. Essays in honour of Toyin Falola. Akinwumi Ogundiran (Ed.) Trenton, NJ, African World Press, 87-104

Alemika, E.E.O. and Chukwuma, I.C. (2010). Summary of 2010 National Crime and safety survey. http://www.cleen.org/crimevictimisationsurvey.html

Allocco, D.J., Song, Q., Gibbons, G.H., Ramoni, M.F., Kohane, I.S. (2007). Geography and genography: prediction of continental origin using randomly selected single nucleotide polymorphisms. *BMC Genomics*, 8. 68

Altschul, S. F., Madden, T.L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J.(1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs *Nucl. Acids Res. 25 (17): 3389-3402*

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

Aremu, D.A. (2005). Change and continuity in metallurgical traditions; origins, technology and social implications: In Pre-colonial Nigeria: Essays in honour of Toyin Falola. Akinwumi Ogundiran (Ed.) Trenton, NJ, African World Press Pp136-141

Arroyo-Pardo, E., Gusmao, L., Lopez-Parra, A.M., Bueza, C., Mesa, M.S. and Antonio, A. (2004). Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea (Central Africa). *Forensic Science International*. 149, 109-113.

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

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Woll-stein, A.W., Ying, C., van Duijn, K., Vermeulen, M., Brauer, S., Decorte, R., Poetsch, M., von Wurmn-Schwark, N., de Kniff, P., Labuda, D., Vezina, H., Knoblauch, H., Lessig, R.,M., Roewer, L., Ploski, R., Dobosz, T., Henke, L., Henke, J., Furtado, M.R. and Kayser, M. (2010). Mutability of Y-chromosomal microsatellites: Rates, characteristic, molecular bases, and forensic implications. *The American Journal of Human Genetics*, 7, 341-353.

Ballantyne, K. N., Keerl, V., Wollstein, A., Choi, Y., Zuniga, S. B., Ralf, A., Vermeulen, M., de Knijff, P. and Kayser, M.(2012). A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet* 6: 208-218.

Ballantyne, K. N., Ralf, A., Aboukhalid, R., Achakzai, N. M., Anjos, M. J., Ayub, Q., ... and Klausriegler, E. (2014). Towards complete male individualization with rapidly mutating Y-chromosomal STRs. *Human Mutation*. doi: 10.1002/humu.22599

Bandelt H-J, Forster P, and Röhl A (1999) Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution* 16, 37-48.

Barbieri, C., Whitten, M., Beyer, K., Schreiber, H., Li, M. and Pakendorf, B. (2012). Contrasting maternal and paternal histories in the linguistic context of Burkina-Faso. *Mol. Biol. Evol*. 29 (4): 1213-1223.

Barbujani, G.V. (2000). Geographic patterns: how to identify them and why. *Hum Biol* 72: 133-153

Batini, C., Ferri, G., Destro-Bisol, G., Brishighelli, F., Luiselli, D., Sanchez-Diz, P., Rocha, J., Simonson, T., Brehm, A., Montano, V., Eiwali, N.E., Spedini, G., D'Amato, M.E., Myres, N., Ebbesen, P., Comas, D. and Capelli, C. (2011). Signature of the pre-agricultural peopling process in sub-Saharan Africa as revealed by the phylogeography of early Y-chromosome lineages. *Mol. Biol. Evol*. 28 (9): 2603-2613.

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., van der Veen, L., Hombert, J. M., Quintana-Murci, L., and Comas, D. (2009) Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular Biology and Evolution*, 26, 1581–1589.

Berniell-Lee, G., Bosch, E., Bertranpetit, J. and Comas, D. (2006). Y-chromosome diversity in Bantu and Pygmy populations from Central Africa. International Congress Series, 1288: 234-236.

Bird, S.C. (2012). Towards improvement in the estimation of the coalescent: Implications for the most effective use of the Y chromosome short tandem repeats mutation rates. PLoS ONE 7 (10) e48638. Doi: 10.1371/journal.pone.0048638

Blench, R (2011). An Atlas of Nigerian Languages. 3rd Edition. http://www.rogerblench.info/Language/Africa/Nigeria/Atlas%20of%20Nigerian%20Languages-%20ed%20III.pdf

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA and Bertranpetit J (2001). High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* 68: 1019-1029

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

Breunig, K., Neumann, K and Neer, W.V (1996) New research on the Holocene settlement and environment of the Chad Basin in Nigeria. *African Archeological Review*, 13 111-145

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

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

Campbell, M.C. And Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 9: 403-33

Cavalli-Sforza, L.L. and Bodmer, W. (1971): The genetics of human populations. Freeman, San Francisco, Calif.

Cavalli-Sforza, L.L. and Herrera, R.J. (2004). The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *American Journal of Human Genetics* 74, 532–544.

Cerny, V. Hajek, M., Cmejla, R., Bruzek, J. and Brdicka, R. (2004). mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. *Ann.Hum.Biol*., 31 (5): 554-569.

Cerny, V., Hajek, M., Bromova, M., Cmejla, R., Diallo, I., and Brdicka, R. (2006). MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum.Biol*., 78 (1): 9-27.

Cerny, V., Salas, A., Hajek, M., Zaloudkova, M., Brdicka, 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. *Ann Hum Genet*. 71 (4):433–452

Chakraborty, R (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Human Biology*. 64 (2), 141-159.

Charchar, F. J., Bloomer, L. D. S., Barnes, T. A., Cowley, M.J., Nelson, C. P., Wang, Y. Z., Denniff, M., Debiec, R., Christofidou, P., Nankervis, S., Dominiczak, A. F., Bani-Mustafa, A. (2012). Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *Lancet* 379:915–922.

Chiaroni, J., Underhill, P. A. and Cavalli-Sforza, L. L. (2009). Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad of Sci* USA 106:20174– 20179.

Coelho, M., Sequeira, F., Luiselli, D., Beleza, S. and Rocha, J. (2009). On the edge of the Bantu expansions: mtDNA, Y chromosome and lactose persistence genetic variation in southwestern Angola. *BMC Evolutionary Biology*. 9: 80

Coia, V., Destro-Bisol, G., Verginelli, F., Battaggia, C., Boschi, I., Cruciani, F., Spedini, G., Comas, D., and Calafell, F. (2005). Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *Am.J.Phys.Anthropol*., 128 (3): 678-681.

Coleman, J (1958). Nigeria: Background to Nationalism. University of California Press.

Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon JM, et al., (2007). Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Ychromosomal haplogroups E-M78 and J-M12. *Mol Biol Evol* 24: 1300-1311

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Homlmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli- Sforza LL, Scozzari R, and Underhill PA (2002). A Back Migration from Asia to Sub- Saharan Africa is Supported by High-Resolution Analysis of Human Y-Chromosome Haplotypes. *American Journal of Human Genetics* 70: 1197-1214.

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

Daniels, D.L., Hall, A.M., Ballantyne, J. (2004). SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework. *J Forensic Sci*. 49: 668–683.

De Filippo, C., Barbieri, C., Whitten, M., Mpoloku, S.W., Gunnarsdottir, E.D., Bostoen, K., Nyambe, T., Beyer, K., Schreiber, H., De Kniff, P., Luiseli, D., Stoneking, M. and Pakendorf, B. (2011). Y chromosomal variation in Sub-Saharan Africa: insights into the history of Niger-Congo groups. *Molecular Biology and Evolution*, 28, 1255–1269.

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

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

Deka, R., Guangyun, S., Wiest, J., Smelser, D., Chunhua, S., Zhong, Y., Chakraborty, R. (1999). Patterns of instability of expanded CAG repeats at the ERDA1 locus in general populations. *Am. J. Hum. Genet*. 65: 192-198

Destro-Bisol, G., Donati, F., Coia, V., Boschi, I., Verginelli, F., Caglia, A., Tofanelli, Spedini, G., and Capelli, C. (2004): Variation of female and male lineages in Sub-Saharan Populations: the Importance of Sociocultural Factors. *Mol. Biol. Evol*. 21 (9):1673-1682.

Dettlaff-Kakol, A., and Pawlowski, R. (2002). First Polish DNA ‘‘manhunt’’—an application of Y-chromosome STRs. *Int. J. Legal Med*. 116, 289–291.

Disotell, T.R. (2000): Human genomic variation. *Genome Biology*. 1; (5) comment 2004.

Ehgosa, E.O. (2002).: Ethnic groups, languages and religions. In: Africa Atlases Nigeria. Pigeonniere, A. L.(ed). Les Editions J.A. Paris.

Elochukwu, U. E. (1997). Worship as a body language. Liturgical Press.

Excoffier L, Laval G and Schneider S (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinfor Online* 1: 47-50.

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

Falola, T and Heaton, M. (2008). A history of Nigeria. Cambridge University Press

Flores, C., Maca-Meyer, N., Perez, J.A., Hernandez, M. and Cabrera, V.M. (2001). Y-chromosome differentiation in Northwest Africa. *Hum.Biol*., 73 (4): 513-524
.
Fluxus-engineering (2008). Fluxus Technology Ltd. 2008-2009.

Foreman, L.A. and Evett, I.W. (2001): Statistical analyses to support forensic interpretation for a new 10-locus STR profiling system. *International Journal of Legal Medicine*, 114, 147-155.

Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L. and Feldman, M. W. (1995)a. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139 (1), 463-471

Goldstein, D.B., Linares, R. A., Cavalli-Sforza, L.L. and Feldman, M.W. (1995)b. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the United States of America* 92, 6723–6727.

Goldstein, D.B., and Pollock, D.D. (1997). Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *J. Heredity* 88: 335-342

Gomes, V., Sanchez-Diz, P., Amorin, A., Carracedo, A. and Gusmao, L. (2010). Digging deeper into East African human Y-chromosome lineages. *Hum Genet*. 127: 602-613.

Hallast, P., Balaresque, P., Bowden, G. R., Ballereau, S., Jobling, M. A. (2013). Recombination dynamics of a human Y-chromosomal palindrome: rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions. *PLoS Genet* 9:e1003666

Hammer, M.F. and Horai, S. (1995). Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet*. 56, 951-962

Hammer, M.F., and Zegura, S.L. (2002). The human Y chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. *Annu. Rev. Anthropol*. 31, 303–321.

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A. R. and Zegura, S.L.(1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 15: 427- 441

Hammer, M.F., Karafet, T.M., Redd, A.J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H. and Zegura, S.L. (2001)a. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18: 1189-203

Hammer, O., Harper, D.A.T. and Ryan, P.D. (2001)b. PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologia Electronica* 4: 9

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

Hartl, D.L. and Clark, A.G. (1997). Principles of Population Genetics, Third Edition. Sunderland, Massachusetts: Sinecure Associates, Inc. Publishers.

Hartle, D.D. (1996) Archeology in eastern Nigeria. Nigerian Magazine 93, 134-143.
Hassan, H.Y, Underhill, P.A., Cavalli-Sforza, L.L. and Ibrahim, M.E (2008). Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history. American *Journal of Physical Anthropology*, 137, 316–323.

Hedman, M., Neuvonen, A.M., Sajantila, A. and Palo, J.U. (2011). Dissecting the Finnish male uniformity: The value of additional Y-STR loci. *Forensic Science International: Genetics*. 5: 199-201.

Helgason, A., Hrafnkelsson, B., Gulcher, J.R., Ward, R. and Stefansson, K. (2003). A population wide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am J Hum Genet*. 72:1370–1388.

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

Hodgkins, T.L. (1984). Nigerian Perspectives, An Historical Anthology. Oxford. UK.

Jobling, M. (2012). The impact of recent events on human genetic diversity. *Phil. Trans. R. Soc. B.* 367, 793-799.

Jobling, M.A., and Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. *Nat. Rev. Genet*. 4, 598–612.

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res*. 18, 830–838.

Kayser, M. (2007). Uni-parental markers in human identity testing including forensic DNA analysis. *Biotechniques* 43, 3042.

Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Krüger, K., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., and Sajantila, A (2000). Characteristics and Frequency of Germline Mutations at Microsatellite Loci from the Human Y Chromosome, as revealed by Direct Observation in Father/Son Pairs. *American Journal of Human Genetics* 66: 1580-1588.

Kayser, M., Vermeulen, M., Knoblauch, H., Schuster H., Krawczak, M., and Roewer, L. (2007). Relating two deep rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping. *Forensic Sci. Int.; Genet*. 1, 125–128.

King, T.E., and Jobling, M.A. (2009). Founders, Drift and Infidelity: The relationship between Y chromosome diversity and patrilineal surnames. *Mol. Biol. Evol*. 26, 1093–1102.

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

Langella, O. (2002). Populations v.1.2.30. 2008.

Laouina, A.,Houate, B.E., Yaha, H., Azeddoug, H., Baulouiz, R. and Chbel, F.(2011). Allele frequencies and population data for 17 Y-STR loci (The AmpFiSTR® Y-filerTM) in Casablance resident population. *Forensic Science International: Genetics* 5, e1-e3.

Larmuseau, M. H. D., Vanderheyden, N., van Geystelen, A., van Oven, M., Kayser, M. and Decorte, R. (2013). Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations. *Forensic Sci Int Genet*. [Epub ahead of print]

Larmuseau, M. H. D., Ottoni, C., Raeymaekers, J. A. M., Vanderheyden, N., Larmuseau, H. F. M. and Decorte, R. (2012). Temporal differentiation across a West-European Ychromosomal cline—genealogy as a tool in human population genetics. *Eur J Hum Genet* 20:434–440.

Liew M., Pryor R., Palais R., Meadows C., Erali M., Lyon E., Wittwer C. (2004) Genotyping of Single-Nucleotide Polymorphisms by High-Resolution Melting of Small Amplicons. *Clin Chem*. 50 (7):1156-1164.

Lovejoy, P.E. (1986). Africans in Bondage: Studies in slavery and the slave trade. Wisconsin Press.

Lovejoy, P.E. (2000): Transformations in slavery: a history of slavery in Africa. Cambridge Press.

Manni, F., Guerard, E. and Heyer, E. (2004). Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. *Human Biology* 76, (2), 173–190.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209-220.

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

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

Michalakis, Y. and Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. *Genetics* 142:1061-1064.

Mizuno, N. , Kitayama, T. , Fujii, K. , Nakahara, H. , Yoshida, K., Sekiguchi, K., , Yonezawa, N., Nakano, M., Kasai, K. (2010). A forensic method for the simultaneous analysis of biallelic markers identifying Y chromosome haplogroups inferred as having originated in Asia and the Japanese archipelago. *Forensic Science International: Genetics* 4, 73–79

Modiano, D., Luoni, G., Petrarca, V., Sodiomon-Sirima,B., De luca, M., Coluzzi, M., Bodmwe, J.G.and Modiano, G. (2001). HLA class I in three West African ethnic groups: genetic distances from sub-Saharan and Caucasoid populations. *Tissue Antigens*, 57 (2): 128-137.

Montano, V., Ferri, G., Marcari, V., Batini, C., Anyaele, O., Destro-Bisol, G. and Comas, D (2011). The bantu expansion revisited: a new analysis of Y chromosome in Central Western Africa. *Molecular Ecology* 20, 2693-2708.

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

Okafor, E.E. (2000) Opi: the earliest iron smelting site in Africa. Nigerian Heritage, 9, 146

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

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., and Stoneking, M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat. Genet.* 29, 20–21.

Otite, O. (1990). Ethnic pluralism and ethnicity in Nigeria, with comparative materials. Shaneson.

Ottoni, C., Larmuseau, M. H. D., Vanderheyden, N., Martinez-Labarga, C., Primativo, G., Biondi, G., Decorte, R. and Rickards, O. (2011). Deep into the roots of the Libyan Tuareg: a genetic survey of their paternal heritage. *Am J Phys Anthropol* 145:118–124.

Oyelaran, P.A. (1998). Early settlement and archaeological sequence of North eastern Yorubaland. *African Archeological Review*, 15, 65-79

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

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M and Santachiara-Benerecetti AS (1998). Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 62: 420-434

Pereira, L., Gusmao, L., Alves, C., Amorim, A., Prata, M.J. (2002). Bantu and European Y-lineages in Sub-Saharan Africa. *Ann.Hum.Genet*. 66 (Pt 5-6): 369-378.

Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Guldermann, B.K., Mpoloka, S.W., Nagawa, H., Naumann, C., Lipson, M., Loh, P., Lachance, J., Mountain, J., Bustamante, C.D., Berger, B., Tishkoff, S.A., Henn, B.M., Stoneking, M., Reich, D., and Pakendorf, B. (2013). The genetic prehistory of southern Africa. *Nature communications*. 3: 1143 (doi: 10.1038/ncomms2140).

Poloni, E. S., Semino, O., Passarino, G., Santachiara-Benerecetti., A.S., Dupanloup, I., Langaney, A., and Excoffier, L. (1997). Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am.J.Hum.Genet*., 61 (5): 1015-1035.

Polzi, T. and Daneshmand, S.V. (2003). On Steiner trees and minimum spanning trees in hypergraphs. *Oper Res Lett* 31: 12–20.

Pruett, C.L. and Winker, K. (2008). The effects of sample size on population genetic diversity estimates in song sparrows Melospiza melodia. *J Avian Biol*. 39: 252-256.

Qamar, R., ayub, Q., Khaliq, S., Mansoor, A., Karafet, T., Mehdi, S.Q. and Hammer, M.F. (1999). African and Levantine origins of Pakistani YAP+ Y chromosomes. *Hum. Biol*. 71, 745-755.

Ravid-Amir, O. and Rosset, S. (2010). Maximum likelihood estimation of locus specific mutation Y-chromosome short tandem repeats. *ECCB*, 26: i440-i445

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. *Am J Hum Genet* 72:1058–1064.

Roewer, L. (2009). Y chromosome STR typing in crime casework. *Forensic Sci. Med. Pathol*. 5, 77–84.

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

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. *Ann.Hum.Genet*., 68 (Pt 4): 340-

Rosa, A., Ornelas, C., Jobling, M.A., Brehm, A. and Villems, R. (2007). Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC.Evol.Biol*. 7: 124.

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.

Rychlik,W., and Rhoads, R.E. (1989) A computer-program for choosing optimal oligonucleotides for filter hybridization, sequencing and in-vitro Amplification of DNA. *Nucleic Acids Research* 17, 8543–8551.

Saitou N, and Nei M (1987) The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* 10: 471-483.

Salas, A., Richards, M., De laFe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P.S., Macaulay, V., Carrecedo, A. (2002).: The Making of the African mtDNA Landscape. *Am. J. Hum. Genet*. 71: 1082-1111.

Sambrook, .J, Fritsch, E.F., and Maniatis, T (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition. Plainview, New York: Cold Spring Harbor Laboratory Press

Schlecht, J., Kaplan, M. E., Barnard, K., Karafet, T., Hammer, M.F., Merchant, N.C. (2008). Machine-Learning Approaches for Classifying Haplogroup from Y Chromosome STR Data. *PLoS Comput Biol* 4(6): e1000093. doi:10.1371/journal.pcbi.1000093

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

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

Scozzari, R., Cruciani, F., Santolamazza, P., Malaspina, P., Torroni, A., Sellitto, D., Arredi, B., et al., (1999). Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 65: 829-846

Seielstad, M.T., Minch, E., and Cavalli-Sforza, L.L. (1998). Genetic evidence for a higher female migration rate in humans. *Nat. Genet*. 20, 278–280.

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.

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., .. and Santachiara-Benerecetti, A. S. (2004). Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *The American Journal of Human Genetics*, *74*(5), 1023-1034.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., ... & Underhill, P. A. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: AY chromosome perspective. *Science,290*(5494), 1155-1159

Semino, O., Santachiara-Benerecetti, A.S., Falaschi, F., Cavalli-Sforza, L.L. and Underhill, P.A. (2002). Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 70: 265-8

Shi, W., Ayub, Q., Vermeulen, M., Shao, R.G., Zuniga, S., van der Gaag, K., de Knijff, P., Kayser, M., Xue, Y., and Tyler-Smith, C. (2010). A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol. Biol. Evol*. 27, 385–393.

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

Smith, A. (2005). The concepts of 'Neolithic' and 'Neolithisation' for Africa? Before Farming 1: 1-6

Sokal, R. R. and Oden, N. L. (1978). Spatial autocorrelation in biology. *Biological Journal of the Linnean Society*, 10, 199-228.

Terreros, M. C., Martinez, L. and Herrera, R. J. 2005. Polymorphic Alu insertions and genetic diversity among African populations. *Hum.Biol*., 77 (5): 675-704

Thomas, M. G., Skorecki, K., Ben-Ami, H., Parfitt, T., Bradman, N., Goldstein, D. B. (1998). Origins of old testament priests. *Nature* 394:138–140

Thomas, M.G., Bradman, N and Flinn, H.M. (1999). High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. *Hum Genet* 105: 577-81

Thomas, M.G., Parfitt, T., Weiss, D.A., Skorecki, K., Wilson, J.F., le Roux, M., Bradman, N., Goldstein, D.B. (2000). Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba, the ‘‘Black Jews of Southern Africa.’’ *Am J Hum Genet* 66: 674–686

Thomson, R., Pritchard, J.K., Shen, P., Oefner, P.J., and Feldman, M.W. (2000) Recent Common Ancestry of Human Y Chromosomes: Evidence from DNA Sequence Data. *Proceedings of the National Academy of Sciences USA* 97: 7360-7365

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

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

Underhill, P.A., and Kivisild, T. (2007). Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet*. 41, 539–564.

Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli- Sforza, L.L. and Oefner, P.J. (1997) Detection of Numerous Y Chromosome Biallelic Polymorphisms by Denaturing High-Performance Liquid Chromatography. *Genome Research* 7: 996-1005.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon, Lahr, M., Foley, R.A., Oefner, P.J., and Cavalli-Sforza, L.L. (2001) The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations. *Annals of Human Genetics* 65: 43-62.

van Oven, M., Geystelen, A., Kayser, M., Decorte, R., and Larmuseau, M. H. (2014). Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Human mutation*, 35(2), 187-191.

van Oven M, Ralf A, Kayser M. 2011. An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *Int J Legal Med* 125:879–885.

van Oven M, Toscani K, van den Tempel N, Ralf A, Kayser M. 2013. Multiplex genotyping assays for fine-resolution subtyping of the major human Y-chromosome haplogroups E, G, I, J and R in anthropological, genealogical, and forensic investigations. *Electrophoresis* 34:3029–3038.

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

Weir, B.S. and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.

Wente-Lukas, R. and Jones, A. (1985). Handbook of Ethnic units in Nigeria. Fanz. Steiner. Verlag. Stuttgart.

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

Wjisman, E.M. (1984): Estimation of genetic admixture in pygmies In L.L.Cavalli-Sforza, ed. African pygmies. Pp 349-358 Academic Press, Orlando, Fla.

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

.

Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348.

Zhivotovsky, L. A., Underhill, P. A., Cinnioğlu, C., Kayser, M., Morar, B., Kivisild, T., and Kalaydjieva, L. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *The American Journal of Human Genetics*, *74*(1), 50-61.

Zhivotovsky, L.A., Underhill, P.A., Feldman, M.W. (2006). Difference between evolutionarily effective and germ line mutation rate due to stochastically varying haplogroup size. *Molecular Biology and Evolution,* 23, 2268-2270.

Zictkiewiicz, E., Motiva, V., Jarnik, M., Korah-Laskowska, M., Kidd, K.K., Modiano, D., Scozzari, R., Stoneking, M., Tishkoff, S., Batzer, M., Labuda, D. (1998). Genetic structure of the Ancestral population of Modern Humans. *J. Mol. Evol*. 17:146-155.

Zuccarelli, G., Alechime, E., Caputo, M., Bobillo, C., Corach, D., and Sala, A. (2011). Rapid screening for Native American mitochondrial and Y-chromosome haplogroups detection in routine DNA analysis. *Forensic Science International: Genetics* 5, 105-108.

7.0 Appendices

Preparation of Solutions

Binding Solution (1.0 mL)
Binding Saline (Promega) 2.5µL
Ethanol (Merck Laboratories Supplies) 50.0µL
Glacial Acetic Acid to fill to 1000.0µL

10X TBE Buffer (2 Litres)

- Tris (Merck Laboratory supplies) 216 gm.
- Boric Acid (Merck Laboratory supplies) 112 gm.
- EDTA (Merck Laboratory supplies) 16 gm.
- Distilled Water to 2 Litres

4 % Polyacrylamide Gel Mix

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

10X AMPS (10ml)

- Ammonium Persulfate (AMPS) 1 gm.
- Distilled Water 10 ml

Loading Buffer

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

Preparation of Gel running Plates

Gel Preparation

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

Plate Set – Up

1. The plates were cleaned using distilled water
2. The plates were set – up according to the manufacturer's instructions.

Pouring of Gel

1. Gel mix, containing AMPS and TEMED, was poured into plates which was previously mounted and left for two hours to solidify
2. The plates were cleaned with distilled water and then dried.
3. Plates were placed onto ABI® 377 DNA Sequencer and set – up was followed according to the manufacturer's instructions.

AGAROSE GEL WORK

2X loading buffer (10mL)

Bromophenol Blue (Merck laboratories Supplies (MLS)) 0.05gm
Xylene Cyanol 0.05gm
NaOH (Sodium Hydroxide) 0.1mM
Absolute Ethanol added to make 10.0mL.

Developing Solution (2Litres)

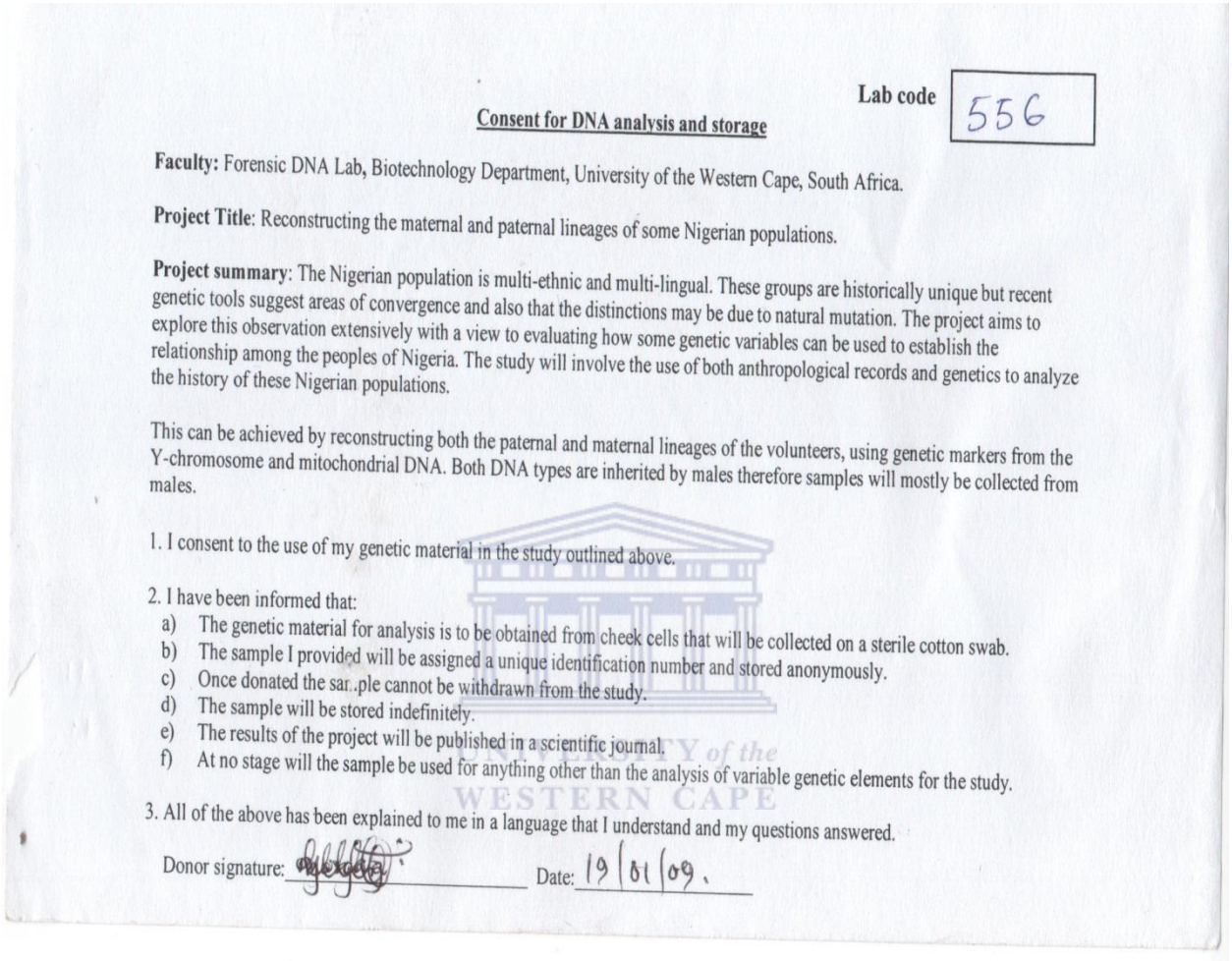
NaOH pellets (MLS) 30.0gms
15% Formaldehyde (MLS) 20mL
Distilled water to make up to 2.0L

Preparation of 1% Agarose Gel (50mL)

Put 0.5gm of agarose in a beaker.
Add 50mL of TBE buffer to the beaker.
Heat up in a microwave oven for 60 seconds.
Cool beaker under tap till lukewarm (not cold).
Add 1.0 µL ethidium bromide and stir till fully dissolved.
Pour in the preset with comb.
Allow to stand for at least 30 minutes.
Remove comb, and the gel is ready for use.

Agarose Electrophoresis of PCR Product

Put 2.0µL of PCR product in a 0.2mL tube.
Add 1.0mL of loading dye to the tube.
Transfer mixture to the well leaving enough space between wells.
Add another sample with 2.0µL distilled water as negative control in another well.
Add columns with DNA standards for comparison.
Align electrodes and apply current till bands separate substantially



Lab code 556

Consent for DNA analysis and storage

Faculty: Forensic DNA Lab, Biotechnology Department, University of the Western Cape, South Africa.

Project Title: Reconstructing the maternal and paternal lineages of some Nigerian populations.

Project summary: The Nigerian population is multi-ethnic and multi-lingual. These groups are historically unique but recent genetic tools suggest areas of convergence and also that the distinctions may be due to natural mutation. The project aims to explore this observation extensively with a view to evaluating how some genetic variables can be used to establish the relationship among the peoples of Nigeria. The study will involve the use of both anthropological records and genetics to analyze the history of these Nigerian populations.

This can be achieved by reconstructing both the paternal and maternal lineages of the volunteers, using genetic markers from the Y-chromosome and mitochondrial DNA. Both DNA types are inherited by males therefore samples will mostly be collected from males.

1. I consent to the use of my genetic material in the study outlined above.

2. I have been informed that:
   a) The genetic material for analysis is to be obtained from cheek cells that will be collected on a sterile cotton swab.
   b) The sample I provided will be assigned a unique identification number and stored anonymously.
   c) Once donated the sar..ple cannot be withdrawn from the study.
   d) The sample will be stored indefinitely.
   e) The results of the project will be published in a scientific journal.
   f) At no stage will the sample be used for anything other than the analysis of variable genetic elements for the study.

3. All of the above has been explained to me in a language that I understand and my questions answered.

Donor signature: _____ Date: 19/01/09.

Appendices Figure 1: Consent Form

Lagos State University
Department of Biochemistry
Lagos - Badagry Expressway, Ojo, Lagos.
P.M.B. 0001 LASU Post Office. LASU Ojo Campus, Ojo,
Lagos - Nigeria.
Tel: Nos:234-1-8547297, 080 33 236191
Website:www.lasunigeria.org. E-mail:bchseries@gmail.com

Office of the Head of Department.

Oladimeji. S. O. (Ph.D)
Senior Lecturer
E-mail:biosearch@yahoo.com

Our Ref: Your Ref: Date:

Date: December 14th, 2008.

Dear Sir,

Re: REPORT OF ETHICS COMMITTEE: ATTENTION Mr. Curtis Lanre Cole-Showers
(Registration Number: LASU/BCH/ EC 0501- 090508)

I hereby certify that the Research Committee of the:
Department of Biochemistry, Faculty of Science, Lagos state University,
has approved the methodology and ethics of the Research Project proposal by Mr. Curtis Lanre Cole-Showers on Research Project title: Population structure and demographic of Nigerian populations with uni-parental markers

Please note that the ethical issues and recommended guide lines for the recruitments and collections of samples apply in this approval

Thanks

14/12/08

Signed: Dr. O.S. OLADIMEJI
Designation: Ag. Head of Dept.

Appendices Figure 2: Ethical Clearance

The Nigerian populations haplotypes and haplogroups profile

S/N	LAB CODE	Ethnicity	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385a,b	DYS438	DYS439	Haplogroup
N001	BINI9	BINI	17	13	30	21	10	11	14	17,21	11	13	E1b1a7
N002	BINI1	BINI	17	13	30	21	10	11	15	17,19	11	12	E1b1a7
N003	BINI10	BINI	17	12	33	21	13	11	16	18,19	13	14	E1b1a7
N004	BINI11	BINI	14	13	30	23	11	11	13	17,18	11	12	E1b1b
N005	BINI12	BINI	16	13	30	21	10	12	15	19,20	11	12	E1b1a7
N006	BINI13	BINI	15	13	30	21	10	11	15	17,18	11	12	E1b1a7
N007	BINI2	BINI	16	13	31	21	10	13	15	16,17	12	12	E1b1a7
N008	BINI3	BINI	16	13	30	21	10	11	15	18,19	10	13	E1b1a7
N009	BINI4	BINI	15	13	31	21	10	11	13	16,17	11	12	E1b1a8
N010	BINI5	BINI	16	14	31	21	10	12	15	17,18	11	12	E1b1a7
N011	BINI6	BINI	17	13	30	21	10	11	14	18,18	11	12	E1b1a7
N012	BINI7	BINI	15	13	30	21	10	11	13	16,18	11	12	E1b1a8
N013	BINI8	BINI	17	13	29	21	9	11	15	18,19	11	12	E1b1a7
N014	HUSA33	HAUSA	17	13	30	21	10	11	15	14,18	11	12	E1b1a7
N015	HUSA78	HAUSA	17	17	33	24	13	15	16	14,16	12	14	E
N016	HUSA27	HAUSA	15	14	31	21	10	11	15	22,23	11	11	E1b1a7
N017	HUSA13	HAUSA	15	13	32	21	10	11	13	16,17	11	13	E1b1a7
N018	HUSA14	HAUSA	17	13	30	21	10	11	15	16,18	11	11	E1b1a7
N019	HUSA34	HAUSA	15	11	29	24	11	11	13	12,12	10	12	B
N020	HUSA54	HAUSA	16	13	31	23	10	14	13	16,17	11	13	E
N021	HUSA60	HAUSA	15	14	30	24	12	14	13	12,13	11	12	R
N022	HUSA66	HAUSA	17	15	30	25	11	14	16	17,18	13	14	R
N023	HUSA74	HAUSA	14	14	32	21	10	11	13	15,17	11	12	E1b1a8
N024	HUSA17	HAUSA	15	13	27	21	10	11	14	17,18	11	11	E1b1a7
N025	HUSA36	HAUSA	16	13	30	21	10	11	14	17,18	11	12	E1b1a7
N026	HUSA16	HAUSA	17	13	30	21	10	11	14	16,17	11	12	E1b1a7
N027	HUSA3	HAUSA	16	12	29	21	10	11	15	19,19	10	14	E1b1a7
N028	HUSA1	HAUSA	15	13	31	21	10	11	13	17,18	11	12	E1b1a7
N029	HUSA10	HAUSA	17	14	33	21	10	11	14	16,18	11	13	E1b1a7
N030	HUSA28	HAUSA	15	13	32	26	10	11	14	22,22	11	12	E
N031	HUSA29	HAUSA	13	14	30	24	9	11	13	13,14	10	10	E1b1b
N032	HUSA38	HAUSA	15	14	30	18	10	12	14	13,13	11	14	E1b1a

N033	HUSA9	HAUSA	16	14	30	24	10	11	13	13,13	11	13	R
N034	HUSA2	HAUSA	15	14	30	24	10	13	14	18,20	12	12	B
N035	HUSA26	HAUSA	14	13	30	21	10	11	13	16,16	10	12	E1b1a
N036	HUSA7	HAUSA	17	13	29	21	10	11	14	12,14	11	11	B
N037	HUSA18	HAUSA	15	14	30	24	10	13	12	14,14	12	12	R
N038	HUSA19	HAUSA	13	13	30	22	10	14	13	14,15	10	12	E
N039	HUSA20	HAUSA	15	13	29	24	13	13	13	15,20	13	12	E
N040	HUSA40	HAUSA	15	13	30	24	9	13	12	13,15	11	12	E
N041	HUSA59	HAUSA	13	14	30	24	10	13	13	13,15	11	12	R
N042	HUSA8	HAUSA	15	14	31	24	12	12	13	13,15	12	13	R
N043	HUSA11	HAUSA	17	14	30	24	10	11	14	13,13	11	13	R
N044	HUSA12	HAUSA	13	14	31	24	10	11	12	17,18	11	11	E1b1b
N045	HUSA22	HAUSA	17	14	30	24	10	12	13	13,13	12	12	R
N046	HUSA23	HAUSA	16	13	30	24	10	13	15	12,13	11	11	R
N047	HUSA25	HAUSA	16	14	30	24	10	11	14	13,13	11	13	R
N048	HUSA31	HAUSA	15	13	32	21	10	10	13	16,18	11	11	E1b1a8
N049	HUSA4	HAUSA	15	13	30	21	10	11	13	16,18	11	12	E1b1a8
N050	HUSA43	HAUSA	17	13	30	21	10	11	14	20,21	10	12	E1b1a7
N051	HUSA47	HAUSA	16	13	32	21	10	11	14	16,17	11	13	E1b1a7
N052	HUSA53	HAUSA	12	14	32	24	10	12	13	17,19	12	11	E1b1b
N053	HUSA41	HAUSA	16	12	30	22	9	12	13	16,16	10	13	E
N054	HUSA49	HAUSA	17	12	29	21	10	11	16	16,18	11	12	E1b1a7
N055	HUSA5	HAUSA	15	13	32	21	10	11	13	16,17	11	11	E1b1a8
N056	HUSA65	HAUSA	14	12	27	22	8	11	13	15,20	11	11	E1b1a
N057	HUSA73	HAUSA	15	13	32	21	10	10	13	16,17	11	11	E1b1a8
N058	HUSA24	HAUSA	15	12	27	22	10	11	15	11,12	8	12	B
N059	HUSA6	HAUSA	15	13	31	21	12	11	13	16,18	12	11	E1b1a8
N060	HUSA15	HAUSA	16	12	29	22	11	11	13	12,16	8	12	B
N061	HUSA21	HAUSA	16	12	30	21	10	11	14	16,16	11	13	E1b1a7
N062	HUSA30	HAUSA	13	11	28	22	10	11	13	15,16	8	12	E
N063	HUSA32	HAUSA	16	13	32	23	12	14	13	18,20	11	14	E
N064	HUSA35	HAUSA	17	16	33	23	12	14	15	20,21	13	14	E
N065	HUSA37	HAUSA	15	13	32	21	10	10	13	16,18	11	11	E1b1a8
N066	HUSA39	HAUSA	16	14	30	24	10	14	13	13,15	12	12	R

N067	HUSA42	HAUSA	15	13	32	21	10	10	13	17,18	11	11	E1b1a7
N068	HUSA44	HAUSA	16	12	29	21	10	12	14	16,18	11	13	E1b1a7
N069	HUSA45	HAUSA	17	13	32	23	10	14	12	17,18	13	11	E
N070	HUSA46	HAUSA	15	13	32	21	10	10	13	16,18	11	11	E1b1a8
N071	HUSA48	HAUSA	15	13	30	20	10	11	13	16,18	11	11	E1b1a8
N072	HUSA50	HAUSA	15	13	30	20	10	11	14	15,16	11	12	E1b1a7
N073	HUSA51	HAUSA	17	12	29	21	11	11	15	16,18	11	12	E1b1a7
N074	HUSA52	HAUSA	15	14	31	24	13	14	13	13,14	12	12	R
N075	HUSA55	HAUSA	13	14	32	24	10	12	13	17,19	12	11	E1b1b
N076	HUSA56	HAUSA	16	12	29	22	11	11	13	15,17	10	12	E1b1a7
N077	HUSA57	HAUSA	16	12	30	22	10	11	13	15,16	10	11	E1b1a
N078	HUSA58	HAUSA	15	13	30	21	10	11	13	16,17	12	13	E1b1a8
N079	HUSA61	HAUSA	15	14	30	18	10	13	11	13,15	9	10	E
N080	HUSA62	HAUSA	15	16	30	24	13	14	16	20,21	11	12	E
N081	HUSA63	HAUSA	15	13	31	21	11	11	13	16,18	11	11	E1b1a8
N082	HUSA64	HAUSA	18	13	30	21	10	11	15	16,18	11	12	E1b1a7
N083	HUSA67	HAUSA	14	12	27	22	10	11	13	15,20	11	11	E
N084	HUSA68	HAUSA	15	12	29	21	10	8	13	15,17	11	11	E1b1a
N085	HUSA69	HAUSA	15	13	32	21	10	10	13	16,18	11	11	E1b1a8
N086	HUSA70	HAUSA	17	13	29	24	10	13	13	14,14	11	11	R
N087	HUSA71	HAUSA	15	13	32	21	10	10	13	16,18	13	11	E
N088	HUSA72	HAUSA	17	13	29	24	10	13	12	11,12	10	14	B
N089	HUSA75	HAUSA	16	14	32	21	10	11	13	15,17	13	12	E
N090	HUSA76	HAUSA	16	14	31	22	11	12	14	13,13	11	13	R
N091	HUSA77	HAUSA	14	14	31	21	10	11	14	11,13	10	12	B
N092	IGBO100	IGBO	17	14	31	21	10	11	14	17,18	11	12	E1b1a7
N093	IGBO102	IGBO	15	13	32	21	10	11	13	17,17	11	11	E1b1a7
N094	IGBO110	IGBO	16	13	32	21	10	11	13	15,17	11	11	E1b1a7
N095	IGBO15	IGBO	15	13	31	21	10	11	14	16,16	11	12	E1b1a7
N096	IGBO19	IGBO	15	12	30	21	11	11	13	17,18	11	11	E1b1a7
N097	IGBO26	IGBO	17	13	30	21	10	11	14	17,18	11	10	E1b1a7
N098	IGBO27	IGBO	16	13	31	22	10	11	13	11,11	10	11	B
N099	IGBO31	IGBO	15	14	30	20	10	12	14	14,16	11	11	E1b1a7
N100	IGBO34	IGBO	17	13	30	21	10	11	13	17,19	11	12	E1b1a7

N101	IGBO52	IGBO	15	13	30	21	10	11	13	17,18	11	12	E1b1a7
N102	IGBO55	IGBO	15	13	32	21	10	11	12	16,16	11	11	E1b1a8
N103	IGBO61	IGBO	17	13	30	21	10	11	13	17,18	11	12	E1b1a7
N104	IGBO67	IGBO	16	14	31	22	10	11	14	16,17	11	11	E1b1a7
N105	IGBO72	IGBO	15	13	31	22	10	11	13	16,18	11	11	E1b1a8
N106	IGBO78	IGBO	15	13	29	21	10	11	12	15,17	9	11	E1b1a8
N107	IGBO82	IGBO	17	13	30	21	10	11	15	16,19	11	12	E1b1a7
N108	IGBO87	IGBO	17	13	30	21	10	11	13	18,18	11	12	E1b1a7
N109	IGBO94	IGBO	15	13	30	21	10	11	13	17,19	11	12	E1b1a7
N110	IGBO99	IGBO	15	13	30	24	12	11	13	11,12	10	13	B
N111	IGBO1	IGBO	15	13	32	21	10	11	13	17,18	11	11	E1b1a7
N112	IGBO108	IGBO	15	13	31	21	10	11	13	15,20	11	12	E1b1a8
N113	IGBO116	IGBO	16	13	31	21	10	11	13	16,18	11	12	E1b1a7
N114	IGBO118	IGBO	16	13	30	21	10	11	14	17,18	11	11	E1b1a7
N115	IGBO119	IGBO	17	13	33	21	10	11	15	17,18	11	12	E1b1a7
N116	IGBO12	IGBO	17	13	31	21	10	11	15	16,18	11	13	E1b1a7
N117	IGBO14	IGBO	15	14	32	21	10	11	13	15,17	11	13	E1b1a8
N118	IGBO21	IGBO	15	13	31	21	9	11	13	16,17	11	11	E1b1a8
N119	IGBO32	IGBO	15	11	29	22	10	11	11	14,17	8	10	E
N120	IGBO36	IGBO	16	13	30	21	10	11	16	16,21	11	12	E1b1a7
N121	IGBO46	IGBO	16	13	30	21	10	11	15	16,18	11	13	E1b1a7
N122	IGBO47	IGBO	17	13	31	21	10	11	14	17,19	12	11	E1b1a7
N123	IGBO49	IGBO	16	13	30	21	10	11	13	17,18	11	12	E1b1a7
N124	IGBO6	IGBO	16	13	30	21	10	12	15	17,18	11	12	E1b1a7
N125	IGBO62	IGBO	16	13	30	21	10	11	14	16,18	11	12	E1b1a7
N126	IGBO70	IGBO	17	13	30	21	10	11	13	17,18	11	12	E1b1a7
N127	IGBO71	IGBO	16	13	30	21	10	11	13	17,18	11	12	E1b1a7
N128	IGBO73	IGBO	15	13	31	21	10	11	13	16,17	11	12	E1b1a8
N129	IGBO74	IGBO	17	13	30	21	10	11	13	18,18	11	12	E1b1a7
N130	IGBO75	IGBO	15	13	30	21	10	11	14	16,17	11	12	E1b1a7
N131	IGBO8	IGBO	15	13	32	21	10	11	13	17,17	11	12	E1b1a7
N132	IGBO84	IGBO	17	13	31	21	10	14	15	17,18	11	11	E1b1a7
N133	IGBO92	IGBO	15	13	30	21	10	11	15	17,18	11	13	E1b1a7
N134	IGBO109	IGBO	17	13	30	21	10	11	14	17,18	11	12	E1b1a7

N135	IGBO13	IGBO	14	12	31	21	10	11	13	16,17	11	12	E1b1a8
N136	IGBO17	IGBO	15	13	31	21	10	11	12	17,17	11	12	E1b1a7
N137	IGBO30	IGBO	16	13	29	21	10	11	14	17,17	11	12	E1b1a7
N138	IGBO50	IGBO	14	12	30	21	10	11	13	13,16	11	12	E1b1a
N139	IGBO104	IGBO	15	13	30	21	10	11	14	15,15	11	13	E1b1a7
N140	IGBO106	IGBO	15	13	31	21	10	11	13	15,16	11	11	E1b1a8
N141	IGBO112	IGBO	17	14	27	21	10	11	14	17,18	11	12	E1b1a7
N142	IGBO16	IGBO	15	13	30	21	10	11	15	17,18	11	12	E1b1a7
N143	IGBO2	IGBO	15	12	29	21	10	11	15	17,19	12	12	E1b1a7
N144	IGBO20	IGBO	16	12	29	21	9	11	15	17,19	11	11	E1b1a7
N145	IGBO40	IGBO	17	13	30	21	10	11	14	17,17	11	11	E1b1a7
N146	IGBO69	IGBO	17	13	31	21	10	11	13	17,19	11	11	E1b1a7
N147	IGBO7	IGBO	14	13	32	21	10	11	13	16,17	11	13	E1b1a8
N148	IGBO90	IGBO	17	13	30	21	10	11	13	17,18	11	12	E1b1a7
N149	IGBO95	IGBO	16	13	30	21	10	11	13	16,17	11	12	E1b1a7
N150	IGBO97	IGBO	16	12	29	24	10	11	13	16,18	13	14	E
N151	IGBO10	IGBO	16	12	31	22	9	13	14	15,16	11	12	E1b1a7
N152	IGBO101	IGBO	14	13	30	21	10	11	16	16,17	11	12	E1b1a7
N153	IGBO103	IGBO	16	13	30	21	10	11	13	17,18	11	13	E1b1a7
N154	IGBO105	IGBO	15	13	31	22	10	11	14	17,19	11	12	E1b1a7
N155	IGBO107	IGBO	15	13	30	21	11	11	15	16,17	11	11	E1b1a7
N156	IGBO11	IGBO	15	13	30	21	9	11	13	18,18	11	11	E1b1a7
N157	IGBO111	IGBO	17	14	31	21	10	11	14	17,18	11	12	E1b1a7
N158	IGBO113	IGBO	17	14	31	21	10	11	15	17,17	11	13	E1b1a7
N159	IGBO114	IGBO	16	14	32	21	11	11	13	17,19	11	11	E1b1a7
N160	IGBO115	IGBO	15	12	30	21	11	11	15	16,18	11	12	E1b1a7
N161	IGBO117	IGBO	15	13	31	21	10	12	12	17,20	11	11	E1b1a7
N162	IGBO18	IGBO	17	13	30	21	10	11	15	16,17	11	13	E1b1a7
N163	IGBO22	IGBO	14	13	31	21	10	11	13	16,18	11	12	E1b1a
N164	IGBO24	IGBO	16	13	30	21	11	11	13	18,18	11	10	E1b1a7
N165	IGBO28	IGBO	17	13	30	21	10	11	15	17,18	11	12	E1b1a7
N166	IGBO29	IGBO	15	13	30	21	10	11	13	15,17	11	11	E1b1a8
N167	IGBO3	IGBO	15	13	31	21	10	11	13	16,17	11	12	E1b1a8
N168	IGBO33	IGBO	17	13	30	21	10	11	13	17,17	11	12	E1b1a7

N169	IGBO35	IGBO	15	12	29	22	11	12	13	13,15	9	11	
N170	IGBO37	IGBO	15	13	30	21	11	11	13	15,16	11	11	E1b1a8
N171	IGBO38	IGBO	15	14	31	21	10	11	13	18,18	11	12	E1b1a7
N172	IGBO39	IGBO	15	12	30	21	11	11	13	16,16	11	13	E1b1a7
N173	IGBO4	IGBO	17	13	31	21	10	11	15	16,17	11	12	E1b1a7
N174	IGBO41	IGBO	16	13	30	21	10	11	14	17,19	11	12	E1b1a7
N175	IGBO42	IGBO	16	13	30	21	10	11	14	17,18	11	12	E1b1a7
N176	IGBO43	IGBO	15	14	31	21	10	11	14	17,18	11	11	E1b1a7
N177	IGBO44	IGBO	15	13	31	21	11	11	12	16,17	11	13	E1b1a8
N178	IGBO45	IGBO	17	13	30	21	10	11	13	18,18	11	12	E1b1a7
N179	IGBO48	IGBO	15	13	31	21	10	11	13	15,17	11	11	E1b1a8
N180	IGBO51	IGBO	14	12	28	25	10	11	13	14,18	11	11	E
N181	IGBO53	IGBO	17	13	30	21	10	11	14	17,20	11	13	E1b1a7
N182	IGBO54	IGBO	16	13	30	21	10	11	13	16,18	11	11	E1b1a7
N183	IGBO56	IGBO	16	13	30	21	10	11	13	16,18	11	11	E1b1a7
N184	IGBO57	IGBO	16	13	30	21	10	11	13	17,18	11	13	E1b1a7
N185	IGBO58	IGBO	17	12	29	21	11	11	15	17,17	11	12	E1b1a7
N186	IGBO59	IGBO	17	13	30	21	10	11	14	17,17	11	12	E1b1a7
N187	IGBO60	IGBO	17	13	31	21	10	11	13	15,17	12	11	E1b1a7
N188	IGBO64	IGBO	17	13	30	21	10	11	14	17,19	11	13	E1b1a7
N189	IGBO65	IGBO	16	13	30	22	10	11	15	11,12	11	13	B
N190	IGBO68	IGBO	17	13	32	21	10	11	13	18,18	11	12	E1b1a7
N191	IGBO76	IGBO	16	13	30	21	10	11	13	17,17	11	12	E1b1a7
N192	IGBO77	IGBO	15	13	29	21	10	11	13	17,18	11	12	E1b1a7
N193	IGBO79	IGBO	15	13	29	21	10	11	12	17,19	11	11	E1b1a7
N194	IGBO80	IGBO	17	13	30	21	10	11	15	18,19	11	13	E1b1a7
N195	IGBO81	IGBO	15	13	30	21	10	11	13	16,18	11	11	E1b1a8
N196	IGBO86	IGBO	17	12	30	20	10	12	15	17,17	11	12	E1b1a7
N197	IGBO88	IGBO	15	13	30	21	10	11	14	17,18	11	11	E1b1a7
N198	IGBO89	IGBO	17	13	31	21	10	11	15	16,17	11	12	E1b1a7
N199	IGBO9	IGBO	15	14	31	21	10	11	14	17,18	11	12	E1b1a7
N200	IGBO91	IGBO	16	13	31	21	10	11	13	15,17	11	12	E1b1a7
N201	IGBO93	IGBO	16	13	30	21	10	11	13	17,17	11	12	E1b1a7
N202	IGBO98	IGBO	16	13	30	21	11	11	13	18,18	11	12	E1b1a7

N203	IGBO23	IGBO	16	12	28	21	10	11	15	15,17	11	12	E1b1a7
N204	IGBO25	IGBO	17	13	30	21	11	11	14	17,19	11	11	E1b1a7
N205	IGBO5	IGBO	16	13	30	21	10	11	15	15,18	11	13	E1b1a7
N206	IGBO63	IGBO	16	14	31	21	10	11	14	17,17	11	12	E1b1a7
N207	IGBO66	IGBO	17	14	32	21	10	11	14	18,18	11	13	E1b1a7
N208	IGBO83	IGBO	16	14	31	21	10	11	14	17,18	11	12	E1b1a7
N209	IGBO96	IGBO	17	13	29	21	10	11	15	15,21	11	11	E1b1a7
N210	IGBO85	IGBO	17	13	30	21	10	12	13	17,19	11	12	E1b1a7
N211	IJAW1	IJAW	15	12	29	21	10	11	13	14,16	10	13	E1b1a8
N212	IJAW11	IJAW	17	14	31	21	10	11	15	17,18	11	12	E1b1a7
N213	IJAW14	IJAW	17	13	30	21	10	11	14	17,18	11	13	E1b1a7
N214	IJAW5	IJAW	16	13	31	21	11	13	15	16,21	12	12	E1b1a7
N215	IJAW6	IJAW	15	14	33	21	10	11	13	15,16	11	12	E1b1a7
N216	IJAW7	IJAW	14	13	30	21	10	11	14	16,17	11	12	E1b1a7
N217	IJAW9	IJAW	17	13	30	21	10	11	14	16,17	11	12	E1b1a7
N218	IJAW10	IJAW	15	12	29	22	10	11	14	14,15	9	13	E
N219	IJAW12	IJAW	17	13	30	21	10	11	15	16,17	11	11	E1b1a7
N220	IJAW13	IJAW	13	13	30	21	10	11	15	20,22	11	11	E1b1a
N221	IJAW15	IJAW	17	13	30	21	10	11	15	16,17	11	11	E1b1a7
N222	IJAW2	IJAW	15	13	33	21	10	11	13	17,17	11	12	E1b1a7
N223	IJAW3	IJAW	15	13	31	21	10	11	15	16,17	11	11	E1b1a7
N224	IJAW4	IJAW	16	13	30	21	10	11	15	17,17	11	11	E1b1a7
N225	IJAW8	IJAW	17	13	30	21	10	11	16	21,22	11	12	E1b1a7
N226	YRBA103	YORUBA	17	16	32	24	12	14	16	19,21	13	13	E
N227	YRBA110	YORUBA	15	13	29	21	10	11	12	16,17	11	11	E1b1a8
N228	YRBA169	YORUBA	17	13	30	21	10	11	14	14,18	11	12	E1b1a7
N229	YRBA187	YORUBA	15	13	32	21	10	11	13	20,22	11	13	E1b1a
N230	YRBA52	YORUBA	16	13	30	21	10	11	14	17,18	11	11	E1b1a7
N231	YRBA80	YORUBA	15	13	32	21	10	11	13	16,17	11	12	E1b1a8
N232	YRBA87	YORUBA	16	13	32	21	10	11	13	15,18	11	12	E1b1a7
N233	YRBA119	YORUBA	17	14	32	21	10	11	15	15,16	11	11	E1b1a7
N234	YRBA136	YORUBA	16	13	30	20	10	11	15	17,17	11	13	E1b1a7
N235	YRBA167	YORUBA	17	13	31	21	11	11	13	17,17	11	12	E1b1a7
N236	YRBA17	YORUBA	15	12	32	21	10	11	13	16,16	12	12	E1b1a7

N237	YRBA173	YORUBA	14	12	30	22	10	11	12	15,15	10	12	E1b1a
N238	YRBA176	YORUBA	17	13	30	22	10	11	14	17,18	11	12	E1b1a7
N239	YRBA178	YORUBA	16	13	32	21	10	10	13	16,18	11	11	E1b1a7
N240	YRBA190	YORUBA	17	13	30	21	10	11	14	17,18	11	12	E1b1a7
N241	YRBA217	YORUBA	17	14	31	21	11	11	15	15,18	11	14	E1b1a7
N242	YRBA226	YORUBA	15	13	31	21	11	11	13	16,17	11	12	E1b1a8
N243	YRBA235	YORUBA	15	14	32	21	10	11	13	16,17	11	12	E1b1a8
N244	YRBA25	YORUBA	16	12	29	21	10	11	15	17,19	10	14	E1b1a7
N245	YRBA50	YORUBA	16	13	30	21	9	12	15	17,19	11	12	E1b1a7
N246	YRBA56	YORUBA	17	13	31	21	10	11	15	16,17	11	11	E1b1a7
N247	YRBA68	YORUBA	15	13	30	21	10	11	13	11,13	11	12	B
N248	YRBA69	YORUBA	15	13	30	21	10	11	14	17,18	11	12	E1b1a7
N249	YRBA76	YORUBA	16	13	30	21	10	12	14	14,17	11	11	E1b1a7
N250	YRBA77	YORUBA	14	13	31	21	10	11	13	16,17	11	12	E1b1a8
N251	YRBA78	YORUBA	15	13	32	22	10	11	13	16,18	10	12	E1b1a8
N252	YRBA79	YORUBA	14	12	31	21	9	11	13	17,17	11	11	E1b1a7
N253	YRBA82	YORUBA	17	12	30	21	10	11	14	17,18	11	12	E1b1a7
N254	YRBA91	YORUBA	15	13	32	21	10	11	13	15,18	11	12	E1b1a8
N255	YRBA153	YORUBA	17	13	31	21	10	11	14	18,18	11	12	E1b1a7
N256	YRBA184	YORUBA	16	13	30	21	10	11	13	16,18	11	11	E1b1a7
N257	YRBA75	YORUBA	17	14	31	21	10	12	15	16,18	11	11	E1b1a7
N258	YRBA10	YORUBA	15	14	30	24	11	13	13	13,16	12	12	R
N259	YRBA126	YORUBA	17	13	30	21	10	11	16	16,18	11	12	E1b1a7
N260	YRBA131	YORUBA	17	13	30	21	9	11	15	16,18	11	12	E1b1a7
N261	YRBA141	YORUBA	16	13	30	21	10	11	14	18,18	11	12	E1b1a7
N262	YRBA150	YORUBA	15	13	30	21	10	11	15	16,18	11	14	E1b1a7
N263	YRBA166	YORUBA	15	13	32	21	10	11	13	17,18	11	12	E1b1a7
N264	YRBA183	YORUBA	13	14	31	24	10	11	12	17,18	11	11	E1b1b
N265	YRBA185	YORUBA	15	12	29	21	10	11	13	16,17	11	11	E1b1a8
N266	YRBA218	YORUBA	15	12	31	21	10	12	13	17,18	11	12	E1b1a7
N267	YRBA219	YORUBA	15	13	31	21	10	11	13	16,16	11	12	E1b1a8
N268	YRBA229	YORUBA	16	13	31	21	10	11	14	16,17	11	12	E1b1a7
N269	YRBA230	YORUBA	16	13	30	21	10	11	15	17,17	11	12	E1b1a7
N270	YRBA231	YORUBA	15	13	30	21	9	11	15	17,18	11	12	E1b1a7

N271	YRBA232	YORUBA	15	13	31	21	10	11	14	16,17	11	12	E1b1a7
N272	YRBA28	YORUBA	15	13	31	25	9	11	13	11,11	10	12	B
N273	YRBA44	YORUBA	16	13	30	21	10	11	15	16,18	11	12	E1b1a7
N274	YRBA6	YORUBA	17	14	32	21	10	11	14	17,18	11	12	E1b1a7
N275	YRBA65	YORUBA	17	13	30	21	10	11	15	17,18	11	12	E1b1a7
N276	YRBA70	YORUBA	15	13	30	21	10	11	15	16,17	11	12	E1b1a7
N277	YRBA81	YORUBA	15	13	30	21	10	11	13	15,17	11	12	E1b1a8
N278	YRBA97	YORUBA	17	13	31	21	10	11	14	17,18	11	12	E1b1a7
N279	YRBA107	YORUBA	16	13	31	21	10	12	15	18,18	11	12	E1b1a7
N280	YRBA109	YORUBA	15	13	30	21	10	11	14	17,18	11	12	E1b1a7
N281	YRBA111	YORUBA	17	13	30	21	10	11	14	15,16	11	12	E1b1a7
N282	YRBA115	YORUBA	15	13	30	21	10	11	14	19,19	11	12	E1b1a7
N283	YRBA117	YORUBA	16	14	31	21	10	11	14	17,17	11	12	E1b1a7
N284	YRBA12	YORUBA	17	13	30	21	10	11	14	17,20	11	12	E1b1a7
N285	YRBA123	YORUBA	16	13	30	21	10	11	13	17,18	11	12	E1b1a7
N286	YRBA129	YORUBA	17	13	30	21	10	11	14	17,17	11	12	E1b1a7
N287	YRBA130	YORUBA	15	13	34	21	10	11	13	15,16	10	11	E1b1a8
N288	YRBA134	YORUBA	15	13	31	21	10	11	13	16,17	11	13	E1b1a8
N289	YRBA135	YORUBA	17	12	28	22	10	10	13	14,14	8	10	E
N290	YRBA138	YORUBA	15	13	31	21	10	11	16	16,16	11	11	E1b1a7
N291	YRBA14	YORUBA	15	13	32	21	10	11	14	16,17	11	12	E1b1a7
N292	YRBA142	YORUBA	14	13	29	21	10	13	14	13,13	11	11	B
N293	YRBA155	YORUBA	15	14	32	21	10	11	13	16,17	12	11	E1b1a8
N294	YRBA156	YORUBA	17	13	31	21	10	11	15	17,18	11	12	E1b1a7
N295	YRBA16	YORUBA	15	13	31	21	10	11	13	16,17	11	12	E1b1a8
N296	YRBA165	YORUBA	17	13	30	21	10	11	14	17,18	11	12	E1b1a7
N297	YRBA174	YORUBA	15	13	30	21	10	11	14	16,18	11	12	E1b1a7
N298	YRBA175	YORUBA	15	13	30	21	10	11	13	16,18	11	11	E1b1a8
N299	YRBA181	YORUBA	17	13	32	21	10	11	13	16,17	11	13	E1b1a7
N300	YRBA182	YORUBA	15	13	29	21	10	11	13	21,22	11	11	E1b1a
N301	YRBA204	YORUBA	17	13	30	21	10	11	14	17,17	11	11	E1b1a7
N302	YRBA206	YORUBA	15	13	30	21	10	11	14	17,18	11	12	E1b1a7
N303	YRBA207	YORUBA	16	13	30	21	10	11	13	16,17	11	12	E1b1a7
N304	YRBA211	YORUBA	15	13	32	21	10	11	14	14,16	11	12	E1b1a7

N305	YRBA22	YORUBA	15	13	31	21	12	11	14	15,17	11	12	E1b1a7
N306	YRBA227	YORUBA	17	13	30	21	10	11	14	17,20	11	12	E1b1a7
N307	YRBA23	YORUBA	15	14	32	21	11	11	13	17,19	11	11	E1b1a7
N308	YRBA234	YORUBA	15	13	31	21	10	11	13	16,17	11	12	E1b1a8
N309	YRBA238	YORUBA	17	13	30	21	10	11	15	17,17	11	12	E1b1a7
N310	YRBA35	YORUBA	17	13	30	21	11	11	15	18,18	11	12	E1b1a7
N311	YRBA4	YORUBA	13	13	30	21	10	11	15	16,17	11	12	E1b1a7
N312	YRBA43	YORUBA	15	13	32	21	10	11	13	16,18	11	13	E1b1a8
N313	YRBA48	YORUBA	14	13	31	21	9	11	13	16,18	11	12	E1b1a8
N314	YRBA49	YORUBA	16	12	28	21	10	11	15	15,17	11	12	E1b1a7
N315	YRBA5	YORUBA	15	14	32	21	10	11	14	16,20	11	13	E1b1a
N316	YRBA51	YORUBA	15	13	31	21	10	11	14	16,16	11	12	E1b1a7
N317	YRBA55	YORUBA	17	13	31	21	10	11	15	16,17	11	11	E1b1a7
N318	YRBA57	YORUBA	17	13	31	21	10	11	15	16,18	11	12	E1b1a7
N319	YRBA58	YORUBA	17	12	29	22	10	12	12	16,17	10	11	E1b1a7
N320	YRBA60	YORUBA	16	13	30	21	10	11	15	17,19	11	12	E1b1a7
N321	YRBA62	YORUBA	16	13	29	21	10	11	15	17,22	11	11	E1b1a7
N322	YRBA67	YORUBA	17	12	28	22	10	11	13	14,14	8	10	E
N323	YRBA71	YORUBA	16	14	31	21	9	11	15	17,17	11	12	E1b1a7
N324	YRBA88	YORUBA	15	13	31	21	10	10	13	16,18	11	13	E1b1a7
N325	YRBA9	YORUBA	17	13	30	21	10	11	14	17,18	11	12	E1b1a7
N326	YRBA93	YORUBA	14	13	27	19	10	10	13	16,16	11	11	E1b1a8
N327	YRBA100	YORUBA	16	13	31	21	10	11	15	17,17	11	13	E1b1a7
N328	YRBA102	YORUBA	17	13	30	24	10	11	12	15,19	13	11	
N329	YRBA108	YORUBA	15	13	30	21	10	11	14	18,19	11	12	E1b1a7
N330	YRBA112	YORUBA	16	13	27	24	10	11	13	11,14	8	12	B
N331	YRBA116	YORUBA	15	13	30	21	10	11	14	17,18	11	13	E1b1a7
N332	YRBA120	YORUBA	15	13	30	21	10	11	14	15,16	11	12	E1b1a7
N333	YRBA122	YORUBA	15	13	29	21	10	11	12	15,17	9	11	E1b1a8
N334	YRBA13	YORUBA	17	13	29	21	10	11	15	17,17	11	12	E1b1a7
N335	YRBA133	YORUBA	15	13	30	21	11	11	15	17,18	11	12	E1b1a7
N336	YRBA137	YORUBA	15	13	30	21	11	11	14	14,17	11	12	E1b1a
N337	YRBA148	YORUBA	15	14	32	21	10	11	14	16,18	11	12	E1b1a7
N338	YRBA15	YORUBA	16	13	32	21	10	11	13	16,16	12	12	E1b1a7

N339	YRBA157	YORUBA	18	13	31	21	10	11	15	17,18	11	11	E1b1a7
N340	YRBA158	YORUBA	17	12	30	22	10	12	12	13,15	10	11	E
N341	YRBA19	YORUBA	16	13	31	21	10	11	14	16,18	11	12	E1b1a7
N342	YRBA191	YORUBA	15	13	30	21	10	11	13	16,18	11	12	E1b1a8
N343	YRBA195	YORUBA	17	13	30	21	11	11	15	17,18	11	12	E1b1a7
N344	YRBA197	YORUBA	16	13	30	21	11	12	16	17,18	9	12	E1b1a7
N345	YRBA202	YORUBA	17	13	30	21	10	11	14	18,18	11	12	E1b1a7
N346	YRBA208	YORUBA	15	12	29	21	10	11	14	14,17	11	12	E1b1a7
N347	YRBA210	YORUBA	15	13	32	21	10	11	13	16,17	11	11	E1b1a8
N348	YRBA213	YORUBA	16	13	30	21	10	12	14	16,18	11	12	E1b1a7
N349	YRBA215	YORUBA	17	12	30	26	10	12	12	13,14	10	11	B
N350	YRBA24	YORUBA	17	13	30	21	9	11	15	17,17	11	13	E1b1a7
N351	YRBA27	YORUBA	15	13	30	21	10	11	13	16,18	11	12	E1b1a8
N352	YRBA29	YORUBA	16	14	32	21	10	11	15	17,17	11	12	E1b1a7
N353	YRBA30	YORUBA	15	14	33	21	10	11	14	16,18	12	12	E1b1a7
N354	YRBA31	YORUBA	17	12	29	21	10	11	14	18,18	11	12	E1b1a7
N355	YRBA59	YORUBA	15	13	31	21	9	11	13	16,17	11	12	E1b1a8
N356	YRBA61	YORUBA	16	13	30	21	10	11	16	16,18	11	12	E1b1a7
N357	YRBA66	YORUBA	15	13	31	21	9	11	13	16,17	11	11	E1b1a8
N358	YRBA7	YORUBA	17	13	30	21	10	11	14	17,18	11	12	E1b1a7
N359	YRBA72	YORUBA	15	11	30	27	10	11	14	17,18	13	11	B
N360	YRBA73	YORUBA	15	12	30	24	10	11	14	11,11	10	11	B
N361	YRBA84	YORUBA	14	13	30	21	10	11	13	16,17	11	13	E1b1a8
N362	YRBA86	YORUBA	15	13	32	21	10	11	13	17,17	11	11	E1b1a7
N363	YRBA90	YORUBA	17	14	31	21	10	11	15	17,18	11	14	E1b1a7
N364	YRBA92	YORUBA	15	14	32	21	10	12	13	16,17	11	12	E1b1a8
N365	YRBA1	YORUBA	15	12	29	21	11	11	14	18,18	11	12	E1b1a7
N366	YRBA121	YORUBA	15	12	29	21	10	11	14	17,20	11	12	E1b1a7
N367	YRBA127	YORUBA	17	12	29	21	10	11	15	16,18	11	14	E1b1a7
N368	YRBA128	YORUBA	14	13	30	20	10	11	13	18,18	11	11	E1b1a7
N369	YRBA146	YORUBA	15	14	32	21	10	11	13	15,17	11	12	E1b1a8
N370	YRBA159	YORUBA	13	12	29	22	10	11	12	14,14	10	13	E
N371	YRBA161	YORUBA	15	14	31	21	10	11	14	17,17	11	11	E1b1a7
N372	YRBA162	YORUBA	17	14	32	22	10	11	14	18,19	11	12	E1b1a7

N373	YRBA168	YORUBA	15	12	28	21	10	11	14	12,12	10	14	B
N374	YRBA196	YORUBA	16	13	30	21	10	11	14	16,17	9	12	E1b1a7
N375	YRBA198	YORUBA	15	14	32	21	10	11	13	16,19	11	12	E1b1a7
N376	YRBA199	YORUBA	16	12	29	21	10	11	13	16,17	11	12	E1b1a7
N377	YRBA220	YORUBA	15	13	30	21	10	11	13	14,16	11	12	E1b1a8
N378	YRBA225	YORUBA	15	12	30	22	10	11	14	16,17	11	13	E1b1a7
N379	YRBA236	YORUBA	17	12	29	21	10	11	14	17,20	11	12	E1b1a7
N380	YRBA237	YORUBA	14	13	31	21	11	11	13	16,17	11	12	E1b1a8
N381	YRBA3	YORUBA	17	14	31	21	10	11	14	16,19	11	13	E1b1a7
N382	YRBA32	YORUBA	15	13	30	21	11	11	13	15,16	11	12	E1b1a8
N383	YRBA33	YORUBA	15	13	32	21	10	12	14	16,17	11	12	E1b1a7
N384	YRBA38	YORUBA	16	12	30	21	10	11	14	14,17	11	11	E1b1a7
N385	YRBA42	YORUBA	17	13	30	21	10	11	14	17,18	11	12	E1b1a7
N386	YRBA45	YORUBA	16	13	30	21	10	11	15	16,18	11	11	E1b1a7
N387	YRBA53	YORUBA	15	14	32	22	10	11	13	15,17	11	13	E1b1a7
N388	YRBA85	YORUBA	15	12	29	21	10	11	13	16,17	11	11	E1b1a8
N389	YRBA143	YORUBA	15	13	30	21	10	10	13	16,17	11	11	E1b1a8
N390	YRBA180	YORUBA	16	13	30	21	10	11	13	20,22	11	11	
N391	YRBA106	YORUBA	15	13	31	21	10	11	14	16,16	11	12	E1b1a7
N392	YRBA114	YORUBA	17	12	29	21	10	11	14	18,18	11	12	E1b1a7
N393	YRBA118	YORUBA	15	14	34	21	10	11	13	16,19	11	11	E1b1a7
N394	YRBA132	YORUBA	17	13	30	21	10	11	16	14,18	11	13	E1b1a7
N395	YRBA139	YORUBA	16	13	31	22	10	11	13	16,19	11	12	E1b1a7
N396	YRBA145	YORUBA	16	13	31	21	10	11	13	15,17	12	14	E1b1a
N397	YRBA152	YORUBA	17	13	30	21	10	11	13	16,17	11	11	E1b1a7
N398	YRBA163	YORUBA	15	14	31	21	10	11	14	11,13	10	12	B
N399	YRBA164	YORUBA	15	12	30	21	11	11	13	17,17	11	11	E1b1a7
N400	YRBA171	YORUBA	15	13	31	21	10	11	13	16,18	11	12	E1b1a8
N401	YRBA172	YORUBA	16	14	31	21	10	11	14	16,17	11	11	E1b1a7
N402	YRBA177	YORUBA	13	13	30	21	10	11	13	15,16	11	12	E1b1a8
N403	YRBA179	YORUBA	18	13	33	21	10	11	16	19,20	12	14	E1b1a7
N404	YRBA18	YORUBA	15	13	31	21	12	11	14	15,17	11	12	E1b1a7
N405	YRBA192	YORUBA	15	13	31	21	11	11	14	15,17	11	12	E1b1a7
N406	YRBA193	YORUBA	15	13	31	21	10	11	13	16,18	11	11	E1b1a8

N407	YRBA201	YORUBA	15	13	31	21	11	11	14	15,17	11	12	E1b1a7
N408	YRBA203	YORUBA	16	14	32	21	10	11	14	16,19	11	12	E1b1a7
N409	YRBA212	YORUBA	15	13	30	21	10	11	14	15,16	11	14	E1b1a
N410	YRBA216	YORUBA	17	13	31	21	10	11	13	17,17	13	12	E1b1a7
N411	YRBA221	YORUBA	17	13	30	21	10	11	14	18,18	11	12	E1b1a7
N412	YRBA222	YORUBA	15	13	30	21	10	11	13	17,18	11	12	E1b1a7
N413	YRBA223	YORUBA	16	14	32	21	10	11	13	12,12	10	11	B
N414	YRBA224	YORUBA	15	14	31	22	10	11	13	16,17	10	12	E1b1a8
N415	YRBA26	YORUBA	15	13	32	21	10	11	13	16,18	11	13	E1b1a8
N416	YRBA37	YORUBA	17	12	28	21	10	11	14	12,12	10	13	B
N417	YRBA40	YORUBA	17	13	31	21	11	11	14	17,17	11	12	E1b1a7
N418	YRBA64	YORUBA	16	13	30	21	11	11	16	17,18	11	12	E1b1a7
N419	YRBA74	YORUBA	15	13	30	21	10	11	13	15,17	11	12	E1b1a8
N420	YRBA83	YORUBA	16	13	32	21	10	10	13	16,18	11	13	E1b1a7
N421	YRBA96	YORUBA	14	13	31	21	10	11	13	16,17	11	12	E1b1a8
N422	YRBA98	YORUBA	17	14	31	21	10	11	15	17,19	11	12	E1b1a7
N423	YRBA99	YORUBA	15	13	30	21	10	11	14	17,18	11	13	E1b1a7
N424	YRBA101	YORUBA	17	12	30	21	10	11	14	17,18	11	13	E1b1a7
N425	YRBA104	YORUBA	17	13	30	21	10	13	14	17,18	9	12	E1b1a7
N426	YRBA105	YORUBA	16	13	30	21	10	11	15	17,18	11	12	E1b1a7
N427	YRBA11	YORUBA	17	13	31	21	10	11	13	17,17	11	12	E1b1a7
N428	YRBA113	YORUBA	16	13	30	21	10	11	13	15,16	9	13	E
N429	YRBA124	YORUBA	17	13	30	21	10	11	14	16,17	11	11	E1b1a7
N430	YRBA125	YORUBA	16	13	31	21	10	11	13	16,17	11	12	E1b1a7
N431	YRBA140	YORUBA	16	14	32	21	10	11	14	16,17	11	13	E1b1a7
N432	YRBA144	YORUBA	17	13	31	21	9	12	15	16,17	11	12	E1b1a7
N433	YRBA147	YORUBA	15	13	33	21	10	11	13	16,19	11	11	E1b1a7
N434	YRBA149	YORUBA	17	13	30	21	10	11	15	18,18	11	12	E1b1a7
N435	YRBA151	YORUBA	15	12	30	21	10	10	13	16,16	11	12	E1b1a8
N436	YRBA154	YORUBA	15	13	32	22	10	11	13	13,16	11	11	B
N437	YRBA160	YORUBA	15	13	30	21	10	11	15	17,18	11	12	E1b1a7
N438	YRBA170	YORUBA	17	16	34	24	13	13	16	18,20	13	14	
N439	YRBA186	YORUBA	15	13	30	21	10	11	12	16,17	11	11	E1b1a8
N440	YRBA188	YORUBA	15	14	32	21	10	11	13	17,21	11	12	E1b1a7

N441	YRBA189	YORUBA	15	14	30	21	11	11	14	16,18	11	12	E1b1a7
N442	YRBA194	YORUBA	17	14	30	22	10	12	10	17,21	10	10	B
N443	YRBA2	YORUBA	15	13	31	26	10	11	13	16,16	11	12	E
N444	YRBA20	YORUBA	16	13	31	21	10	11	15	17,17	11	12	E1b1a7
N445	YRBA200	YORUBA	15	12	30	21	10	10	13	15,18	11	12	E1b1a8
N446	YRBA205	YORUBA	16	13	30	21	10	11	14	17,17	11	11	E1b1a7
N447	YRBA209	YORUBA	15	13	30	21	10	10	13	16,17	11	11	E1b1a8
N448	YRBA21	YORUBA	15	13	32	21	10	11	13	16,17	11	11	E1b1a8
N449	YRBA214	YORUBA	15	14	31	21	10	11	14	17,18	11	13	E1b1a7
N450	YRBA228	YORUBA	16	13	32	21	10	11	13	16,16	12	12	E1b1a7
N451	YRBA233	YORUBA	15	13	30	21	10	11	14	14,18	11	12	E1b1a7
N452	YRBA34	YORUBA	16	13	29	21	10	11	15	17,18	11	11	E1b1a7
N453	YRBA36	YORUBA	15	13	30	22	10	11	13	16,17	10	11	E1b1a8
N454	YRBA39	YORUBA	17	13	30	21	10	11	13	17,18	11	12	E1b1a7
N455	YRBA41	YORUBA	15	12	29	22	10	11	13	15,15	9	13	E1b1a
N456	YRBA46	YORUBA	15	15	33	21	10	11	14	14,17	11	12	E1b1a7
N457	YRBA47	YORUBA	16	13	30	21	11	11	16	17,18	11	12	E1b1a7
N458	YRBA54	YORUBA	15	13	33	21	10	11	13	17,17	11	12	E1b1a7
N459	YRBA63	YORUBA	15	13	31	21	10	11	12	16,17	11	11	E1b1a8
N460	YRBA8	YORUBA	16	13	30	21	11	11	13	17,17	11	12	E1b1a7
N461	YRBA89	YORUBA	15	13	31	21	10	11	13	16,17	11	12	E1b1a8
N462	YRBA94	YORUBA	15	13	31	21	10	11	14	17,17	11	11	E1b1a7
N463	YRBA95	YORUBA	17	14	31	21	10	11	14	16,20	11	12	E1b1a7

UNIVERSITY of the
WESTERN CAPE