
# Population Genetic Analysis of Forensic DNA Markers in the Nigerian Population

**By**

**Daniel Azubuike Awomukwu**

**June 2025**

# Population Genetic Analysis of Forensic DNA Markers in the Nigerian Population

**By**


## Daniel Azubuike Awomukwu

**Centre for Forensic Science**

**Department of Pure and Applied Science**

**University of Strathclyde**


**A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy**

**Centre for Forensic Science**

**Department of Pure and Applied Science**

**University of Strathclyde**

**PhD Thesis**

**Daniel Azubuike Awomukwu**

**2025**

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.



Date: 28/06/2025

Sign: 

**ABSTRACT**

Nigeria, located in West Africa within sub-Saharan Africa, is the continent's most populous Black nation and ranks as the seventh most populated country globally. It is home to three major ethnic groups: Igbo, Yoruba, and Hausa-Fulani. A thorough study of population genetics is essential for accurately interpreting forensic genetic evidence in criminal investigations. The genetic composition of these ethnic groups has not been extensively investigated compared to global reference populations. This project collected blood samples from 303 unrelated individuals from Nigeria's three major ethnic groups using FTA® cards. The genetic analysis utilised three kits: the 17-locus QIAGEN™ Investigator® ESSplex SE QS Kit and the 21-locus GlobalFiler™ Express Kit for autosomal STRs, along with the 23-locus Promega PowerPlex® Y23 System Kit for Y-chromosomal STRs. Of the 303 blood samples, 167 male samples were analysed for Y-STRs, while all 303 were analysed for autosomal STRs. The autosomal STR data was then analysed to determine forensic parameters, F-statistics, and Hardy-Weinberg equilibrium. Population structure analysis was also conducted to assess inter-subpopulation differentiation using *STRUCTURE*, principal component analysis, and neighbour-joining methods. A total of 606 allele calls were obtained from 16 autosomal STR loci using the QIAGEN™ Investigator® ESSplex SE QS Kit and 21 autosomal STR loci using the GlobalFiler™ Express Kit. The allele frequencies were calculated for the entire population and for each ethnic group individually. The QIAGEN™ Investigator® ESSplex SE QS Kit analysis showed SE33 as the most informative, followed by D2S1338, while TH01 had the fewest allelic variants. Similarly, the GlobalFiler™ Express Kit indicated SE33 as the most informative, with D2S1338 next, but D13S317 had the lowest allelic variants. Both kits had a combined power of discrimination and exclusion exceeding 99.999%. The 21-locus GlobalFiler™ Express Kit, which had no off-ladder alleles and more markers, offered higher discriminatory power, improved likelihood ratios, and lower random match probabilities than the 16-locus QIAGEN™ Investigator® ESSplex SE QS Kit. The *STRUCTURE*, principal component analysis, neighbour-joining, and $F_{ST}$ analyses revealed no genetic differences among the Igbo, Yoruba, and Hausa-Fulani ethnic groups based on autosomal STR data from the QIAGEN™ Investigator ESSplex SE QS Kit and the GlobalFiler™ Express Kit. The Promega PowerPlex® Y23 System Kit analysis identified the DYS385a/b locus as the most informative, while the DYS391 locus had the lowest levels of polymorphic information content, gene diversity, and discrimination capacity. Overall, haplotype diversity and discrimination capacity across all 23 loci were found to be 99.4% and 99.9%, respectively. Principal component and neighbour-joining analyses suggest that the Igbo and Yoruba ethnic groups share a paternal lineage, while a subset of the Hausa-Fulani group appears distinct, indicating admixture. The three subpopulations also showed differences in linkage disequilibrium across specific locus pairs. The study analysed inter-

population genetic distances across five continents, involving 7,664 unrelated people from 24 populations using autosomal STR markers and 7,348 unrelated individuals from 50 populations using Y-STR markers. Y-STR results show Nigerian populations are genetically closely aligned with other West African groups (Niger-Congo) at the continental level but genetically distinct from North African (Afroasiatic), East African (Nilotic), and Indigenous Central African (Pygmy, Khoisan) populations. Autosomal STR findings placed Nigeria within the African clade, with closer ties to diaspora populations in North and South America. This research highlights the forensic implications of population differences and the importance of establishing a national DNA repository and allele frequency data for Nigeria.

## ACKNOWLEDGEMENTS

I am deeply indebted to my supportive supervisor, Dr Penelope Haddrill, for her expert guidance, reassurance and immense care throughout my research. Her dedication to helping me complete this project, no matter how lengthy it became, is unquantifiable. I will always be thankful for her support. I also extend my thanks to Dr. Christine Davidson for her consistent support and motivation, which greatly contributed to my progress.

Furthermore, I am deeply thankful to my family, especially my wife, for encouraging me when I needed it most. Her love, kindness, loyalty, and faith helped me have the courage to follow my goals during the critical stages of this work. Throughout this project, I will always appreciate having my beautiful children by my side.

My immense gratitude goes to my parents, particularly my father, who continuously encouraged me to complete my degree promptly. I am grateful to everyone who contributed in any way to the completion of this project; their support has made it possible to navigate the challenges I encountered along the way.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| **AMOVA** | Analysis of Molecular Variance |
| **BMGF** | Bill and Melinda Gates Foundation (USA) |
| **CND** | Chinese National Database |
| **CODIS** | Combined DNA Index System (USA) |
| **Cq** | Quantification Cycle |
| **CSI** | Crime Scene Investigation |
| **CLUMPAK** | Clustering Markov Packager Across K |
| **°C** | degree Celsius |
| **DC** | Discrimination Capacity |
| **DHS** | Demographic and Health Surveys (Nigeria) |
| **DNA** | Deoxyribonucleic Acid |
| **DVI** | Disaster Victim Identification |
| **ECHR** | European Convention on Human Rights |
| **$F_{IS}$** | Fixation Index - Individual within the Subpopulation |
| **$F_{IT}$** | Fixation Index - Individual within the Total population |
| **FORDEC** | Forensic Research and Development Centre |
| **FSS** | Forensic Science Service (UK) |
| **$F_{ST}$** | Fixation Index - **S**ubpopulation within the Total population |
| **GD** | Gene Diversity |
| **GPI** | Global Peace Index |
| **HD** | Haplotype Diversity |
| **$H_{om\_exp}$** | Expected Homozygosity ($1-H_e$) |
| **$H_{exp}$** | Expected Heterozygosity |
| **$H_I$** | Individual Heterozygosity |
| **HMP** | Haplotype Match Probability |
| **$H_{oms}$** | Observed Homozygosity **(**$1-H_o$) |
| **$H_{obs}$** | Observed Heterozygosity |
| **$H_p$** | Prosecution Hypothesis |
| **$H_S$** | Average Heterozygosity in the **S**ubpopulation |
| **$H_T$** | Average Heterozygosity in an overall **T**otal Population |
| **HWE** | Hardy-Weinberg equilibrium |
| **ICF** | Inner City Fund International, Inc. (USA) |
| **INDEL(S)** | Insertion(s)/ Deletion(s) |
| **IPSA** | International Police Science Association |
| **ISO** | International Organisation for Standardization |
| **K** | Number of Genetic Clusters Being Tested |
| **LD** | Linkage Disequilibrium |
| **LE** | Linkage Equilibrium |
| **LnP(K)** | Logarithm of the Probability of the Data Given a Specific Number of K |
| **LR** | Likelihood Ratio |
| **Mb** | Megabases |
| **MCMC** | Markov Chain Monte Carlo |
| **MDS** | Multidimensional Scaling |
| **mg** | Milligram |
| **Min** | Minutes |
| **MSY** | Male-Specific Region of the Y-chromosome |
| **mtDNA** | Mitochondrial DNA |
| **NDNAD** | National DNA Database (UK) |
| **ng** | Nanogram |
| **NGO** | Non-Governmental Organisation |

| | |
|---|---|
| **NGS** | Next Generation Sequencing |
| **NJ** | Neighbour Joining |
| **NPC** | National Population Commission (Nigeria) |
| **NRY** | Non-recombinant Region of the Y-chromosome |
| **NSD** | National Security Determination |
| **PCA** | Principal Component Analysis |
| **PCR** | Polymerase Chain Reaction |
| **PD** | Power of Discrimination |
| **PE** | Power of Exclusion |
| **PI** | Paternity Index |
| **PIC** | Polymorphic Information Content |
| **PND** | Penalty Notice for Disorder |
| **PAR** | Pseudo Autosomal Regions of the Y-chromosome |
| **(Φ)** | Phi coefficient |
| **RFLP** | Restriction Fragment Length Polymorphism |
| **RFU** | Relative Fluorescence Unit |
| **RMP** | Random Match Probability |
| **RM-YSTRs** | Rapidly Mutating Y-STRs |
| **RNA** | Ribonucleic Acid |
| **RPM** | Revolution Per Minute |
| **$R_{ST}$** | Slatkin's R-statistic |
| **S** | Seconds |
| **SSA** | Sub-Saharan Africa |
| **SD** | Standard Deviation |
| **SGM** | Second-Generation Multiplex |
| **SGM+** | SGM Plus™ |
| **SNPs** | Single Nucleotide Polymorphisms |
| **SSR** | Simple Sequence Repeats |
| **STR** | Short Tandem Repeats |
| **STRider** | STR for Identity ENFSI Reference database |
| **SRY** | Sex-determining Region Y |
| ***Taq*** | *Thermus aquaticus* Super Therm DNA Polymerase |
| **TWGDAM** | Technical Working Group on DNA Analysis Method |
| **UNFPA** | United Nations Population Fund |
| **USAID** | United States Agency for International Development |
| **WHO** | World Health Organisation |
| **WISPI** | World Internal Security and Police Index |
| **µL** | Microlitre |
| **YHRD** | Y-STR Haplotypic Reference Database |

# CHAPTER ONE

# INTRODUCTION

## 1.1 OVERVIEW OF FORENSIC GENETICS

Forensic genetics is a specialised discipline within forensic science that analyses biological samples, such as skin cells, semen, hair, blood, saliva, and sweat. This discipline involves extracting and characterising biological variations using DNA and RNA, particularly for law enforcement agencies that require these samples as evidence for identification in criminal investigations (Carey and Mitnik, 2002; Bauer, 2007; Elkins, 2012). Forensic science is an applied science that implements empirical scientific analysis in supporting civil and criminal law (Plourd, 2010; Tiwari and Kusum, 2024). When addressing serious and high-volume criminal cases, a nation's government utilizes forensic science services to aid in prosecuting offenders and ensure that justice is served (Ward *et al*., 2017). Forensic science is a broad discipline that integrates many disciplines, including genetics, biology, chemistry, toxicology, botany, entomology, geology, anthropology, archaeology, fingerprint examination, and questioned document analysis (Delémont et al., 2017). While forensic science primarily focuses on laboratory practices, the specialised work conducted at a crime scene are referred to as crime scene investigation (CSI) (James and Nordby, 2002; NicDaeid and White, 2024). Forensic scientists specialise in obtaining information through collecting, preserving, and analysing valuable evidence during an investigation (Millo *et al*., 2008). Forensic scientists also use their skills to work together with the police during a crime scene investigation (Durnal, 2010; Kelty *et al.,* 2024). Also, a forensic scientist is unbiased and can be an expert witness, objectively reporting regardless of supporting a prosecutor or defendant while testifying on the evidence presented in court. (Shuman and Greenberg, 2003).

Nowadays, the field of forensic genetics has been accepted and widely applied in criminal cases involving rape, murder, terrorism, and disaster investigations (Henke *et al*., 2001; Sobrino *et al*., 2005; Børsting and Morling, 2015; Malhotra and Jamir, 2024). Analysis of genetic materials transmitted from generation to generation can be utilized to either implicate or exonerate individuals under investigation (Parks, 2000). Deoxyribonucleic acid (DNA) is the primary focus of forensic genetics due to its unique and ubiquitous presence in every human, except for monozygotic twins, who produce identical DNA profiles (Kobilinsky *et al*., 2005). The examination of DNA from crime scene samples is a powerful tool for offender identification due to its high discriminatory capacity (Mehar *et al.*, 2024). However, the probative value of DNA evidence is substantially enhanced when a reference profile is available for comparison, such as through a known suspect or a database match (Malhotra and Jamir, 2024). In many jurisdictions,

this is achieved by searching national forensic DNA databases, and in limited circumstances through international data-sharing frameworks (Greely *et al*., 2006). Where an offender has previously been convicted and their profile is stored in a database, a match may be obtained However, for justice to be served, the significance of the offender's DNA profile match must be statistically evaluated from a separate population DNA database containing allele frequencies (Foreman and Evett, 2001). Allele frequency represents the proportion of a particular allele that appears at a specific genetic locus within a population's gene pool, compared to other alleles at that locus (Butler, 2014). The allele frequency database of a nation's population serves as the reference for calculating the statistical weight of DNA profile rarity when a match is observed between a crime scene sample and a reference profile. Assessing DNA profile evidence rarity is crucial, as it is a standard procedure for expert witnesses to convey the significance of genetic matches in court cases, playing a vital role in legal proceedings (Bakhtiar, 2024). DNA databases are valuable in various fields, particularly healthcare and security (Foreman *et al*., 2003).

## 1.2 BRIEF HISTORY OF FORENSIC GENETICS

Forensic genetics has been a long-term practice that progressively evolved from forensic serology when the ABO human blood type for identification in 1900 by Karl Landsteiner (Yamamoto and Hakomori, 1990; Li, 2018). In 1910, French criminologist Edmond Locard established modern forensic science by introducing Locard's exchange principle, which asserts that "every contact leaves a trace" (Byard *et al*., 2016; Shaw and Sandiford, 2024). Forensic genetics progressed further with the proposal of the "theory of genes" by Thomas Hunt Morgan in 1926. However, the molecular application of forensic genetics was developed after DNA was discovered in 1953 (Reich *et al*., 2002; Bhardwaj *et al.*, 2025). Deoxyribonucleic acid (DNA) functions as the genetic blueprint that carries the hereditary instructions vital for the development and functioning of all living beings, including humans. Within cells, DNA is predominantly located in the nucleus (nuclear DNA), while a lesser portion is found in the mitochondria (mitochondrial DNA). DNA is structurally arranged in a double helix and comprises four nitrogen-containing bases: thymine (T), guanine (G), adenine (A), and cytosine (C). Among these, adenine (A) and guanine (G) are known as purines, whereas cytosine (C) and thymine (T) are classified as pyrimidines. Base pairing occurs between purines and pyrimidines via hydrogen bonding—adenine pairs with thymine through two hydrogen bonds (A=T), and guanine pairs with cytosine via three hydrogen bonds (G≡C). The human genome contains roughly 3 billion base pairs, with more than 99.9% of these sequences being identical among all individuals (Collins and Mansoura, 2001). The sequences or order of the nitrogenous bases determines the information for an organism's establishment, functionality, and preservation. (Watson and Crick, 1953; Alberts *et al*., 2015; Sood and Singh, 2024). Figure 1.1

shows the DNA structure, which includes deoxyribose sugar, nitrogenous bases, and phosphate groups. The human genome comprises approximately 19,000 to 20,000 genes that encode proteins, with exons accounting for around 1.5% and protein-coding regions representing approximately 1.1% (International Human Genome Sequencing Consortium, 2004; Rand *et al*., 2014).

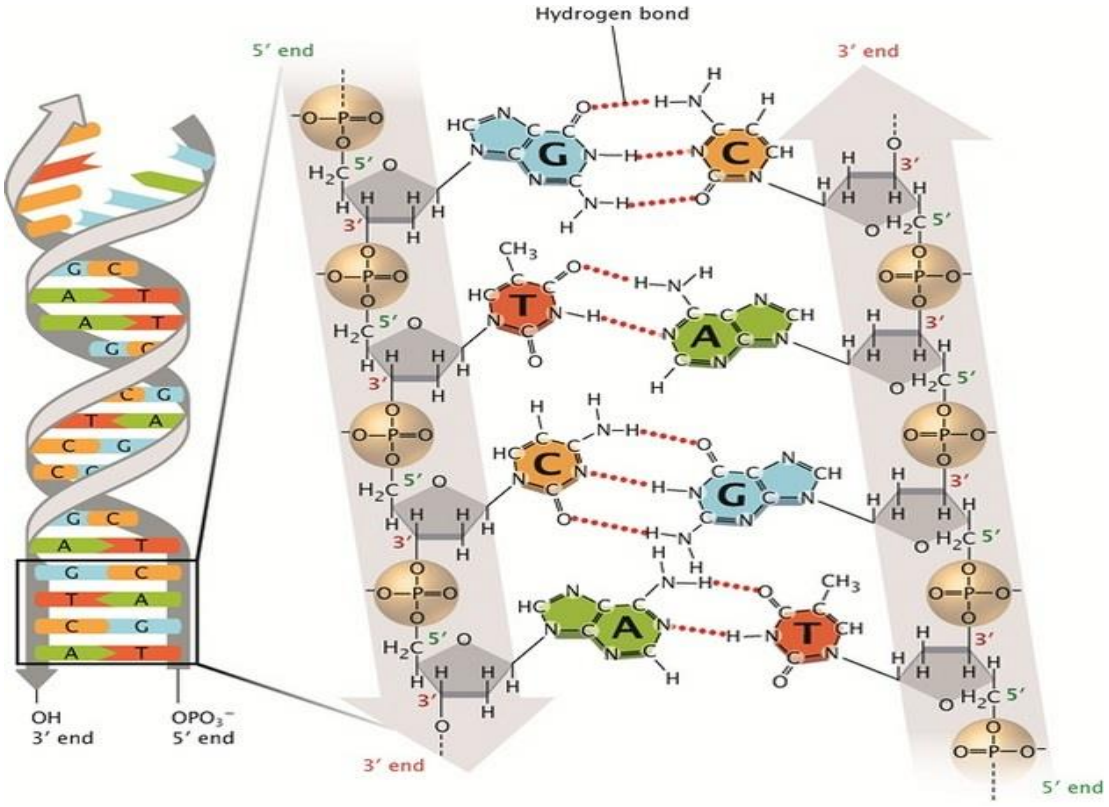



**Figure 1.1. Chemical structure of DNA illustrating antiparallel alignment of the two strands (5′→3′ and 3′→5′), complementary base pairing via hydrogen bonds between purines—adenine (A) and guanine (G)—and pyrimidines—cytosine (C) and thymine (T)—and the sugar–phosphate backbone, as described by Pray (2008).**

The human genome consists of 46 chromosomes, comprising 22 pairs of autosomes and one pair of sex chromosomes, X and Y (XX in females and XY in males). **Figure 1.2** demonstrates autosomal chromosome inheritance and meiotic recombination. Genes are positioned at loci on homologous chromosomes, with each carrying a specific allele. During meiosis, recombination reshuffles parental alleles, producing genetically distinct gametes. Crossover events mainly occur in recombination hotspots, while coldspots experience minimal crossover. This variability in crossover breakpoints contributes to genetic diversity and affects allele inheritance patterns across generations (Hartl and Clark, 2007; Neuvonen *et al*., 2017; Cooper and Adams, 2022).

Forensic genetics examines the variable regions of DNA that distinguish individuals, which constitute approximately 0.1 per cent of the human genome's 3 billion base pairs (Romeika and Yan, 2013). Using multilocus autosomal STR profiling, the probability that two unrelated

individuals share an identical DNA profile is extremely low, often estimated to be on the order of one in hundreds of trillions, depending on the loci analysed and the population considered. This estimate is derived from analyses using a limited set of short tandem repeat (STR) markers, specifically 13 CODIS loci, before the 2017 expansion to 20 loci (McDonald and Lehman, 2012; Margoliash, 2024).

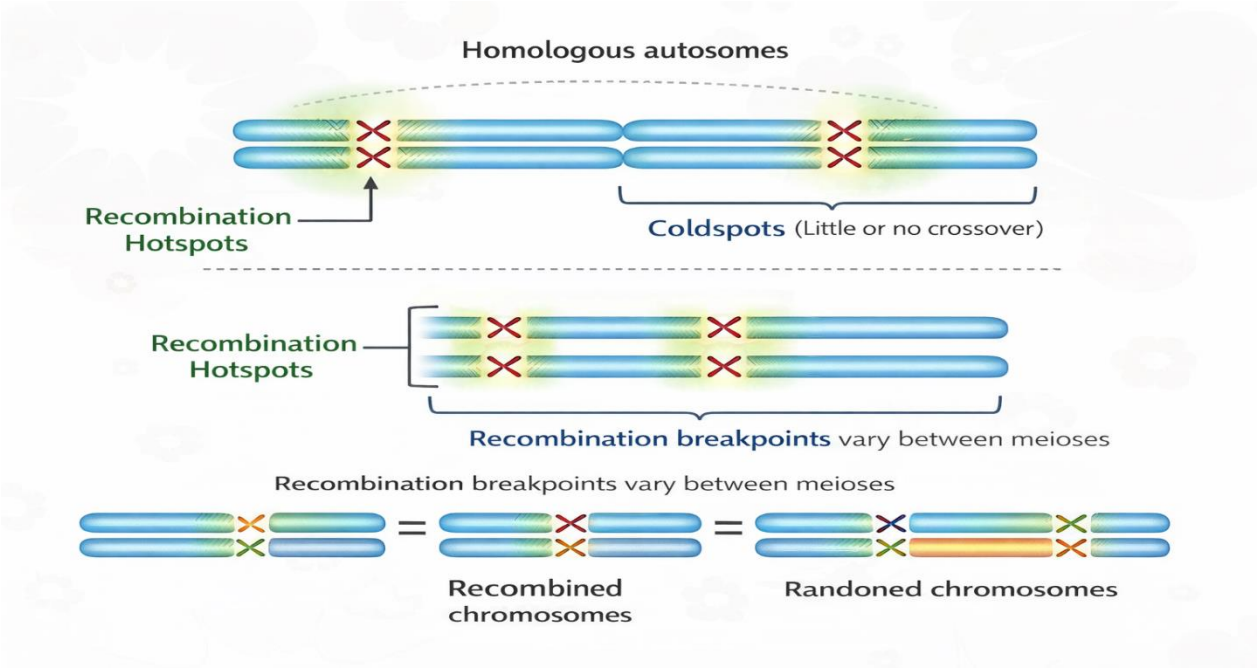



**Figure 1.2. Schematic representation of meiotic recombination along homologous autosomes. Recombination events are non-uniformly distributed along chromosomes and are concentrated within recombination hotspots, while other regions (coldspots) experience little or no crossover. Consequently, although crossovers occur within hotspot regions, their precise breakpoints vary between meiosis and across generations, contributing to genetic diversity.**

Genetic markers are specific sequences of DNA found in particular locations on chromosomes. They can help identify individuals, groups, or species and track how genes are passed down through generations. They are essential tools in studies of genetic linkage, association mapping, and evolutionary biology (Kumar, 1999; Hedrich, 2012; Doveri *et al.,* 2024). Li (2018) noted that the evolution of the use of genetic markers has developed over time in phases distinguished by their applications. The reliability of DNA fingerprinting was established through extensive validation studies demonstrating high polymorphism, reproducibility, Mendelian inheritance, and extremely low probabilities of coincidental matches, together with successful application in exclusion and inclusion casework (Butler, 2009; Jobling and Gill, 2004). The evolution of molecular markers paved the way for Alec Jeffreys's development of DNA fingerprinting in the mid-1980s at the University of Leicester in the UK. He applied variable number tandem repeats (VNTRs) and coined the term 'DNA fingerprinting' or 'genetic profiling'—a technique used to identify individuals by their distinct DNA patterns. He built upon the principles of the Southern blot technique developed

by Edwin Southern in 1975, which transfers DNA fragments to a nylon membrane for hybridization to distinguish unique patterns in DNA for individual identification (Southern, 1975). He and his colleagues discovered that DNA hybridised with multi-locus probes displayed high levels of variation and heritable motifs (Roewer, 2013). The approach utilised the restriction fragment length polymorphism (RFLP) method, examining minisatellite regions that are distinguishable based on their fragment sizes after separation. DNA fingerprinting was first used in a paternity dispute in March 1985 to prevent the deportation of a young British boy of Ghanaian origin (Jeffreys *et al*., 1985a; 1985b).

In 1986, DNA fingerprinting was first applied in a criminal investigation to exonerate an initial suspect, Richard Buckland, in the Narborough murder case, demonstrating its power not only for identification but also for exclusion (Butler, 2009; Roewer, 2013). The next application occurred in 1987, marking the police's first use of this scenario. The precise reliability of DNA fingerprinting was instrumental in securing the conviction of Colin Pitchfork for the sexual assault and homicide of two teenage girls by matching his DNA to the semen samples found on the victims (Butler, 2009). Many scientific companies and law enforcement agencies worldwide became interested in DNA fingerprinting for criminal investigations during this period, which led to improvements and new techniques in the field. The technique offered significantly greater discriminatory capacity than blood group typing, as the chance of a coincidental match was lower for minisatellite markers compared to blood groups, though still higher than that of the more advanced STR markers. This approach was efficient because a single expert could apply a unified method across samples (Gill *et al*., 1985; Butler, 2015a). However, the technique had significant limitations: it required large quantities of high-quality, undegraded DNA; interpretation of complex multi-band patterns could be subjective and demanded highly trained analysts; the process was lengthy, lacked automation, and involved exposure to radioactive materials. DNA fingerprinting was not very effective for analysing degraded forensic samples because the RLFP method requires high-quality samples and fresh, large DNA fragments. (Mnookin, 2001; Johnson, 2024).

In the mid-1980s, American biochemist Kary Mullis developed a laboratory method for amplifying DNA called the "polymerase chain reaction" (PCR). This method employs the *Taq* DNA polymerase enzyme to synthesise new DNA strands utilising existing strands as templates. *Taq* DNA polymerase is an enzyme that can withstand high temperatures. It comes from a bacterium, *Thermus aquaticus,* discovered in Yellowstone National Park's hot springs (Chien *et al*., 1976). This enzyme can handle the heat needed for DNA denaturation, around 94–96 °C. This ability allows the PCR process to be automated without adding new enzymes after each heating cycle. The

innovative technique was used by scientists to target different repetitive regions in the human genome, known as microsatellites, and only required nanograms (ng) rather than milligrams (mg) of DNA (Butler, 2011; Watson, 2012).

The early 1990s marked the golden era for DNA fingerprinting. The original technique of Jeffreys underwent significant improvements. The Southern blot technique was supplanted by the PCR technique, fluorescent labels replaced radioactive labels, and slab gels replaced by capillary electrophoresis around mid to late 1990s (Butler, 2009; Roewer, 2013).

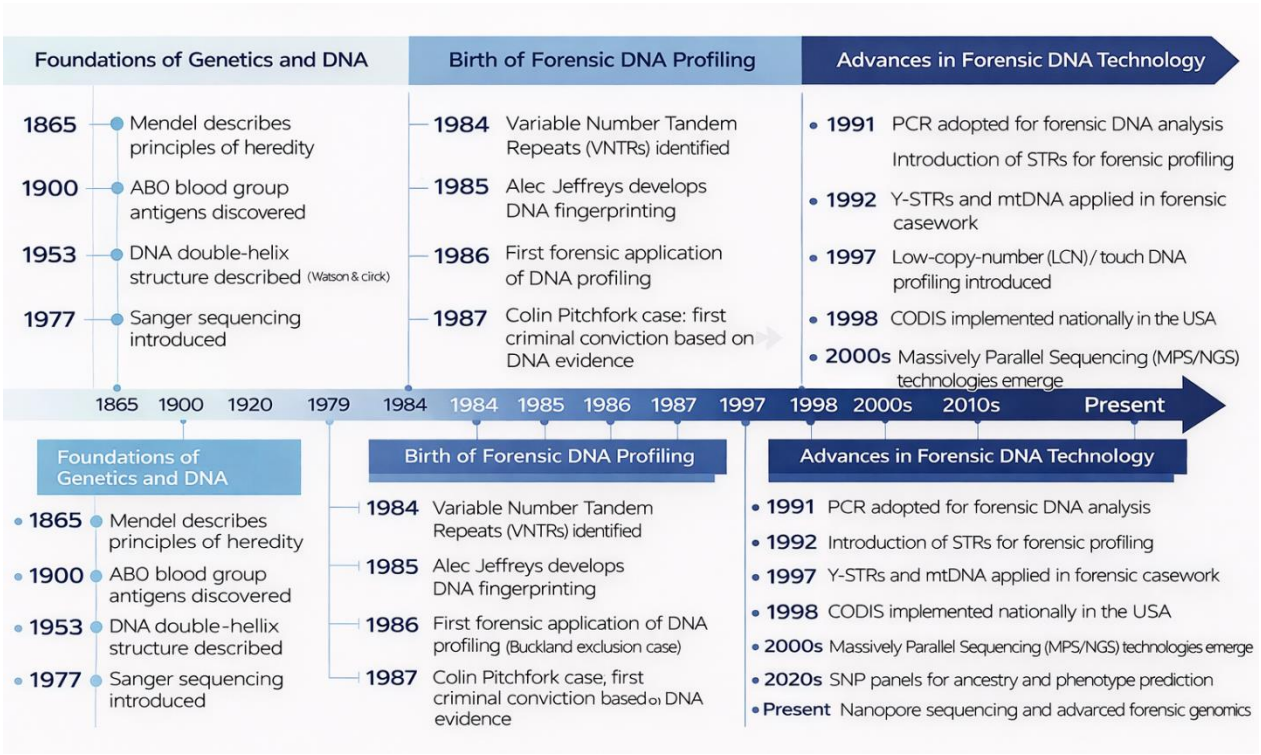



**Figure 1.3. Timeline of forensic genetics development**

A summary of the development of forensic genetics over time is presented in **Figure 1.3**. The PCR-based technique of STR profiling that replaced RFLP-based analysis of microsatellite markers was universally applied because of the genotyping precision, speed, and sensitivity (Roewer, 2013). During this period, the UK Forensic Science Service (FSS) achieved significant progress by creating the first forensic multiplex STR amplification kit, which included four loci: vWA, TH01, F13A1, and FES/FSP. The move to STR analysis with polymerase chain reaction (PCR) from RFLP techniques was an essential improvement in forensic DNA profiling. Multiplex PCR allowed the simultaneous analysis of multiple STR loci in a single reaction, saving time and increasing productivity. Over time, these multiplex systems have improved, enabling the analysis of larger areas and improving the clarity and accuracy of forensic DNA profiling (Butler, 2015; Huang *et al.*, 2022).

## 1.3 AUTOSOMAL SHORT TANDEM REPEATS

### 1.3.1 DESCRIPTION AND OVERVIEW

Short tandem repeats (STRs) are short, tandemly repeated DNA sequences that are predominantly located in non-coding regions of the genome, often situated between genes that code for proteins (Ellegren, 2004). They are sometimes called 'simple sequence repeats' (SSRs) or microsatellites. STRs are polymorphic, hypervariable, and ubiquitous in thousands of locations across the chromosomes. The core sequences vary in length, ranging from 2-6 bps and can exist in sequence motifs of di, tri, tetra, penta, and hexanucleotide repeats. The sequence motifs may contain repeated elements in a simple, compound, or complex structure (Ellegren, 2004; Fandade *et al*., 2024). **Table 1.1** displays the categories of repeats and sequence motifs identified at a standard short tandem repeat (STR) locus. 'Simple repeats' are composed of a single repeat unit that occurs multiple times. 'Compound repeats' comprise two or more neighbouring simple repeat units. Conversely, 'complex repeats' feature multiple repeat blocks of differing lengths, along with intervening sequences within a single unit (Gill *et al*., 2020).

**Table 1.1. An example of STR variations illustrating simple, compound, and complex repeats with an example marker and sequence motifs.**

| Repeat | Locus/Marker | Sequence Motif |
|---|---|---|
| **Simple** | TH01 | $[AATG]_n$ |
| **Compound** | vWA | $[TCTA]_a$ $[TCTG]_b$ $[TCTA]_c$ TCCA $[TCTA]_d$ |
| **Complex** | D21S11 | $[TCTA]_a$ $[TCTG]_b$ $[TCTA]_c$ TA$[TCTA]_e$ TCCATA$[TCTA]_f$ |

Short tandem repeats (STRs) are prevalent throughout the human genome, representing approximately 1–3% of genomic DNA (Shi *et al*., 2023). These repeats occur in both non-coding and coding regions and typically consist of repeat units ranging from one to six base pairs. STRs exhibit very high mutation rates relative to most other genomic regions, contributing to their extensive polymorphism (Xia *et al*., 2024). Because of their high mutation rates, STRs show significant variations among individuals and are highly effective for discrimination in genetic analysis. The high mutation rates in STRs occur because of errors during DNA replication, known as DNA polymerase slippage. This slippage occurs when the DNA strands temporarily separate and misalign, resulting in the insertion or deletion of repeat units (Fan and Chu, 2007). These mutations involve the loss or gain of repeat units rather than single nucleotide changes and often occur stepwise. While most mutations occur as single-step changes, multi-step alterations can happen less frequently (Verbiest *et al*., 2023). Repetitive sequences, such as STRs, are more prone to slippage events. During these events, the mutation rate can vary from $1 \times 10^{-5}$ to $1 \times 10^{-3}$ per

locus per generation (Shi *et al*., 2023). As a result, repetitive DNA tends to exhibit higher mutation rates compared to non-repetitive DNA. Factors such as the repeat unit size and the surrounding sequence influence the specific range of mutations (Steely *et al*., 2022). Mutation rates vary based on repeat unit length and sequence composition; dinucleotide repeats show higher mutation rates than longer repeat units like trinucleotide and tetranucleotide (Doss *et al*., 2025). Microvariants can also increase STR variation, e.g. Allele 14.2 at locus D19S433 (Butler, 2015). Microvariants are STR alleles that vary slightly from standard alleles because they have partial repeat units. These are often shown with decimals, like 10.2 (Butler, 2014). Small insertions, deletions, or point mutations in or near the repeat region can cause changes. These changes usually happen due to an error during DNA replication, known as replication slippage (Gettings *et al*., 2015). Microvariants are crucial in forensic DNA analysis because they enhance the variability in STR profiles, helping to distinguish between individuals. The naming of these microvariants follows international guidelines set by the ISFG (International Society for Forensic Genetics) (Parson *et al*., 2016).

The standard workflow for DNA profiling begins with DNA extraction from biospecimens such as saliva, blood, and touch or trace DNA obtained from individuals or crime scenes (Budowle *et al*., 2005). This process may include purification. The method of isolating DNA from cells using chemicals is particularly well-suited for degraded or low-quality samples. It ensures greater reliability for complex samples, especially in forensic casework, because it removes inhibitors like heme, phenols, proteases, bile salts, proteins, and urea during extraction. The purified DNA utilised as a template for PCR is cleaner and contains little to no inhibitors that could interfere with the amplification process. However, standard PCR can be time-consuming and requires specific kits or chemicals (Gill, 2001). The extracted DNA sample can be quantified to ensure it is adequate for analysis. During PCR amplification, particular DNA segments—typically short tandem repeats (STRs)—are amplified to produce multiple copies. The resulting amplified PCR products containing fragments of DNA are then separated using capillary electrophoresis. Fluorescent dyes help in detecting the STR markers during this process. The generated DNA profiles are interpreted as numerical representations of the sizes of the STR fragments on programmes like the GeneMapper software. The DNA profiles collected at a crime scene can be matched against to known profiles from victims, suspects or DNA databases to confirm or eliminate individuals (Butler, 2015b).

The direct PCR technique is an STR typing method that bypasses the extraction and quantification steps allowing for immediate progression to PCR amplification. This approach adds the biological sample directly to the PCR mix without any extraction process. Though the direct PCR technique

is not an automated method for analysing DNA in forensics, it is faster and cheaper than the standard procedure because it saves both time and consumables (Shrivastava *et al*., 2021). Direct PCR amplification is an approach that bypasses DNA extraction and quantification before STR profiling, thereby reducing processing time and accelerating the generation of DNA profiles. In contrast, conventional DNA profiling workflows involving extraction and quantification are comparatively laborious and time-consuming. Direct PCR is therefore preferred when appropriate, particularly for reference samples such as buccal swabs, which typically contain sufficient quantities of high-quality DNA and minimal inhibitors (Sim *et al*., 2013). However, this technique may face inhibition when applied to samples like blood, as the crude samples can introduce PCR inhibitors such as haemoglobin and contaminants, which reduces the efficiency of PCR amplification (Flores *et al*., 2014; McKiernan and Danielson, 2017; Myers *et al.*, 2012; Wang *et al*., 2024).

The first-generation multiplex system to be widely utilised in forensic science was a simple STR multiplex system (Gill, 2002). The UK Forensic Science Service (FSS) first designed and validated a four-locus "quadruplex" in 1994, which included the vWA, TH01, F13A1, and FES/FPS loci for forensic analysis (Lygo *et al*., 1994). This system was potent and sensitive for its time but had a relatively high match probability (1 in 100,000) due to the number of loci (Kimpton *et al.,* 1993). In 1996, the UK FSS introduced the Second-Generation Multiplex (SGM), which contained six STR loci: D21S1, D18S51, D8S1179, FGA, vWA and TH01. It was accompanied by the Amelogenin (sex-typing) marker, which amplifies X- and Y-chromosome homologs (Sparkes *et al*., 1996a). The SGM system had higher discriminatory power than the quadruplex and demonstrated a lower match probability (1 in 50 million) (Gill, 2002). It also performed better on degraded and aged samples, mainly if mini-STRs were applied (Sparkes *et al.,* 1996b). Concordance studies of PCR kits provide information to determine whether the outcome of two or more kits that assay the same loci generates the same genotype (Hill *et al.,* 2007). Null alleles are a type of gene that, due to mutations or other changes, fails to produce a functional product or one that can be detected at the molecular level. Allelic drop-out occurs when one or both gene copies are absent from a DNA profile, even though they are expected to be present. These issues and underlying genetic variations can be identified by using different STR kits, which contain various PCR primers for the same markers, to analyse a DNA sample (Hill *et al*., 2011). Information from genotype discrepancies between PCR kits can help raise awareness of their occurrence and necessitate a secondary method/kit for verification. In 1995, the UK National DNA Database (NDNAD) was established, coinciding with the introduction of the SGM marker system as the preferred DNA profiling system (Amankwaa and McCartney, 2019). The SGM marker system

added over a million DNA profiles to the database. The SGM marker system was replaced by the AmpFℓSTR® SGM Plus™ (SGM+) in 1998, as the preferred marker system for the UK National DNA Database. SGM+ was an updated version of SGM, developed commercially by Applied Biosystems (Cotton *et al*., 2000). The multiplex system was a ten-locus multiplex comprising the six original SGM loci and four extra STR loci: D19S433, D2S1338, D16S539, and D3S1358, in conjunction with the Amelogenin locus (Foreman and Evett, 2001). The Technical Working Group on DNA Analysis Method (TWGDAM) recommended instructions for validating the system for forensic work (Budowle, 1995). With the new improvement, the match probability for a complete profile between two unrelated people decreased to 1 in more than 1 billion (Foreman and Evett, 2001). Over time, other European countries, the USA, Australia, Asia, and Africa began establishing and expanding their national DNA databases. Shortly after launching the UK NDNAD, the USA established the National DNA Index System (NDIS) in 1998 using the Combined DNA Index System (CODIS) software (Linacre, 2003).

### 1.3.2 APPLICATIONS AND RELEVANCE OF AUTOSOMAL STR TYPING

STR typing has been established as a standard DNA profiling technique for forensic investigation in many countries (Butler, 2015a). It is mainly applied in the court of law during sexual offences, murder, physical assault, burglary, and terrorism cases. STR profiles are interpreted by comparing the alleles and genotypes present at each STR locus between two or more samples: a questioned sample (evidence sample) and a known reference sample (suspect sample) (National Research Council, 1992). For example, **Table 1.2.** shows a fictional case of crime scene evidence. An evidence sample can be compared with two or more suspects for a match.

**Table 1.2. Fictional DNA profile from blood stain evidence collected from a crime scene and from two suspects, with allele and genotype frequencies for Suspect B**

| STR Locus | Blood Stain Evidence | Suspect A Alleles | Suspect B Alleles | Frequencies of Alleles found in Suspect B's Profile | Hardy-Weinberg Formula | Genotype Frequencies of Suspect B |
|---|---|---|---|---|---|---|
| **TH01** | 7,8 | 9.3,9.3 | 7,8 | 7(0.3883), 8(0.2104) | 2pq | 0.1634 |
| **D3S1358** | 15,16 | 17,17 | 15,16 | 15(0.2977), 16(0.3123) | 2pq | 0.1860 |
| **vWA** | 17,17 | 18,19 | 17,17 | 17(0.2217) | $p^2$ | 0.0492 |
| **D21S11** | 30,33 | 28, 30 | 30,33 | 30(0.1861), 33(0.0194) | 2pq | 0.0072 |
| **D16S1656** | 16,16.3 | 15,15 | 16,16.3 | 16(0.1036), 16.3(0.0809) | 2pq | 0.0168 |
| **D16S539** | 9,11 | 12,13 | 9,11 | 9(0.2104), 11(0.2799) | 2pq | 0.1178 |
| **D8S1179** | 12,12 | 13,14 | 12,12 | 12(0.1505) | $p^2$ | 0.0227 |
| **D7S820** | 10,10 | 9,11 | 10,10 | 10(0.1683) | $p^2$ | 0.0283 |
| **FGA** | 22,22 | 21,23 | 22,22 | 22(0.1926) | $p^2$ | 0.0371 |
| **RANDOM MATCH PROBABILTY (Product of Genotype Frequencies for Suspect B) = $5.078 \times 10^{-13}$** | | | | | | |

If the STR alleles from the evidence do not match any suspects' STR alleles, those suspects can be quickly excluded as potential sources of the evidence suspect's scene. In cases where a suspect's

STR alleles match all the alleles in the evidence profile, a statistical evaluation is to determine the probability of such a match occurring in the population. This calculation first considers the frequency of the relevant alleles within the suspect's ethnic group. At each locus, genotype frequency is calculated by applying the Hardy-Weinberg principle, using the product of allele frequencies for homozygous and heterozygous genotypes. (Norrgard, 2008).

## 1.4 Y-SHORT TANDEM REPEATS

### 1.4.1 DESCRIPTION AND OVERVIEW OF Y-CHROMOSOME AND Y-STR

The Y chromosome, which constitutes one of the two sex chromosomes (X and Y) in mammals, is one of the smallest (the third smallest) of the 46 chromosomes in the human genome **(Figure 1.4)**. All biological human males possess Y chromosomes, which are transmitted as a single haplotype block from father to son across generations **(Figure 1.5).** However, the *SRY* gene, responsible for male sex determination, can be translocated from the Y chromosome to another chromosome (for example, the X chromosome). This means a person may develop phenotypically as male but be genotyped as female if they do not have a Y chromosome (Calafell and Comas, 2021).

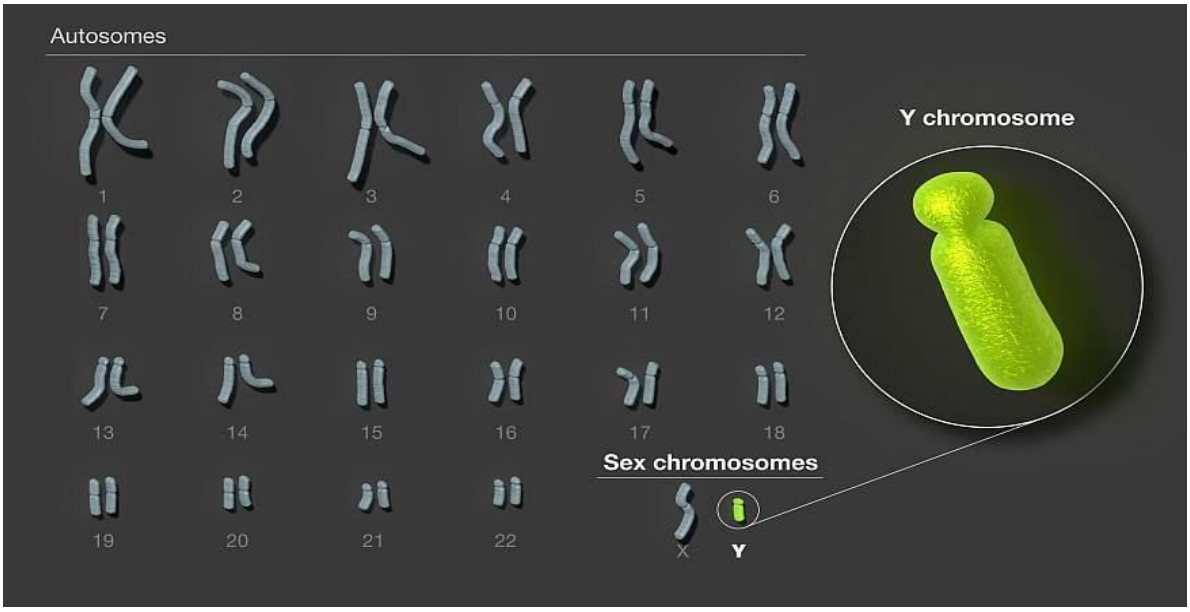



**Figure 1.4. The human male chromosomal karyotype consists of 46 chromosomes, 22 pairs of autosomes and one pair of sex chromosomes (XY), with the Y chromosome fully depicted. The human Y chromosome, recently fully sequenced, reveals essential features related to fertility and sperm production (https://www.nih.gov/news-events/news-releases/researchers-assemble-first-complete-sequence-human-y-chromosome).**

In the reference human genome, the Y chromosome spans approximately 60 megabases (Mb), though its size can range from 45.2 to 84.9 million base pairs among individuals (Rhie *et al*., 2023). This makes it significantly smaller than the X chromosome, which is about 160 megabases (Mb)

long and contains approximately 156 million base pairs. The Y chromosome exhibits more significant genetic and structural variation among males than any other chromosome in the human genome (Hallast *et al*., 2023). This characteristic is uniquely associated with the genetic makeup of human males. The Y chromosome is essential for determining male sex, passing on traits from fathers, and producing sperm, which is essential for fertility. It has about 55 genes that code for proteins, which is much fewer than the number found on the X chromosome and most other chromosomes. The Y chromosome has around 78 protein-coding genes, many encoding male-specific functions related to testis development and spermatogenesis. Among these, the *SRY* gene plays a central role in initiating male sex determination and the development of male characteristics (Westemeier-Rice *et al.*, 2024). The AMELY variant of the Amelogenin gene is found on the short arm of the Y chromosome (Yp11.2). In forensic DNA profiling, STR kits typically amplify a segment of approximately 112 base pairs from the AMELY locus. Its X chromosome counterpart, AMELX, yields a fragment of about 106 base pairs. The size difference between these amplicons allows for reliable sex determination in forensic analyses (Keefe, 2024).

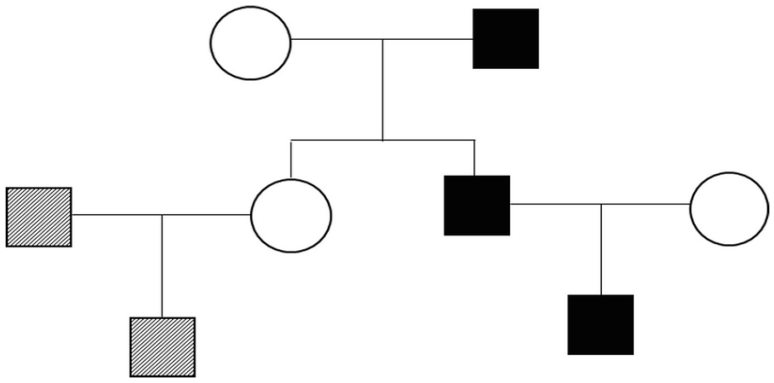


**Figure 1.5. The inheritance pattern of Y chromosome DNA is illustrated by the black boxes, which represent a patrilineal transfer from father to son across generations without genetic recombination. Females are depicted by circles (Dash *et al*., 2020).**

The Y chromosome is known to contain the Azoospermia Factor (AZF) regions, which are critical for sperm development. Deletions within these regions can result in male infertility (Pazoki *et al*., 2024). The Y chromosome is primarily composed of a section known as the "non-recombining region of the Y chromosome" (NRY), also known to as the "male-specific region of the Y chromosome" (MSY) **(Figure 1.6)**. Unlike the autosome and X chromosome, this region does not participate in genetic recombination. The MSY constitutes about 95% of humans' total Y chromosome length (Jain *et al*., 2017). The entire MSY region exhibits linkage disequilibrium, which is transmitted along the paternal lineage from father to son without the occurrence of recombination (Prokop and Deschepper, 2015). The remaining regions of the Y chromosome include two pseudo autosomal regions (PAR1 and PAR2) **(Figure 1.6)**, which together make up

about 5% of the entire Y chromosome. These regions are located at the distal end of the Y chromosome and contain at least 29 genes, spanning a combined length of less than 3 megabases (Mb) out of the total 60 Mb length of the chromosome.

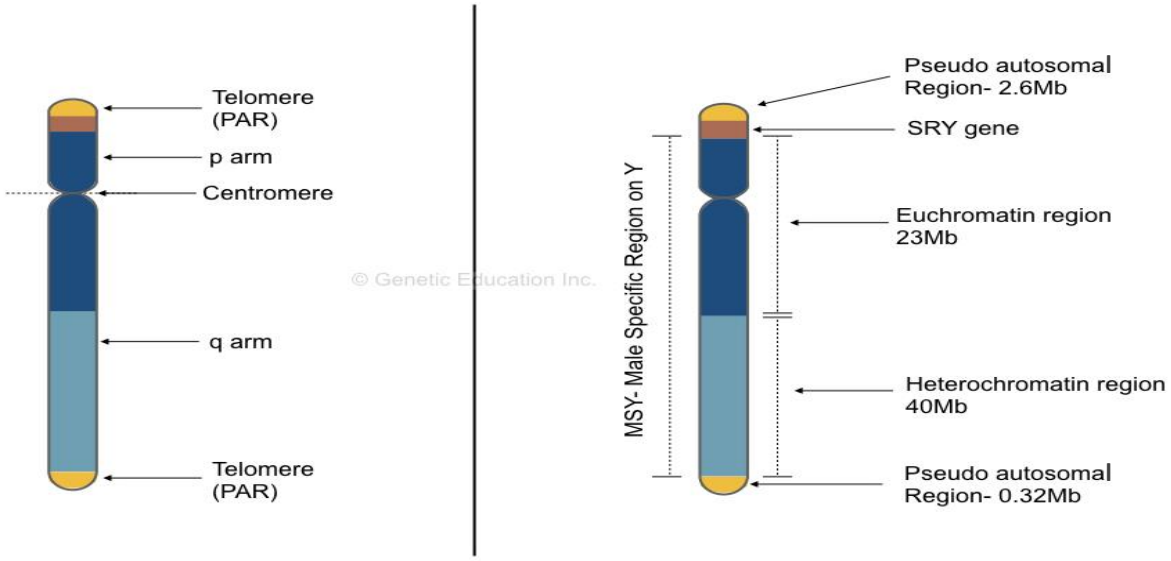



**Figure 1.6. The Y chromosome structure, highlighting the long arm (q arm or Yq) and the short arm (p arm or Yp). It also displays the locations of the euchromatin and heterochromatin regions along the arms and the structural locations of the male-specific region (MSY) and the pseudo-autosomal regions (PARs) (https://geneticeducation.co.in/y-chromosome-structure-and-function/)**

The PAR regions can undergo recombination and are homologous to the pseudo autosomal areas found on the X chromosome. This pairing allows the X chromosome to align with the Y chromosome during meiosis, facilitating autosomal rather than sex-linked inheritance. Additionally, Veerappa *et al*. (2013) identified a third pseudo autosomal region (PA3) located on the Y chromosome short arm at Yp11.2. This region might be found in about 2% of the general population; however, its functional role is yet to be known (Colaco and Modi, 2018). The Y chromosome is useful in tracing paternal lineage due to its lack of recombination. The distinctive characteristics of the Y chromosome, including the non-recombing pattern and variations introduced by mutations through generations along the paternal lineage, have sparked significant interest among scientists in studying polymorphisms in the MSY regions (Petr *et al.*, 2020).

Cytogenetically, the Y chromosome comprises two types of chromatins (a mixture of DNA and proteins), euchromatin and heterochromatin **(Figure 1.6)**, which have distinct structures and functions. Euchromatin is 23Mb, rich in genes, less condensed and more readily transcribed than heterochromatin. It contains all the Y chromosome protein regions and several hundred Y-STR loci and is less condensed than heterochromatin. It initiates transcription by allowing gene regulatory proteins to bond with DNA, modifying histone tails to make chromatin more open. In contrast,

heterochromatin, comprising about 40 Mb, is gene-poor, highly condensed, and generally transcriptionally inactive. It consists mainly of repetitive sequences, contributes to centromere stability, and plays a role in genome organization and repression of transposable elements. Heterochromatin is established during early embryogenesis and is critical for maintaining genome integrity and regulating chromatin architecture and gene expression through histone modifications and epigenetic mechanisms (Rhie *et al*., 2023).

Y-chromosomal short tandem repeats (Y-STRs) are distinct genetic markers located on the Y chromosome, composed of repeating DNA sequences. These Y-STR markers are inherited exclusively from father to son, which makes them valuable for forensic, anthropological, and genealogical research (Butler 2015b). Y-STRs are specific to males and are used solely to identify male lineages because Y chromosomes are present only in males. Y-STRs are relatively stable within patrilineal lineages because the Y chromosome non-recombinant across generations due to haploid inheritance (Roewer, 2013). The non-recombinant nature of Y-STR made it a valuable tool for tracking paternal ancestry over many generations, thereby creating a stable lineage. The Y-STR uniparental inheritance pattern made the markers ideal for studying male-line ancestry (Kayser, 2017; Jobling and Tyler-Smith, 2003).

Researchers discovered that particular STR loci are unique to the Y chromosome. Early studies aimed to detect polymorphic Y-linked STR markers for forensics and human population genetics applications. The first polymorphic STR on the MSY was recognised in 1992, and the initial application of Y-STR analysis was introduced in 1995 (Roewer and Epplen, 1992; Roewer *et al*., 1992). Following the discovery of Y-STRs, research scientists realised that Y-STR markers were handy in male-specific profiling in sexual assault cases involving mixed male and female DNA (Kayser, 2017). By the late 1990s, scientific researchers established the first standard panel of Y-STR markers (Keyser *et al*., 1997). The Y-STR kits advancement has progressed through multiple stages, beginning in 1997 when an international consortium of laboratories focused on Y-STR technology established a foundational set of nine markers. These markers include DYS385a/b, DYS393, DYS392, DYS391, DYS390, DYS389II, DYS89I, and DYS19, which constituted the European minimal core haplotype of Y-STRs for forensic applications. Y-STR markers gained prominence, validation, and widespread use in forensic casework (Keyser *et al*., 1997). In 1999, the Y-Chromosome Haplotype Reference Database (YHRD) was created to collect and analyse Y-STR haplotypes from various populations worldwide (Willuweit *et al*., 2007). The database became crucial for forensic casework, genetic genealogy, and population studies (Roewer, 2019). In 2003, the FBI and the Scientific Working Group on DNA Analysis Methods (SWGDAM) established

guidelines for the forensic use of Y-STRs. Following this, SWGDAM announced a panel that includes 11 Y-STRs: a European set of nine, plus two extra loci, designated as DYS438 and DYS439 (Lee, 2004).

Specific collections of Y-STR markers that are analysed in a DNA profile are referred to as Y-STR panels. Y-STR panels can be analysed using commercially available Y-STR kits. These kits contain all the necessary chemicals and primers to amplify the Y-STR markers, producing a Y-STR haplotype inherited as a single block. Multiple Y-STR panels are utilised in forensic and genealogical research studies **(Table 1.3)**. Some of the most commonly used marker panels include PowerPlex® Y - which includes: 12 Y-STR loci; PowerPlex® Y23 System (Promega) - this expands contains 23 Y-STR loci; Yfiler™ (Applied Biosystems) - which contains 17 Y-STR loci; Yfiler™ Plus (Applied Biosystems) - this expanded version features 27 Y-STR loci, offering improved discrimination, and YHRD - a global reference database for Y-STR haplotypes (Willuweit *et al*., 2007). These panels enable researchers to conduct more accurate analyses and comparisons (Alghafri, 2020).

**Table 1.3. Y-STR Panel and corresponding Y-STR markers in the different dye colours**

| Y-STR Panel | Y-STR Marker |
|---|---|
| **Minimal** | **DYS385, DYS393, DYS392, DYS391, DYS390, DYS389II, DYS389I, DYS19,** |
| **PowerPlex Y** | **DYS385, DYS393, DYS390, DYS392, DYS19, DYS437, DYS438, DYS389II, DYS439, DYS389I, DYS391** |
| **PowerPlex Y23** | **YGATAH4, DYS456, DYS385, DYS458, DYS393, DYS643, DYS392, DYS439, DYS390, DYS635, DYS570, DYS437, DYS438, DYS533, DYS549, DYS481, DYS391, DYS19, DYS389II, DYS448, DYS389I, DYS576** |
| **Yfiler** | **DYS448, DYS438, DYS437, YGATAH4, DYS392, DYS635, DYS439, DYS391, DYS393, DYS385, DYS19, DYS458, DYS389II, DYS390, DYS389I, DYS456** |
| **Yfiler Plus** | **DYS389II, DYS576, DYS627, sDYS635, DYS460, DYS458, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS518, DYS570, DYS437, DYS385, DYS449, DYS393, DYS439, DYS481, DYF38751, DYS533** |

**Figure 1.7** shows the Y-STR loci frequently used in forensic genetics, indicating their approximate positions on the Y chromosome. Since 2015, significant advancements have been made in using Next Generation Sequencing (NGS) for Y-STR analysis, allowing for the simultaneous analysis of multiple Y-STR loci and improving sensitivity and resolution. NGS-based Y-STR typing has greatly enhanced forensic analysis, particularly in low-template or degraded DNA cases (Sobiah *et al*., 2018; Silva, 2024).

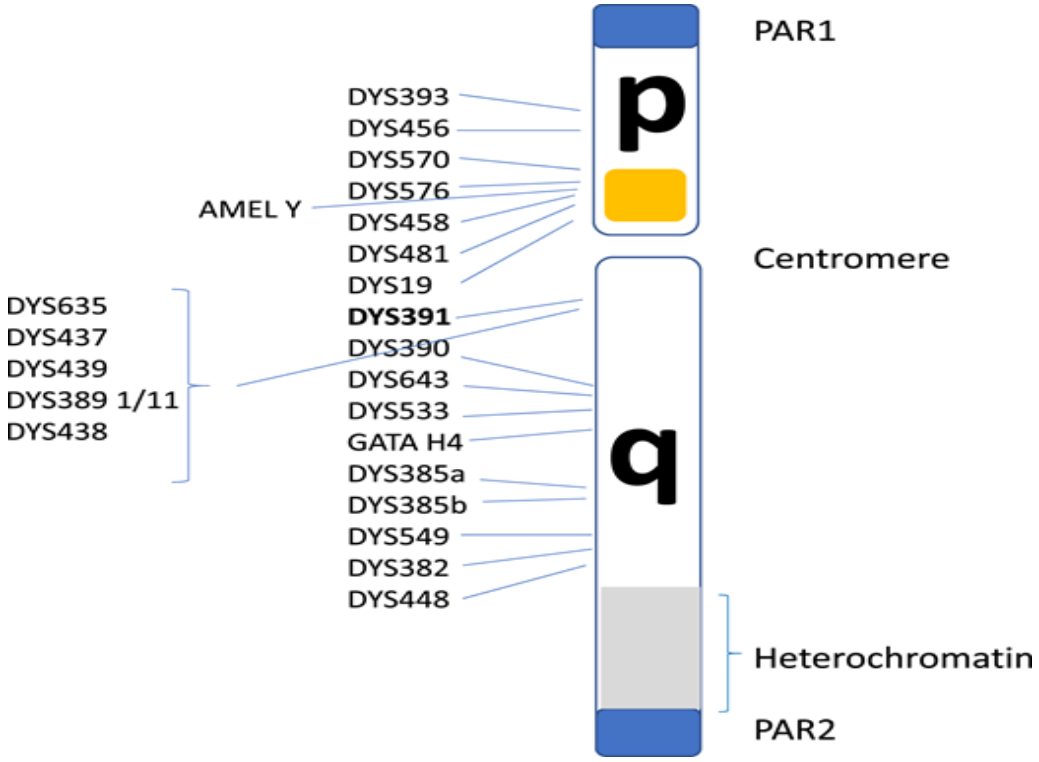



**Figure 1.7. The Y chromosome, highlighting the approximate locations of commonly used Y-STR markers in forensic genetics (Syndercombe, 2021).**

### 1.4.2 APPLICATION AND RELEVANCE OF Y-STR TYPING

Y-STR analysis in forensic DNA profiling helps identify male DNA in criminal investigations, missing person cases, and paternity testing (for sons only) when standard STR profiling is inconclusive. Y-STR typing has been crucial in sexual assault cases, mainly when mixed samples from female victims make it challenging to identify the male components, especially when standard STR profiling fails or shows weak amplification (Jobling and Gill, 2004; Butler, 2014). Y-STR typing enables the recovery of robust male DNA profiles without the need for differential DNA extraction, particularly in samples dominated by female DNA (Roewer, 2019). This approach has proven especially valuable in multiple-perpetrator sexual assault cases, where the female contributor is the major DNA donor and several male contributors are present as minor components, resulting in highly complex mixed autosomal STR profiles (Syndercombe, 2021). In such scenarios, Y-STR profiling isolates male-specific genetic information and can assist in narrowing the pool of potential male contributors (Forouzesh *et al*., 2022). Y-STR markers are also important beyond forensic work. Y-chromosomal haplogroups are widely used to reconstruct paternal lineages and to map historical human migration patterns, providing insights into population structure and biogeography (Roewer, 2019). Y-STR analysis is commonly used in genealogy, such as surname studies, because men in the same family line usually share similar Y-STR haplotypes (Syndercombe, 2021). In addition, Y-chromosome analysis has applications in

medical genetics, where it has been used to investigate conditions such as male infertility, Turner syndrome (45, XO), and other Y-linked genetic disorders (Forouzesh *et al*., 2022).

The development of Y-STR analysis has revolutionised forensic sciences, genetic genealogy, and population genetics. Since its initial discovery in the 1990s, Y-STR profiling has emerged as a vital method for male-specific DNA analysis, paternal lineage studies, and criminal investigations, particularly with the adoption of commercial kits and next-generation sequencing (NGS) technology (Bozzo *et al*., 2019; Roewer, 2019). Y-STRs have been helpful in genetic genealogy. With the rise of direct-to-consumer genetic testing (e.g., AncestryDNA, 23andMe), Y-STR analysis has become a popular tool for tracing paternal ancestry. Identifying Y-chromosomal haplogroups has helped individuals connect with historical and geographical lineage data (Jobling and Tyler-Smith, 2017).

## 1.5 FORENSIC STATISTICS AND PARAMETERS

### 1.5.1 ALLELE FREQUENCY

Alleles are alternative forms or variants of a gene located at a specific position (locus) on a chromosome. As diploid organisms, humans receive one allele from each parent, meaning they have a pair of alleles for each genetic position on a chromosome (Pierce, 2012). These alleles can occur in a population with either similar or different frequencies. Allele frequency is important in the analysis of population genetics as it reflects commonness of specific gene variants within a population and helps in understanding gene diversity and evolution of a population (Vizmanos *et al.,* 2020). The frequency of a particular allele (denoted as $p$) in a population can be calculated by dividing the number of times that allele appears ($i$) by the total number of alleles at that locus ($N$). This is represented as p =i/N. This formula helps determine how common a specific variant (allele) is at a particular genetic locus. In a biallelic system (e.g., alleles A and a), the frequencies are often represented as $p$ and $q$, where $p + q = 1$. Allele frequencies are essential for understanding genetic differences. They help make wise choices in forensic science, medicine, conservation, and breeding. Allele frequencies also play a role in studying how populations change over time (Park *et al*., 2021).

### 1.5.2 GENOTYPE FREQUENCY

A genotype is an individual's genetic makeup, specifically characterized by the distinct combination of alleles inherited from both parents at one or more loci on their chromosomes (Kockum, *et al*., 2023). In any population, alleles will combine in the genotype frequencies; pp

(homozygote), pq (heterozygote), and qq (homozygote). In a randomly mating population with no evolutionary forces acting on it, allele frequencies are expected to remain constant across generations. Genotype frequency is crucial because it reflects the proportion of individuals with a particular genetic makeup. An understanding of genotype frequencies assists in examining genetic diversity and inheritance patterns. The calculation of genotype frequencies can be achieved by multiplying the corresponding allele frequencies. The genotype frequencies for homozygous loci are computed as a square of the allele frequencies {(p2 = A x A) and (q2 = a x a)}. In contrast, the frequencies of heterozygous genotypes are calculated using the formula 2pq, where p represents the frequency of the A-allele and q represents the frequency of the a-allele.

### 1.5.3 HARDY-WEINBERG EQUILIBRIUM (HWE)

In 1908, two scientists, Godfrey Harold Hardy and Wilhelm Weinberg, first introduced the concept of Hardy-Weinberg equilibrium (HWE). This principle states that a population's allele and genotype frequencies are bound to remain constant from generation to generation unless disruptive evolutionary forces influence them (Wright, 1931; Crow, 2017). The population genetics module of HWE is based on a series of five assumptions: no natural selection, no mutation, random mating, no migration (gene flow), and a sufficiently large population size to prevent genetic drift. The Hardy–Weinberg principle assumes that allele and genotype frequencies are bound to remain constant from generation to generation in a large population of diploid individuals with random mating, no overlapping generations, and no evolutionary influences. However, it is practically impossible for a population to meet all these assumptions, as factors such as inbreeding, gene flow, mutation, natural selection, migration, and genetic drift frequently affect genetic variation in natural populations. Hence, HWE is a theoretical concept that describes the expected relationship between allele and genotype frequencies in a population not undergoing evolutionary change. It provides a foundational baseline in population genetics for assessing whether allele and genotype frequencies remain stable from generation to generation. STR loci should be tested for conformity with HWE before being used in further statistical analyses, particularly in forensic and population genetic studies. New populations must also be tested for HWE as data are collected. Deviations from HWE suggest that evolutionary forces such as selection, mutation, migration, or non-random mating may be acting on the population. Importantly, the sum of allele frequencies and genotype frequencies at a given locus are equal, whether or not the population is in equilibrium (Wier, 1996; Hartl and Clark, 1997).

$$\text{Sum of all allele frequencies} \rightarrow p + q = 1$$

$$\text{Sum of all genotype frequencies} \rightarrow p^2 + 2pq + q^2 = 1$$

### 1.5.4 RANDOM MATCH PROBABILITY (RMP)

Random Match Probability (RMP) refers to the likelihood that a randomly selected, unrelated individual from a specific population would coincidentally have a DNA profile matching that obtained from an evidence sample. RMP is typically applied when the DNA profile originates from a single source and a suspect has already been identified. RMP is distinct from "match probability", which refers to the likelihood of a match between random individuals in a general population or profile frequency (Butler, 2014). RMP can be determined by multiplying the genotype frequencies derived from allele frequencies across all loci. This calculation estimates the frequency of a particular DNA profile within a given population. The RMP is used when there is a match between a DNA profile obtained from evidence and a reference DNA profile from a suspect. This helps to determine the likelihood that the match occurred by chance. Calculating the RMP is crucial for evaluating the importance of the matches observed between individuals. It evaluates the strength of evidence to support conclusions about a suspect's likelihood of being associated with a crime scene, thereby equipping forensic scientists with more informed judgment about the observed data and aiding in the interpretation of the analysis. To calculate a complete profile for multiple loci, assuming that the loci are independent, the formula is expressed as follows:

$$\mathrm{RMP} = \prod_{i=1}^{n} \mathrm{GF}_i$$

Where $n$ represents the number of STR loci in the profile (e.g., 15 or more), each genotype frequency is calculated using allele frequencies specific to the relevant population (Balding and Nichols, 1994; Butler, 2014).

The probability of a match (PM) at a given locus in the population, as reported by Fisher (1951) gave:

$$PM = \sum_{i} G_i^2$$

Where $G_i$ is the genotype frequency at a particular locus in the population, which can be squared and summed across loci when dealing with multiple loci because each locus contributes independently to the overall probability of a random match.

RMP is primarily used to calculate match probabilities in autosomal STR analysis. This includes factors such as gene segregation, independent assortment, and recombination. In contrast, haplotype match probability (HMP) is used to analyse Y-STR and other haplotype groups. HMP is the likelihood that two randomly chosen individuals from a population share the same haplotype.

Unlike RMP, which considers recombination, HMP is derived from haplotype frequency databases since Y-STRs do not undergo recombination. HMP is typically estimated as:

$$HMP = 1/N$$

Where $N$ represents the total number of unique haplotypes in the reference database.

### 1.5.5 LIKELIHOOD RATIO (LR)

This is another statistical framework for presenting DNA profile evidence, different from the RMP. The Likelihood Ratio (LR) plays a central role in the Bayesian approach to evaluating forensic evidence. This method helps evaluate forensic DNA evidence clearly and logically by comparing how likely the evidence is under two different explanations. It involves measuring evidential strength under two competing hypotheses: that the recovered DNA originated from the suspect (prosecution hypothesis—$H_p$) or an unrelated individual in the population (defence hypothesis—$H_d$). LR is evaluated by considering the probability of observing the DNA evidence under two competing hypotheses: that the DNA profile originated from the suspect $P(E|\ H_p,\ I)$ (prosecution hypothesis) or that it originated from an unknown, unrelated individual $P(E|\ H_d,\ I)$ (defence hypothesis), given relevant conditioning information (I) (Evett & Weir, 1998). For DNA evidence, this can be straightforwardly carried out by calculating the reciprocal of the RMP under most circumstances, but not universally.

$$\mathrm{LR} = \frac{1}{\mathrm{RMP}}$$

The lower the RMP, the higher the LR, the stronger the evidential strength. (Norrgard, 2008). The equation (LR = 1/RMP) holds when the numerator $P(E|\ H_p,)$=1, assuming the DNA evidence perfectly matches the suspect's profile, and the laboratory analysis is accurate. This reflects the idea that if the suspect is the trustworthy source of the DNA, then observing the matching profile is almost inevitable. The denominator $P(E|\ H_d,)$ represents the probability that the same DNA profile could occur by chance in a random, unrelated individual from the population. This probability is referred to as the RMP (Butler, 2014).

DNA evidence can be reported in court as the RMP, LR or as a verbal expression describing the evidential strength (Perlin, 2010). For example, a statistical calculation, which gives an RMP estimate of one in one billion or $1x10^{-9}$; this can be expressed as "the chances of finding an individual with the same profile as the one obtained in the same population is one in one billion" (Thompson *et al*., 2003). The LR can be reported as "the matching profiles are at least one billion times more likely to be obtained if the DNA came from the suspect than an unknown unrelated individual" (Perlin, 2010), while a verbal equivalence of evidential strength can be expressed as

"the result provides extremely strong support for the prosecution proposition" (Martire *et al*., 2013). The statistical assessment of a matching profile furnishes the jury with an estimate of the evidence's importance in court.

### 1.5.6 POWER OF DISCRIMINATION (PD)

This essential forensic parameter helps convey information in forensic investigations and paternity testing; however, it is not typically reported in court. Power of discrimination (PD) specifically indicates the usefulness of certain markers and serves as an assessment of how these markers can be utilised in the future. It describes the effectiveness of a genetic marker in distinguishing between individuals within a population. PD represents the probability that two randomly selected, unrelated individuals will not share the same genotype at a given locus. A higher value of discriminatory power indicates a reliable accuracy in distinguishing individuals in a population. PD calculation is important in assessing the informativeness or discriminatory capacity of a genetic marker or a set of markers. This is crucial in forensic investigations to determine whether the selected markers are capable of reliably linking a victim or suspect to a crime scene, and of effectively excluding unrelated individuals from suspicion. PD is a fundamental step in the selection and validation of genetic markers because markers with high discrimination power provide are preferred as they provide more reliable and accurate identification of individuals.

Power of discrimination can be determined by deducting the value of a locus's random match probability (RMP) from 1, which increases if more markers are added (Jeffrey *et al.*, 1985c).

$$PD = 1 - RMP$$

### 1.5.7 PATERNITY INDEX (PI)

A specialised form of the likelihood ratio is the Paternity Index (PI), which calculates the parenthood probability. It considers the likelihood that a particular man is the biological father of a child (under the paternity hypothesis) against the possibility that a random, unrelated man from the population is the father (under the non-paternity hypothesis). Typically, PI is calculated for each locus individually, and the Combined Paternity Index (CPI) is obtained by multiplying the individual PI values across all tested loci. Assessing these forensic parameters will demand the cognisance of allele frequencies and their distribution in the population (Butler, 2006). A higher paternity index indicates a more substantial paternity likelihood of an alleged father, thereby strengthening the evidence supporting the existence of a biological relationship. PI is calculated by comparing the probability of the child's genotype given the alleged father and mother under the

paternity hypothesis $\{H_{pi}\ Pr(G_C, G_M, G_{AF} \mid H_p)\}$ to the probability of the child's genotype given the mother and a random, unrelated man from the population under the non-paternity hypothesis $\{H_{di}\ Pr(G_C, G_M, G_{AF} \mid H_d)\}$. Typically, allele frequencies from the relevant population are used to calculate these probabilities.

$$PI = \frac{\Pr(G_C, G_M, G_{AF} \mid H_p)}{\Pr(G_C, G_M, G_{AF} \mid H_d)}$$

Where $H_p$ is the prosecution hypothesis (i.e. the father of the child is the alleged father); $H_d$ is the defence hypothesis (i.e. the father of the child is a random man unrelated to the alleged father); $G_C$ is the child's DNA profile; $G_M$ is the mother's DNA profile; $G_{AF}$ is the alleged father's DNA profile (Gjertson *et al.,* 2007; Tillmar, 2010).

### 1.5.8 POWER OF EXCLUSION (PE)

Power of Exclusion (PE) is a forensic parameter that assesses the probability of excluding an unrelated individual as the biological source of a DNA profile at a specific locus. It is critical to rule out innocent individuals as potential sources of DNA samples found at a crime scene. The PE is also applied in eliminating relatives (e.g. brothers) of the true father of a child as the biological father in a paternity investigation. A higher power of exclusion enhances the dependability of forensic DNA examination by reducing the possibility of false inclusion (Hameed *et al.*, 2015). The formula below is applied in the calculation.

$$PE = h^2 \cdot (1 - 2 \cdot h \cdot H^2)$$

Where $h$ and $H$ are the proportions of heterozygous and homozygous individuals in the population, respectively.

### 1.5.9 POLYMORPHISM INFORMATION CONTENT (PIC)

The polymorphism information content (PIC) is a forensic parameter that conveys the informativeness and variability of a genetic marker by measuring the polymorphism, i.e. the multiple forms of genes or alleles among individuals in a population, particularly in the context of population studies and genetic diversity. The higher the polymorphic content in a specific gene locus, the greater the genetic differences among individuals, thus indicating higher diversity, which makes the gene locus useful for forensic analysis, genetic research, and studying populations. PIC is essential because it measures how informative a marker is. Linkage mapping helps identify when genes recombine and makes the genetic map more accurate. In the analysis of markers, it helps select markers that show a lot of variation, which helps to distinguish between different individuals or genotypes clearly. It is valuable in marker selection, genetic diversity comparison, and marker

comparisons (Botstein *et al.,* 1980; Shete *et al*., 2000). The formula for the calculation is stated below:

$$\text{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

Where $p_i$ and $p_j$ are allele frequencies; $n$ is the number of alleles.

### 1.5.10 OBSERVED HETEROZYGOSITY ($H_o$) AND EXPECTED HETEROZYGOSITY ($H_{exp}$)

Other important forensic parameters are the observed heterozygosity ($H_o$) and expected heterozygosity ($H_{exp}$). Observed heterozygosity ($H_o$) is the proportion of heterozygous individuals (those with two different alleles, e.g., Aa) observed at a specific locus within a sampled population. It is calculated by dividing the number of individuals who are heterozygous at that locus by the total number of individuals sampled ($N$) (Hameed *et al*., 2015).

$$Ho = \frac{Aa}{N}$$

Expected heterozygosity ($H_{exp}$) at a given locus is the probability that two randomly chosen alleles from the population are different, assuming Hardy-Weinberg equilibrium (HWE). It is calculated by subtracting the sum of the squared allele frequencies (representing the expected homozygote genotype frequencies) from one.

$$H_{exp} = 1 - \sum p_i^2$$

Where $p_i$ is the frequency of the $i^{\text{th}}$ allele at the locus, calculated under Hardy-Weinberg equilibrium based on the observed genotype frequencies in the sampled population, the summation is taken over all alleles at that locus (Butler, 2015a). The expected heterozygosity could also be called the **gene diversity (GD).** The GD is the probability that estimates the genetic differences when two alleles are randomly drawn from a population (Tillmar, 2010). An unbiased estimate formula for the calculation of GD was described by Nei (1987) and Edwards *et al*. (1992).

$$GD = \frac{n}{n-1}\left(1 - \sum_i p_i^2\right)$$

Where $n$ is the sampled allele counts at a locus; $p_i$ is the allele frequency in the population.

### 1.5.11 WRIGHT'S F-STATISTICS ($F_{ST}$, $F_{IS}$, and $F_{IT}$)

Wright's F-statistics ($F_{ST}$, $F_{IS}$, and $F_{IT}$) are essential population genetic parameters utilised in population genetics to quantify genetic differentiation, inbreeding, and population structure, with important applications in fields such as evolutionary biology, conservation, ecology, and forensic genetics. The F-statistics are applied in estimating a population's genetic structure and measuring

the deviation from HWE in a subdivided population. Sewall Wright developed this statistical method alongside Ronald Fisher and John Haldane (Wright, 1922; Fisher, 1949; Holsinger and Weir, 2009). The letters ***F, $F_{ST}$, $F_{IS}$,*** and ***$F_{IT}$*** typify the "fixation index", specifically for the "**S**ubpopulation within the **T**otal population", "**I**ndividual within the **S**ubpopulation", and "**I**ndividual within the **T**otal population" respectively. They help in forensic genetics by showing how genetic differences are distributed. They ensure accurate estimates of allele frequencies and improve DNA evidence interpretation when there is population substructure (Weir, 2012).

The ***$F_{ST}$*** statistic shows the genetic difference between subpopulations and the entire population. It considers the reduction in heterozygosity due to the variation in the population structure. ***$F_{ST}$*** values range from 0 to 1. 0 means no genetic difference (the subpopulations are genetically identical), while 1 means a complete genetic difference between the subpopulations. The calculation is based on the expected heterozygosity (Weir and Cockerham, 1984; Weir, 1996). The ***$F_{ST}$*** statistics can be calculated using the formula below.

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Where $H_T$ is the total genetic variation, measured as the average expected heterozygosity in the overall population (considering all subpopulations as one); $H_S$ is the within-subpopulation variation, measured as the average expected heterozygosity within individual subpopulations.

The ***$F_{IS}$* statistic** assesses genetic variation correlation within individuals and their subpopulations, specifically measuring deviation from Hardy-Weinberg equilibrium in subpopulations, often referred to as the **inbreeding coefficient.** The value ranges from -1 to +1. A high positive value ($F_{IS} > 0$) indicates inbreeding, which reflects a deficit of heterozygotes and an excess of homozygotes within the subpopulation. Conversely, a negative value ($F_{IS} < 0$) suggests outbreeding or disassortative mating, indicating an excess of heterozygotes. A value of zero ($F_{IS} = 0$) implies random mating under Hardy-Weinberg equilibrium. The $F_{IS}$ statistic is calculated based on the observed and expected heterozygosity proposed by Weir and Cockerham (1984) and further described by Weir (1996). $F_{IS}$ can be calculated using the formula below.

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

Where $H_I$ is the variation of the individual (average heterozygosity of the individual); $H_S$ is the variation within the subpopulation (average heterozygosity in the subpopulation);

The ***$F_{IT}$* statistic** represents the correlation of genetic variation within an individual relative to the total population. The calculation is based on the expected heterozygosity. The ***$F_{IT}$*** statistic considers the overall inbreeding of the individual within and between the components of the population ($F_{IT} = F_{IS} + F_{ST}$) – ($F_{IS}$ x $F_{ST}$). The ***$F_{IT}$*** statistic is less frequently utilized (Weir and Cockerham, 1984; Weir, 1996). ***$F_{IT}$*** can be calculated using the formula below.

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

Where $H_I$ is the variation of the individuals (average heterozygosity of the individual); $H_T$ is the variation between populations (average heterozygosity in the total population).

### 1.5.12 BONFERRONI CORRECTION

The Bonferroni correction is a statistical adjustment used to control the familywise Type I error rate when multiple statistical tests are performed on the same dataset. When multiple hypotheses are tested simultaneously, the probability of obtaining at least one false-positive result increases beyond the nominal significance level (α = 0.05). The Bonferroni correction addresses this by applying a more stringent significance threshold to each individual test (Weir, 1996).

The familywise error rate (αFW) is defined as the probability that at least one Type I error occurs across a set of multiple statistical tests and can be expressed as:

$$\alpha_{FW} = 1 - (1 - \alpha_{PC})^{C}$$

where αPC is the per-comparison error rate and *c* is the number of comparisons performed. The Bonferroni correction provides a conservative approximation to control αFW by dividing the desired overall significance level by the number of tests conducted:

$$\alpha_{PC} = \alpha/c$$

Thus, when an overall significance level of α = 0.05 is used, the adjusted significance threshold for each test becomes 0.05/c. This adjustment reduces the likelihood of false-positive results arising from multiple testing and ensures that the familywise error rate remains controlled (Weir, 1996; Rice 1989).

## 1.6 BACKGROUND OF FORENSIC DNA DATABASES

### 1.6.1 OVERVIEW

Forensic DNA investigation does not rely on the intelligence of the offender DNA database alone before prosecuting a case in court. A DNA profile must be evaluated statistically for profile matches

within the population to be accepted in court. The statistical calculations of probabilities, such as RMP or LR, first consider the prevalence of alleles and genotypes within the population. In forensic investigations, there are two types of DNA databases: population and criminal. The rationale behind creating a national criminal DNA database of offender profiles is to identify any criminals becoming repeat offenders. The first and oldest national forensic DNA database to be launched worldwide for crimes punishable by imprisonment was the UK National DNA Database (NDNAD), established on 10th April 1995. (Graham, 2007; Maguire *et al*., 2014). Over 60% of convicts of past crimes in the United Kingdom get rearrested after three years of being set free (Wallace, 2006). Also, about 90% of convicted sex offenders are serial perpetrators, and more than 50% of robbery case perpetrators have faced one-time or more convictions in the past. Law enforcement agencies regularly analyse DNA profiles within the national DNA database to identify suspects, verify links to past or current cases, and solve crimes. It also provides the police with intelligence about potential suspects by checking a profile against the NDNAD when no known suspect exists in a case (Amankwaa and McCartney, 2019; Dash *et al.,* 2023).

Research has revealed that DNA databases of criminal offenders decrease criminal activities in a nation (Roewer, 2013). A strong and reliable DNA database system helps a country fight and reduce crime more effectively (Doleac, 2017). Much controversy arose due to varying policies across countries regarding information storage, suspect inclusion, storage duration, and usage purpose during the creation and operation of national DNA databases (Roman-Santos, 2010). The concerns about ethical privacy have been a subject of debate (Wallace, 2006). Nonetheless, the implementation of the Protection of Freedoms Act 2012 has addressed the matter of retaining and regulating DNA profiles in the UK National DNA Database (Amankwaa and McCartney, 2018; Ferrara, 2024). The UK National DNA Database for criminal offenders has successfully identified suspects in many criminal investigations by linking them to a crime scene when no known suspect (Maguire *et al*., 2014). A “cold hit” is the terminology applied when crime scene's DNA profile matches a DNA profile held in a national DNA database for criminal offenders (Wallace, 2006; Doleac, 2017). This is very much attainable if the suspect has previously committed a crime and had their DNA profile captured in the criminal DNA database. DNA technology can show a suspect's innocence by proving their DNA does not match the DNA collected from a crime scene (Johnson and Williams, 2004a). In 2024, the UK National DNA Database (NDNAD) achieved a 64% hit rate from 7.2 million subjects and 692,365 crime scene profiles, with 1,377,140 females, 5,851,779 males, and 40,608 of unknown gender (Stanciu *et al*., 2024; Uberoi *et al.,* 2024). Also, in the same year, the US Combined DNA Index System (CODIS) assisted in more than 562,412 cases. The hit rate from the UK National DNA Database in 2024 identified up to 26,955

perpetrators, who were further convicted for an additional 1,593 violent crimes. These crimes involved 618 murders, 629 rapes, 138 other related sexual offences, 432 thefts, 58 frauds, 11,493 burglaries, and 1,034 robbery cases. (Stanciu *et al*., 2024).

A survey carried out by the international policing agency, Interpol, and the Council for Responsible Genetics in 2022 revealed that 70 of its 89 countries possessed national DNA databases while more than 26 countries were making plans to establish a national DNA database among 120 countries that applied DNA profiling in the investigation of crime (Butler, 2023). Among the countries with forensic DNA databases, nine out of 89 collected DNA profiles representing over 5% of their total populations **(Figure 1.8)**. These countries include Australia, Austria, China, Estonia, France, Latvia, Lithuania, the United Kingdom, and the United States of America (Uberoi *et al*., 2024).

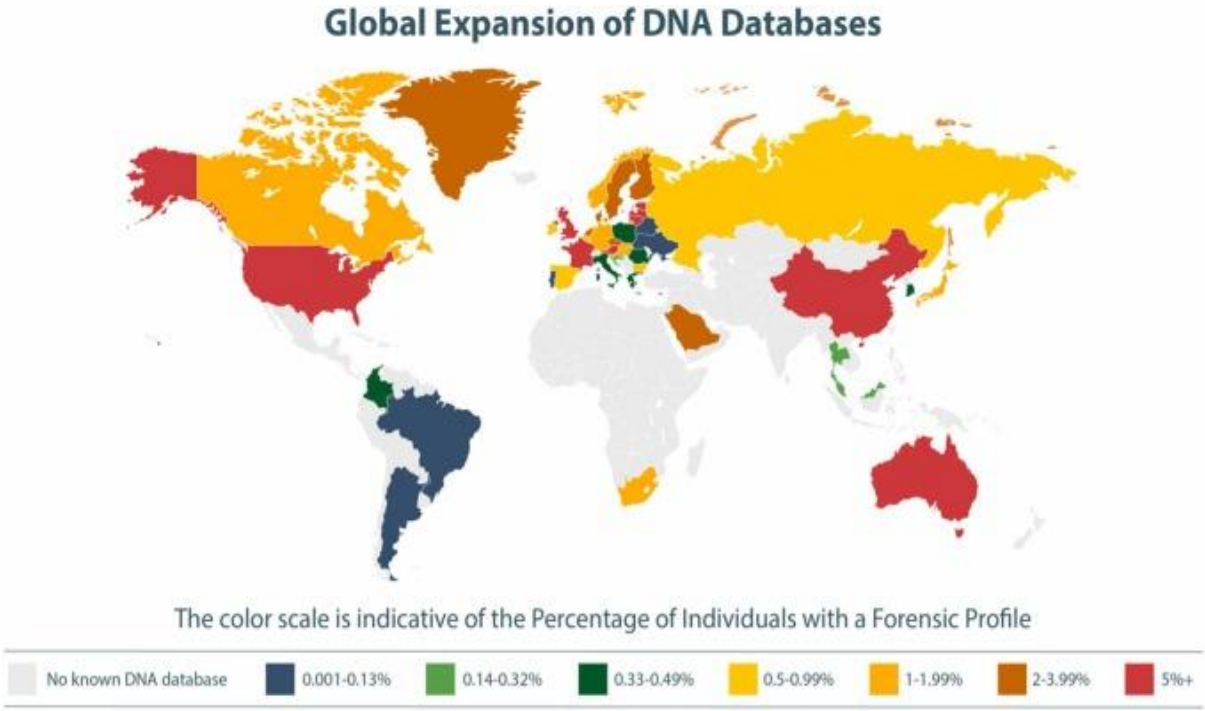



**Figure 1.8. Percentage of individuals with forensic profiles in countries operating forensic DNA databases (https://gdo.global/en/dna-profiles-in-G20-and-EU-countries).**

The world's largest national DNA database currently is the "China's DNA database. It contains between 100 and 140 million offender profiles in it (Dirks and Leibold, 2020). The database constantly changes, so its size is estimated rather than precisely known. Exact numbers are not publicly available, and different sources provide varying estimates due to a lack of transparency. The United States (US) National DNA Index System (NDIS) is the second-largest national DNA database, containing more than 19 million offender profiles (Federal Bureau of Investigation, 2024). In comparison, the third largest, the United Kingdom's (UK) National DNA Database, holds more than 7 million subject sample profiles (Home Office, 2024a; 2024b).

### 1.6.2 TYPES OF FORENSIC DNA DATABASES

#### 1.6.2.1 NATIONAL CRIMINAL DATABASES

The national criminal DNA databases are sometimes referred to as the "intelligence databases" or "offenders' DNA databases" (Gill, 2002). National criminal DNA databases are government-run databases of DNA profiles that law enforcement agencies use to identify suspects during a criminal investigation. The national database assists in curtailing criminals from advancing to more severe crimes by allowing the connection of separate crimes to a single offender and granting the re-evaluation of cold cases (Linacre, 2003; Wallace, 2006). A national database is structured to contain two main types of DNA profiles: i. Crime scene DNA profiles, which are generated from evidence recovered at crime scenes, and ii. Reference DNA profiles, which consist of known offender profiles from individuals who have been convicted of crimes (McCartney, 2005). Databases contain DNA data obtained from individuals' profiles, which are created using a variety of kits to amplify STR regions through PCR. Different countries and regions, like states or provinces, use various DNA testing kits that contain different numbers of genetic markers (Bodner *et al*., 2016). Each place has its laws and rules for managing forensic DNA databases. They choose specific kits for Short Tandem Repeat (STR) testing based on what they need to meet international standards (Tan *et al*., 2017). Factors influencing this choice include how many markers they use (which affects how well they can identify individuals), differences in genetic makeup among populations, local laws, and available technology. As a result, the panels of STR markers can vary a bit, which affects how well databases can match DNA from different places (Jacewicz, 2018). In the United Kingdom, the number of STR loci used in forensic DNA profiling increased from 1995 to 2014 as result of the lower power of discrimination provided by a small number of loci. Increased loci numbers increase the discriminatory power (Johnson and William, 2004a; Hopwood *et al.,* 2012). Consequent to the issues on expanding the national DNA database, it became apparent that the chances of getting adventitious matches will also increase as the database increases. Due to this development, an increment in the number of loci was introduced. In 1999, the UK National DNA Database (NDNAD) upgraded the SGM system, which initially contained six loci, to the AmpF*l*STR SGM Plus system, which had an additional four loci, making it up to 10 loci in total, including Amelogenin (the sex marker) (Gill, 2002). The successful upgrade of the SGM loci for expanding UKDNAD reduced the chances of adventitious matches and lowered the matching probability ($10\text{-}^{13}$) (Foreman and Evett, 2001). Further investigation also revealed that additional loci would reduce the match probabilities and chances of adventitious matches. This benefits countries with a more extensive national DNA database than the United Kingdom. DNA kits such as the "Applied Biosystems 16-plex (AmpF*l*STR® Identifiler Plus™)" and the "Promega 16-plex (Powerplex® 16 BIO System)" provide very low match probabilities down to $10^{-20} – 10^{-25}$ (Gill,

2002). Recent technological advancements have enhanced Short Tandem Repeat (STR) kits by incorporating additional loci. In 2014, England and Wales adopted a 17-STR-locus system, while Scotland has been using a 24-STR-locus system for DNA profiling since 2015 (Scottish Legal News, 2015; Puch-Solis and Pope, 2021). Including more STR loci increases the ability to differentiate between samples, even when a DNA sample contains only a partial profile (Johnson and William, 2004b).

In 2018, the UK National DNA Database (NDNAD) was the largest and most inclusive forensic DNA database, with a significant proportion of citizens included. It maintains DNA profiles from police forces across England and Wales, as well as from the Northern Ireland DNA Database and Scottish DNA database. Additionally, it includes DNA data from the Crown Dependencies, which comprise the Isle of Man, the Bailiwick of Jersey, and the Bailiwick of Guernsey (Amankwa and McCartney, 2018). The legal foundation for the offenders' DNA database was established in 1994, and by 2024, it had grown to include 7,045,155 subject-reference DNA samples (Home Office, 2024). The UK National DNA Database (NDNAD) is currently governed by the Home Office and overseen by the Forensic Information Databases Strategy Board. Access to the National DNA Database (NDNAD) and the ability to contribute are confined to a select group of authorised personnel. Only accredited laboratories are authorised to submit DNA profiles to the NDNAD. In England and Wales, examples of companies that supply DNA profiles to NDNAD include the Key Forensic Services, Cellmark Forensic Services, and Eurofins Forensic Services. The Scottish DNA Databases and the UK DNA Databases have different but overlapping legislation guiding the retention of STR profiles and have different DNA STR marker systems. The Scottish DNA Database also operates under a stricter law in retaining DNA profiles and samples, although it is similar now. Scotland has strict regulations regarding the deletion of DNA from individuals who have not been convicted of crimes. It also has a body that oversees this process independently. After reform, Northern Ireland created laws similar to those in England and Wales and added safeguards in 2012 (Wallace, 2006; Amankwaa and McCartney, 2018).

In the UK, DNA samples submitted to the national DNA database are generally stored for future comparison with potential crimes, as outlined in the Protection of Freedoms Act 2012 (Siegel and Saukko, 2012). The retention of DNA profiles in the UK National DNA Database has been controversial since the database was established. The legislation guiding the retention of DNA profiles has undergone several changes which include the Police and Criminal Evidence Act (1984), Criminal Justice and Public Act (1994), Criminal Procedure and Investigation Act (1996), Criminal Evidence {Amendment Act} (1997), DNA Expansion Programme (2000), Criminal

Justice and Police Act (2001) and the Criminal Justice Act (2003), Serious Organised Crime and Police Act (2005) and Counter-Terrorism Act (2008) (Williams and Johnson, 2005; Blakemore and Blake, 2012; McCartney, 2013). As a result, issues surrounding government surveillance, human rights, and privacy of the DNA profiles stored in the national DNA database have been raised (McCartney, 2004; Wallace, 2006). The debate focused on the duration and regulation of DNA profile retention. Similar issues have also affected other world nations as concerns about developing and enlarging forensic DNA databases are raised (William and Johnson, 2005). **Table 1.4** summarizes the conditions for retaining and deleting a DNA profile in various European countries. Issues concerning the retention of DNA profiles started in March 2006 in a case involving a supply teacher from Birmingham, Philippa Jones, who was arrested in 2005 for the alleged assault of a child. Her DNA was taken while she was still in custody, and the charges were later dropped when the Crown Prosecution Service decided not to pursue the case further. She later filed a case in the High Court accusing the police authority of DNA and fingerprints being collected unlawfully without appropriate authority. She obtained the legal right to remove her photograph, DNA, and fingerprint data from her records. Subsequently, she received compensation for £250 for damages (Kaye, 2006; Amelung and Machado, 2019). The debate on the retention of DNA further escalated in 2008 in a case involving Michael Marper and Mr S, which resulted in the abolishment of the indefinite retention of DNA under prosecution. The case of "S & Marper v United Kingdom [2008] ECHR 1581" involved Mr S, a minor (11 years old) charged with attempted robbery and Michael Marper, a young man arrested for the assault of his girlfriend on 19 January 2001 and 13 March 2001 respectively. They both had their DNA and fingerprint profiles taken and submitted to the UK NDNAD but were not convicted as they were lawfully acquitted. Michael Marper never had any charges pressed due to the pretrial reconciliation with his girlfriend. Application by S. and Marper to seek the removal of their DNA profiles from the national DNA database was futile as the High Court, Court of Appeal, and House of Lords upheld the legality of retaining DNA samples. Subsequently, an appeal was filed with the European Court of Human Rights on February 27, 2008, which resulted in a ruling by 17 judges on December 14, 2008, that such retention violated Article 8 of the European Convention on Human Rights (ECHR), protecting the "right to respect an individual's private or family life". Each appellant was awarded €42,000, and the ruling led to the Protection of Freedoms Act 2012, which revised DNA retention laws in the UK NDNAD. However, it allowed the indefinite retention of those convicted of any crimes but mandated the deletion of DNA profiles generated from those not convicted at the end of a trial (Nydick, 2009; Cole, 2013; Amelung and Machado, 2019). **Table 1.4** summarizes the final decision of the **Protection of Freedom Act 2012.**

**Table 1.4. Summary of the conditions for the retention and deletion of DNA profiles in the national databases of some European countries (Santos *et al.*, 2013; Obleščuk *et al.*, 2024)**

| Country | Condition for Retention | Condition for Deletion |
|---|---|---|
| **Austria** | Convicted lawbreakers or Suspects of dangerous/violent crimes. | Below 18 or Minor: DNA profiles can be deleted if the individual does not have any previous record of forensic identification for the past three years.<br>Convicted adult: DNA profiles can be deleted after five years of demise or at age 80 if the individual does not have any previous record of forensic identification for the past five years.<br>Acquitted suspect: An application needs to be made for DNA profile deletion, and the decision for the removal lies on the authorities to determine if the retention of the acquitted suspect's profile is unnecessary. |
| **Belgium** | Convicted lawbreakers or suspects of serious crimes (based on crimes listed) | Convicted criminal: DNA profiles can be deleted after 30 years of inclusion. When inessential, authorities can delete retained profiles in the "criminal investigation" database. |
| **Denmark** | Convicted lawbreakers of crimes punishable by more than one year and six months sentence or suspects of similar crimes | Convicted criminal: DNA profile can be deleted after two years of demise or at age 80.<br>Suspect: DNA profile can be deleted after two years of decease, after ten years of acquittal or at age 70. |
| **Estonia** | Convicts and suspects | Convicted criminal and suspect: DNA profile can be deleted after ten years of demise. |
| **Finland** | Convicted lawbreakers with more than three years' sentence and suspects of crimes punishable by more than six months' sentence. | Convicted criminals: DNA profile can be deleted after ten years of demise.<br>Suspects: DNA profile can be deleted after one year of acquittal (based on the legal officer's order) or after ten years of demise. |
| **France** | Convicted lawbreakers or suspects of serious crimes. (based on crimes listed) | Convicted criminal: DNA profile can be deleted after 40 years of the termination of a served sentence or at age 80.<br>Suspect: DNA profiles can be deleted at the request of the concerned parties or when law authorities consider the retention unnecessary. |
| **Germany** | Convicted lawbreakers of serious crimes or individuals indicted with more than one crime and suspects officially charged with a criminal offence. | Convicted criminal: Deletion of DNA profiles is at the discretion of the court.<br>Adult: DNA profiles are reviewed after ten years of inclusion.<br>Young individual: DNA profiles examined after five years of inclusion.<br>Children: DNA profiles reviewed after five years of inclusion; |
| **Hungary** | Convicted lawbreakers and suspects of crimes punishable by more than five years' sentences (or criminal offences | Convicted criminal: DNA profile can be deleted after 20 years of sentence service termination.<br>Suspects: DNA profile can be removed after acquittal. |

| | | |
|---|---|---|
| | listed for lower sentences, e.g., drug trafficking). | |
| **Ireland** | Convicted lawbreakers and suspects of crimes punishable by more than five years sentence (or specific criminal offences listed for lower sentences) and ex-convicts. | Convicted criminal: DNA profile is retained indefinitely.<br>Suspects: DNA profile can be deleted after ten years for an adult or five years for a minor if the individual is acquitted or not charged. |
| **Italy** | Suspects arrested, detained in custody, and convicted of preconceived criminal offences. | Convicted criminal: DNA profile can be deleted after 20 years of the incident but never held more than 40 years.<br>Suspects: DNA profile can be deleted after acquittal. |
| **Latvia** | Convicted lawbreakers and suspects of any crime. | Convicted criminal: DNA profile can be deleted if convict turns 75 years of age.<br>Suspects: DNA profile can be deleted after ten years of the verdict if acquitted. |
| **Lithuania** | Convicted lawbreakers, suspects of any crime and any temporarily detained individual. | A DNA profile can be deleted after 100 years of inclusion and ten years after the demise of the convict or suspect. |
| **Luxemburg** | Convicted lawbreakers and suspects of any crime (carried out by the court handling the case). | Convicted criminal: DNA profile can be deleted after ten years of the individual's death.<br>Suspects: DNA profiles can be deleted after acquittal, ten years after death, or after a crime prescription. |
| **Poland** | Convicts or suspects (based on crimes listed) | Convicted criminal: DNA profile can be deleted after 35 years of the incident.<br>Suspects: DNA profile can be deleted after acquittal. |
| **Portugal** | Suspects convicted by court order of preconceived criminal offences punishable by more than three years sentence. | Convicted criminal: DNA profile can be deleted when the criminal record is nullified. |
| **Romania** | Convicts and suspects (based on crimes listed) | Convicted criminal: DNA profiles can be deleted after the individual turns 60 years of age or after five years in the event of death.<br>Suspect: Public Prosecution or court can delete DNA profiles when considered unnecessary. |
| **Scotland** | Convicts and suspects remanded for any crime. | Convicted criminal: DNA profile is retained indefinitely.<br>Suspects: DNA profile can be deleted after acquittal but can be extended in cases involving violent or sexual offences. |
| **Slovakia** | Convicted lawbreakers and suspects of any crime. | Convicted criminal: DNA profiles can be deleted after the individual is 100 years old.<br>Suspect: DNA profiles can be deleted after acquittal. |
| **Spain** | Convicted lawbreakers or suspects detained for severe crimes (based on crimes listed). | Convicted criminal: DNA profile can be deleted on the prescription date of the criminal record (except if a different court order is stated).<br>Suspects detained: DNA profile can be deleted after a crime prescription. |
| **Sweden** | Convicted lawbreakers are condemned to more than two | Convicted criminal: DNA profiles can be deleted after ten years of the served sentence. |

| | | |
|---|---|---|
| | years of non-financial sentences. | Suspect: DNA profiles can be deleted after acquittal. |
| **The Netherlands** | Convicted lawbreakers and suspects. | Convicted criminal: DNA profiles can be deleted after 30 years of the served sentence if the offence is punishable by more than a year's sentence or 20 years after the individual's demise. The profile can also be removed after 20 years of the served sentence if the offence is punishable by less than a year's sentence or 12 years after the individual's demise. Minors convicted of sexual assaults may have their profiles retained for 80 years and maybe extended in case of a new conviction.<br><br>Suspects: DNA profile can be deleted if the individual is acquitted or not charged but can be retained if a match is discovered in the DNA database. |
| **United Kingdom (England, Wales)** | Convicted lawbreakers or suspects detained for any qualifying offences. | The DNA profile is retained indefinitely. |

The Protection of Freedoms Act further led to the deletion and destruction of 7.75 million unnecessarily retained DNA profiles and samples. These included those who were arrested but not charged with an offence and those who were acquitted (Erbaş, 2017). The ruling of the European Convention on Human Rights cited the Scottish DNA Database approach (as shown in **Table 1.5**) as "particularly significant" when compared to the practices in England and Wales, due to the stricter policies regarding the retention of DNA profiles (William and Johnson, 2013). On May 27, 2005, several European Union countries agreed to sign a treaty to establish a network of European DNA databases for exchanging forensic data. The agreement was made during a conference held in Prüm, Germany. This treaty was called the **Prüm Convention** (Kierkegaard, 2008). The Prüm Treaty permitted members of the European Union to exchange biometric data and enable cooperation of database access for vehicle registration, fingerprints, and DNA profiles information within the seven member states, namely Australia, Belgium, France, Germany, Luxembourg, Netherlands, and Spain (Bellanova, 2008). The "Prüm Convention" was initiated by the German Minister at that time, Otto Schilly, in 2003 (Servant and MacKenzie, 2017). The agreement of the treaty was aimed at combating international crimes, fighting terrorism, strengthening security at the border, and curtailing illegal immigration. This also permitted a step of cooperation in exchanging demographic data after scientific verification, which was conducted as a "hit/ or no-hit" query. The treaty also improved European integration to transpose the agreement into the European Framework. However, the main provisions were later incorporated into the European Union framework through "Council Decision 2008/615/JHA and 2008/616/JHA" (Fiodorova, 2014; Muñoz and Fiodorova, 2014). In subsequent years, some European nations joined the treaty

to exchange forensic data. These countries included Hungary, Bulgaria, Slovenia, Romania, Estonia, Finland, and Slovakia. Other European nations, including Sweden, Italy, Portugal, and Greece, have expressed their interest in the Council of the European Union and are seeking to comply with the Prüm Convention's decision (Bellanova, 2009).

**Table 1.5. Summary of the Protection of Freedoms Act 2012 for convicts and non-convicts in the United Kingdom (Amankwaa and McCartney, 2018)**

| **Convictions *** | |
|---|---|
| **Circumstance** | **DNA and Fingerprint Retention** |
| Adult – all crimes | Indefinite |
| Under 18 – Serious offences ** | Indefinite |
| Under 18 – Minor offence | Five years + any custodial sentence for the first conviction; if the sentence is five years or more, retention is indefinite.. |
| Under 18 – Second conviction | Indefinite |
| **Non-convictions** | |
| Arrested and charged with serious offence** | Three years + possible 2-year retention continuation by the court District Judge. |
| Arrested but not charged with serious offence** | None, but the Biometric Commissioner may grant a three-year retention continuation in uncommon cases on application or a possible 2-year retention continuation by the District Judge of the court. |
| Minor offence – Penalty Notice for Disorder (PND) | 2 years |
| Minor offence – arrested or charged | None – but conjecturally searched |
| Suspected of terrorism | The legislation determines the retention + multiple two years with National Security Determination (NSD) by the chief constable. |

* ***Convictions include cautions, reprimands, and final warnings.***
*** Serious offences are crimes such as sexual or violent offences, burglary, and terrorism offences.***

In 2009, Iceland and Norway agreed to share their biometric data with the Commission and comply with specific provisions of the Prüm decision (Pedro and Campos, 2013; Zarza, 2015). Denmark opted out of the Prüm Treaty and has never been part of the agreement as the rules do not apply to the nation (Joannin, 2017). In June 2016, the Council of the European Union initiated an agreement for Liechtenstein and Switzerland to join the Prüm Treaty (Santos, 2016). The proposal agreement between both parties was concluded in January 2019, leading to the endorsement of the treaty for forensic data exchange on June 6, 2019 (Council of the European Union, 2019). Denmark decided not to join the Prüm Treaty because it had a broader opt-out from Justice and Home Affairs policies since 1992. This choice was confirmed by a public vote in the 2015 referendum, where the Danish people rejected a proposal for a flexible opt-in system. On the other hand, Ireland uses a case-by-case opt-in approach and has chosen to participate in the Prüm framework. On December 1, 2014,

the United Kingdom chose to opt out of the treaty (Council of the European Union, 2016). However, the large number of profiles retained within the UK National DNA Database increases the likelihood of obtaining investigative DNA matches (Home Office, 2024a). The UK law enforcement agencies recommended that the treaty agreement may be notably helpful for combating crime in the United Kingdom (Amankwaa, 2020). An experimental analysis was conducted in 2015 in which 2500 DNA profiles from the UK police were searched across Germany, France, Spain, and the Netherlands. Consequently, the investigation produced 181 hits for various crimes, with some profiles reporting multiple hits across countries. Following this discovery, the United Kingdom voted to be part of the Prüm Treaty on December 1, 2015, despite the Brexit scheme (Home Office, 2015). On May 20, 2016, the Commission approved the UK's decision to join the Prüm Treaty (Curtin, 2017). In June 2019, the UK agreed to share DNA and fingerprint data with the European Union; however, they left the European Union after concluding negotiations for the withdrawal but remained part of the Prüm Treaty. The UK has maintained close cooperation with the European Union on police matters through its trade agreement. Dating from September 2019, only 25 European Union States were active in the Prüm Treaty. Greece, Ireland, and Italy faced delays in fully operationalising the Prüm framework due to a combination of legislative inertia, financial constraints, and technical or policy considerations following the 2008 crisis. The levels of connection differ within the operational states. Recent data revealed that Bulgaria shares DNA data with nine countries, while Austria and Netherlands exchange with 23 countries (Toom *et al*., 2019; Machado and Granja, 2020). By 2024, 31 countries have actively implemented the Prüm framework. This includes all 27 European Union (EU) members and 4 non-EU Schengen-associated countries. The 27 EU Member States are Sweden, Spain, Slovenia, Slovakia, Romania, Portugal, Poland, Netherlands, Malta, Luxembourg, Lithuania, Latvia, Italy, Ireland, Hungary, Greece, Germany, France, Finland, Estonia, Denmark, Czech Republic, Cyprus, Bulgaria, Belgium, and Austria. The non-EU Schengen-associated countries are Iceland, Liechtenstein, Norway, and Switzerland. In 2025, the United Kingdom is no longer part of the Prüm Treaty or its framework due to its departure from the EU and its lack of integration into the mechanism for sharing EU data under the Prüm framework (UK Parliament, 2025). However, the United Kingdom relies on international and bilateral agreements to share data across borders and manage criminal investigations across EU member states through the Trade and Cooperation Agreement (TCA) (Bullock *et al*., 2025).

#### 1.6.2.2 POPULATION DNA DATABASES

Population DNA databases are sometimes referred to as allele frequency databases or non-offender DNA databases (Gill, 2002). A population DNA database is different from a national criminal DNA

database. It holds DNA profiles of non-offenders within a community and provides genetic information about the frequencies of alleles and genotypes in the general population (Butler, 2015). The collection of the DNA samples for setting up such a database is voluntary, with strict confidentiality in keeping the donor's identity anonymous and does not involve any form of coercion during sample donation or collection. The sampling population may comprise hundreds to thousands of unrelated individuals, a range that is generally considered sufficient to allow reliable estimation of allele frequencies in forensic population databases (Chakraborty, 1992; Gill, 2002; Butler, 2015). Each individual's DNA is genotyped using PCR-based kits to analyse specific genetic loci, specifically Short Tandem Repeat (STR) markers. Researchers count the number of times each allele appears in a sample for each locus. They then create a database to store the allele frequencies for all the loci. This database is usually specific to the population or subpopulation being studied. These allele frequencies are then used to calculate each locus's Random Match Probability (RMP). This is done by multiplying the genotype frequencies for both homozygous and heterozygous genotypes across all loci to obtain the overall profile frequency, as explained in Chapter 1.5.4 (Balding and Nichols, 1994; Butler, 2014). The RMP shows how rare a DNA profile is in a population, calculated from the genotype frequencies' product (National Research Council, 1997; Zeller and Elkins, 2020).

Population DNA databases show how often a specific DNA profile can be found in the general population. This information helps us understand how likely it is that the same DNA profile could appear in an unrelated person chosen randomly from that population. Instead of focusing on a specific individual, it looks at the chance of a random match happening (Balding, 1999). A DNA profile is unique, and experts in court express this uniqueness using a "random match probability." This probability helps to explain how strong the evidence is. It is usually calculated when there is a complete match between a suspect's DNA and the evidence from a crime scene. This probability shows how likely it is to find someone with the same DNA profile in the population. A lower random match probability means more substantial evidence that the suspect's DNA matches the sample from the crime scene. Conversely, a higher match probability weakens the evidence (Balding and Donnelly, 1995a; b). When a DNA match is found between a crime scene and a suspect, experts calculate the rarity of that DNA in the population using a population DNA database as a reference. This process is standard in the United States, the United Kingdom, and worldwide to show how strong the DNA evidence is (Foreman and Evett, 2001).

Allele frequencies can vary between populations with different biogeographical ancestries due to historical population structure and migration (Prohaska *et al*., 2019). Due to this fact, it is always ideal to collect databases from individuals that form most of the everyday racial or ethnic groups

in a nation (Gill, 2002). Occasionally, distinct sub-populations may arise in a population due to incomplete mixing due to non-random mating. In the United Kingdom population database, the England and Wales forensic DNA database classifies profiles into five ethnic categories: White North European (EA1), White South European (EA2), African/Caribbean (EA3), Asian (EA4), and Chinese/Southeast Asian (EA5). In some cases, data for African and Caribbean backgrounds are shown separately, but they are combined as one group (EA3) when used in investigations. (Foreman and Evett, 2001; Steele, 2016). Minor variations also occur between groups of the same race residing in places of different locations. An ongoing discussion on the utility of population DNA database begs the question of whether the database is the actual representation of the population's genetic diversity given that most forensic allele frequency databases are developed based on a random assemblage of the racial population without taking consideration into the sub-population that exist within the groups (Buckleton *et al*., 2016; Sirugo *et al*., 2019). An example is the Asian population (of Indo-Pakistani descent), with individuals from diverse cultural and geographical origins (Foreman *et al*., 1998). However, studies on the genetic variations between these sub-population groups have revealed that low differences exist (Gill and Evett, 1995; Balding *et al*., 1996; Budowole *et al.,* 1999). Balding and Nichols (1994) explained that inferences obtained from the allele frequencies of a subpopulation can be adjusted for inbreeding/coancestry within a population using a correction factor ($F_{ST}$) (Balding and Nichol, 1994).

The RMP or LR is calculated from an acceptable population DNA database, e.g. STRider (**STR** for **Id**entity **E**NFSI **R**eference database [https://strider.online]). In court, this format is accepted to present the strength of a DNA match already found in a criminal DNA database (Gill, 2002).

The population DNA database is forensically significant because a matched DNA profile of a suspect from the evidence may have occurred by chance, or the perpetrator of the crime emerged from another population not the same as the suspect (Schneider, 2007). Allele and genotype frequencies for a national population database provide information on the genetic variations between races, ethnic groups, or tribes (Gill, 2002). Such genetic information enables statistical calculations to evaluate DNA profiling in a forensic context within a population and sub-population using the Hardy-Weinberg Equilibrium (HWE) (Evett and Weir, 1998). Also, allele and genotype frequencies of sub-groups or ethnic groups within a population can be very distinct from other significant populations (Bär *et al*., 1993). The **Wahlund effect** is a reduction in observed heterozygosity and a corresponding increase in homozygosity that arises when genetically distinct subpopulations are combined and analysed as a single population. In contrast, an increase in observed heterozygosity and a reduction in homozygosity resulting from increased gene flow or

interbreeding between previously isolated subpopulations is referred to as **Isolate breaking** (Holland, 2000).

Individuals within a subpopulation are more likely to share alleles that are identical by descent due to shared ancestry (Weir, 1994). For instance, some communities, practice consanguineous or restricted marriages. Such practice can reduce genetic diversity in the gene pool and change the distribution of alleles (Foreman and Lambert, 2000). Consequently, analysing an allele frequency without considering the coancestry level is erroneous (Balding and Nichols, 1994). Therefore, it is recommended to use an allele frequency database that is particular to the individual's sub-population when calculating the random match probability (Weir, 1992). As explained earlier, the issue of coancestry can be accommodated using the Balding-Nichols Correction Factor, which incorporates an estimated or standard $F_{ST}$ value for a population depending on the level of inbreeding. This increases the match probability and provides more advantage to the suspect. The Balding-Nichols Correction Factor does not always lower the match probability below the 1 in 1 billion thresholds. When calculating match probability using STR kits with many markers, it might not have a noticeable effect. In these cases, the reported value can still be 1 in 1 billion or higher, even though it may significantly affect the match probability based on Hardy-Weinberg principles (Foreman and Evett, 2001). The rationale is to lower the likelihood ratio to benefit the suspect and presume their innocence by biasing the DNA statistics in their favour and maintaining conservation (Balding and Nichols, 1994).

## 1.7 BACKGROUND OF THE NIGERIAN REPUBLIC

### 1.7.1 LAND AND POLITICAL DESCRIPTION

The Federal Republic of Nigeria is a sovereign country in West Africa, geographically situated in the Gulf of Guinea. The nation has plenty of wildlife reserves and natural landmarks and two major rivers, the Niger and the Benue, which form a confluence at Lokoja town. Other vital rivers that affect the climate and vegetation include Anambra, Bonny, Cross, Delta, Imo, Kaduna, Katsina-Ala, Kwa Ibo, Ogun, Osun, Sokoto, and Taraba (Ogbaa, 2003).

Several countries and a body of water border the country. To the north lies the Republic of Niger, with an approximate border length of 1,608 km. To the west is the Republic of Benin, which has an approximate border length of 809 km. On the east, the nation shares its borders with the Republic of Chad, approximately 85 km long, and Cameroon, which has a border length of about 1,975 km. Finally, to the south, the nation has an Atlantic Ocean coastline measuring approximately 853 km.

In total, the land border perimeter is approximately 4,477 km (Bello *et al*., 2012). The country has 36 states and a Federal Capital Territory, Abuja as the capital city (Onimisi, 2014).

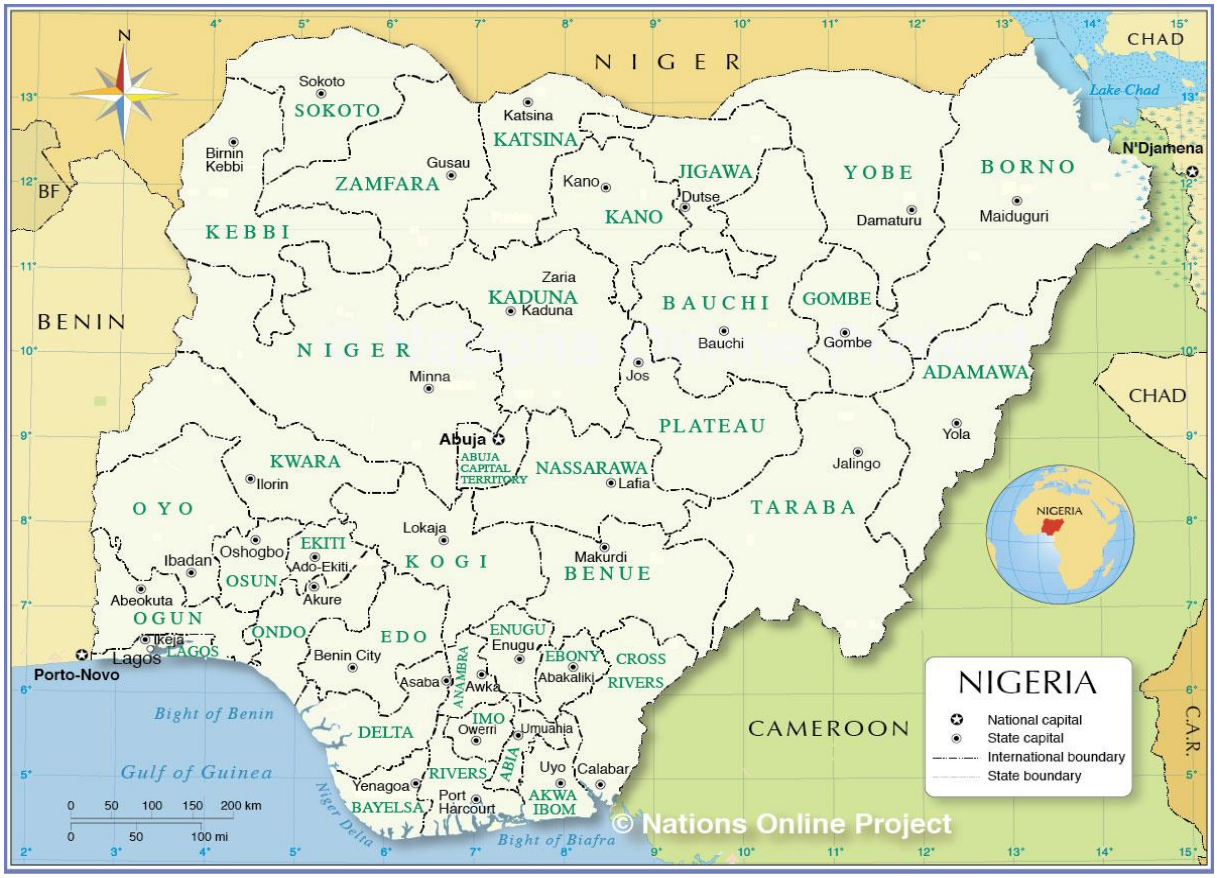



**Figure 1.9. Nigerian map displaying the states and capitals with the coordinates and the land borders (https://www.nationsonline.org/oneworld/map/nigeria-administrative-map.htm)**

The zones and their respective states are as follows:

**The North-Central** – Plateau, Niger, Nassarawa, Kwara, Kogi, and Benue.

**The North-West** – Zamfara, Sokoto, Kebbi, Katsina, Kano, Kaduna, and Jigawa.

**The North-East** – Yobe, Taraba, Gombe, Borno, Bauchi, and Adamawa

**The South-South** – Rivers, Edo, Delta, Cross-River, Bayelsa, and Akwa-Ibom.

**The South-West** – Oyo, Osun, Ondo, Ogun, Lagos, and Ekiti.

**The South-East** – Imo, Enugu, Ebonyi, Anambra, and Abia.

(Bakare, 2015; Adebisi and Oni, 2012).

**Figure 1.9** shows the geographical description of the Nigerian landmass. The Nigerian Republic also recognises six geopolitical zones, namely: the North-Central (also known as the Middle Belt), the North-West, the North-East, the South-South (also known as the Niger Delta), the South-West, and the South-East (Binuomoyo, 2016). **Figure 1.10** describes the locations of geopolitical zones

in the Nigeria map. Each geopolitical zone is occupied by the 36 states that house the various ethnic groups (Igiede, 2013).

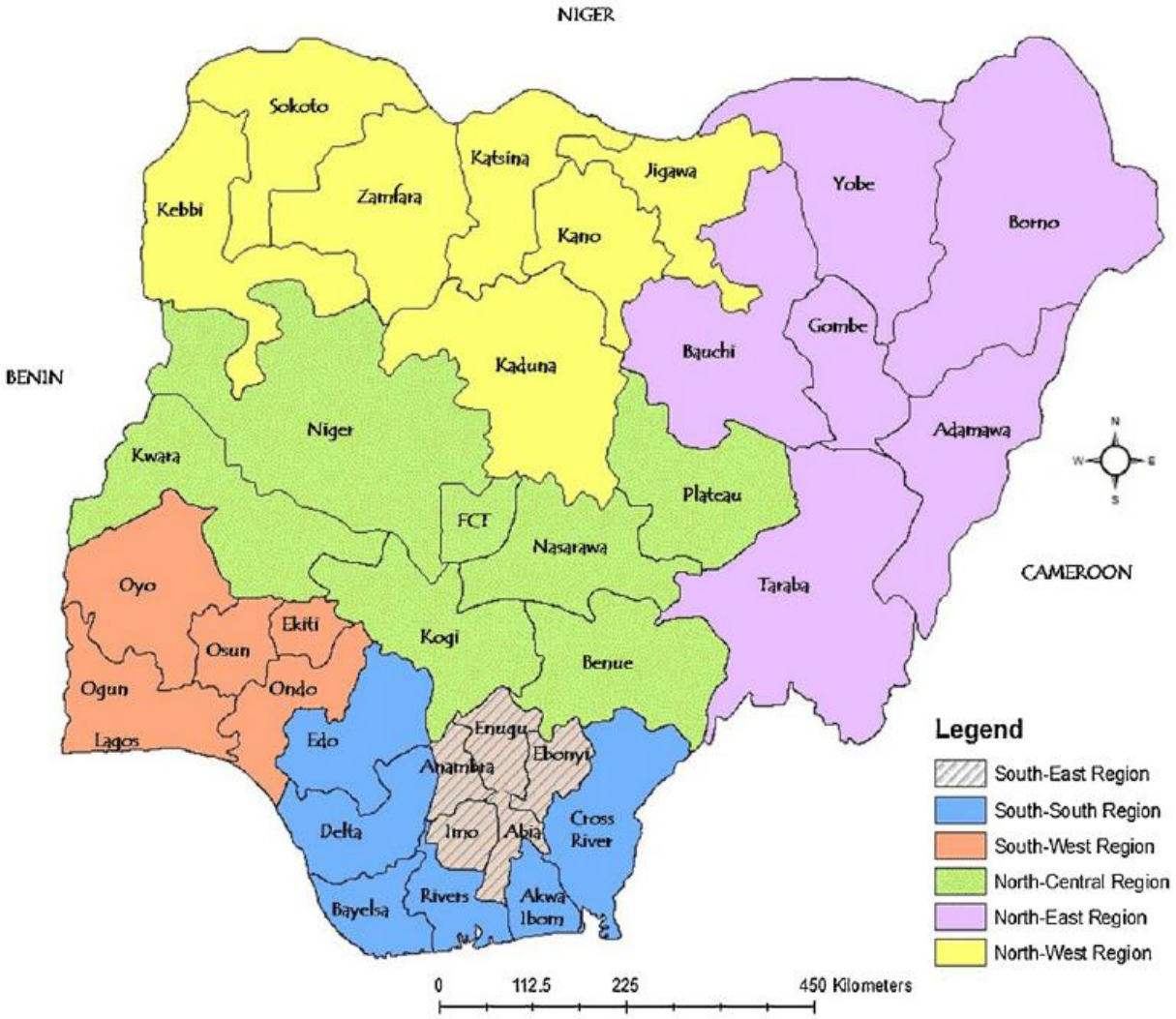



**Figure 1.10. Nigerian map displaying the six geopolitical zones with the 36 states (https://consulter-voyant.com/wear_rm.php)**

Geographically, the country lies north of the equator and has coordinates between 4°N to 14°N latitude and 2°E to 15°E longitude. It covers a total area of 923,768 km² (356,669 square miles). The country has a land area of 910,768 km² and a water area of 13,000 km² (Fadare, 2009; Ajibade *et al*., 2019). The country holds the position of the 32nd largest country in the world (Uthman *et al*., 2011).

### 1.7.2 ORIGIN AND HISTORICAL OVERVIEW

Human occupation in the region now known as Nigeria dates back to the Stone Age, with archaeological evidence indicating continuous settlement for tens of thousands of years. This history continues through the Common Era and includes the period of the Transatlantic Slave Trade, when European invaders were involved (Falola and Heaton, 2008). Evidence from archaeological studies revealed that people have dwelled in every region of the country for

thousands of years. Initially, these human societies were localised in nature, existing in smaller settlements and villages until the emergence of European colonisation (Falola and Aderinto, 2010). Before the British colonisation, these civilisations existed: Sokoto Caliphate (1809-1903), Oyo (1603-1800), Hausa States (1500-1808), Benin (1440-1897), Kwararafa (1400-1800), Kanem-Bornu (1068-1900), and the Nri Kingdom (948-1911) (Abdurrahman, 2012). Many of these societies within today's middle-belt and eastern regions, such as the Kingdom of Nri (Igbo tribes of today), retained their localised settlements but became united and centralised in the 10th century until the emergence of European colonisation in the 20th century. However, some societies in the southern regions, such as the Benin Empire and the Ile-Ife Kingdom, were more centralised during the first millennium, attracting economic, political, and cultural importance to the urban hub for the development of state structures via the purpose of kingship (Bondarenko and Roese, 1999). A civilisation known as the "Nok culture", noted for their terracotta sculptures, appeared, and vanished from 500 BCE to 200 CE in northern Nigeria. The circumstances leading to their extinction were unknown (Rupp *et al*., 2005). From the eleventh to fifteenth century CE, the empires of Kanem and Born in the northern regions of Nigeria began their ascendancy, producing the Hausa states, which occupy today's North-East and North-West geopolitical zones (Falola and Heaton, 2008). Trading, agriculture, craftwork, fishing, and hunting were the main occupations of the indigenous Nigerians living during the pre-colonial period (Fafunwa, 2018). Migrants from other locations also exchanged ideas, goods, and services that contributed to the economies of the centralised and decentralised societies, which impacted their culture, politics, and religion. The trans-Saharan trade from the North African countries to sub-Saharan Africa towards the south expanded due to the spread of Islam from the northern region down to the southern region (Frederiks, 2010). During 1500CE, the trading relationship created societies that significantly developed into complex communities of people with integrated culture, politics, and economy (Falola and Heaton, 2008).

**The Kingdom of Nri**, which is accepted to be the foundation of the **Igbo civilization**, became united and centralized between the 10th to the 20th centuries (948 -1911) but was overpowered by the British in 1911 (Abdurrahman, 2012). The “Eze Nri” ruled the kingdom and was seen as a priest-king bestowed with royal and divine powers. The Igbo tribe is the oldest existing ethnic group in present-day Nigeria and forms one of the major ethnic groups in the federation (Arowolo, 2010).

**The Benin Empire**, sometimes called the **Edo Kingdom**, flourished as a civilisation over its region from the 15th to the 19th centuries (1440-1897). The Edo Kingdom, ruled by the "Oba of Benin",

was so powerful that it extended its territory from the south-south geographical region to the south-eastern parts of the Yorubaland and parts of the present-day Delta State. The supremacy of the kingdom made it the first civilisation to be recognised by foreign traders. The Benin Empire is currently occupied by the Benin ethnic group, which is situated in the present-day Edo State (Bradbury, 2017).

**The Oyo Empire**, which formed a significant part of the **Yoruba ethnic group** of today, rose to its pinnacle between the 17th to the 19th centuries (1603-1800), influencing the South-western region of Nigeria, stretching out to the Fon Kingdom of Dahomey in the Benin Republic, and parts of the Togolese Republic, Ghana, and the Republic of Côte d'Ivoire. The supreme overlord of the medieval Oyo Empire was ruled by the "Alaafin of Oyo" (Ogundiran, 2009).

**The Hausa states** existed from 1500 to 1808, long before the **Fulani Empire** (1808-1903), which is currently the Sokoto Caliphate of today. In 1809, the states of northern Nigeria were re-established by Sheikh Usman dan Fodio after his successful jihad conquest of the Hausa Kingdom of Gobir, Kano, and Katsina. The Sokoto Caliphate became a merger of the modified Hausas and the Islamic Fulani. Over a long period of religious association and intermarriage between the Hausas and Fulani, they became racially indistinguishable. The Hausa-Fulani ethnic groups largely dominate the northern regions of Nigeria and are traditionally ruled by supreme rulers, notably the “Emir of Kano” and the “Sultan of Sokoto” (Hopen, 2018; Smith, 2018).

### 1.7.3 DEMOGRAPHICS AND POPULATION OVERVIEW

Due to Nigeria’s enormous population, rich heritage, and diverse ethnicity, it is often called the “Giant of Africa” (Okonkwo and Smith, 2018). Nigeria was once Africa's largest economy but is now the fourth largest (International Monetary Funds, 2024). Nigeria is Africa's most populous black nation and the seventh most populous country globally, with about 232.6 million people (Babatunde, 2019; Oweibia *et al*., 2024). Nigeria's population is expected to reach 239 million in 2025 and 440 million in 2050, notwithstanding the decreasing fertility (Etebong, 2018). According to the 2006 population census report, the population grew by 3.02%. It is expected to double in 22 years. This results from the high birth rate, the primary factor leading to unprecedented population growth. Some established customs, religions, and superstitions promote the high birth rate in the country. These cultural and religious components encourage early marriage, polygamous family systems and sometimes the prevention of the use of birth control measures such as the use of contraceptives and family planning, which consequently encourage a large population (Akinyemi *et al*., 2015). Another factor leading to the high population growth in Nigeria is the

illiteracy rate (Etebong, 2018; Onyeiwu, 2024). Reports of 2023-2024 Demographic and Household Surveys (DHS) which was implemented by the National Population Commission (NPC) with technical support from the Inner City Fund (ICF) International, Inc. and funded by the World Health Organisation (WHO), United Nations Population Fund (UNFPA), Global Fund, Bill and Melinda Gates Foundation (BMGF), and United States Agency for International Development (USAID) revealed that the nation has population illiteracy of 71.3% for males and 52.7% for females. Poor education policy will inadvertently lead to early marriage, low-income family planning and poverty (National Population Commission - NPC and ICF, 2024). On the other hand, comparing the level of education between urban and rural residents, the urban dwellers are well educated, with about 69% of them possessing secondary or higher-level education. In contrast, in rural settlements, only about 42% have formal education. The country is 54.28% urbanised, with a 5.63% urbanisation rate projected from 2015 to 2023. However, population distribution showed that the South-West, South-East, and South-South, geopolitical zones had more population density than the northern regions due to the high rate of urban development.

The Nigerian population constitutes principally young people between the ages of 0-14, which make up about 42% of the population, with other age groups ranging from 15-64 years (55%) and 65 years – above (3%), making up the rest of the population. The median age that divided half of the population into younger and older people, as estimated in 2024, is 17.9 years, with an estimated total sex ratio of 1.06 males to one female at birth. **Figure 1.11** describes the sex ratio with the age structure from 0-100. In 2024, the crude birth rate is 35.68 births per 1,000 people, the death rate is 12.43 deaths per 1,000 people, and the net migration rate shows a decrease of 0.27 migrants per 1,000 people. However, the infant mortality rate in 2024 is high at an estimated total of 53.7 deaths per 1000 persons, with an average life expectancy of 62.2 years at birth. The nation also recorded an estimated total fertility rate (FTR) of 5.01 children born per woman, indicating a high FTR and a low contraceptive prevalence rate of 27% in 2024. The total dependency ratio (individuals relying on government, parents, and guardians for daily needs) as estimated in 2024 was 78.8 dependants for every 100 non-dependants, showing a high dependency rate (National Population Commission - NPC and ICF, 2024).

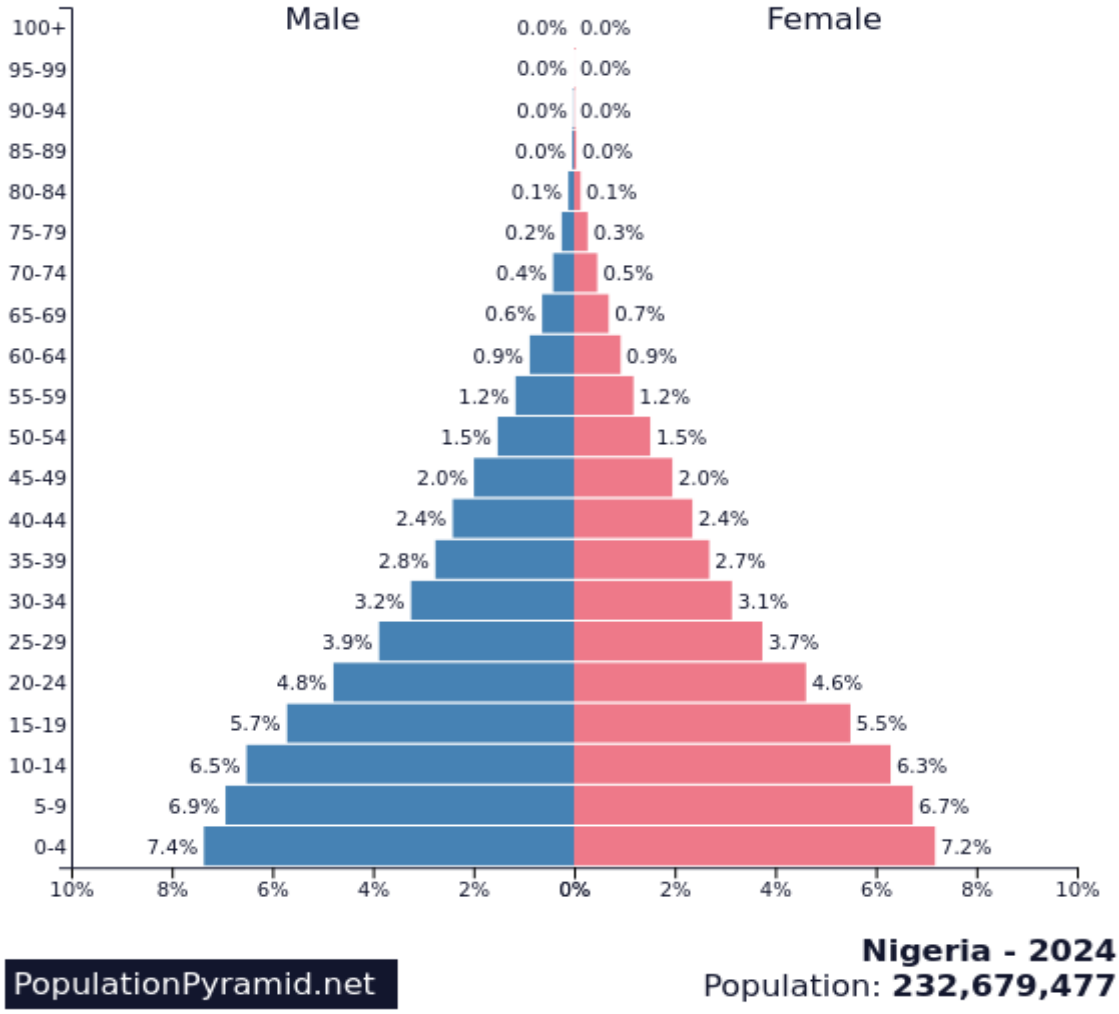



**Figure 1.11. A population pyramid summarising the age structure and sex ratio in Nigeria for the year 2024. (https://www.populationpyramid.net/nigeria/2024/)**

The Nigerian population is diverse, with more than 300 ethnic groups and over than 500 languages spoken, each reflecting distinct customs and traditions. Out of 300 ethnic groups, three main groups make up a large part of the population: the Igbo at 18%, the Yoruba at 21%, and the Hausa-Fulani at 29% (Agbaire and Dunne, 2024). The remaining ethnic groups make up a minority. They include the Bini (1.05%), Tiv (2.5%), Efik-Ibibio (3.5%), Kanuri (4.0%), and Ijaw (10%). Many other groups exist, totalling around 300 different ethnicities. The ethnic groups mentioned are the most populous and prominent among the minority groups in Nigeria. **Figure 1.12** gives an ensemble of the most well-known ethnic groups in Nigeria based on their language and geographical location. The Hausa-Fulanis are the most prominent ethnic group in the North, the Yorubas are numerous in the West, and the Igbos are prominent in the East. However, based on the geopolitical zones, the North-West is mainly occupied by the Hausa-Fulanis, with some presence in the Northeast and North-Central. The Yorubas dominate the South-West, and the Igbos are primarily in the South-East.

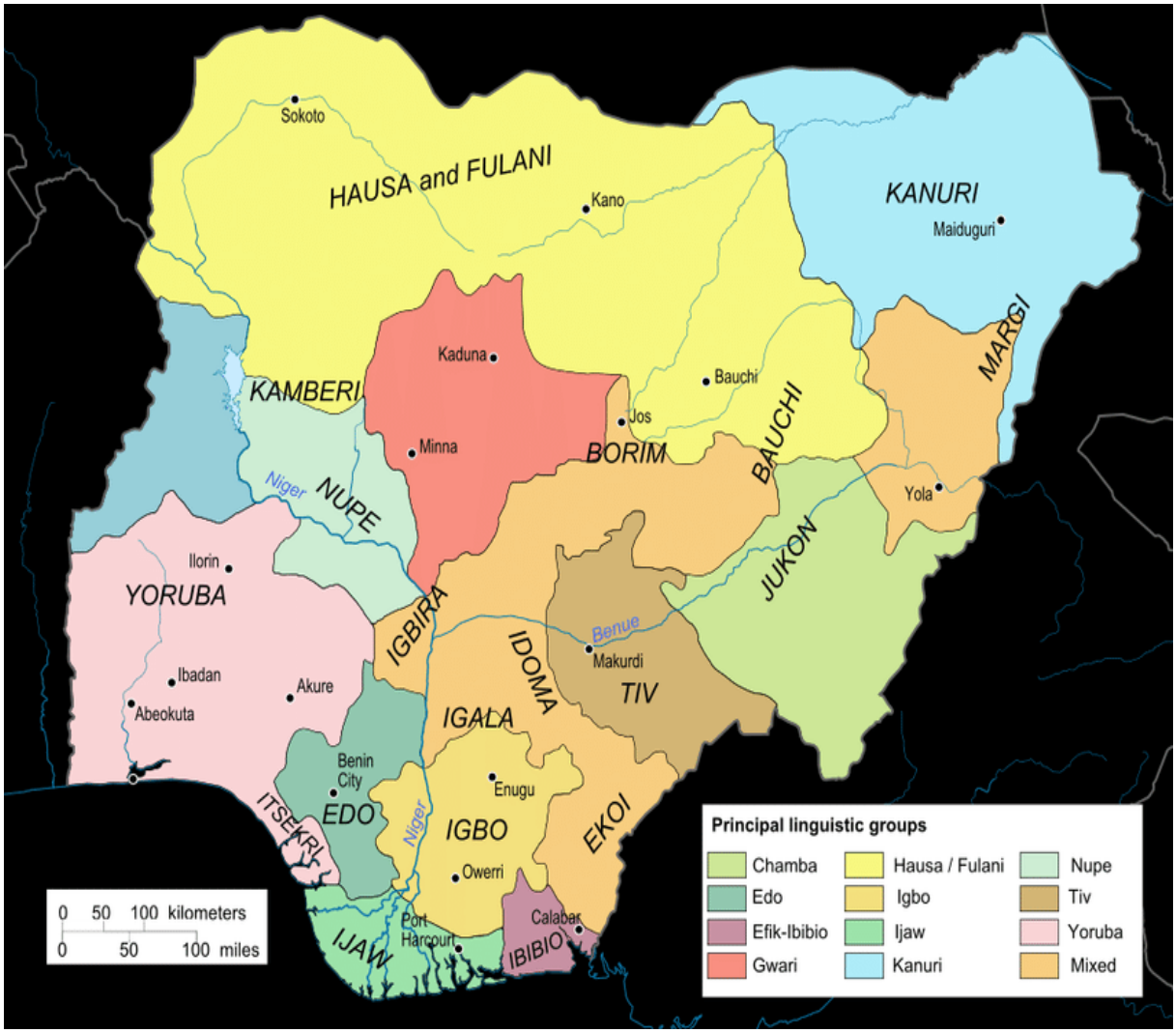



**Figure 1.12. Nigerian map displaying the regions dominated by the various ethnic groups (https://www.legit.ng/1142953-major-tribes-nigeria-states.html)**

**The Hausa-Fulani tribe** is the largest major ethnic group in Nigeria and Africa. The Hausa language is the second most spoken language in Africa, following Arabic. The Hausa tribe remained a distinct ethnic group until their conquest by the Fulani tribe during the Jihadist war led by Usman Dan Fodio, leading to them mixing to form an inseparable ethnic group over a long period (Smith, 2018). The Hausa-speaking people are also found in other parts of West Africa, which include the following countries: Senegal, Ghana, Gambia, Gabon, Equatorial Guinea, Eritrea, the Democratic Republic of Congo, Côte d'Ivoire, Chad, Central African Republic, Cameroon, and Benin Republic. The states predominately occupied by the Hausa-Fulanis include these states but are not limited to the following: Zamfara, Sokoto, Kebbi, Katsina, Kano, Kaduna, and Jigawa. Satellite settlements of the Hausa-Fulani can also be found in Adamawa, Bauchi, Borno, Niger, Plateau, and Taraba states (Batibo, 2005). A summary of the most predominant states occupied by the Hausa-Fulanis is shown in **Figure 1.13.**


**Figure 1.13. Nigerian map displaying the predominant Hausa-Fulani states (https://www.legit.ng/1176164-major-hausa-states-nigeria.html)**

**The Yoruba tribe** is Nigeria's second-largest ethnic group. Many Yoruba-speaking people live in other parts of West Africa, including Benin and Togo. The states predominantly occupied by the Yorubas include these states but not limited to the following: Oyo, Osun, Ondo, Ogun, Lagos, and Ekiti. **Figure 1.14** shows the various states predominantly occupied by the Yoruba tribe. Satellite settlements of the Yorubas can also be found in Kogi, Kwara, and Plateau states (Abimbola, 2006).

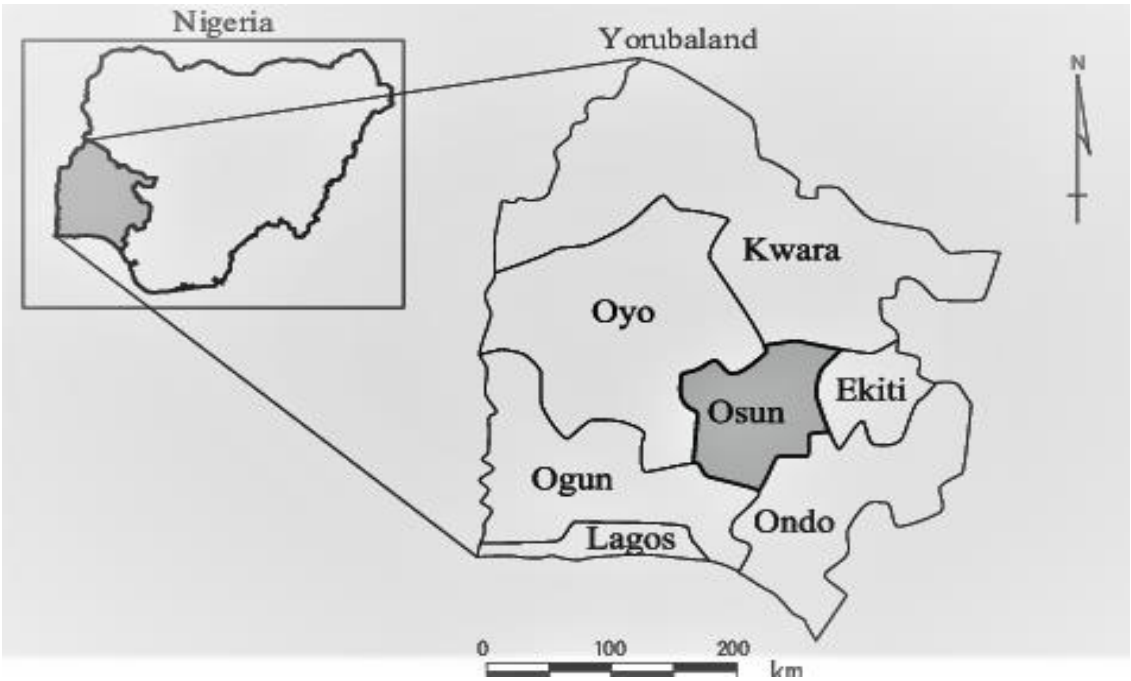



**Figure 1.14. Nigerian map showing the predominant Yoruba states (https://www.naijahomebased.com/list-of-the-yoruba-states-in-nigeria-their-sizes/)**

**The Igbo tribe** is Nigeria's third-largest ethnic group. The Igbos mainly live in the following states Imo, Enugu, Ebonyi, Anambra, and Abia. **Figure 1.15** shows the homeland of the Igbo tribe, describing the states and locations on the Nigerian map. Satellite settlements of the Igbos can also be found in Delta, Rivers, and Benue (Ogbaa, 2003).

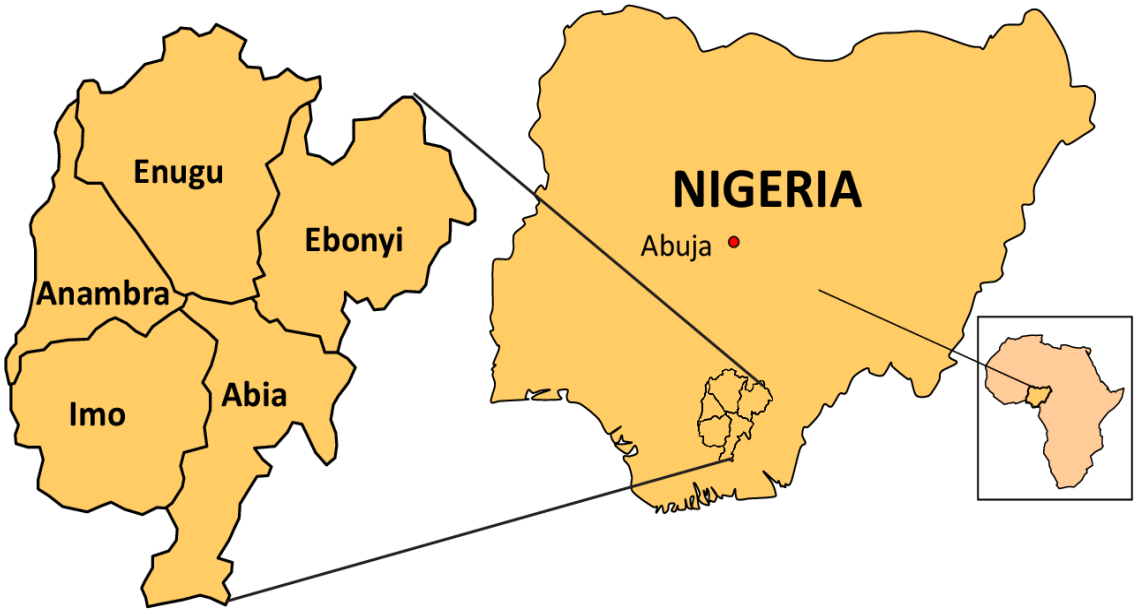



**Figure 1.15. Nigerian map showing the predominant Igbo states (https://www.semanticscholar.org/paper/)**

**The Ijaw tribe** is Nigeria's fourth-largest ethnic group and the predominant ethnic group in the country's South-South region. The Ijaw-speaking people have also been discovered as migrant fishermen in riverine areas of Gabon and Sierra Leone. They are very predominant in Bayelsa State but are also found scattered along the coastlines and intermixed with other minor ethnic groups in Akwa-Ibom, Delta, Edo, Rivers, Ogun, and Ondo states (Ukeje and Adebanwi, 2008).

**The Kanuri tribe** is a minor ethnic group occupying the northeast of Nigeria and is currently the fifth largest in the nation. The Kanuri people speak the Kanuri language, which is a dialect chain expressed by northern Cameroon, western Chad, and the southeast Niger Republic. The Kanuri people predominantly populate Borno State and are spread across Adamawa, Taraba, and Yobe (Umar *et al*., 2011).

**Efik-Ibibio** is a common language spoken among the Annang, Ibibio, and Efik tribes, which share similar culture and heritage. The Efik and Ibibio people live together peacefully along the coast of Nigeria's South-South region. The Efik-Ibibio-speaking people are the sixth-largest ethnic group in Nigeria, although other minor separate tribes claim to exist within this region or share a common language. Some of them have been discovered as migrant fishermen in riverine areas of Gabon and

Sierra Leone. They predominantly occupy two distinct states in Nigeria, viz. Akwa-Ibom and Cross River (Omaka, 2014).

**The Tiv tribe** is among the minor ethnolinguistic groups populating the North-Central geopolitical region of Nigeria. The Tiv or Tivoid languages, spoken by the Tiv ethnic group, are also spoken in some parts of Cameroon. However, they are classified as a sub-family of the Southern Bantoid language. The Tiv indigenes are farmers and depend on agriculture for commerce and livelihood. They are intermixed with other minor ethnic groups in Benue and Taraba, with a few satellite settlements in Plateau and Nasarawa (Dagba *et al.,* 2013).

**The Bini or Edo tribe** is a significant ethnic group of the South-South geopolitical zone due to its unique cultural importance in Nigeria. The Bini-speaking people are native to Edo State in Nigeria, although a few settlements also speak the language in Bayelsa, Ekiti, Rivers and some neighbouring West African countries, which include Togo, Ghana, and Benin Republic (Bradbury, 2017).

Nigeria is also internationally recognised as a religious country with three main recognised religious groups being Christianity, Islam, and Traditional. The Christian community makes up about 45% of the entire religious group, while the Islamic community makes up about 45%. The Traditional religion accounts for the remaining 10%. Islam mostly dominates the northern regions, with a few indigenous peoples practising Christianity, while the southern regions are primarily Christian, with a handful practising Islam. Very few individuals still practice indigenous beliefs such as the traditional religion or atheism (Kitause and Achunike, 2013). Among the three major Nigerian ethnic groups, the Hausa-Fulanis are ethnoreligious and identify their ethnic group with their religious affiliation. They are predominantly Muslims. The Yorubas define themselves based on ethnicity but do not affiliate with a particular religion. They divide themselves among the Christian and Muslim communities. The Igbos are ethnoreligious and affiliate their ethnicity with the Christian religion (Osaghae and Suberu, 2005).

## 1.8 DNA DATABASES IN NIGERIA

### 1.8.1 NATIONAL DNA DATABASES

Nigerian legislators are considering the establishment of forensic DNA databases, which would include a national and a population DNA database. These databases would help better use DNA evidence in investigations involving convicts and non-convicts in the country (Nte, 2012). The Federation of Nigeria has not established a fully operational national criminal DNA database for forensic purposes, although legislative discussions and policy proposals have been ongoing.

Legislative efforts to tackle sexual offences have led to the passing of important bills and the creation of the Violence Against Persons (Prohibition) Act. This law sets clear guidelines for prosecuting sexual offences and helps establish a system for registering sex offenders. (Osinuga, 2015). Some authors argue that the lack of a national forensic DNA database limits police investigations, delays crime-solving, and makes it harder to achieve justice in Nigeria's criminal justice system. (Etim-Osa and Etim-Osa, 2019). The World Population Review ranked the country 14th in 2025 for having the highest annual crime rate. It received a crime index score of 66.20 and a safety index of 33.80 (World Population Review, 2025). Additionally, in 2024, the Global Peace Index (GPI) ranked Nigeria 147th out of 163 independent nations for peacefulness (Institute for Economics and Peace, 2024). The Nigerian Federal Government has not been active in setting up a working national DNA database. This delay in progress is due to a lack of political commitment, skilled workers, and funding (Etim-Osa and Etim-Osa, 2019; Aborisade *et al*., 2024).

Legal and institutional frameworks make it harder to create a national forensic DNA database. When standards for collecting, packaging, documenting, and handling forensic evidence are inadequate or inconsistently applied, it can reduce the usefulness of biological material. This, in turn, can complicate its effective use in criminal investigations and prosecutions. These limitations can make it harder for law enforcement and the courts to use forensic evidence effectively. This may cause investigative delays and lower the reliability of the evidence. Without a national forensic DNA database, these problems could make forensic science less effective in the criminal justice system. Nevertheless, the Nigerian court is the last hope of regular people and still admits authentic and reliable evidence. However, another major challenge lies with the resourcefulness of the Nigerian Police, the country's chief prosecutor of crime (Iorliam, 2018). The Nigerian Police still lack adequate skills, capability, competence, and sophisticated training to conduct a proper forensic investigation (Nte, 2012). Moreover, the Nigerian Police Force was ranked as the worst and most corrupt Police in the world globally in 2016, according to the reports of the World Internal Security and Police Index (WISPI) and International Police Science Association (IPSA) (Esoimeme, 2019). The lack of forensic evidence in many criminal cases has led to over-reliance on eyewitness testimonies, which are sometimes biased and unreliable when the lawyers and prosecutors may have been compromised with bribery and corruption (Akhiwu and Obaseki, 2014). Furthermore, the Nigerian Police Academy in Kano State, which serves as the nation's central degree-awarding institution for forensic training, is handled by basic science lecturers with little or no basic foundation and training in forensic practice. This limitation has affected the creation a national forensic DNA database for forensic practice in Nigeria (Otu and Elechi, 2018; Etim-Osa and Etim-Osa, 2019; Aborisade *et al*., 2024).

Along with the adverse dearth of the offenders' DNA database, the nation currently lacks a Disaster Victim Identification (DVI) system. Most of the disasters experienced in the country are not natural but artificial. The government has suffered from prison breaks, fire outbreaks, genocides, gas inhalation, plane crashes, building collapses, derailments, explosions, and road accidents. A lot of challenges were faced when identifying the victims because there was no DNA database for the identification, resulting in most of them being given a mass burial (Obafunwa *et al*., 2015). A handful of non-governmental organisations (NGOs) have advocated for establishing a national forensic DNA database to facilitate standard practice in Nigeria. These NGOs include First Digital and Tecno Law Forensic Company, Abuja; Sentinel Forensics, Abuja; Forensic Research and Development Centre (FORDEC), Edo State; and Centre for Forensic Criminology and Legal Research, Lagos State (Etim-Osa and Etim-Osa, 2019; Aborisade *et al*., 2024).

### 1.8.2 ALLELE FREQUENCY DATABASES

Several research studies have explored the short tandem repeat allele frequency for the autosomal and sex chromosomes (Y-chromosome) in Nigeria. The data generated so far has established a starting and reference point for analysing mutation rate and allele frequency databases in forensic casework involving the Nigerian population. Most of the data provided is centred on the population study of three major Nigerian ethnic groups: Igbo, Yoruba, and Hausa. However, limited data has been recorded for other ethnic minorities in the country.

Hohoff *et al*. (2009) were the first to analyse samples collected from 337 Nigerian immigrants in the United States. They used 16 STR loci in the Power ES Kit from Promega and the AmpF*l*STR Identifiler, Profiler and SEFiler PCR amplification kits from Applied Biosystems. They also discovered that the Nigerian population's genetic makeup did not differ from expected under the Hardy-Weinberg Equilibrium (HWE). Hohoff *et al.* (2009) also pointed out that the FGA, D18S51, and ACTBP2 (SE33) loci were the most informative STR markers with significant power of discrimination (PD). Based on the calculated statistical parameters, the authors recommended the 16 STR loci for use in forensic casework and parentage testing within the Nigerian population.

Similar work was done by Agbo *et al*. (2017) using a lower number of STR loci. The research involved investigating 315 unrelated individuals within the populations of the foremost Nigerian ethnic-linguistic groups involving twenty different communities across six distinct states: Igbo (7 communities), Yoruba (7 communities) and (6 communities) using the 15-loci in the AmpF*l*STR ™ Identifiler™ Direct PCR Amplification Kit from the Applied Biosystem Inc. containing core CODIS loci. The results revealed that seven loci conformed with the Hardy-Weinberg Equilibrium

expectation. Furthermore, the pairwise comparison results from the population variance showed that the Yoruba and Igbo ethnic-linguistic groups are much closer but very distinct from the Hausa ethnic group.

Okolie *et al.* (2018) conducted research by sampling and analysing 364 unrelated people from the three main ethnic groups in the country, of which 134 were Yorubas, 128 were Igbos, and 102 were Hausas, using the primers provided by the 21 autosomal STR loci of the Qiagen™ Investigator 24plexGO! Kit for PCR amplification. The 21 STR loci containing SE33 and the 20 USA CODIS loci revealed that these loci were very varied in the three ethnic groups and provided fundamental information for studying gene diversity, human identification, and forensic testing in Nigeria. The researchers noted that the SE33 locus significantly differed from the expectation of the Hardy-Weinberg Equilibrium and attributed the result to sampling error. The Nigerian population showed no significant difference from other neighbouring African populations in Kenya, Namibia, Democratic Republic of Congo (DRC), Central African Republic (CAR), and Senegal using the observed heterozygosity statistical parameter. Also, three novel variants were discovered in the Hausa population, viz. allele 15.1 in marker D12S391 and alleles 16.1 and 29.1 in marker SE33 (Okolie *et al*., 2018).

Currently, a few researchers have studied the genetic diversity of Y-chromosome STR loci among Indigenous individuals of specific ethnic groups in the Nigerian population to some extent. The demographics and population structure of Nigeria have been analysed by Cole-Showers (2014) using Y-chromosome markers (11 loci STR). The study collected 463 samples from unrelated male grandfathers across five ethnolinguistic groups: Ijaw, Bini, Yoruba, Igbo, and Hausa. The study compared the genetic diversity of the Nigerian population with that of over 2,000 people from the Middle East and Africa. The findings indicated that the Nigerian population shares a closer genetic relationship with other populations in sub-Saharan Africa than those in the Middle East and North Africa. This evidence challenges the hypothesis that most indigenous people from northern Nigeria have origins in Middle East or the North Africa.

Likewise, Chiedozie and Isaac (2015) conducted a study to explore the genetic differences between two men from the Igbo tribe in the southeastern region of Nigeria, specifically in Owerri's senatorial province. They utilised the 16 Y-STR loci from the AmpFlSTR® Y-filer kit, along with the Amelogenin loci from the multiplex AmpFlSTR® Identifiler kit and analysed the Hyper-Variable Segment 1 of the D-loop mitochondrial DNA (mtDNA HVS-1). The findings indicated a

significant ability to distinguish between the two Igbo males. Furthermore, the Y-STR data suggested they likely share a common paternal ancestor from approximately 90 generations ago.

Similarly, Martinez *et al*. (2017) conducted a forensic study on 27 Y-STR markers found in the Yfiler® Plus kit, using 142 samples from unrelated individuals representing the three prominent Nigerian ethnic groups: Igbo, Yoruba, and Hausa. The study found that the haplotype diversity from the 17 Y-STRs analysed with the Yfiler kit was slightly lower than that from the 27 Y-STRs. They also observed a significant difference when comparing these results with those from five East African populations, especially in Djibouti, Ethiopia, Eritrea, Kenya, and Somalia, regarding the Yfiler® Plus markers.

Fakorede *et al*. (2024) evaluated 461 saliva samples collected from unrelated males, which included 96 Igbos, 139 Hausas, and 226 Yorubas. They utilised 10 Y-STR markers in the UniQTyper™ Y-10 system, comprising DYS626, DYS710, DYS612, DYS644, DYS518, DYS504, DYS481, DYS449, DYS447, and DYS385ab. The analysis revealed that DYS626 exhibited the highest allelic richness while DYS504 displayed the lowest. The study also reported the presence of duplication and microvariant alleles. Additionally, the results identified 403 unique singletons, 430 haplotypes, and 27 shared alleles. Fakorede *et al.* (2019) also analysed blood samples collected from 110 unrelated Yoruba males residing in Lagos in the southwestern region of Nigeria using five Y-STR markers: DYS576, DYS490, DYS393, DYS392. The experiment showed that the markers can successfully tell apart males of Yoruba descent. Also, these markers can be used in paternity testing and forensic investigations.

A few other researchers have conducted forensic DNA investigations using markers other than short tandem repeats (STRs), but there is limited population data available in Nigeria for insertion/deletion polymorphisms (indels) and mitochondrial DNA (mtDNA). Furthermore, next-generation sequencing has not been extensively researched in the Nigerian population. A study on INDEL polymorphisms in Nigeria was conducted by Du *et al.* (2017) using 30 INDEL loci from the Qiagen™ Investigator DIPplex kit. The findings indicated that the population adhered to the Hardy-Weinberg Equilibrium for nearly all markers except for the HDL97 locus. However, the INDEL dataset is not ideal for paternity testing in Nigeria, especially when excluding individuals. However, it can still be used as an additional tool for paternity testing. The findings on combined discrimination power and average observed heterozygosity indicate that indels may work well for forensic identification in Nigeria.

## 1.9 AIMS AND OBJECTIVES OF THE STUDY

The universal aim of this project is to evaluate the effectiveness of various forensic genetic markers within the Nigerian population. This will involve the analysis of two autosomal Short Tandem Repeat (STR) kits: the GlobalFiler™ Express PCR Amplification Kit and the QIAGEN™ Investigator ESSplex SE QS Kit. The QIAGEN™ Investigator ESSplex SE QS Kit was selected because it is a well-validated forensic STR multiplex system that provides sensitive, reliable amplification of autosomal STR loci, ideal for forensic and population genetics applications. The GlobalFiler™ Express PCR Amplification Kit was selected for its inclusion of all CODIS core loci, and compatibility with international forensic DNA databases, allowing for global dataset comparisons and supporting forensic casework and genetic studies. Additionally, Y-STR data will be generated using the Promega PowerPlex® Y23 System Kit to study the population genetics of unrelated Nigerian individuals from three distinct ethnic groups: Igbo, Yoruba, and Hausa-Fulani. The results from analysing the autosomal STR kits will be used to recommend the most appropriate STR kit for forensic DNA analysis in Nigeria. Furthermore, this research will assist in selecting the best kit for creating an offender DNA database and provide essential allele frequency data for the diverse population groups involved.

The comprehensive aim and objectives for each chapter in the project are outlined below:

**Chapter 3**

This research aims to analyse the population genetics of 303 unrelated Nigerian individuals from three ethnic groups: Yoruba, and Hausa-Fulani, and Igbo using 16 autosomal STR markers from the 17-locus Qiagen™ Investigator ESSplex SE QS Kit. Notably, this kit's evaluation has not been reported previously for the Nigerian population.

The specific objectives are as follows:

1. Perform STR genotyping for this kit and evaluate any STR off ladders allele calls.
2. Generate STR allele frequency data to create an allele frequency database for each subpopulation and the entire population.
3. Evaluate forensic parameters for each subpopulation and the entire population, assess the effectiveness of the kit used, and determine its suitability for forensic work in Nigeria.
4. Compare the genetic structure within populations and examine the relationships between the different subpopulations.

**Chapter 4**

This study aims to evaluate the population genetics of 303 unrelated Nigerian individuals from three ethnic groups: Yoruba, and Hausa-Fulani, and Igbo. The evaluation will utilise 21 autosomal STR markers from the 24-locus GlobalFiler™ Express PCR Amplification Kit (Life Technologies), which will be reported for the first time for the Nigerian population. The second aim is to access the genetic information and examine how it aligns with the data presented in Chapter 3, contributing to the existing genetic information for the Nigerian population.

The research objectives are as follows:

1. To generate allele frequency data that will contribute to developing an allele frequency database for each sub-population group and the overall population.
2. To determine forensic parameters for each subpopulation and the overall population and assess the kit's effectiveness and forensic utility.
3. To compare the inter-subpopulation genetic structure and examine the relationships between the sub-populations.
4. To compare the genetic data generated by the GlobalFiler™ Express PCR Amplification Kit with that from the Qiagen™ Investigator ESSplex SE QS Kit, and to assess the concordance of results for loci common to both kits
5. Recommend the most suitable STR kit for establishing a forensic DNA database in Nigeria.

**Chapter 5**

This research aims to evaluate the population genetics of 167 unrelated male individuals from three Nigerian ethnic groups: Yoruba, and Hausa-Fulani, and Igbo. The study utilized 23 Y-STR markers from the Promega PowerPlex® Y23 System Kit. It is noteworthy that the evaluation of this kit has not been previously reported for the Nigerian population.

The research objectives are as follows:

1. Generate Y-STR profiles and allele frequency data to create a database for each ethnic group and the overall population.
2. Ascertain forensic parameters for each ethnic group and the overall population while evaluating the kit's effectiveness.
3. Evaluate the genetic structure within each population and examine the relationships between the different ethnic groups.

**Chapter 6**

This investigation aims to analyse genetic relationships between the Nigerian population (Igbo, Yoruba, and Hausa-Fulani sub-populations) and diverse populations worldwide, utilising autosomal STR and Y-STR data. To our knowledge, no prior study has comprehensively compared the Nigerian population—comprising the Igbo, Yoruba, and Hausa-Fulani ethnic groups—with other global populations: South America, North America, Europe, Asia, and Africa. This study uniquely employs autosomal STR and Y-STR data within a unified framework that includes *STRUCTURE* analysis, $F_{ST,}$ $R_{ST}$, Nei's genetic distance, Multi-Dimensional Scaling (MDS), Analysis of Molecular Variance (AMOVA), and Neighbour-Joining (NJ) methods.

The research objectives are as follows:

1. To examine how the Nigerian population differs from other populations worldwide using autosomal STR data with *STRUCTURE* analysis.
2. To calculate the genetic divergence and differentiation of the autosomal STR using Nei's genetic distance and $F_{ST}$ based genetic distance method, respectively.
3. To assess the genetic differentiation of Y-STR data among populations using Slatkin's $R_{ST}$ based genetic distance method.
4. To conduct a comparison of the Nigerian population with other populations using MDS analysis based on autosomal STR and Y-STR data.
5. To analyse and visualise the relationships between the Nigerian population and other global populations using NJ analysis based on autosomal STR and Y-STR data.
6. To quantify and partition genetic variation within and among populations in the autosomal STR and Y-STR datasets using AMOVA, and to assess their genetic structure and differentiation.

# CHAPTER TWO

# MATERIALS AND METHODS

## 2.1 OVERVIEW

Processes involved in the methodology of this project are detailed in this chapter. The method consists of sample collection, experimental procedures, and statistical evaluation. Before lab work, personal protective equipment (nitrile gloves, safety glasses, face masks, Howie style lab coats) were worn to minimise the risk of contamination. Distel disinfectant was used to clean benches and surfaces before and after each session to decontaminate the laboratory area of human DNA. Equipment, like adjustable micropipettes, was cleaned with 75% alcohol to avoid cross-contamination from exogenous DNA.

## 2.2 ETHICAL APPROVAL, SAMPLE COLLECTION AND PRESERVATION

The Departmental Ethics Committee (DEC) of the Department of Pure and Applied Chemistry at the University of Strathclyde in Glasgow granted ethical approval for the study, with Approval Reference Number DEC19/PAC01. The blood samples of 303 unrelated individuals (167 males and 136 females) from the Nigerian population were collected in Nigeria and the United Kingdom. An initial collection involving 150 unrelated individuals was conducted in various organisations (details below) across the three geopolitical zones: North-Central, South-East, and South-West in Nigeria to capture individuals of Hausa-Fulani, Yoruba, and Igbo, respectively.

Before sample collection, permission was obtained from the organisation's management via a sample collection request letter (**Appendix 3**). Volunteers were provided with participant information sheets to understand the importance and nature of the experiment and consent forms to sign (**Appendices 1** and **2**) for the blood sample collection. The information of the sample donors was anonymised by labelling each sample with a numeric code during the experiment. In the South-East, volunteers were recruited from the Federal University of Technology in Owerri, Imo State and from the Federal Medical Centre in Umuahia, Abia State. In the South-West, volunteers were recruited from students and patients from Yaba College of Technology, Yaba, Lagos State and Health Centre Laboratory, Osogbo, and Osun State, respectively. In the North Central, donors came from members and students of the Central Mosque, Abuja, Federal Capital Territory and Nasarawa State University, Keffi, Nasarawa State, respectively. Participants were also given verbal explanations of the procedures for collecting the blood samples.

In the United Kingdom, 153 blood samples were collected from Nigerian students studying at the University of Strathclyde, University of Glasgow, Glasgow Caledonian University, and University of Leicester. Volunteers were issued participant information sheets to read and understand the importance and nature of the experiment. Consent forms were also issued afterwards for the donors to sign and return. Participants were given verbal explanations of the procedures for collecting the blood samples.

During blood sample collections, donors' fingertips were cleaned and sterilised with an alcohol wipe before and after a puncture to reduce infection hazards. Sterile plasters were also administered to dress the punctured area and to stop bleeding. Sterile lancets (Sinocare Sinodraw® Soft Twist Lancets, SteriLance Medical Inc, Suzhou, China) were used to prick fingers, and four to six drops of blood were placed in the printed circle area of an FTA® card (QIAcard FTA Classic, QIAGEN GmbH, Germany). FTA® cards were left at room temperature for two to three hours to dry.

In total, 303 blood samples were collected from unrelated individuals within the three major ethnic populations in Nigeria: 102 Igbos, 101 Yorubas, and 100 Hausa-Fulanis. Blood samples preserved on FTA cards are considered safe to transport as the chemicals impregnated into the card inactivate any pathogens on the card, stabilising the DNA and protecting it from degradation (Krambrich *et al*., 2022). The cards were therefore transferred to the Centre for Forensic Science, University of Strathclyde, Glasgow, for DNA profiling.

## 2.3 DNA PROFILING USING SHORT TANDEM REPEAT MARKER

Blood preserved on the FTA® Cards was prepared using a Harris Uni-Core Micro-Punch® (1.2mm) on a sterile cutting mat (GE Healthcare Life Science, UK). The punch cutting edge and the cutting mat were cleaned and sterilised using 75% alcohol (Fisher Scientific, International Inc.) to avoid cross-contamination from blood samples before and after a punched disc was taken from a blood spot. During the processes of DNA extraction, quantification, and amplification, reaction mixtures were prepared in areas distinct from those used for DNA extraction and the analysis of post-PCR products. Disposable tips (Elkay Manufacturing Company, USA) containing hydrophobic filters were used to minimise risks of cross-contamination.

This project used two methods for DNA profiling: the direct PCR amplification technique and the standard DNA extraction, quantification, and PCR amplification technique. The amplified PCR products of each method were profiled using capillary electrophoresis on an Applied Biosystems 3500 Genetic Analyzer (Thermo Fisher Scientific, WA, USA) for the detection and separation of

the DNA fragments. The generated capillary electrophoresis data was interpreted using GeneMapper® ID-X software version 1.6 (ThermoFisher Scientific, WA, USA).

### 2.3.1 DNA EXTRACTION

DNA was extracted from blood samples using a solid-phase extraction technique, following the extraction protocol from the Qiagen QIAamp DNA Investigator Kit (QIAGEN N.V., Hilden, Germany). The process can be summarised as follows: Five to six punches were taken from blood samples on FTA cards and transferred into appropriately labelled 1.5 mL microcentrifuge tubes (Elkay Manufacturing Company, USA) and in each tube added 400 µL of ATL buffer and 20 µL of Proteinase K. The samples were vortexed for 10 s and then incubated at 56°C for 60 min, followed by a brief centrifugation. Subsequently, 400 µL of AL buffer was added to the samples, which were then vortexed for another 10 s and incubated at 70°C for 10 min. After this, 200 µL of 95-100% ethanol (Fisher Scientific, International Inc.) was added, followed by vortexing and brief centrifugation. Finally, the supernatant was pipetted into a spin column and centrifuged at 8,000 rpm (≈ 5,700 × g) for 1 min (QIAamp® DNA Investigator Handbook, 2020).

The spin column was placed into a new collection tube, the previous collection tube was discarded, and 500 µL of AW1 buffer was added before centrifugation at 8,000 rpm (≈ 5,700 × g) for 1 min. The spin column was then transferred to a fresh collection tube, and 700 µL of AW2 buffer was added before centrifugation for 1 min at 8000 rpm (≈ 5,700 × g). The spin column was moved to a new collection tube again, and 700 µL of 95-100% ethanol (Fisher Scientific, International Inc.) was added, followed by another centrifugation for 1 minute at 8000 rpm. The spin column was then placed in a fresh collection tube and dried by centrifugation for 3 min at 13000 rpm. After that, the spin column was transferred to a new 1.5 mL tube, and 100 µL of elution (ATE) buffer was added. This mixture was incubated for 3 min at 56°C. Following the incubation, the extract was centrifuged for 1 minute at 13000 rpm and then stored in a refrigerator at 4°C until the next step, according to the QIAamp® DNA Investigator Handbook (2020).

### 2.3.2 DNA QUANTIFICATION

DNA was quantified using the Qiagen Investigator Quantiplex Kit (QIAGEN N.V., Hilden, Germany) and amplified through quantitative real-time PCR with the Qiagen Rotor-Gene Q 5-plex thermocycler, following the manufacturer's instructions with some modifications outlined below. The Qiagen Investigator Quantiplex system accurately quantifies DNA within the 0 to 20 ng/µL concentration range. To achieve this range, DNA extracts were diluted 1:10 with deionized water. Fresh serial dilutions of the Male Control DNA M1 sample (provided at a concentration of 50

ng/µL with the kit) were prepared, as detailed in **Table 2.1**, to create a standard curve for determining the concentration of the test samples.

**Table 2.1. Dilution series of the Male Control DNA M1 for plotting standard curve using the QIAGEN™ Investigator Quantiplex kit**

| Serial dilution of control DNA M1 (ng/µL) | Control DNA M1 (µL) | QuantiTect Nucleic Acid Dilution buffer (µL) |
|---|---|---|
| **20** | Undiluted DNA | - |
| **5** | 10 | 30 |
| **1.25** | 10 | 30 |
| **0.3125** | 10 | 30 |
| **0.078125** | 10 | 30 |
| **0.01953125** | 10 | 30 |
| **0.0048828125** | 10 | 30 |
| **No Template Control** | - | Only buffer |

The Male Control DNA M1 was serially diluted 1:4, starting from an undiluted concentration of 20 ng/µL to a no-template control (zero DNA concentration). A master mix of 11.5 µL was prepared, consisting of 5.75 µL of Reaction Mix FQ and 5.75 µL of Primer Mix IC FQ. This master mix was combined with 1 µL of the test sample (or the dilution series of Male Control DNA M1) in a 0.1 mL tube (QIAGEN N.V., Hilden, Germany) and sealed. The tubes were then placed in the QIAGEN Rotor GeneQ qPCR thermocycler to amplify the DNA according to the conditions outlined in **Table 2.2** below.

**Table 2.2. Cycling conditions of QIAGEN Rotor GeneQ real time PCR thermocycler**

| Step | Temperature (℃) | Time (seconds) | Number of cycles |
|---|---|---|---|
| ***Taq polymerase activation*** | 95 | 60 | 1 |
| ***Denaturation*** | 95 | 5 | 40 |
| ***Combined annealing/ extension*** | 60 | 10 | |

The fluorescence level was measured during each cycle of the qPCR run. Using Q-Rex software, a standard curve was generated by plotting the quantification cycle (Cq) values against the logarithm of known DNA concentrations. This curve enabled the estimation of DNA concentrations in unknown samples by comparing their Cq values to the standard curve. The Cq represents the number of cycles the fluorescent signal needs to surpass the background level or cross the threshold (Investigator® Quantiplex Handbook, 2018).

## 2.4 AUTOSOMAL STR PROFILING USING QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT

Using standard and direct PCR amplification methods, this experiment employed STR markers from the 17-locus Qiagen™ Investigator ESSplex SE QS Kit to create DNA profiles from blood samples. Blood samples that posed challenges for direct PCR amplification were processed using standard PCR, including DNA extraction and quantification, to obtain complete STR profiles.

### 2.4.1 STANDARD PCR AMPLIFICATION

The STR loci were amplified using multiplex PCR kits, which allowed for the simultaneous analysis of multiple loci in one reaction. The extracted DNA was first amplified with the QIAGEN™ Investigator ESSplex SE QS Kit (QIAGEN N.V., Hilden, Germany) **(Table 2.3)**. This kit targets 16 STR loci along with the Amelogenin sex chromosome marker, employing four-colour fluorescent detection (6-FAM, BTG, BTY, and BTR) (Investigator® ESSplex SE QS Handbook (2021).

**Table 2.3. List of loci incorporated in the QIAGEN™ Investigator ESSplex SE QS Kit with the dye labels and alleles present in the allelic ladder**

| Locus | Matrix Standard | Dye Colour | Allelic Ladder Repeat Numbers |
|---|---|---|---|
| QS1 | 6-FAM | Blue | S, Q |
| Amelogenin | 6-FAM | | X, Y |
| TH01 | 6-FAM | | 4, 5, 6, 7, 8, 9, 9.3, 10, 10.3, 11, 13, 13.3 |
| D3S1358 | 6-FAM | | 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 |
| vWA | 6-FAM | | 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 |
| D21S11 | 6-FAM | | 24, 24.2, 25, 26, 26.2, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 36.2, 37, 38 |
| QS2 | 6-FAM | | Q, S |
| D16S539 | BTG | Green | 5, 8, 9, 10, 11, 12, 13, 14, 15 |
| D1S1656 | BTG | | 9, 10, 11, 12, 13, 14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18, 18.3, 19.3, 20.3 |
| D19S433 | BTG | | 6.2, 8, 9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2 |
| SE33 | BTG | | 3, 4.2, 6.3, 8, 9, 10, 11, 12, 13, 13.2, 14, 14.2, 15, 15.2, 16, 17, 18, 18.2, 19, 19.2, 20, 20.2, 21, 21.2, 22, 22.2, 23.2, 24.2, 25, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 36, 36.2, 37, 38, 39, 42 |
| D10S1248 | BTY | Yellow/ Black | 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |
| D22S1045 | BTY | | 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |
| D12S391 | BTY | | 14, 15, 16, 17, 17.3, 18, 18.3, 19, 20, 21, 22, 23, 24, 25, 26, 27 |
| D8S1179 | BTY | | 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |
| D2S1338 | BTY | | 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 |
| D2S441 | BTR | Red | 8, 9, 10, 11, 11.3, 12, 13, 14, 15, 16, 17, |
| D18S51 | BTR | | 8, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16, 17, 17.2, 18, 18.2, 19, 20 |
| FGA | BTR | | 14, 16, 17, 18, 18.2, 19, 19.2, 20, 20.2, 21, 21.2, 22, 22.2, 23, 23.2, 24, 24.2, 25, 25.2, 26, 27, 28, 29, 30, 31.2, 33, 34, 37.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2 |
| - | BTO | Orange | Internal Size Standard |

The Investigator ESSplex SE QS Kit includes internal PCR controls, known as "Quality Sensors" (QS1 and QS2), which help monitor assay performance, ensure that quality standards are met, and detect the presence of PCR inhibitors in the sample. Details on the STR kits' loci, accession numbers on GenBank®, repeat sequence motifs, and cytogenetic locations are in **Table 2.4** (Investigator® ESSplex SE QS Handbook, 2021).

**Table 2.4. Genetic marker-specific details for the QIAGEN™ Investigator ESSplex SE QS Kit with GenBank® accession numbers, repeat sequence motifs, and cytogenetic location**

| Locus/ Maker | GenBank® Accession Number | Core Repeat Motif | Chromosomal Location |
|---|---|---|---|
| Amelogenin X | M55418 | Non-STR (length polymorphism) | Xp22.1-22. |
| Amelogenin Y | M55419 | Non-STR (length polymorphism) | Yp11.2 |
| TH01 [TC11] | D00269 | [AATG] | 11p15.5 |
| D3S1358 | 11449919 | [TCTA] | 3p25.3 |
| vWA | M25858 | [TCTA] | 12p13.31 |
| D21S11 | AP000433 | [TCTA] | 21q21.1 |
| D16S539 | G07925 | [GATA] | 16q24.1 |
| D1S1656 | NC_000001.9 | [TAGA] | 1q42 |
| D19S433 | G08036 | [AAGG] | 19q12 |
| SE33 [ACTBP2] | NG000840 | [AAAG] | 6q14. |
| D10S1248 | AL391869 | [GGAA] | 10q26.3 |
| D22S1045 | AL022314 | [GGAA] | 22q12.3 |
| D12S391 | G08921 | [AGAT] | 12p13.2 |
| D8S1179 | G08710 | [TCTA] | 8q23.1-23.2 |
| D2S1338 | G08202 | [TGCC] | 2q35 |
| D2S441 | AL079112 | [TCTA] | 2p14 |
| D18S51 | L18333 | [AGAA] | 18q21.3 |
| FGA [FIBRA] | M64982 | [TTTC] | 4q28.2 |

PCR amplification was conducted using the Sure Cycler 8800 thermal cycler (Agilent Technologies, Santa Clara, USA). The DNA amplification was performed according to the manufacturer's recommended protocol, with some adaptations. The PCR kits have an optimal input quantity of 0.25 ng (ranging from 0.1 to 1 ng). The optimal DNA input concentration for a 5 µL reaction is 0.05 ng/µL (ranging from 0.02 to 0.2 ng/µL). Based on the qPCR results, diluting the DNA extracts with nuclease-free water may be necessary before amplifying. Nuclease-free water was used as a negative control, while Control DNA 9948 (included with the kit) was a positive control. For each sample, a PCR master mix was prepared in half-volume reactions (12.5 µL) by mixing 3.75 µL of Fast Reaction Mix, 1.25 µL of Primer Mix, and 7.5 µL of a mixture containing the DNA extract and water, according to the appropriate dilution. The STR amplification was

executed following the protocols specified in **Table 2.5** below (Investigator® ESSplex SE QS Handbook, 2021).

**Table 2.5. DNA amplification cycling protocol for samples taken from blood deposited on FTA cards, using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Step | Temperature (℃) | Time (seconds) | Number of cycles |
|---|---|---|---|
| ***Denaturation and 'hot start' for Taq polymerase*** | 98 | 30 | 3 |
| ***Primer annealing*** | 64 | 55 | |
| ***Primer extension*** | 72 | 5 | |
| ***Denaturation*** | 96 | 10 | 27 |
| ***Primer annealing*** | 61 | 55 | |
| ***Primer extension*** | 72 | 5 | |
| ***Adenylation*** | 68 | 120 | 1 |

### 2.4.2 DIRECT PCR AMPLICATION

The manufacturer's recommended protocol for this project was specific for directly amplifying the STR markers from blood sample punches stored on the FTA card (Direct Amplification of DNA using the Investigator ESSplex SE QS Kit, 2021). The master mix was prepared using reduced-volume (half-volume) PCR reactions as specified in **Table 2.6** (Direct Amplification of DNA using the Investigator ESSplex SE QS Kit, 2021).

**Table 2.6. Direct PCR master mix formulation using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Component | Volume per reaction (μL) |
|---|---|
| **Fast Reaction Mix 2.0** | 3.75 |
| **Primer Mix** | 1.25 |
| **Investigator STR GO! Punch Buffer** | 2.00 |
| **Nuclease-Free Water** | 5.00 |
| **Total reaction volume per sample** | **12.00** |

The reason for using reduced volumes was practical and economical, as there was a limited supply of the kit for many samples. In addition, it was cost-efficient as using reduced volume allowed the researcher to minimise the overall expenditure. The PCR master mix (12.0 μL) was prepared for each sample as described in **Table 2.6**. The master mix was combined thoroughly, and 12.0 μL of the mixture was carefully dispensed into 0.2 mL PCR tubes (Elkay Manufacturing Company, USA)

for each sample. The Investigator STR GO! Punch Buffer was incorporated into the master mix to facilitate the cells' lysis and the DNA release into the PCR mixture. Additionally, it creates a chemical environment that allows DNA polymerase to effectively access and amplify the DNA, even in the presence of potential inhibitors from the sample matrix, such as paper stains. The Harris Uni-Core Micro-Punch® 1.2 mm was used to take a 1.2 mm disc from the centre of the blood sample spot on the Whatman® FTA® Cards and transferred to each PCR reaction. The positive control, a 2.0 $\mu$L template DNA solution containing Control DNA (5ng/L), was added in a separate PCR reaction with the same volume as the sample PCR. The negative control (PCR reaction with no blank FTA Card disc, water, or template DNA) was also prepared in separate PCR tubes. The reactions in the PCR tubes were briefly centrifuged for 20 seconds to ensure that the punch discs were fully submerged. PCR cycling was performed using an Agilent SureCycler 8800 thermal cycler (Agilent Technologies, Santa Clara, USA), following the manufacturer's instructions as outlined in **Table 2.7** (Direct Amplification of DNA using the Investigator ESSplex SE QS Kit, 2021).

**Table 2.7. DNA amplification cycling protocol for samples taken from blood deposited on FTA cards, using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Step | Temperature (°C) | Time (seconds) | Number of cycles |
|---|---|---|---|
| ***Denaturation and 'hot start' for Taq polymerase*** | 98 | 30 | 3 |
| ***Primer annealing*** | 64 | 55 | |
| ***Primer extension*** | 72 | 5 | |
| ***Denaturation*** | 96 | 10 | 22 |
| ***Primer annealing*** | 61 | 55 | |
| ***Primer extension*** | 72 | 5 | |
| ***Adenylation*** | 68 | 120 | 1 |
| ***Final Adenylation*** | 60 | 120 | 1 |

After completion of the protocol, the PCR products were stored in the freezer at -15°C to -30°C and protected from light. The reaction was left in the freezer for at least 24 hours and thawed before proceeding to capillary electrophoresis. According to the manufacturer's protocol, the PCR products can be maintained at 10°C in the thermal cycler before proceeding directly to capillary electrophoresis or stored at -20°C if the electrophoresis is not conducted immediately. However, it was discovered that the longer the direct PCR product remained in the thermal cycler after the programme ended seemed to have negatively impacted the quality of the resulting DNA profiles produced after capillary electrophoresis. The reason could be that PCR amplicons are optimally

stable at lower temperatures ranging from -20°C to 4°C (Aslam *et al*., 2023). At these lower temperatures, stability is preserved by minimising enzymatic activity and reducing the risk of degradation. Therefore, instead of leaving the PCR products in the thermal cycler, they were promptly removed and snap-frozen at -15°C to -30°C, then thawed before proceeding to capillary electrophoresis (Direct Amplification of DNA using the Investigator ESSplex SE QS Kit, 2021).

### 2.4.3 CAPILLARY ELECTROPHORESIS

After the STR amplification, the resulting products, known as amplicons, were analysed using an Applied Biosystem 3500 Genetic Analyzer (Thermo Fisher Scientific) capillary electrophoresis (CE) instrument. Detection was performed through fluorescence, utilizing a fluorescent dye attached to the amplicons' PCR primers. The alleles were separated based on their size, which was determined using an internal lane standard, DNA Size Standard 550 (BTO) (QIAGEN N.V.), added to each PCR product. The allelic ladder ESSplex SE QS (QIAGEN N.V.) was included in each run/injection and was processed concurrently as a separate sample alongside the PCR products to facilitate accurate allele calls.

A formamide-internal lane standards master mix (10 $\mu$L) was prepared by mixing 9.6 $\mu$L of Hi-Di™ Formamide (Thermo Fisher Scientific) with 0.4 $\mu$L of Size Standard 550 (BTO). This master mix was added to a 1 $\mu$L amplified DNA sample (or allelic ladder) in a half-skirted 96-well plate (Thermo Fisher Scientific). The DNA was denatured by incubating it in a 2720 Thermal Cycler (Thermo Fisher Scientific) at 96°C for 3 minutes. After denaturation, the mixture was snap-frozen in a freezer for another 3 minutes before being placed in the autosampler of the CE instrument, where it was automatically injected (Investigator® ESSplex SE QS Handbook, 2021).

### 2.4.4 DATA INTERPRETATION

The generated capillary electrophoresis data was interpreted by the GeneMapper® ID-X software version 1.6 (Thermo Fisher Scientific) for genotyping, following the manufacturer's instructions. This process involved calculating the sizes of PCR amplicons in base pairs with the help of the relevant internal size standard and determining allele calls using the allelic ladder, as outlined in the Investigator® ESSplex SE QS Handbook (2021). **Figure 2.1** displays the electropherogram of the Allelic Ladder ESSplex SE QS. A calibration curve produced from the internal lane standard fragment containing 24 fragments of known length labelled **(**see **Figure 2.2)** using an orange dye was applied to determine the fragment sizes for all amplicons (Investigator® ESSplex SE QS Handbook, 2021).

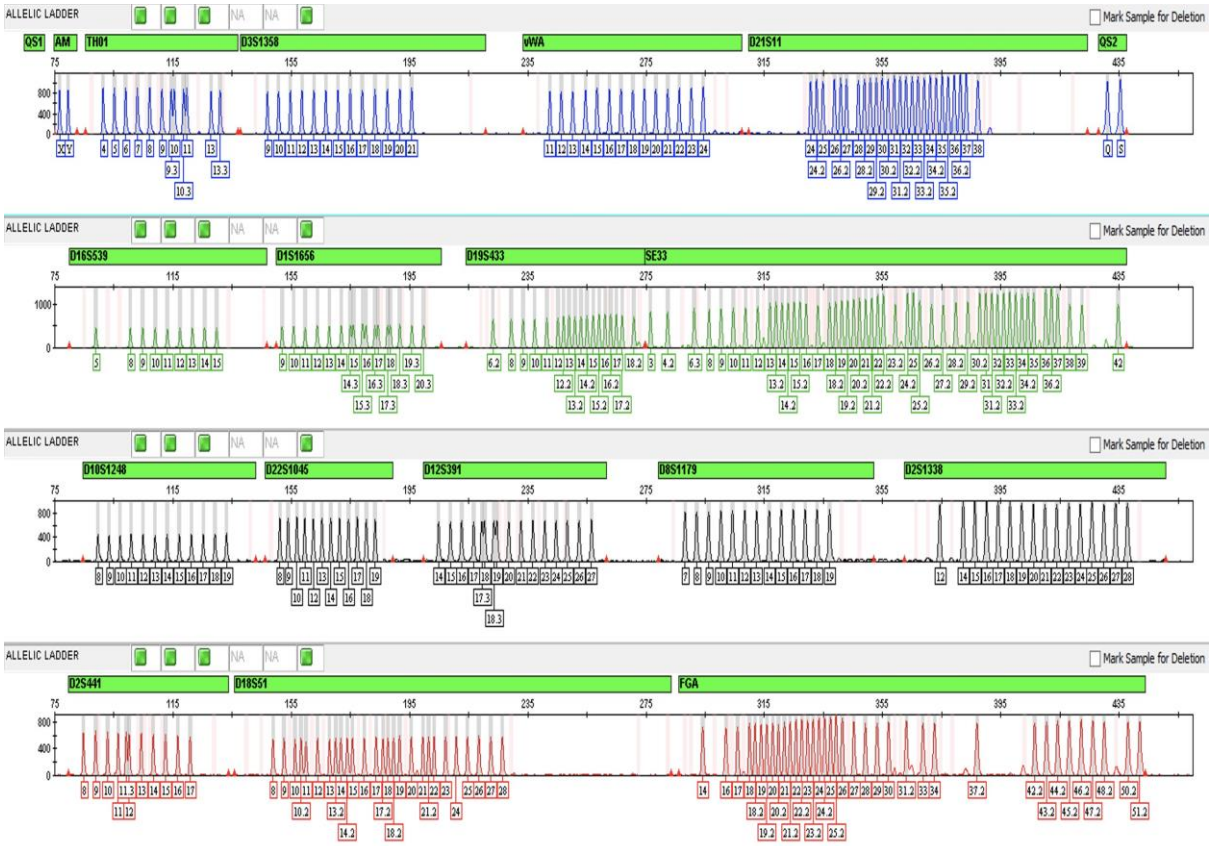



**Figure 2.1. The electropherogram presents the Allelic Ladder ESSplex SE QS, which was analysed using the Applied Biosystems 3500 Genetic Analyzer.**

**Figure 2.2.** displays the electropherogram of the DNA Size Standard 550 (BTO). It shows the data point plotted against the known fragment length. The analytical threshold was fixed at 150 RFU, with a 75% heterozygote peak balance threshold applied. Peaks that fell below this analytical threshold were not included in the analysis. Stutter peaks were automatically identified and flagged by the GeneMapper software using allele- and locus-specific stutter thresholds built into the analysis method. Peaks that measured lower than 15% of the associated parent allele height were classified as stutter and excluded from allele calling. Any peaks detected during the environmental DNA background check or identified as artefacts (e.g., spikes) were discarded from the final DNA profile. (Investigator® ESSplex SE QS Handbook, 2021).

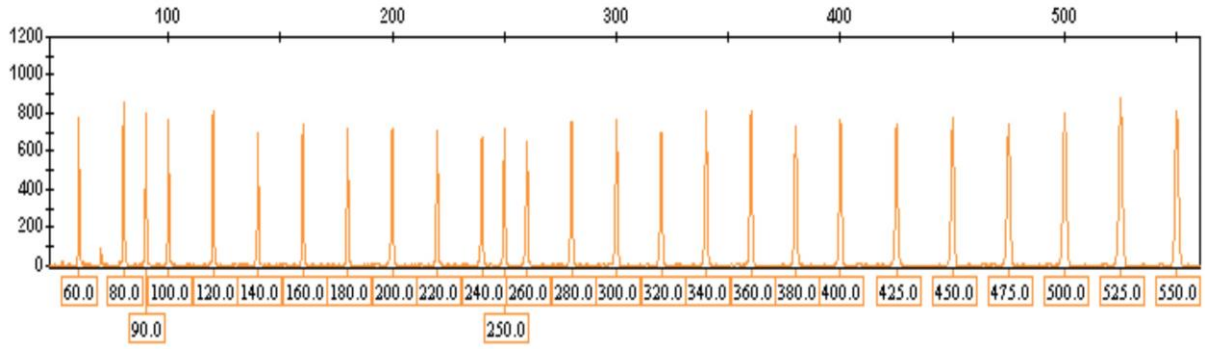



**Figure 2.2. The electropherogram presents the DNA Size Standard 550 (BTO), which indicates the lengths of fragments in base pairs (bp).**

## 2.5 AUTOSOMAL STR PROFILING USING GLOBALFILER™ EXPRESS PCR AMPLIFICATION KIT

Using standard and direct PCR amplification, this experiment used STR markers from the 24-locus GlobalFiler™ Express PCR Amplification Kit to generate DNA profiles from blood samples. Blood samples that were difficult to amplify with direct PCR were processed with standard PCR, which included DNA extraction and quantification, to obtain complete STR profiles.

### 2.5.1 STANDARD PCR AMPLIFICATION

GlobalFiler™ Express PCR Amplification Kit (Thermo Fisher Scientific) is a STR multiplex PCR kit used to amplify 21 autosomal STR markers, a Y-linked indel locus, a Y-STR, and the Amelogenin sex chromosome locus **(Table 2.8)**. The PCR kit allowed for simultaneous amplification of multiple loci using five-colour fluorescent detection (6-FAM™ , VIC™, NED™, TAZ™, and SID™) (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020).

**Table 2.8. Loci in the GlobalFiler™ Express PCR Amplification Kit with dye colours and alleles present in the allelic ladder**

| Locus | Dye Label | Dye Colour | Alleles incorporated in the Allelic Ladder |
|---|---|---|---|
| D3S1358 | 6-FAM™ | Blue | 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 |
| vWA | 6-FAM™ | | 11,12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 |
| D16S539 | 6-FAM™ | | 5, 8, 9, 10, 11, 12, 13, 14, 15 |
| CSF1PO | 6-FAM™ | | 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 |
| TPOX | 6-FAM™ | | 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 |
| Y indel | VIC™ | Green | 1, 2 |
| Amelogenin | VIC™ | | X, Y |
| D8S1179 | VIC™ | | 5, 6, 7, 8, 9 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |
| D21S11 | VIC™ | | 24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 37, 38 |
| D18S51 | VIC™ | | 7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16,17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 |
| DYS391 | VIC™ | | 7, 8, 9, 10, 11, 12, 13 |
| D2S441 | NED™ | Yellow | 8, 9, 10, 11, 11.3, 12, 13, 14, 15, 16, 17 |
| D19S433 | NED™ | | 6, 7, 8, 9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2, 18.2, 19.2 |
| TH01 | NED™ | | 4, 5, 6, 7, 8, 9, 9.3, 10, 11, 13.3 |
| FGA | NED™ | | 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2 |
| D22S1045 | TAZ™ | Red | 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |
| D5S818 | TAZ™ | | 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 |
| D13S317 | TAZ™ | | 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 |
| D7S820 | TAZ™ | | 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 |
| SE33 | TAZ™ | | 4.2, 6.3, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20.2, 21, 21.2, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35, 35.2, 36, 37 |
| D10S1248 | SID™ | Purple | 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |
| D1S1656 | SID™ | | 9, 10, 11, 12, 13, 14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3 |
| D12S391 | SID™ | | 14, 15, 16, 17, 18, 19, 19.3, 20, 21, 22, 23, 24, 25, 26, 27 |
| D2S1338 | SID™ | | 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 |
| | LIZ™ | Orange | GeneScan™ 600 LIZ™ Size Standard v2.0 |

**Table 2.9** contains the loci of the GlobalFiler Express kits, GenBank® accession numbers, repeat sequence motifs, and chromosomal location, for the GlobalFiler™ Express PCR Amplification Kit (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020).

**Table 2.9. Locus-specific information for the GlobalFiler™ Express PCR Amplification Kit with GenBank accession numbers, repeat sequence motifs, and cytogenetic location.**

| Locus Designation | GenBank® Accession Number | Repeat Motif of the Reference Allele | Chromosomal Location |
|---|---|---|---|
| D3S1358 | 11449919 | [TCTA] | 3p21.31 |
| vWA | M25858 | [TCTA] | 12p13.31 |
| D16S539 | G07925 | [GATA] | 16q24.1 |
| CSF1PO | X14720 | [AGAT] | 5q33.3-34 |
| TPOX | M68651 | [AATG] | 2p23–2pter |
| Y indel | N/A | Non-STR (indel marker) | Yq11.221 |
| Amelogenin X | M55418 | Non-STR (length polymorphism) | Xp22.1-22.3 |
| Amelogenin Y | M55419 | Non-STR (length polymorphism) | Yp11.2 |
| D8S1179 | G08710 | [TCTA] | 8q24.13 |
| D21S11 | AP000433 | [TCTA] | 21q1.2-q21 |
| D18S51 | L18333 | [AGAA] | 18q21.33 |
| DYS391 | AC011302 | [TCTA] | Yq11.21 |
| D2S441 | AL079112 | [TCTA] | 2p14 |
| D19S433 | G08036 | [AAGG] | 19q12 |
| TH01 | D00269 | [TCAT] | 11p15.5 |
| FGA | M64982 | [TTTC] | 4q28 |
| D22S1045 | AL022314 | [ATT] | 22q12.3 |
| D5S818 | G08446 | [AGAT] | 5q21.31 |
| D13S317 | G09017 | [TATC] | 13q22-31 |
| D7S820 | G08616 | [GATA] | 7q11.21-22 |
| SE33 | NG000840 | [AAAG] | 6q14 |
| D10S1248 | AL391869 | [GGAA] | 10q26.3 |
| D1S1656 | NC_000001.9 | [TAGA] | 1q42.2 |
| D12S391 | G08921 | [AGAT] | 12p13.2 |
| D2S1338 | G08202 | [TGCC] | 2q35 |

The Sure Cycler 8800 thermal cycler (Agilent Technologies, Santa Clara, USA) was used to perform standard PCR amplification using the manufacturer's recommended protocol for DNA amplification with modifications (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020). The PCR kit has an optimum input quantity of 0.25 ng (0.1-1 ng), of which, for a 3 $\mu$L input, an optimum DNA input concentration will be 0.083 ng/$\mu$L (0.02-0.2 ng/$\mu$L). DNA extracts that required dilution based on qPCR quantification results were diluted with nuclease-free water prior to amplification. Before preparing the PCR reaction mix, all components stored at -20°C were

thawed and vortexed. After their initial use, these PCR components were placed in a refrigerator at 2°C to 8°C. The master mix additive reagent was then added to the master mix and vortexed for 4 to 7 seconds before being discarded. A blank disc was used as a negative control and 2 $\mu$L DNA Control 007 (provided with the kit) was used as a positive control. The PCR reaction mix (half volume reactions) (7.5$\mu$L) was prepared for each kit by mixing 3$\mu$L Master Mix, 3$\mu L$ Primer Set, and a 1.5$\mu L$ mixture of the sample and water according to the appropriate dilution. The STR amplification was executed following the protocols specified in **Table 2.10** below (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020).

**Table 2.10. DNA amplification cycling protocol for samples taken from blood deposited on FTA cards, using the GlobalFiler™ Express PCR Amplification Kit**

| Step | Temperature (°C) | Time (seconds) | Number of cycles |
|---|---|---|---|
| ***Initial Incubation*** | 95 | 60 | 1 |
| ***Denaturation*** | 94 | 3 | 27 |
| ***Anneal/Extension*** | 60 | 30 | |
| ***Final Extension*** | 60 | 480 | 1 |
| ***Final Hold/ Storage*** | 4 | Up to 24 hours | - |

### 2.5.2 DIRECT PCR AMPLIFICATION

The User Guide for the GlobalFiler™ Express PCR Amplification Kit (2020) outlines the manufacturer's protocol for directly amplifying STR loci from blood samples collected on FTA and other paper substrates (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020). The PCR master mix was prepared using half-volumes (7.5 µL), as specified in **Table 2.11**.

**Table 2.11. Master mix setup for direct PCR reactions using the GlobalFiler™ Express PCR Amplification Kit**

| Reaction Component | Volume per reaction ($\mu$L) |
|---|---|
| **Master Mix** | 3.0 |
| **Primer Set** | 3.0 |
| **Nuclease-Free Water** | 1.5 |
| **Total reaction volume per sample** | **7.5** |

The master mix was thoroughly combined and then dispensed into 0.2 mL PCR tubes, each receiving 7.5 µL of the mixture (Elkay Manufacturing Company, USA). A Harris Uni-Core Micro-Punch® with a diameter of 1.2 mm was used to extract a 1.2 mm disc from the centre of the blood

sample spot on the Whatman® FTA® Cards. This disc was transferred to each PCR reaction. Following the disc transfer, the reactions were not mixed. The positive control, consisting of a 2.0µL DNA Control 007, was added in a separate PCR reaction with the same volume as the sample PCR. The negative control containing a blank disc was also prepared. The PCR reactions in the tubes were briefly centrifuged for 20 seconds to ensure the discs were fully submerged. The PCR cycle was performed using an Agilent SureCycler 8800 thermal cycler (Agilent Technologies, Santa Clara, USA), following the manufacturer's protocol outlined in Table 2.12. Once the protocol was completed, the PCR products were stored in a refrigerator at 2 °C to 8 °C if the amplified DNA was to be used within two weeks. If the amplified DNA was not used within this timeframe, it was stored at -20 °C and protected from light before proceeding with capillary electrophoresis (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020).

**Table 2.12. DNA amplification cycling protocol for samples taken from blood deposited on FTA cards, using the GlobalFiler™ Express PCR Amplification Kit.**

| Step | Temperature (°C) | Time (seconds) | Number of cycles |
|---|---|---|---|
| ***Initial Incubation*** | 95 | 60 | 1 |
| ***Denaturation*** | 94 | 3 | 27 |
| ***Anneal/Extension*** | 60 | 30 | |
| ***Final Extension*** | 60 | 480 | 1 |
| ***Final Hold/ Storage*** | 4 | Up to 24 hours | - |

### 2.5.3 CAPILLARY ELECTROPHORESIS

Following the completion of the STR amplification, amplicons were analysed using an Applied Biosystem 3500 Genetic Analyzer (Thermo Fisher Scientific) capillary electrophoresis (CE) instrument. The alleles were separated based on their size, which was determined using an internal lane standard, GeneScan™ 600 LIZ™ Size Standard v2.0, added to each PCR product. The allelic ladder (GlobalFiler™ Express Allelic Ladder) was included in each run/injection and was processed concurrently as a separate sample alongside the PCR products to facilitate accurate allele calls.

A standard master mix for the formamide-internal lane standard was prepared by mixing 9.5 $\mu$L Hi-Di™ Formamide with 0.5 $\mu$L GeneScan™ 600 LIZ™ Size Standard v2.0, totalling a volume of 10.0 $\mu$L. This master mix was added to a 1.0 $\mu$L amplified DNA sample or allelic ladder in a 96-well plate. The DNA was denatured by incubating in a 2720 Thermal Cycler (ThermoFisher Scientific) at 96 °C for 3 min. After this, it was snap-frozen for an additional 3 minutes before

being placed in the autosampler of the capillary electrophoresis (CE) instrument, where it was automatically injected into the system, as described in the GlobalFiler™ Express PCR Amplification Kit User Guide (2020).

### 2.5.4 DATA INTERPRETATION

The fragment analysis data generated was interpreted by the GeneMapper® ID-X software version 1.6 (Thermo Fisher Scientific) for genotyping following the manufacturer's recommended protocol. The software performed an automated analysis of output from the Genetic Analyzer to calculate the size of PCR amplicons in base pairs using an internal size standard. It also determined allele calls in an unknown sample by utilising the allelic ladder provided with the kit (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020). **Figure 2.3** displays the electropherogram of the GlobalFiler™ Express Allelic Ladder analysed on an Applied Biosystems 3500 Genetic Analyzer. An established calibration curve using an internal lane standard fragment contained 36 DNA fragments of known lengths in base pairs **(**see **Figure 2.4).** These were labelled as orange peaks and were used to determine the sizes of the peaks (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020).

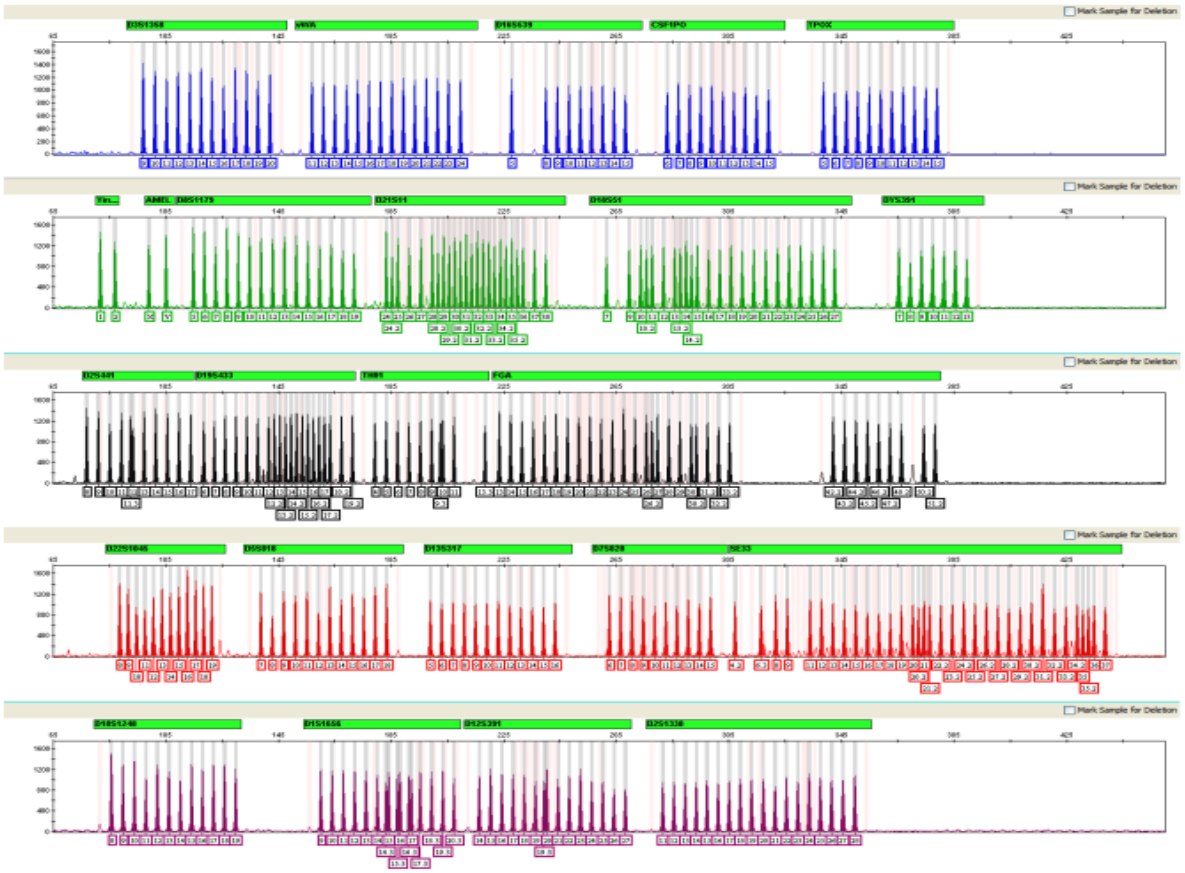


**Figure 2.3. The electropherogram presents the GlobalFiler™ Express Allelic Ladder, which was analysed using the Applied Biosystems 3500 Genetic Analyzer.**

**Figure 2.4** displays the electropherogram of the GeneScan™ 600 LIZ™ Size Standard v2.0, as referenced in the GlobalFiler™ Express PCR Amplification Kit User Guide (2020). The analytical threshold was fixed at 150 RFU, with a heterozygote peak balance threshold applied. Peaks that fell below this analytical threshold were not included in the analysis. Stutter peaks were automatically identified and flagged by the GeneMapper software using allele- and locus-specific stutter thresholds built into the analysis method. Peaks that measured lower than 15% of the associated parent allele height were classified as stutter and excluded from allele calling. Any peaks detected during the environmental DNA background check or identified as artefacts (e.g., spikes) were discarded from the final DNA profile (GlobalFiler™ Express PCR Amplification Kit User Guide, 2020).

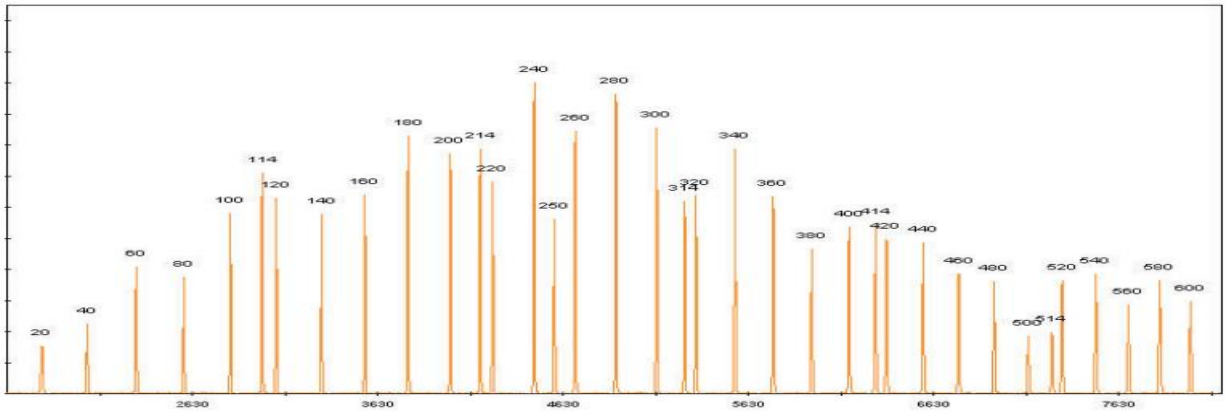



**Figure 2.4. Electropherogram presents the GeneScan™ 600 LIZ™ Size Standard v2.0, which shows fragments in base pairs (bp)**

## 2.6 Y-CHROMOSOMAL STR PROFILING USING PROMEGA POWERPLEX® Y23 SYSTEM KIT

This experiment used Y-STR loci from the 23-locus Promega PowerPlex® Y23 System Kit to generate DNA profiles from blood samples via standard and direct PCR amplification. Blood samples that were challenging to amplify with direct PCR were amplified using standard PCR, which involved DNA extraction and quantification, to obtain complete STR profiles.

### 2.6.1 STANDARD PCR AMPLIFICATION

The Promega PowerPlex® Y23 System Kit (Promega Corporation, Madison, Wisconsin, United States) is a multiplex PCR kit used to amplify 23 Y-STR markers from extracted DNA **(Table 2.13)**. The kit allowed simultaneous amplification of multiple loci using four-colour fluorescent detection (Fluorescein, JOE, TMR-ET, and CXR-ET) in a single reaction (Promega Technical Manual, 2023).

**Table 2.13. Loci in the Promega PowerPlex® Y23 System Kit with dye colours and alleles present in the allelic ladder.**

| Locus | Dye Label | Dye Colour | Alleles incorporated in the Allelic Ladder |
|---|---|---|---|
| **DYS576** | Fluorescein | Blue | 11, 12, 13. 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 |
| **DYS389I** | Fluorescein | Blue | 9, 10, 11, 12, 13, 14, 15, 16, 17 |
| **DYS389II** | Fluorescein | Blue | 24, 25, 26, 27, 28, 29, 30, 31,32, 33, 34, 35 |
| **DYS448** | Fluorescein | Blue | 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 |
| **DYS19** | Fluorescein | Blue | 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 |
| **DYS391** | JOE | Green | 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 |
| **DYS481** | JOE | Green | 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 |
| **DYS549** | JOE | Green | 7, 8, 9, 10, 11, 12, 13, 14, 15,16, 17 |
| **DYS533** | JOE | Green | 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 |
| **DYS438** | JOE | Green | 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 |
| **DYS437** | JOE | Green | 11, 12, 13, 14, 15, 16, 17, 18 |
| **DYS570** | TMR-ET | Black | 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 |
| **DYS635** | TMR-ET | Black | 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 |
| **DYS390** | TMR-ET | Black | 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 |
| **DYS439** | TMR-ET | Black | 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 |
| **DYS392** | TMR-ET | Black | 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 |
| **DYS643** | TMR-ET | Black | 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 |
| **DYS393** | CXR-ET | Red | 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 |
| **DYS458** | CXR-ET | Red | 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 |
| **DYS385a/b** | CXR-ET | Red | 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 |
| **DYS456** | CXR-ET | Red | 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 |
| **Y-GATA-H4** | CXR-ET | Red | 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 |
| | WEN ILS | Orange | WEN Internal Lane Standard 500 Y23 |

Standard PCR amplification was performed using the Agilent Sure Cycler 8800 thermal cycler (Agilent Technologies, Santa Clara, CA, USA), according to the manufacturer's recommended protocol. The kit has an optimum input quantity of 0.5 ng (0.1-1 ng); an optimum DNA input concentration for a 17.5 µL input is 0.0286 ng/µL (0.02-0.2 ng/µL). Based on the qPCR results, the DNA extract may require dilution using nuclease-free water before amplification. Before preparing the PCR reaction mix, all components were protected from light, stored at 20°C, thawed and vortexed for 15 seconds. After initial use, these PCR components were placed in a refrigerator at 2°C to 8°C. A blank disc punch was used as a negative control, and 1.0 µL 2800M Control DNA (provided with the kit) was used as a positive control. The PCR reaction mix (for half-volume reactions, totalling 12.5 µL) was prepared for each kit by combining 2.5 µL of PowerPlex® Y23 5X Master Mix, 1.25 µL of PowerPlex® Y23 10X Primer Set, and 8.75 µL of a mixture that included the sample (0.5 ng of DNA template) and water (amplification grade) provided in the kit, according to the appropriate dilution (Promega Technical Manual, 2023). **Table 2.14** lists the kit's loci, the size range of allelic ladder components, the core repeat sequence motif, and chromosomal location (Promega Technical Manual, 2023).

**Table 2.14. Locus-specific information of the PowerPlex® Y23 System Kit with the size ranges of allelic ladder, core repeat sequence motif, and chromosomal location.**

| Y-STR Locus | Allelic Ladder Components (bases) Size Range | Repeat Sequence 5′ - 3′ | Chromosomal Location |
|---|---|---|---|
| DYS576 | 97-145 | [AAAG] | Y |
| DYS389I | 147-179 | [TCTA] | Y |
| DYS389II | 196-256 | [TCTA] | Y |
| DYS448 | 259-303 | [AGAGAT] | Y |
| DYS19 | 312-352 | [TAGA] | Y |
| DYS391 | 86-130 | [TCTA] | Y |
| DYS481 | 139-184 | [CTT] | Y |
| DYS549 | 198-238 | [GATA] | Y |
| DYS533 | 245-285 | [ATCT] | Y |
| DYS438 | 293-343 | [TTTTC] | Y |
| DYS437 | 344-380 | [TCTA] | Y |
| DYS570 | 90-150 | [TTTC] | Y |
| DYS635 | 150-202 | [TCTA] | Y |
| DYS390 | 207-255 | [TCTA] | Y |
| DYS439 | 263-307 | [AGAT] | Y |
| DYS392 | 314-362 | [TAT] | Y |
| DYS643 | 368-423 | [CTTTT] | Y |
| DYS393 | 101-145 | [AGAT] | Y |
| DYS458 | 159-215 | [GAAA] | Y |
| DYS385a/b | 223-307 | [GAAA] | Y |
| DYS456 | 316-364 | [AGAT] | Y |
| Y-GATA-H4 | 374-414 | [TAGA] | Y |

The STR amplification for the extracted and quantified DNA was executed following the protocol specified in **Table 2.15** below (Promega Technical Manual, 2023).

**Table 2.15. DNA amplification cycling protocol for samples taken from blood deposited on FTA cards, using the PowerPlex® Y23 System Kit**

| Step | Temperature (°C) | Time (seconds) | Number of cycles |
|---|---|---|---|
| ***Initial Incubation*** | 96 | 120 | 1 |
| ***Denaturation*** | 94 | 10 | 30 |
| ***Anneal*** | 61 | 60 | |
| ***Extension*** | 72 | 30 | |
| ***Final Extension*** | 60 | 1,200 | 1 |
| ***Final Hold/ Storage*** | 4 | ∞ | - |

### 2.6.2 DIRECT PCR AMPLICATION

The technical manual for the PowerPlex® Y23 System, designed for use with the Applied Biosystems® Genetic Analyzers (2023), provided procedures for the direct amplification of Y-STR loci from storage card punches. This process utilized a 12.5 µL reaction volume prepared as specified in **Table 2.16** by combining 2.5 µL PowerPlex® Y23 5X Master Mix, 1.25 µL PowerPlex® Y23 10X Primer Set, 6.25 µL water (amplification grade), and 2.5 µL 5X AmpSolution™ Reagent. The master mix was thoroughly mixed before dispensing 12.5 µL into each PCR tube. (Promega Technical Manual, 2023).

**Table 2.16. Master mix setup for direct PCR reactions using the PowerPlex® Y23 System Kit**

| Component | Volume per reaction (µL) |
|---|---|
| **PowerPlex® Y23 5X Master Mix** | 2.50 |
| **PowerPlex® Y23 10X Primer Set** | 1.25 |
| **5X AmpSolution™ Reagent** | 2.50 |
| **Water, Amplification Grade** | 6.25 |
| **Total reaction volume per sample** | **12.50** |

A Harris Uni-Core Micro-Punch® with a diameter of 1.2 mm was used to extract a 1.2 mm disc from the centre of the blood sample spot on Whatman® FTA® Cards. This disc was then transferred into each PCR reaction in a PCR tube. A 1 µL of 2800M Control DNA solution was prepared to create a positive control reaction. A negative control, consisting of a blank disc in a PCR reaction, was also prepared. The PCR reactions in the tubes were briefly centrifuged for 20 seconds to ensure the punched discs were fully submerged. (Promega Technical Manual, 2023).

**Table 2.17. DNA amplification cycling protocol for samples taken from blood deposited on FTA cards, using the PowerPlex® Y23 System Kit.**

| Step | Temperature (°C) | Time (seconds) | Number of cycles |
|---|---|---|---|
| ***Initial Incubation*** | 96 | 120 | 1 |
| ***Denaturation*** | 94 | 10 | 26 |
| ***Anneal*** | 61 | 60 | |
| ***Extension*** | 72 | 30 | |
| ***Final Extension*** | 60 | 1,200 | 1 |
| ***Final Hold/ Storage*** | 4 | ∞ | - |

The PCR was conducted using an Agilent Sure Cycler 8800 thermal cycler. The thermal cycler was programmed according to the conditions specified in **Table 2.17**. Following PCR amplification, the products were stored at −20 °C in accordance with the manufacturer's protocol and protected from light prior to capillary electrophoresis. (Promega Technical Manual, 2023).

### 2.6.3 CAPILLARY ELECTROPHORESIS

After completing the STR amplification, the resulting amplicons were analysed using an Applied Biosystems 3500 Genetic Analyser (Thermo Fisher Scientific) capillary electrophoresis (CE) instrument. The alleles were separated by size with the help of an internal lane standard known as the WEN Internal Lane Standard (WEN ILS) 500 Y23, supplied with the kit. Additionally, an allelic ladder, the PowerPlex® Y23 System Allelic Ladder, was run alongside the samples to aid in making accurate allele calls.

A master mix of 10 µL was prepared by combining 9.5 µL of Hi-Di™ Formamide with 0.5 µL of WEN Internal Lane Standard 500 Y23. This mix was then added to 1 µL of either the amplified DNA sample or an allelic ladder. The DNA was denatured by incubating it in a 2720 Thermal Cycler (Applied Biosystems) at 96°C for 3 minutes. Afterwards, it was snap-frozen for an additional 3 minutes before being placed into the autosampler of the capillary electrophoresis (CE) instrument. The sample was automatically injected into the CE system, as outlined in the Promega Technical Manual (2023).

### 2.6.4 DATA INTERPRETATION

GeneMapper® ID-X software version 1.6 (Thermo Fisher Scientific) was utilised following the manufacturer's protocol to analyse the fragment data generated for genotyping (Promega Technical Manual, 2023). The software performed an automated analysis of the output from the Genetic Analyzer Capillary Electrophoresis instrument, calculating the size in base pairs of PCR amplicons using an internal size standard. It also determined allele calls in an unknown sample by referencing the allelic ladder. **Figure 2.5** displays the electropherogram of the PowerPlex® Y23 System allelic ladder analysed on an Applied Biosystems 3500 Genetic Analyzer. A calibration curve was established using an internal lane standard fragment that included 21 DNA fragments of known lengths in base pairs **(**see **Figure 2.6)**. These fragments were labelled as orange peaks and were used to determine the sizes of the observed peaks (Promega Technical Manual, 2023).

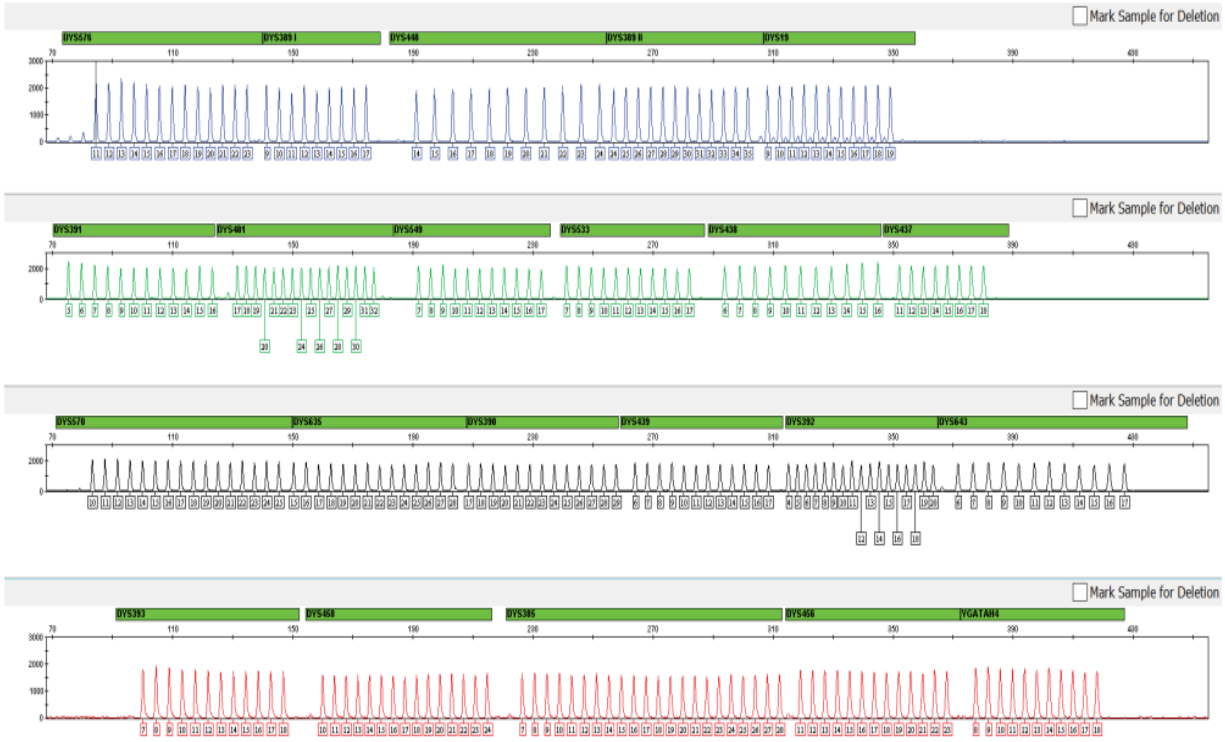



**Figure 2.5. Electropherogram of the PowerPlex® Y23 System Allelic Ladder analysed on an Applied Biosystems 3500 Genetic Analyzer**

**Figure 2.6** presents the electropherogram of the WEN Internal Lane Standard 500 Y23, as specified in the Promega Technical Manual (2023). It shows the datapoint plotted against known fragment length. The analytical threshold was fixed at 150 RFU, with a 75% heterozygote peak balance threshold applied. Peaks that fell below this analytical threshold were not included in the analysis. Stutter peaks were automatically identified and flagged by the GeneMapper software using allele- and locus-specific stutter thresholds built into the analysis method. Peaks that measured lower than 15% of the associated parent allele height were classified as stutter and excluded from allele calling. Any peaks detected during the environmental DNA background check or identified as artefacts (e.g., spikes) were discarded from the final DNA profile (Promega Technical Manual, 2023).

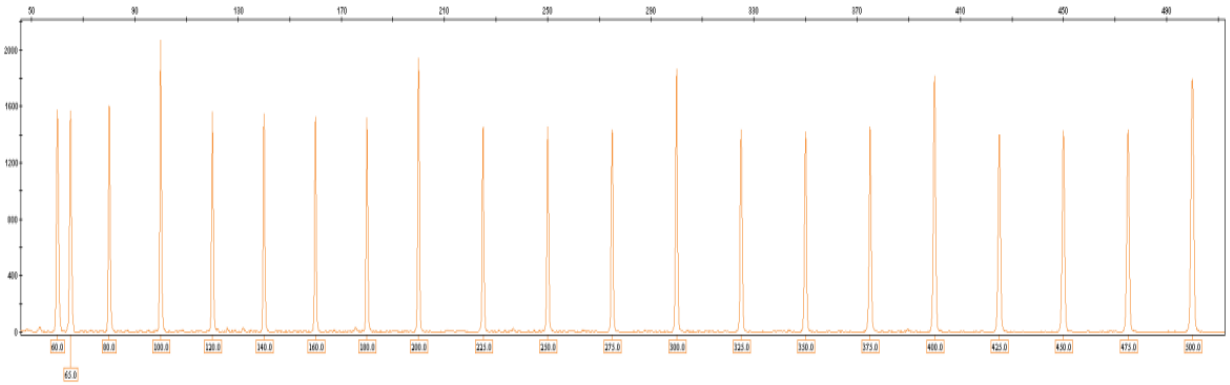


**Figure 2.6. Electropherogram of the WEN Internal Lane Standard 500 Y23, showing fragment lengths in base pairs (bp).**

## 2.7 STATISTICAL ANALYSIS

### 2.7.1 GENETIC AND FORENSIC PARAMETERS

The DNA data, tables, graphs, and charts were organized in Microsoft Excel spreadsheets. They were statistically analysed using the STR Analysis for Forensic (*STRAF*-A 2.2.2) online tool for short tandem repeat (STR) data analysis (Gouy & Zieger, 2017) and the Genetic Analysis in Excel (GenAIEx 6.5) software package (Peakall and Smouse, 2006, 2012) to calculate allele and genotype frequencies, respectively. The GenAIEx 6.5 platform was also utilised to assess the private alleles within populations.

Arlequin version 3.5 (Excoffier & Lischer, 2010) was used to test each locus for deviation from Hardy–Weinberg equilibrium (HWE), employing the Exact Test for multi-allelic loci to assess the conformity of observed genotype frequencies with HWE expectations. Wright's F-statistics ($F_{ST}$, $F_{IS}$, and $F_{IT}$) were calculated to evaluate inter-sub-population genetic structure within the Nigerian population using GenAIEx version 6.5. GenAIEx 6.5 was also used to identify private alleles and to generate pairwise matrices of Nei's genetic distance and genetic identity, thereby assessing genetic differentiation and relatedness among populations.

The forensic parameters, including the Polymorphic Information Content (PIC), Typical Paternity Index (TPI), Power of Discrimination (PD), Random Match Probability (RMP), Power of Exclusion (PE), Expected Heterozygosity ($H_{exp}$), Observed Heterozygosity ($H_o$), and Expected Homozygosity ($H_{oms\_exp}$), were calculated and evaluated using the *STRAF*-A 2.2.2 online tool for STR data analysis.

### 2.7.2 POPULATION GENETIC AND GEOGRAPHIC ANALYSIS

**POPULATION STRUCTURE**

*STRUCTURE* software version 2.3.4 was employed to evaluate genetic structure of the population. *STRUCTURE* was developed by Pritchard *et al*. (2000) and is available for free download online (https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html) (Falush *et al*., 2003; Hubisz *et al*., 2009). *STRUCTURE* is a population analysis tool that utilises a Bayesian clustering technique to assign individuals to populations based on their genotypes, utilising multi-locus genotype data. Through maximum likelihood analysis, *STRUCTURE* can identify entire subsets or sub-populations by detecting differences in allele frequencies within the data and assigning individuals accordingly. The software's Bayesian iterative algorithm groups individuals with the same variation patterns and evaluates the differences in population genetic variants while considering their geographical distributions. This process is

achieved using Markov Chain Monte Carlo (MCMC) estimation, which randomly assigns individuals to a predetermined number of groups. The frequency of genetic variants is then assessed within each group, and individuals are re-assigned based on the analysis of those frequencies. This procedure was repeated multiple times, involving a total of 100,000 iterations along with a burnin period of 50,000 iterations. The burnin period is the first phase of the Markov Chain Monte Carlo (MCMC) process. In this phase, the software discards early results that may not be reliable. As a result, reliable estimates of allele frequencies and the probability of an individual's association with a specific population were obtained for each group. *STRUCTURE* analyses various population sizes (denoted as *K*), which are set by the user, to estimate the proportion of each individual's genome derived from these assumed populations. It calculates the likelihood of the observed data (X) for different values of *K* by establishing posterior probabilities for each *K*. The analysis explored *K* values ranging from 1 to 10 iterations and was programmed to repeat five runs. The parameters used were as follows: number of individuals = 303, ploidy = 2, number of loci = 16, missing values coded as -9. The optimal *K* value was determined using the Structure Selector software (https://lmme.ac.cn/StructureSelector/) (Evanno *et al*., 2005) and the results of the *K* dataset were visualised using CLUMPACK (Clustering Markov Packager Across K) (https://clumpak.tau.ac.il/) (Kopelman *et al*., 2015). The highest value of Ln P(X | K) indicates the valid *K* number. Additionally, the maximum rate of change (Delta *K*) in the Ln P(X | K) values between successive *K* values helps identify the appropriate number of genetic clusters in the data (Porras-Hurtado *et al*., 2013).

**PRINCIPAL COMPONENT ANALYSIS**

Principal Component Analysis (PCA) was performed using the PAleontological Statistics (PAST 4.03) software. Developed by Øyvind Hammer (Deinert *et al*., 2024), PAST 4.03 is a free scientific tool for download online (https://past.en.lo4d.com/download). Performing PCA in the software involved inputting the numeric data with individuals as rows and variables (such as STR data, allele frequencies, and pairwise genetic distance matrix data) as columns. The relevant data were highlighted, and the "Multivariate" section was navigated to select "Ordination" to access the principal components analysis. The software creates plots and eigenvalues to help visualise population structure using a covariance or correlation matrix. Group labels can also be added to the plot visualisations to get better clarity. PCA is a statistical technique that employs an unsupervised clustering method to analyse and visualise population structure more effectively (Paschou *et al*., 2007; 2008). It has been very frequently utilised in genetic and geographic data analysis. It accomplishes this by reducing large datasets into smaller sets of “summary indices” while accounting for variation in genome-wide relationships without significant data loss. As a

multivariate data analysis method, PCA reduces the dimensionality of complex datasets, which are often difficult to interpret, and increases interpretability while minimising the loss of information. It successively maximises variance by creating new uncorrelated variables. The PCA technique retains variables that exhibit the most significant variation while eliminating those genetic markers with no variation. In this multivariate analysis, 303 unrelated individuals from the Igbo, Yoruba, and Hausa-Fulani sub-populations were compared using loci data from 16 STRs to visualize genetic differentiation. The procedure involved comparing the Principal Component (PC1), which captures the highest variance, on the X-axis and the Principal Component (PC2), which represents the second highest variance, on the Y-axis. This setup provides the best low-dimensional representation of the data. This process helps to reveal hidden relationships between samples. Consequently, the PCA plot presents similar samples clustered closely while dissimilar samples are positioned further apart (Patterson *et al*., 2006).

**NEIGHBOUR-JOINING**

Neighbour-Joining (NJ) was performed using Molecular Evolutionary Genetics Analysis (MEGA-11) software. The MEGA 11 software program is powerful for analysing DNA and protein sequence evolution studies (Kumar *et al.*, 1994; 2018). The software provides a complete set of tools and methods for understanding how molecules evolve. Developed by Koichiro Tamura, Glen Stecher, and Sudhir Kumar (Tamura *et al.*, 2021), MEGA 11 is a free scientific tool for download online (https://www.megasoftware.net/). It is frequently utilised to analyse and study evolutionary relationships and phylogenetic relatedness using DNA and protein sequences Performing NJ analysis for STR and Y-STR data in the software involved creating a pairwise genetic distance matrix of Nei genetic distance. This matrix was then loaded into MEGA, where the "Phylogeny" section was accessed to select "Construct/Test Neighbour-Joining Tree." The analysis preferences were configured, and the computation was executed. The software generated a visual Neighbour-Joining tree, illustrating the population's phylogenetic relationships. The branch lengths—calculated based on phylogenetic distances using the number of differences method—were computed by MEGA 11 to create accurately scaled trees. The NJ tree can be edited and exported for further interpretation. (Tamura *et al.*, 2021).

# CHAPTER THREE

# POPULATION GENETICS STUDY OF THE NIGERIAN POPULATION USING THE QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT

## 3.1 INRODUCTION

The Qiagen™ Investigator ESSplex SE QS Kit (QIAGEN N.V., Hilden, Germany) is a forensic STR multiplex kit designed for the simultaneous amplification of 17 STR markers covering 16 autosomal STR loci. It is designed to comply with the European Standard Set (ESS) loci. This ensures that it meets the DNA analysis standards for European databases. The kit also includes two quality sensors (QS1 and QS2) as internal PCR controls, amplified alongside the polymorphic markers. These controls provide critical details about the PCR process efficiency, DNA degradation status, and the presence of PCR inhibitors (Barbaro *et al*., 2024). Highly sensitive and optimized for difficult forensic samples, such as low template and degraded DNA by employing short amplicon targets, quality sensors, and robust PCR chemistry, the Qiagen™ Investigator ESSplex SE QS Kit is compatible with the ABI 3130 and 3500 Genetic Analyzer (ThermoFisher Scientific, WA, USA). It has been successfully applied in forensic casework, including body fluid identification and crime scene sample analysis (Harrel *et al*., 2021; Rath *et al*., 2024).

Nigeria has a diverse society comprising approximately 300 ethnic groups. The three largest groups that make up significant parts of the population are the Hausa-Fulanis (29%), Yorubas (21%), and Igbos (18%). The remaining ethnic groups, while classified as minorities, include notable populations such as the Ijaw (10%), Kanuri (4.0%), Efik-Ibibio (3.5%), Tiv (2.5%), Bini (1.05%), and among others. These groups are some of the largest and most prominent minority ethnic groups in Nigeria (Agbaire and Dunne, 2024). It is essential to highlight that the evaluation of the Qiagen™ Investigator ESSplex SE QS Kit has not been previously documented for the Nigerian population.

## 3.2 AIM AND OBJECTIVES

This research aims to analyse the population genetics of 303 unrelated individuals from Nigeria, specifically from the Igbo, Yoruba, and Hausa-Fulani ethnic groups. This study utilizes 16 autosomal STR markers from the 17-locus Qiagen™ Investigator ESSplex SE QS Kit (QIAGEN N.V., Hilden, Germany).

The specific objectives for the work detailed in this chapter are as follows:

1. Perform STR genotyping for this kit and evaluate any STR off ladders allele calls.

2. Generate STR allele frequency data to create an allele frequency database for each subpopulation and the entire population.
3. Evaluate forensic parameters for each subpopulation and the entire population, assess the effectiveness of the kit used, and determine its suitability for forensic work in Nigeria.
4. Compare the genetic structure within populations and examine the relationships between the different subpopulations.

## 3.3 MATERIALS AND METHODS

Blood samples for the study on the Nigerian population were collected from different locations in Nigeria and the United Kingdom, involving 303 unrelated individuals (167 males and 136 females), as described in **Chapter 2.2**. Donors received participant information sheets and consent forms to sign and return. Participants additionally received verbal explanations regarding the procedures for collecting the blood samples. STR profiling, data analysis, and statistical calculations were conducted at the University of Strathclyde in Glasgow. Detailed procedures for sample collection can be found in **Chapter 2.2**, and the participation information sheet and consent form are in **Appendices 1** and **2**. The DNA profiling technique, utilising a 1.2mm punch of blood sample from the FTA® card, involved standard PCR and direct PCR with the Qiagen™ Investigator ESSplex SE QS Kit (QIAGEN N.V., Hilden, Germany), as described in **Chapters 2.4.1** and **2.4.2**, respectively. The capillary electrophoresis method, which detects and separates DNA fragments, was performed using the ABI Genetic Analyzer 3500 with the internal lane standard, Size Standard 550 (BTO) and ESSplex SE QS Allelic Ladder, as detailed in **Chapter 2.4.3.** Data interpretation, including allele calling and genotype determination for the 16 autosomal STR loci, was conducted using GeneMapper® ID-X software version 1.6 (Thermo Fisher Scientific), as described in **Chapter 2.4.4.** The procedure for calculating the allele frequencies and various forensic parameters using the STR Analysis for Forensic (*STRAF*-A 2.2.2) online tool and the Genetic Analysis in Excel (GenAIEx 6.5) platform was explained in **Chapter 2.7.1**. The Wright's F-statistics ($F_{ST}$, $F_{IS}$, and $F_{IT}$) were evaluated using the Genetic Analysis in Excel (GenAIEx 6.5) platform to assess genetic variation among the Igbo, Yoruba, and Hausa-Fulani sub-populations compared to the total population, as explained in **Chapter 2.7.1**. To evaluate the inter-subpopulation structure, *STRUCTURE* software version 2.3.4 was employed, as described in **Chapter 2.7.2**. The software PAleontological Statistics (PAST 4.03) conducted Principal Component Analysis (PCA) to assess sub-population differentiation, as explained in **Chapter 2.7.2**. Consequently, the PCA scatter plot presents similar samples clustered closely while dissimilar samples are positioned further apart. As explained in **Chapter 2.7.2**, the software Molecular Evolutionary Genetics Analysis (MEGA-11)

was used to conduct a Neighbour-Joining analysis of the Igbo, Yoruba, and Hausa-Fulani sub-populations.

## 3.4 RESULTS

### 3.4.1 STR GENOTYPING

The QIAGEN™ Investigator ESSplex SE QS Kit successfully generated complete DNA profiles from blood spots on FTA cards using the direct PCR amplification approach (254 samples) and the standard PCR method after DNA extraction and quantification from the samples on FTA cards (49 samples). Blood samples that were challenging to amplify using direct PCR were successfully amplified using standard PCR, which included DNA extraction and quantification to obtain complete STR profiles. The DNA extraction and quantification processes overcame PCR inhibition, likely caused by matrix components in the samples that hindered the effectiveness of the direct PCR method. Each DNA profile contained genotypes at 16 polymorphic STR markers, plus the sex chromosome-specific marker Amelogenin. All the negative controls included in every single run detected no sign of contamination. The DNA profile produced from the positive control sample (**Figure 3.1**) was in concordance with the expected profile. The Quality Sensors peaks were identified as amplicons.

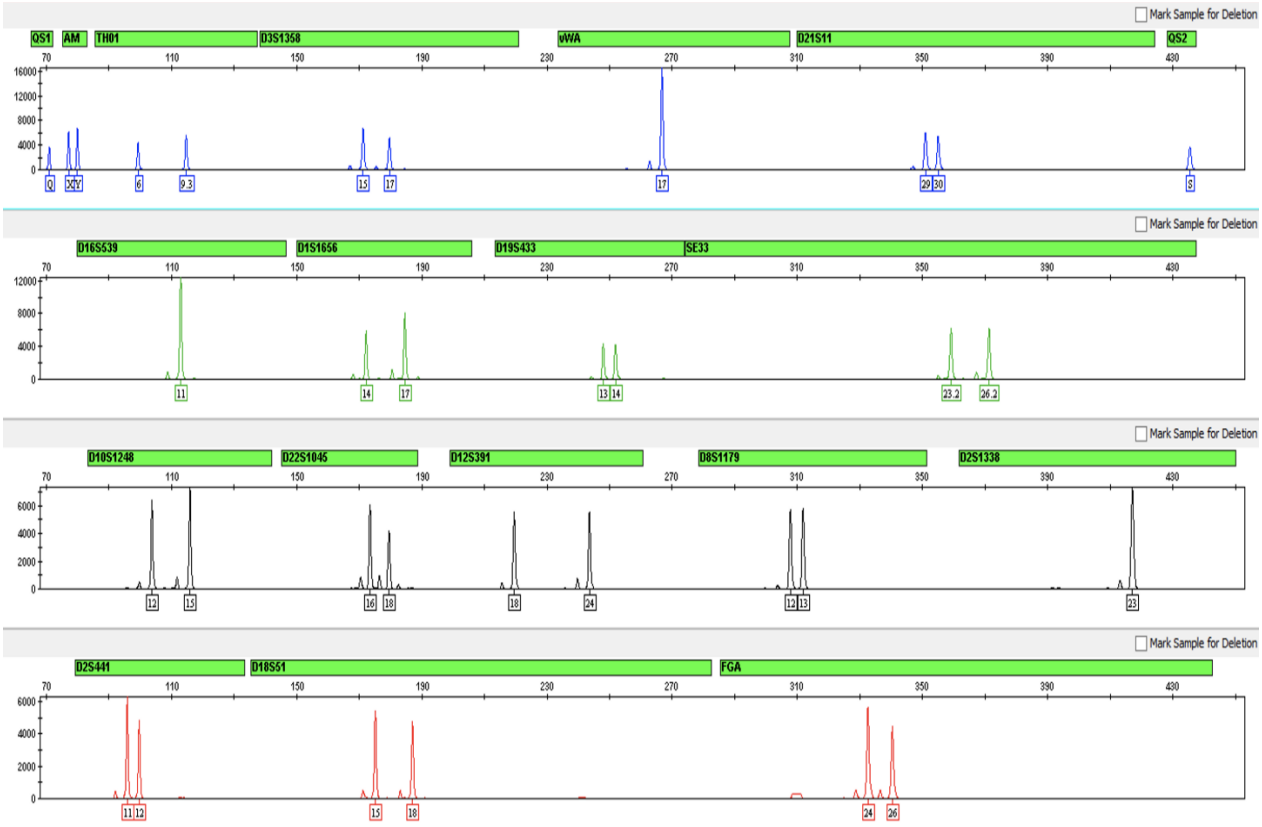



**Figure 3.1. Electropherogram of the QIAGEN™ Investigator ESSplex SE QS Control DNA showing the loci labelled across four colour channels.**

### 3.4.2 OFF-LADDER ALLELES

GeneMapper® ID-X v1.6 software classified unidentified alleles as off-ladder (OL). To address the OL alleles, the ladder alignment of the affected samples was first checked and confirmed that the peaks were not artefacts. These checks were conducted before undergoing a second capillary electrophoresis run. If the OL allele issues persisted, a second PCR amplification was performed using the same optimised conditions, which included cycle number, reagent concentration, temperature profile, and input DNA quantity to confirm the reproducibility or authenticity of the alleles. The results were then reviewed using the allelic ladder bins generated by the GeneMapper® ID-X software, checking the alignment of the peaks with the allelic ladder to ensure no misalignment occurred.

Three distinct OL alleles were identified at the D12S391 locus (see **Figure 3.2**), and one allele was noted at both the D3S1358 (see **Figure 3.3**) and D1S1656 (see **Figure 3.4**) loci. The OL alleles were found in positions where no bins were available, resulting in their classification as off-ladder in the QIAGEN™ Investigator ESSplex SE QS Kit.

During the concordance analysis (**Chapter 4.4.2**), the results from the QIAGEN™ Investigator ESSplex SE QS Kit samples were compared to those from the GlobalFiler™ Express Kit, based on the data generated for the same sample. It was discovered that alleles not identified in the QIAGEN™ Investigator ESSplex SE QS Kit due to the absence of designated bins in the allelic ladders were correctly identified in the GlobalFiler™ Express Kit, which included the missing bins and ladders in its allelic ladder.

Further investigation confirmed that alleles generated from the QIAGEN™ Investigator ESSplex SE QS Kit were concordant with those generated from the GlobalFiler™ Express Kit. The identified off-ladder alleles are reported in **Chapter 4.4.2**. The loci were genotyped based on the GlobalFiler™ Express Kit result.

Off-ladder alleles observed with the QIAGEN™ Investigator ESSplex SE QS Kit correspond to previously described STR variants documented in forensic STR databases, including STRBase, and were correctly designated by the GlobalFiler™ Express Kit allelic ladder. These alleles were therefore treated as concordant and genotyped based on the GlobalFiler™ Express Kit results.

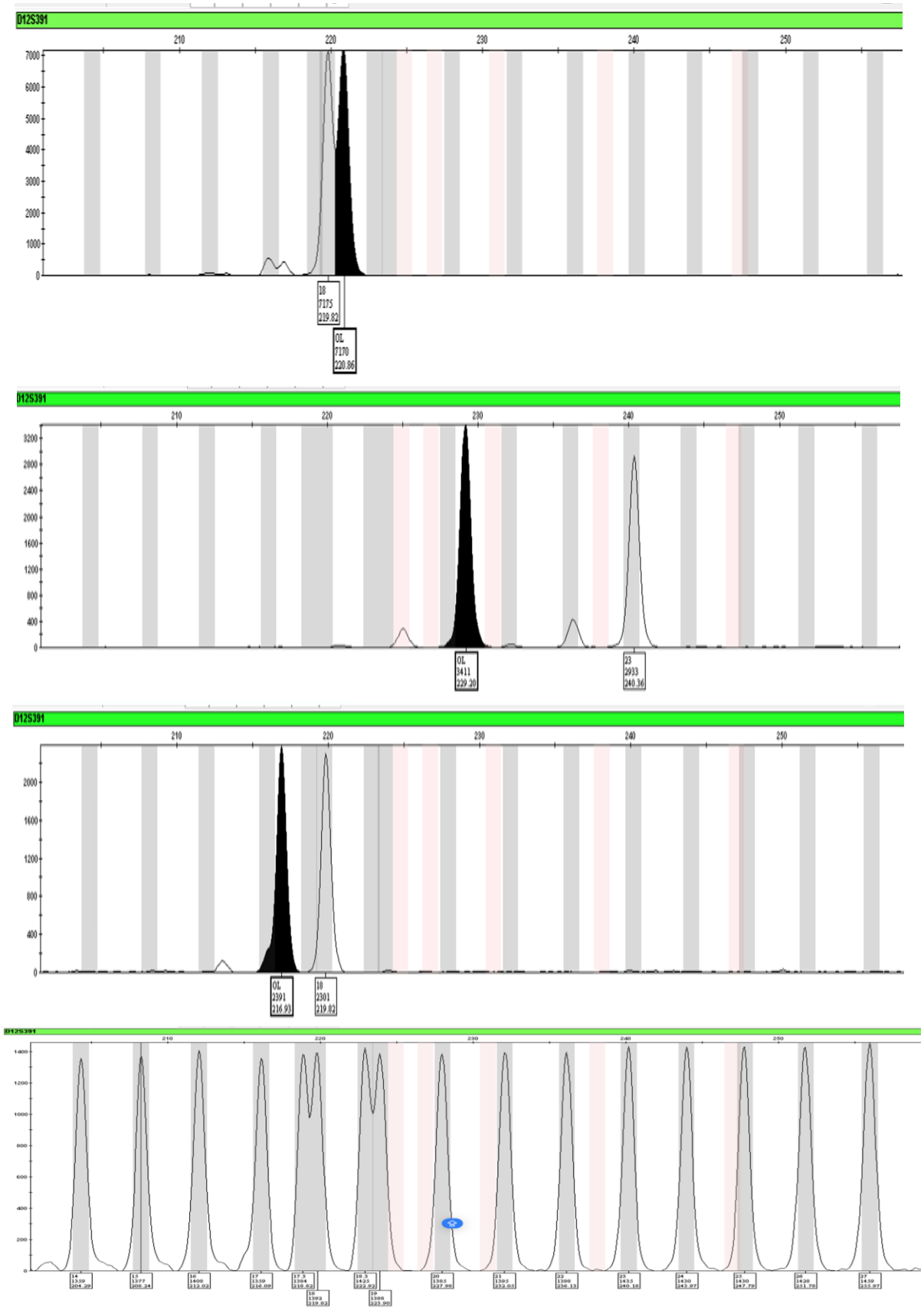



**Figure 3.2. D12S391 off-ladder alleles observed on the electropherogram from the QIAGEN™ Investigator ESSplex SE QS Kit. The three off-ladder (OL) alleles are indicated shaded in black and compared to the allelic ladder of the associated locus in the lower panel.**

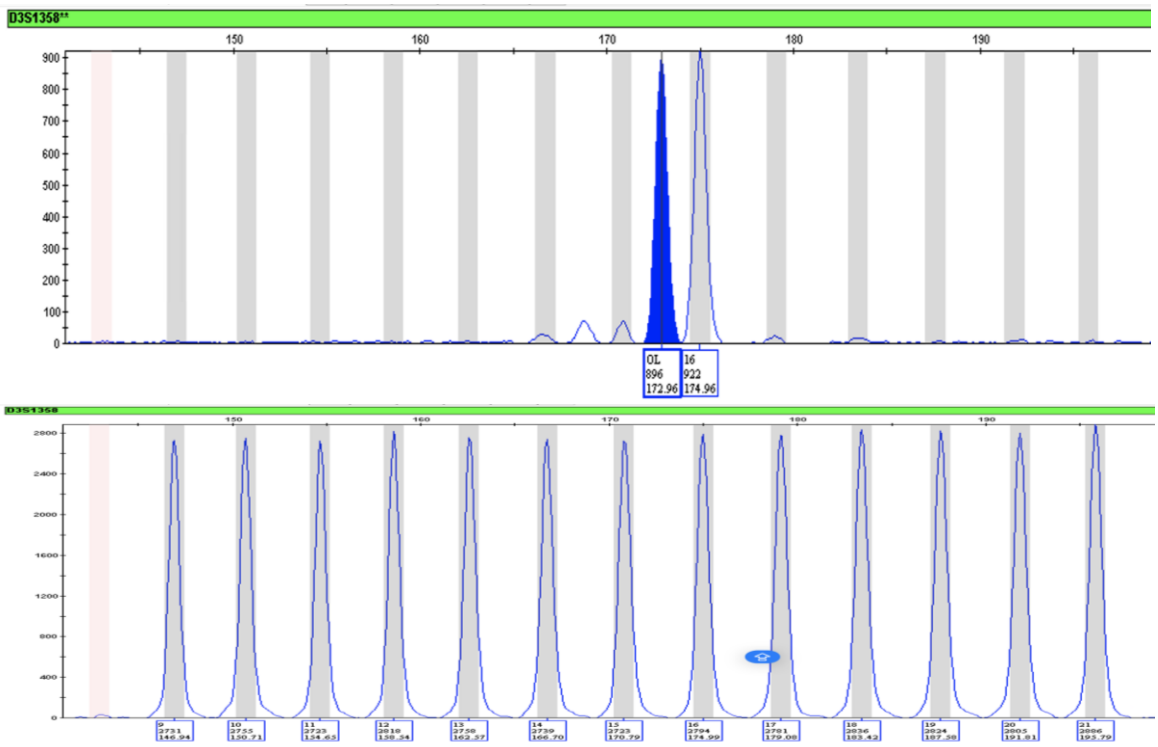



**Figure 3.3. D3S1358 off-ladder alleles observed on the electropherogram from the QIAGEN™ Investigator ESSplex SE QS Kit. The off-ladder (OL) allele is indicated shaded in blue and compared to the allelic ladder of the associated locus in the lower panel.**

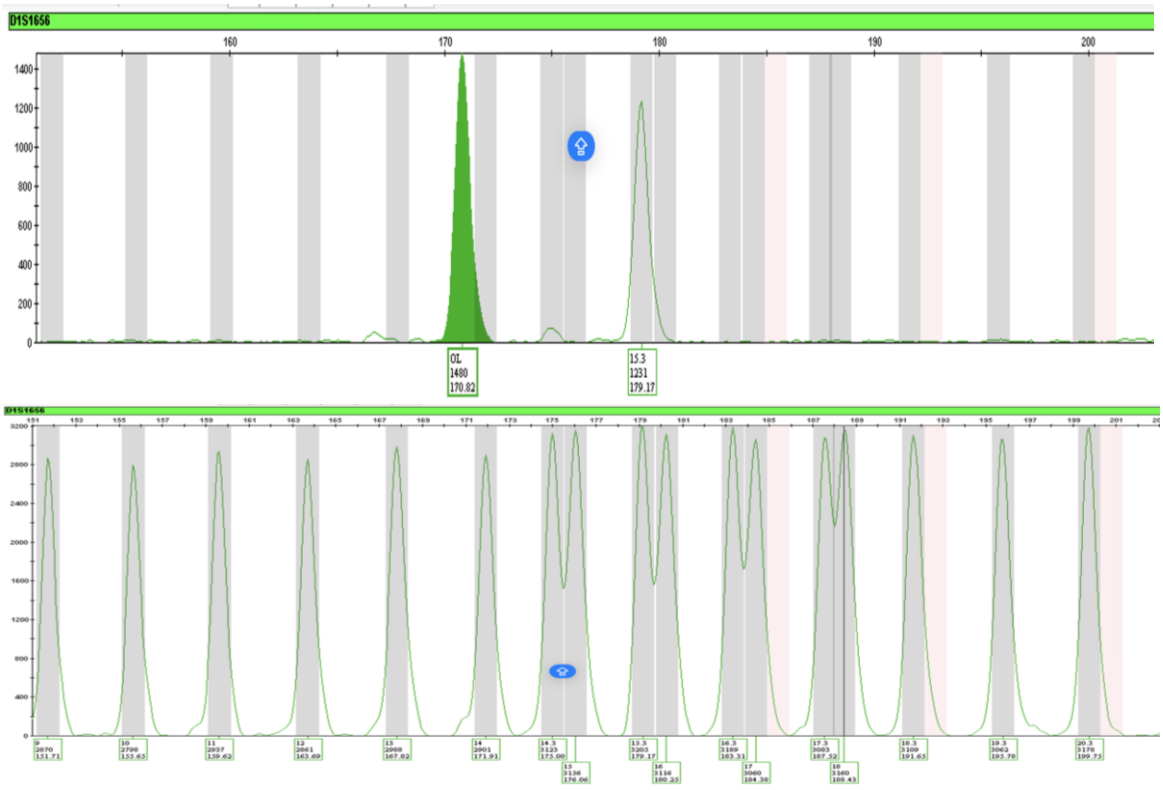



**Figure 3.4. D1S1656 off-ladder alleles observed on the electropherogram from the QIAGEN™ Investigator ESSplex SE QS Kit. The off-ladder (OL) allele is indicated shaded in green and compared to the allelic ladder of the associated locus in the lower panel.**

### 3.4.3 ALLELE FREQUENCIES AND FORENSIC PARAMETERS

Complete STR profiles were generated from blood samples on FTA cards collected from 303 unrelated Nigerians using the QIAGEN™ Investigator ESSplex SE QS Kit except for the three samples with off-ladder alleles. The off-ladder loci detected using the QIAGEN™ kit were genotyped by referencing the corresponding allele calls from the GlobalFiler™ Express Kit. The allele frequencies and forensic parameters were calculated based on the total collections of alleles surveyed in the samples collected from the Nigerian population and separately calculated for the sub-populations; Igbo, Yoruba, and Hausa-Fulani. **Table 3.1** and **Table 3.2** show the allele frequencies and forensic parameters, respectively, calculated for 303 unrelated individuals of the Nigerian population for the total sample with all individuals from all three sub-populations pooled. Among the 303 Nigerian population were 102, 101 and 100 Igbos, Yorubas and Hausa-Fulanis subpopulations, respectively. **Tables 3.3** and **Table 3.4, Tables 3.5** and **Table 3.6, and Tables 3.7** and **Table 3.8** display the allele frequencies and forensic parameters generated for the Igbo, Yoruba, and Hausa-Fulani sub-populations, respectively. **Table 3.9** presents the private alleles observed among the sub-populations.

The two software programs, STR Analysis for Forensic (STRAF-A 2.15), an online tool for analysing STR data, and Genetic Analysis in Excel (GenAIEx 6.5), used for statistical data analysis, facilitated quicker and more practical outputs. These software tools also generated graphical representations (see **Appendices 4 - 19**) after calculating various allele frequencies and forensic parameters.

**Table 3.1. Distribution of allele frequencies across 16 autosomal STR loci in the QIAGEN™ Investigator ESSplex SE QS Kit, analysed within 303 individuals from the Nigerian population**

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | | | | | | | | | | | | | | | 0.127 | |
| **7** | | | | | | | | | | | | | | | 0.493 | |
| **8** | 0.003 | | | 0.020 | | | | 0.005 | | | | | | | 0.203 | |
| **9** | 0.002 | | | 0.249 | 0.002 | 0.002 | | | | | 0.003 | | | | 0.134 | |
| **9.3** | | | | | | | | | | | | | | | 0.041 | |
| **10** | 0.018 | | 0.015 | 0.122 | | 0.010 | | 0.048 | | 0.066 | | 0.015 | | | 0.002 | |
| **10.2** | | | | | 0.003 | | | | | | | | | 0.003 | | |
| **11** | 0.040 | | 0.069 | 0.302 | 0.008 | 0.089 | | 0.158 | | 0.403 | | 0.036 | | | | |
| **11.3** | | | | | | | | | | 0.043 | | | | | | |
| **12** | 0.185 | | 0.041 | 0.153 | 0.051 | 0.130 | | 0.053 | | 0.172 | 0.010 | 0.119 | | 0.005 | | |
| **12.2** | | | | | | 0.048 | | | | | | | | 0.008 | | |
| **12.3** | | | | | | | | | | 0.005 | | | | | | |
| **13** | 0.224 | | 0.142 | 0.145 | 0.020 | 0.259 | | | | 0.041 | 0.003 | 0.195 | | 0.003 | | 0.026 |
| **13.2** | | | | | 0.018 | 0.066 | | | | | | | | | | |
| **14** | 0.256 | | 0.254 | 0.008 | 0.040 | 0.177 | | 0.096 | | 0.239 | 0.111 | 0.366 | | 0.046 | | 0.071 |
| **14.2** | | | | | 0.002 | 0.079 | | | | | | | | 0.005 | | |
| **14.3** | | | 0.002 | | | | | | | | | | | 0.002 | | |
| **15** | 0.173 | 0.076 | 0.175 | | 0.188 | 0.041 | | 0.213 | 0.002 | 0.036 | 0.274 | 0.228 | | 0.035 | | 0.188 |
| **15.2** | | | | | | 0.053 | | | | | | | | 0.005 | | |
| **15.3** | | | 0.008 | | | | | | | | | | | | | |
| **16** | 0.086 | 0.054 | 0.114 | | 0.200 | 0.007 | | 0.168 | 0.073 | | 0.318 | 0.031 | | 0.079 | | 0.290 |
| **16.1** | | | | | | | | | | | | | 0.002 | | | |
| **16.2** | | | | | | 0.038 | | | | | | | | | | |
| **16.3** | | | 0.104 | | | | | | | | | | | | | |
| **17** | 0.010 | 0.134 | 0.015 | | 0.170 | | | 0.221 | 0.092 | | 0.229 | 0.010 | 0.002 | 0.099 | | 0.186 |
| **17.2** | | | | | | 0.002 | | | | | | | | | | |
| **17.3** | | 0.003 | 0.031 | | | | | | | | | | | | | |
| **18.2** | | | | | | | | | | | | | 0.021 | | | |
| **18.3** | | 0.002 | 0.015 | | | | | | | | | | | | | |
| **19** | 0.002 | 0.193 | 0.002 | | 0.116 | | | 0.003 | 0.191 | | 0.002 | | 0.066 | 0.140 | | 0.074 |
| **19.1** | | 0.008 | | | | | | | | | | | | | | |
| **19.2** | | | | | 0.002 | | | | | | | | 0.010 | 0.003 | | |
| **19.3** | | | 0.012 | | | | | | | | | | | | | |
| **20** | | 0.135 | | | 0.040 | | | | 0.066 | | | | 0.040 | 0.109 | | 0.033 |

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **20.1** | | | | | | | | | | | | | | 0.002 | | |
| **20.2** | | | | | | | | | | | | | 0.005 | 0.005 | | |
| **21** | | 0.045 | | | 0.026 | | | | 0.150 | | | | 0.112 | 0.084 | | 0.005 |
| **21.2** | | | | | | | | | | | | | | 0.002 | | |
| **22** | | 0.018 | | | 0.008 | | | | 0.167 | | | | 0.216 | 0.025 | | 0.002 |
| **22.2** | | | | | | | | | | | | | 0.002 | 0.008 | | |
| **23** | | 0.021 | | | 0.002 | | | | 0.078 | | | | 0.172 | | | |
| **23.2** | | | | | | | | | | | | | | 0.008 | | |
| **24** | | 0.003 | | | | | | | 0.066 | | | | 0.168 | | | |
| **24.2** | | | | | | | | | | | | | | 0.010 | | |
| **25** | | 0.002 | | | | | | | 0.064 | | | | 0.073 | 0.003 | | |
| **25.2** | | | | | | | | | | | | | | 0.025 | | |
| **26** | | | | | | | 0.005 | | 0.015 | | | | 0.045 | | | |
| **26.2** | | | | | | | | | | | | | | 0.058 | | |
| **27** | | | | | | | 0.079 | | 0.007 | | | | 0.030 | 0.002 | | |
| **27.2** | | | | | | | | | | | | | | 0.050 | | |
| **28** | | | | | | | 0.297 | | | | | | 0.020 | 0.002 | | |
| **28.2** | | | | | | | | | | | | | | 0.040 | | |
| **29** | | | | | | | 0.175 | | | | | | 0.005 | | | |
| **29.2** | | | | | | | | | | | | | | 0.017 | | |
| **30** | | | | | | | 0.117 | | | | | | | | | |
| **30.2** | | | | | | | 0.031 | | | | | | | 0.017 | | |
| **31** | | | | | | | 0.079 | | | | | | | | | |
| **31.2** | | | | | | | 0.050 | | | | | | 0.002 | 0.002 | | |
| **32** | | | | | | | 0.015 | | | | | | | | | |
| **32.2** | | | | | | | 0.053 | | | | | | | | | |
| **33** | | | | | | | 0.007 | | | | | | | | | |
| **33.2** | | | | | | | 0.031 | | | | | | | 0.002 | | |
| **34** | | | | | | | 0.015 | | | | | | | | | |
| **34.2** | | | | | | | 0.002 | | | | | | | | | |
| **35** | | | | | | | 0.038 | | | | | | | | | |
| **36** | | | | | | | 0.007 | | | | | | | | | |

**Table 3.1. continued**

**Table 3.2. Forensic parameters across 303 Nigerian individuals typed at 16 autosomal STR loci using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_{exp}$ | $H_o$ | HWE | RMP | PIC | PD | PE | TPI | $F_{ST}$ | $F_{IT}$ | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **D10S1248** | 303 | 606 | 12 | 5.276 | 0.828 | 0.812 | 0.188 | 0.764 | 0.067 | 0.784 | 0.933 | 0.653 | 2.913 | 0.004 | -0.022 | -0.026 |
| **D12S391** | 303 | 606 | 14 | 5.352 | 0.835 | 0.819 | 0.181 | 0.839 | 0.057 | 0.797 | 0.943 | 0.665 | 3.030 | 0.002 | -0.027 | -0.029 |
| **D1S1656** | 303 | 606 | 16 | 5.680 | 0.858 | 0.854 | 0.146 | 0.444 | 0.041 | 0.836 | 0.959 | 0.711 | 3.523 | 0.004 | -0.042 | -0.045 |
| **D16S539** | 303 | 606 | 7 | 4.682 | 0.762 | 0.788 | 0.212 | 0.998 | 0.077 | 0.754 | 0.923 | 0.531 | 2.104 | 0.005 | 0.030 | 0.025 |
| **D18S51** | 303 | 606 | 18 | 7.290 | 0.884 | 0.865 | 0.135 | 0.829 | 0.035 | 0.850 | 0.965 | 0.764 | 4.329 | 0.004 | -0.025 | -0.029 |
| **D19S433** | 303 | 606 | 14 | 4.510 | 0.848 | 0.859 | 0.141 | 0.443 | 0.036 | 0.843 | 0.964 | 0.691 | 3.293 | 0.002 | -0.090 | -0.092 |
| **D21S11** | 303 | 606 | 16 | 5.830 | 0.848 | 0.847 | 0.153 | 0.063 | 0.046 | 0.831 | 0.954 | 0.691 | 3.293 | 0.006 | -0.024 | -0.029 |
| **D22S1045** | 303 | 606 | 10 | 6.131 | 0.838 | 0.838 | 0.162 | 0.786 | 0.050 | 0.816 | 0.950 | 0.672 | 3.092 | 0.004 | -0.002 | -0.006 |
| **D2S1338** | 303 | 606 | 13 | 8.307 | 0.871 | 0.881 | 0.119 | 0.098 | 0.031 | 0.868 | 0.969 | 0.737 | 3.885 | 0.002 | 0.010 | 0.006 |
| **D2S441** | 303 | 606 | 9 | 3.418 | 0.736 | 0.745 | 0.255 | 0.887 | 0.101 | 0.708 | 0.899 | 0.486 | 1.894 | 0.002 | -0.041 | -0.043 |
| **D3S1358** | 303 | 606 | 9 | 4.100 | 0.759 | 0.757 | 0.243 | 0.995 | 0.105 | 0.715 | 0.895 | 0.525 | 2.075 | 0.004 | -0.004 | -0.007 |
| **D8S1179** | 303 | 606 | 8 | 4.154 | 0.776 | 0.761 | 0.239 | 0.989 | 0.094 | 0.724 | 0.906 | 0.555 | 2.228 | 0.004 | -0.022 | -0.026 |
| **FGA** | 303 | 606 | 19 | 7.395 | 0.855 | 0.869 | 0.131 | 0.459 | 0.034 | 0.854 | 0.966 | 0.704 | 3.443 | 0.005 | 0.011 | 0.007 |
| **SE33** | 303 | 606 | 34 | 12.151 | 0.921 | 0.924 | 0.076 | 0.152 | 0.015 | 0.917 | 0.985 | 0.838 | 6.313 | 0.003 | -0.004 | -0.007 |
| **TH01** | 303 | 606 | 6 | 3.027 | 0.700 | 0.682 | 0.318 | 0.977 | 0.145 | 0.640 | 0.855 | 0.428 | 1.665 | 0.003 | -0.045 | -0.048 |
| **vWA** | 303 | 606 | 10 | 5.489 | 0.812 | 0.819 | 0.181 | 0.784 | 0.059 | 0.794 | 0.941 | 0.621 | 2.658 | 0.003 | 0.007 | 0.004 |

N - Number of unrelated individuals
Allele (n) - Number of alleles surveyed
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
$H_o$ - Observed heterozygosity
$H_{exp}$ - Expected heterozygosity
$H_{om\_exp}$ - Expected homozygosity
RMP - Random Match Probability
HWE - Hardy-Weinberg Equilibrium probability
PIC - Polymorphism Information Content
PD - Power of Discrimination
PE - Power of Exclusion
TPI - Typical Paternity Index
$F_{ST}$ - Fixation Index - **S**ubpopulation within the **T**otal population
$F_{IT}$ - Fixation Index - **I**ndividual within the **T**otal population
$F_{IS}$ - Fixation Index - **I**ndividual within the **S**ubpopulation

**Table 3.3. Distribution of allele frequencies across 16 autosomal STR loci in the QIAGEN™ Investigator ESSplex SE QS Kit, analysed within 102 individuals from the Igbo subpopulation**

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | | | | | | | | | | | | | | | 0.127 | |
| **7** | | | | | | | | | | | | | | | 0.495 | |
| **8** | 0.010 | | | 0.015 | | | | 0.005 | | | | | | | 0.225 | |
| **9** | | | | 0.201 | 0.005 | 0.005 | | | | | | | | | 0.113 | |
| **9.3** | | | | | | | | | | | | | | | 0.034 | |
| **10** | 0.010 | | 0.010 | 0.132 | | | | 0.044 | | 0.049 | | 0.010 | | | 0.005 | |
| **10.2** | | | | | 0.005 | | | | | | | | | | | |
| **11** | 0.020 | | 0.054 | 0.358 | 0.010 | 0.113 | | 0.191 | | 0.402 | | 0.034 | | | | |
| **11.3** | | | | | | | | | | 0.029 | | | | | | |
| **12** | 0.157 | | 0.044 | 0.162 | 0.044 | 0.118 | | 0.059 | | 0.191 | | 0.113 | | 0.020 | | |
| **12.2** | | | | | | 0.054 | | | | | | | | | | |
| **12.3** | | | | | | | | | | 0.010 | | | | | | |
| **13** | 0.221 | | 0.113 | 0.123 | 0.020 | 0.240 | | | | 0.044 | 0.005 | 0.167 | | 0.010 | | 0.034 |
| **13.2** | | | | | 0.039 | 0.059 | | | | | | | | | | |
| **14** | 0.294 | | 0.294 | 0.010 | 0.034 | 0.181 | | 0.088 | | 0.240 | 0.118 | 0.417 | | 0.054 | | 0.088 |
| **14.2** | | | | | 0.005 | 0.083 | | | | | | | | 0.005 | | |
| **15** | 0.201 | 0.059 | 0.162 | | 0.186 | 0.020 | | 0.206 | | 0.044 | 0.279 | 0.221 | | 0.054 | | 0.157 |
| **15.2** | | | | | | 0.059 | | | | | | | | | | |
| **15.3** | | | 0.010 | | | | | | | | | | | | | |
| **16** | 0.064 | 0.059 | 0.132 | | 0.245 | | | 0.157 | 0.074 | | 0.348 | 0.029 | | 0.059 | | 0.333 |
| **16.2** | | | | | | 0.069 | | | | | | | | | | |
| **16.3** | | | 0.108 | | | | | | | | | | | | | |
| **17** | 0.015 | 0.118 | 0.025 | | 0.147 | | | 0.206 | 0.098 | | 0.225 | 0.010 | | 0.108 | | 0.181 |
| **17.3** | | | 0.039 | | | | | | | | | | | | | |
| **18** | 0.005 | 0.324 | | | 0.108 | | | 0.044 | 0.029 | | 0.020 | | | 0.108 | | 0.103 |
| **18.2** | | | | | | | | | | | | | 0.025 | | | |
| **18.3** | | | 0.005 | | | | | | | | | | | | | |
| **19** | 0.005 | 0.216 | | | 0.093 | | | | 0.181 | | 0.005 | | 0.113 | 0.157 | | 0.069 |
| **19.1** | | 0.004 | | | | | | | | | | | | | | |
| **19.2** | | | | | 0.002 | | | | | | | | 0.005 | 0.005 | | |
| **19.3** | | | 0.005 | | | | | | | | | | | | | |
| **20** | | 0.147 | | | 0.029 | | | | 0.059 | | | | 0.059 | 0.078 | | 0.029 |
| **20.1** | | | | | | | | | | | | | 0.002 | | | |
| **20.2** | | | | | | | | | | | | | 0.005 | 0.010 | | |
| **21** | | 0.044 | | | 0.029 | | | | 0.172 | | | | 0.093 | 0.078 | | 0.005 |
| **22** | | 0.025 | | | | | | | 0.142 | | | | 0.230 | 0.020 | | |

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **22.2** | | | | | | | | | | | | | | 0.010 | | |
| **23** | | 0.010 | | | | | | | 0.074 | | | | 0.176 | | | |
| **23.2** | | | | | | | | | | | | | | 0.010 | | |
| **24** | | | | | | | | | 0.093 | | | | 0.167 | | | |
| **24.2** | | | | | | | | | | | | | | 0.005 | | |
| **25** | | | | | | | | | 0.049 | | | | 0.069 | | | |
| **25.2** | | | | | | | | | | | | | | 0.029 | | |
| **26** | | | | | | | | | 0.025 | | | | 0.020 | | | |
| **26.2** | | | | | | | | | | | | | | 0.059 | | |
| **27** | | | | | | | 0.069 | | 0.005 | | | | 0.025 | | | |
| **27.2** | | | | | | | | | | | | | | 0.069 | | |
| **28** | | | | | | | 0.319 | | | | | | 0.010 | | | |
| **28.2** | | | | | | | | | | | | | | 0.034 | | |
| **29** | | | | | | | 0.147 | | | | | | 0.005 | | | |
| **29.2** | | | | | | | | | | | | | | 0.005 | | |
| **30** | | | | | | | 0.137 | | | | | | | | | |
| **30.2** | | | | | | | 0.025 | | | | | | | 0.015 | | |
| **31** | | | | | | | 0.098 | | | | | | | | | |
| **31.2** | | | | | | | 0.044 | | | | | | | | | |
| **32** | | | | | | | 0.010 | | | | | | | | | |
| **32.2** | | | | | | | 0.039 | | | | | | | | | |
| **33** | | | | | | | 0.005 | | | | | | | | | |
| **33.2** | | | | | | | 0.025 | | | | | | | | | |
| **34** | | | | | | | 0.015 | | | | | | | | | |
| **35** | | | | | | | 0.064 | | | | | | | | | |
| **36** | | | | | | | 0.005 | | | | | | | | | |

**Table 3.3. continued**

**Table 3.4. Forensic parameters across 102 individuals from the Igbo subpopulation typed at 16 autosomal STR loci using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_{exp}$ | $H_o$ | HWE | RMP | PIC | PD | PE | TPI | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **D10S1248** | 102 | 204 | 11 | 4.876 | 0.794 | 0.799 | 0.201 | 0.096 | 0.085 | 0.765 | 0.915 | 0.588 | 2.429 | 0.001 |
| **D12S391** | 102 | 204 | 10 | 5.096 | 0.833 | 0.814 | 0.186 | 0.690 | 0.069 | 0.787 | 0.931 | 0.662 | 3.000 | -0.037 |
| **D1S1656** | 102 | 204 | 13 | 5.118 | 0.824 | 0.842 | 0.158 | 0.241 | 0.052 | 0.821 | 0.948 | 0.643 | 2.833 | -0.024 |
| **D16S539** | 102 | 204 | 7 | 4.335 | 0.725 | 0.773 | 0.227 | 0.604 | 0.087 | 0.736 | 0.913 | 0.469 | 1.821 | 0.057 |
| **D18S51** | 102 | 204 | 16 | 6.872 | 0.843 | 0.861 | 0.139 | 0.243 | 0.043 | 0.842 | 0.957 | 0.681 | 3.188 | 0.013 |
| **D19S433** | 102 | 204 | 11 | 4.706 | 0.873 | 0.865 | 0.135 | 0.995 | 0.036 | 0.847 | 0.964 | 0.740 | 3.923 | -0.108 |
| **D21S11** | 102 | 204 | 14 | 5.500 | 0.833 | 0.839 | 0.161 | 0.054 | 0.052 | 0.818 | 0.948 | 0.662 | 3.000 | -0.019 |
| **D22S1045** | 102 | 204 | 9 | 6.208 | 0.843 | 0.843 | 0.157 | 0.599 | 0.053 | 0.819 | 0.947 | 0.681 | 3.188 | -0.005 |
| **D2S1338** | 102 | 204 | 12 | 8.581 | 0.922 | 0.888 | 0.112 | 0.176 | 0.037 | 0.872 | 0.963 | 0.840 | 6.375 | -0.043 |
| **D2S441** | 102 | 204 | 8 | 3.489 | 0.716 | 0.744 | 0.256 | 0.422 | 0.110 | 0.704 | 0.890 | 0.453 | 1.759 | -0.003 |
| **D3S1358** | 102 | 204 | 7 | 3.783 | 0.745 | 0.739 | 0.261 | 0.870 | 0.121 | 0.689 | 0.879 | 0.501 | 1.962 | -0.013 |
| **D8S1179** | 102 | 204 | 8 | 3.774 | 0.745 | 0.739 | 0.261 | 0.681 | 0.110 | 0.698 | 0.890 | 0.501 | 1.962 | -0.014 |
| **FGA** | 102 | 204 | 14 | 6.897 | 0.814 | 0.861 | 0.139 | 0.227 | 0.045 | 0.841 | 0.955 | 0.625 | 2.684 | 0.048 |
| **SE33** | 102 | 204 | 24 | 11.749 | 0.882 | 0.923 | 0.077 | 0.212 | 0.020 | 0.913 | 0.980 | 0.760 | 4.250 | 0.036 |
| **TH01** | 102 | 204 | 6 | 3.026 | 0.686 | 0.681 | 0.319 | 0.674 | 0.158 | 0.634 | 0.842 | 0.407 | 1.594 | -0.025 |
| **vWA** | 102 | 204 | 9 | 5.161 | 0.843 | 0.810 | 0.19 | 0.330 | 0.070 | 0.783 | 0.930 | 0.681 | 3.188 | -0.046 |

| | | |
|---|---|---|
| N | - | Number of unrelated individuals |
| Allele (n) | - | Number of alleles surveyed |
| Allele (v) | - | Number of different allelic variants |
| Allele (e) | - | Number of effective alleles |
| $H_o$ | - | Observed heterozygosity |
| $H_{exp}$ | - | Expected heterozygosity |
| $H_{om\_exp}$ | - | Expected homozygosity |
| HWE | - | Hardy-Weinberg Equilibrium probability |
| RMP | - | Random Match Probability |
| PIC | - | Polymorphism Information Content |
| PD | - | Power of Discrimination |
| PE | - | Power of Exclusion |
| TPI | - | Typical Paternity Index |
| $F_{IS}$ | - | Fixation Index - **I**ndividual within the **S**ubpopulation |

**Table 3.5. Distribution of allele frequencies across 16 autosomal STR loci in the QIAGEN™ Investigator ESSplex SE QS Kit, analysed within 101 individuals from the Yoruba subpopulation.**

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | | | | | | | | | | | | | | | 0.149 | |
| **7** | | | | | | | | | | | | | | | 0.505 | |
| **8** | | | | 0.020 | | | | 0.010 | | | | | | | 0.193 | |
| **9** | | | | 0.292 | | | | | | | 0.005 | | | | 0.109 | |
| **9.3** | | | | | | | | | | | | | | | 0.045 | |
| **10** | 0.035 | | 0.005 | 0.109 | | 0.020 | | 0.059 | | 0.059 | | 0.020 | | 0.005 | | |
| **11** | 0.050 | | 0.079 | 0.292 | 0.010 | 0.064 | | 0.144 | | 0.391 | | 0.059 | | | | |
| **11.3** | | | | | | | | | | 0.050 | | | | | | |
| **12** | 0.218 | | 0.050 | 0.129 | 0.045 | 0.144 | | 0.054 | | 0.173 | 0.025 | 0.144 | | 0.005 | | |
| **12.2** | | | | | | 0.045 | | | | | | | | | | |
| **12.3** | | | | | | | | | | 0.005 | | | | | | |
| **13** | 0.238 | | 0.153 | 0.153 | 0.020 | 0.282 | | | | 0.025 | | 0.198 | | | | 0.035 |
| **13.2** | | | | | 0.010 | 0.059 | | | | | | | | | | |
| **14** | 0.208 | | 0.248 | 0.005 | 0.025 | 0.153 | | 0.149 | | 0.272 | 0.129 | 0.312 | | 0.040 | | 0.079 |
| **14.2** | | | | | | 0.084 | | | | | | | | 0.005 | | |
| **15** | 0.149 | 0.104 | 0.188 | | 0.198 | 0.059 | | 0.208 | 0.005 | 0.030 | 0.282 | 0.233 | | 0.035 | | 0.193 |
| **15.2** | | | | | | 0.059 | | | | | | | | 0.005 | | |
| **15.3** | | | 0.005 | | | | | | | | | | | | | |
| **16** | 0.099 | 0.030 | 0.099 | | 0.158 | 0.005 | | 0.144 | 0.089 | | 0.287 | 0.025 | | 0.094 | | 0.262 |
| **16.2** | | | | | | 0.025 | | | | | | | | | | |
| **16.3** | | | 0.084 | | | | | | | | | | | | | |
| **17** | 0.005 | 0.153 | 0.010 | | 0.183 | | | 0.203 | 0.059 | | 0.223 | 0.010 | 0.005 | 0.104 | | |
| **17.3** | | | 0.020 | | | | | | | | | | | | | |
| **18** | | 0.292 | 0.005 | | 0.109 | | | 0.025 | 0.020 | | 0.050 | | 0.015 | 0.109 | | 0.119 |
| **18.2** | | | | | | | | | | | | | 0.010 | | | |
| **18.3** | | | 0.035 | | | | | | | | | | | | | |
| **19** | | 0.193 | 0.002 | | 0.158 | | | 0.005 | 0.198 | | | | 0.035 | 0.149 | | 0.079 |
| **19.1** | | 0.004 | | | | | | | | | | | | | | |
| **19.2** | | | | | | | | | | | | | 0.015 | 0.005 | | |
| **19.3** | | | 0.020 | | | | | | | | | | | | | |
| **20** | | 0.144 | | | 0.045 | | | | 0.094 | | | | 0.045 | 0.149 | | 0.040 |
| **20.2** | | | | | | | | | | | | | 0.010 | 0.005 | | |
| **21** | | 0.030 | | | 0.030 | | | | 0.124 | | | | 0.084 | 0.084 | | 0.005 |
| **21.2** | | | | | | | | | | | | | | 0.005 | | |
| **22** | | 0.020 | | | 0.010 | | | | 0.168 | | | | 0.213 | 0.020 | | 0.005 |

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **22.2** | | | | | | | | | | | | | 0.005 | 0.010 | | |
| **23** | | 0.030 | | | | | | | 0.104 | | | | 0.168 | | | |
| **23.2** | | | | | | | | | | | | | | 0.010 | | |
| **24** | | | | | | | | | 0.050 | | | | 0.203 | | | |
| **24.2** | | | | | | | | | | | | | | 0.010 | | |
| **25** | | 0.005 | | | | | | | 0.074 | | | | 0.059 | 0.005 | | |
| **26** | | | | | | | 0.010 | | 0.010 | | | | 0.069 | 0.005 | | |
| **26.2** | | | | | | | | | | | | | | 0.050 | | |
| **27** | | | | | | | 0.109 | | 0.005 | | | | 0.035 | 0.005 | | |
| **27.2** | | | | | | | | | | | | | | 0.025 | | |
| **28** | | | | | | | 0.267 | | | | | | 0.020 | | | |
| **28.2** | | | | | | | | | | | | | | 0.020 | | |
| **29** | | | | | | | 0.223 | | | | | | 0.010 | | | |
| **29.2** | | | | | | | | | | | | | | 0.030 | | |
| **30** | | | | | | | 0.064 | | | | | | | | | |
| **30.2** | | | | | | | 0.030 | | | | | | | 0.015 | | |
| **31** | | | | | | | 0.059 | | | | | | | | | |
| **31.2** | | | | | | | 0.069 | | | | | | | | | |
| **32** | | | | | | | 0.010 | | | | | | | | | |
| **32.2** | | | | | | | 0.079 | | | | | | | | | |
| **33** | | | | | | | 0.005 | | | | | | | | | |
| **33.2** | | | | | | | 0.020 | | | | | | | | | |
| **34** | | | | | | | 0.015 | | | | | | | | | |
| **35** | | | | | | | 0.030 | | | | | | | | | |
| **36** | | | | | | | 0.010 | | | | | | | | | |

**Table 3.5. continued**

**Table 3.6. Forensic parameters across 101 individuals from the Yoruba subpopulation typed at 16 autosomal STR loci using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_{exp}$ | $H_o$ | HWE | RMP | PIC | PD | PE | PI | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **D10S1248** | 101 | 202 | 8 | 5.433 | 0.842 | 0.820 | 0.18 | 0.882 | 0.065 | 0.790 | 0.935 | 0.678 | 3.156 | -0.031 |
| **D12S391** | 101 | 202 | 11 | 5.441 | 0.812 | 0.824 | 0.176 | 0.518 | 0.058 | 0.798 | 0.942 | 0.621 | 2.658 | 0.005 |
| **D1S1656** | 101 | 202 | 15 | 5.943 | 0.842 | 0.855 | 0.145 | 0.154 | 0.047 | 0.834 | 0.953 | 0.678 | 3.156 | -0.012 |
| **D16S539** | 101 | 202 | 7 | 4.484 | 0.752 | 0.781 | 0.219 | 0.408 | 0.090 | 0.743 | 0.910 | 0.514 | 2.020 | 0.032 |
| **D18S51** | 101 | 202 | 13 | 7.074 | 0.891 | 0.863 | 0.137 | 0.925 | 0.042 | 0.843 | 0.958 | 0.777 | 4.591 | -0.038 |
| **D19S433** | 101 | 202 | 12 | 4.385 | 0.851 | 0.856 | 0.144 | 0.400 | 0.043 | 0.837 | 0.957 | 0.698 | 3.367 | -0.103 |
| **D21S11** | 101 | 202 | 15 | 5.945 | 0.931 | 0.850 | 0.15 | 0.010 | 0.067 | 0.830 | 0.933 | 0.858 | 7.214 | -0.119 |
| **D22S1045** | 101 | 202 | 10 | 6.454 | 0.881 | 0.849 | 0.151 | 0.091 | 0.059 | 0.826 | 0.941 | 0.757 | 4.208 | -0.043 |
| **D2S1338** | 101 | 202 | 13 | 8.093 | 0.901 | 0.883 | 0.117 | 0.390 | 0.037 | 0.866 | 0.963 | 0.797 | 5.050 | -0.028 |
| **D2S441** | 101 | 202 | 8 | 3.297 | 0.703 | 0.742 | 0.258 | 0.904 | 0.102 | 0.700 | 0.898 | 0.433 | 1.683 | -0.009 |
| **D3S1358** | 101 | 202 | 7 | 4.322 | 0.733 | 0.772 | 0.228 | 0.586 | 0.094 | 0.731 | 0.906 | 0.481 | 1.870 | 0.047 |
| **D8S1179** | 101 | 202 | 8 | 4.633 | 0.802 | 0.788 | 0.212 | 0.955 | 0.081 | 0.752 | 0.919 | 0.603 | 2.525 | -0.023 |
| **FGA** | 101 | 202 | 17 | 7.136 | 0.871 | 0.868 | 0.132 | 0.579 | 0.041 | 0.850 | 0.959 | 0.737 | 3.885 | -0.013 |
| **SE33** | 101 | 202 | 27 | 10.351 | 0.921 | 0.914 | 0.086 | 0.401 | 0.024 | 0.902 | 0.976 | 0.838 | 6.313 | -0.019 |
| **TH01** | 101 | 202 | 5 | 2.960 | 0.683 | 0.675 | 0.325 | 0.897 | 0.150 | 0.631 | 0.850 | 0.403 | 1.578 | -0.032 |
| **vWA** | 101 | 202 | 10 | 5.912 | 0.752 | 0.835 | 0.165 | 0.238 | 0.054 | 0.810 | 0.946 | 0.514 | 2.020 | 0.094 |

N - Number of unrelated individuals
Allele (n) - Number of alleles surveyed
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
$H_o$ - Observed heterozygosity
$H_{exp}$ - Expected heterozygosity
$H_{om\_exp}$ - Expected homozygosity
HWE - Hardy-Weinberg Equilibrium probability
RMP - Random Match Probability
PIC - Polymorphism Information Content
PD - Power of Discrimination
PE - Power of Exclusion
TPI - Typical Paternity Index
$F_{IS}$ - Fixation Index - **I**ndividual within the **S**ubpopulation

**Table 3.7. Distribution of allele frequencies across 16 autosomal STR loci in the QIAGEN™ Investigator ESSplex SE QS Kit, analysed within 100 individuals from the Hausa-Fulani subpopulation sample.**

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | | | | | | | | | | | | | | | 0.105 | |
| **7** | | | | | | | | | | | | | | | 0.48 | |
| **8** | | | | 0.025 | | | | | | | | | | | 0.19 | |
| **9** | 0.005 | | | 0.255 | | | | | | | 0.005 | | | | 0.18 | |
| **9.3** | | | | | | | | | | | | | | | 0.045 | |
| **10** | 0.01 | | 0.03 | 0.125 | | 0.01 | | 0.04 | | 0.09 | | 0.015 | | | | |
| **10.2** | | | | | 0.005 | | | | | | | | | 0.005 | | |
| **11** | 0.05 | | 0.075 | 0.255 | 0.005 | 0.09 | | 0.14 | | 0.415 | | 0.015 | | | | |
| **11.3** | | | | | | | | | | 0.05 | | | | | | |
| **12** | 0.18 | | 0.03 | 0.17 | 0.065 | 0.13 | | 0.045 | | 0.15 | 0.005 | 0.1 | | 0.005 | | |
| **12.2** | | | | | | 0.045 | | | | | | | | 0.01 | | |
| **13** | 0.215 | | 0.16 | 0.16 | 0.02 | 0.255 | | | | 0.055 | 0.005 | 0.22 | | | | 0.01 |
| **13.2** | | | | | 0.005 | 0.08 | | | | | | | | | | |
| **13.3** | | | | | | | | | | 0.005 | | | | | | |
| **14** | 0.265 | | 0.22 | 0.01 | 0.06 | 0.195 | | 0.05 | | 0.205 | 0.085 | 0.37 | | 0.045 | | 0.045 |
| **14.2** | | | | | | 0.07 | | | | | | | | 0.01 | | |
| **14.3** | | | 0.005 | | | | | | | | | | | 0.005 | | |
| **15** | 0.17 | 0.065 | 0.175 | | 0.18 | 0.045 | | 0.225 | | 0.035 | 0.26 | 0.23 | | 0.015 | | 0.215 |
| **15.2** | | | | | | 0.04 | | | | | | | | 0.01 | | |
| **15.3** | | | 0.01 | | | | | | | | | | | | | |
| **16** | 0.095 | 0.075 | 0.11 | | 0.195 | 0.015 | | 0.205 | 0.055 | | 0.32 | 0.04 | | 0.085 | | 0.275 |
| **16.1** | | | | | | | | | | | | | 0.005 | | | |
| **16.2** | | | | | | 0.02 | | | | | | | | | | |
| **16.3** | | | 0.12 | | | | | | | | | | | | | |
| **17** | 0.01 | 0.13 | 0.01 | | 0.18 | | | 0.255 | 0.12 | | 0.24 | 0.01 | | 0.085 | | 0.195 |
| **17.2** | | | | | | 0.005 | | | | | | | | | | |
| **17.3** | | 0.01 | 0.035 | | | | | | | | | | | | | |
| **18** | | 0.325 | 0.005 | | 0.105 | | | 0.035 | 0.04 | | 0.08 | | 0.02 | 0.09 | | 0.15 |
| **18.2** | | | | | | | | | | | | | 0.03 | | | |
| **18.3** | | 0.005 | 0.005 | | | | | | | | | | | | | |
| **19** | | 0.17 | | | 0.095 | | | 0.005 | 0.195 | | | | 0.05 | 0.115 | | 0.075 |
| **19.2** | | | | | | | | | | | | | 0.01 | | | |
| **19.3** | | | 0.01 | | | | | | | | | | | | | |
| **20** | | 0.115 | | | 0.045 | | | | 0.045 | | | | 0.015 | 0.1 | | 0.03 |
| **21** | | 0.06 | | | 0.02 | | | | 0.155 | | | | 0.16 | 0.09 | | 0.005 |

| Allele | D10S1248 | D12S391 | D1S1656 | D16S539 | D18S51 | D19S433 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D8S1179 | FGA | SE33 | TH01 | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **22** | | 0.01 | | | 0.015 | | | | 0.19 | | | | 0.205 | 0.035 | | |
| **22.2** | | | | | | | | | | | | | | 0.005 | | |
| **23** | | 0.025 | | | 0.005 | | | | 0.055 | | | | 0.17 | | | |
| **23.2** | | | | | | | | | | | | | | 0.005 | | |
| **24** | | 0.01 | | | | | | | 0.055 | | | | 0.135 | | | |
| **24.2** | | | | | | | | | | | | | | 0.015 | | |
| **25** | | | | | | | | | 0.07 | | | | 0.09 | 0.005 | | |
| **25.2** | | | | | | | | | | | | | | 0.045 | | |
| **26** | | | | | | | 0.005 | | 0.01 | | | | 0.045 | | | |
| **26.2** | | | | | | | | | | | | | | 0.06 | | |
| **27** | | | | | | | 0.06 | | 0.01 | | | | 0.03 | | | |
| **27.2** | | | | | | | | | | | | | | 0.05 | | |
| **28** | | | | | | | 0.305 | | | | | | 0.03 | 0.005 | | |
| **28.2** | | | | | | | | | | | | | | 0.065 | | |
| **29** | | | | | | | 0.155 | | | | | | | | | |
| **29.2** | | | | | | | | | | | | | | 0.015 | | |
| **30** | | | | | | | 0.15 | | | | | | | | | |
| **30.2** | | | | | | | 0.04 | | | | | | | 0.02 | | |
| **31** | | | | | | | 0.08 | | | | | | | | | |
| **31.2** | | | | | | | 0.035 | | | | | | 0.005 | 0.005 | | |
| **32** | | | | | | | 0.025 | | | | | | | | | |
| **32.2** | | | | | | | 0.04 | | | | | | | | | |
| **33** | | | | | | | 0.01 | | | | | | | | | |
| **33.2** | | | | | | | 0.05 | | | | | | | 0.005 | | |
| **34** | | | | | | | 0.015 | | | | | | | | | |
| **34.2** | | | | | | | 0.005 | | | | | | | | | |
| **35** | | | | | | | 0.02 | | | | | | | | | |
| **36** | | | | | | | 0.005 | | | | | | | | | |

**Table 3.7. continued**

**Table 3.8. Forensic parameters across 100 individuals from the Hausa-Fulani subpopulation typed at 16 autosomal STR loci using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_e$ | $H_o$ | HWE | RMP | PIC | PD | PE | TPI | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **D10S1248** | 100 | 200 | 9 | 5.323 | 0.850 | 0.816 | 0.184 | 0.830 | 0.069 | 0.786 | 0.931 | 0.695 | 3.333 | -0.047 |
| **D12S391** | 100 | 200 | 12 | 5.379 | 0.860 | 0.824 | 0.176 | 0.882 | 0.058 | 0.800 | 0.942 | 0.715 | 3.571 | -0.056 |
| **D1S1656** | 100 | 200 | 16 | 5.780 | 0.910 | 0.864 | 0.136 | 0.269 | 0.047 | 0.845 | 0.953 | 0.816 | 5.556 | -0.100 |
| **D16S539** | 100 | 200 | 7 | 5.009 | 0.810 | 0.806 | 0.194 | 0.911 | 0.073 | 0.772 | 0.927 | 0.618 | 2.632 | -0.012 |
| **D18S51** | 100 | 200 | 15 | 7.460 | 0.920 | 0.871 | 0.129 | 0.733 | 0.042 | 0.852 | 0.958 | 0.836 | 6.250 | -0.062 |
| **D19S433** | 100 | 200 | 13 | 4.353 | 0.820 | 0.859 | 0.141 | 0.177 | 0.045 | 0.839 | 0.955 | 0.637 | 2.778 | -0.065 |
| **D21S11** | 100 | 200 | 17 | 5.598 | 0.780 | 0.847 | 0.153 | 0.024 | 0.051 | 0.828 | 0.949 | 0.562 | 2.273 | 0.050 |
| **D22S1045** | 100 | 200 | 9 | 5.416 | 0.790 | 0.819 | 0.181 | 0.495 | 0.063 | 0.790 | 0.937 | 0.581 | 2.381 | 0.031 |
| **D2S1338** | 100 | 200 | 12 | 7.590 | 0.790 | 0.873 | 0.127 | 0.034 | 0.044 | 0.856 | 0.956 | 0.581 | 2.381 | 0.090 |
| **D2S441** | 100 | 200 | 8 | 3.412 | 0.790 | 0.753 | 0.247 | 0.877 | 0.098 | 0.718 | 0.902 | 0.581 | 2.381 | -0.118 |
| **D3S1358** | 100 | 200 | 8 | 4.144 | 0.800 | 0.763 | 0.237 | 0.544 | 0.115 | 0.719 | 0.885 | 0.599 | 2.500 | -0.054 |
| **D8S1179** | 100 | 200 | 8 | 3.994 | 0.780 | 0.753 | 0.247 | 0.991 | 0.104 | 0.711 | 0.896 | 0.562 | 2.273 | -0.040 |
| **FGA** | 100 | 200 | 15 | 7.516 | 0.880 | 0.874 | 0.126 | 0.614 | 0.037 | 0.855 | 0.963 | 0.755 | 4.167 | -0.015 |
| **SE33** | 100 | 200 | 28 | 13.450 | 0.960 | 0.934 | 0.066 | 0.943 | 0.017 | 0.925 | 0.983 | 0.919 | 12.500 | -0.037 |
| **TH01** | 100 | 200 | 5 | 3.047 | 0.730 | 0.692 | 0.308 | 0.992 | 0.140 | 0.646 | 0.860 | 0.476 | 1.852 | -0.087 |
| **vWA** | 100 | 200 | 9 | 5.234 | 0.840 | 0.813 | 0.187 | 0.337 | 0.080 | 0.782 | 0.920 | 0.675 | 3.125 | -0.038 |

N - Number of unrelated individuals
Allele (n) - Number of alleles surveyed
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
$H_o$ - Observed heterozygosity
$H_{exp}$ - Expected heterozygosity
$H_{om\_exp}$ - Expected homozygosity
HWE - Hardy-Weinberg Equilibrium probability
RMP - Random Match Probability
PIC - Polymorphism Information Content
PD - Power of Discrimination
PE - Power of Exclusion
TPI - Typical Paternity Index
$F_{IS}$ - Fixation Index - **I**ndividual within the **S**ubpopulation

**Table 3.9. Summary of private alleles by subpopulation using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Subpopulation/ Ethnic group | Locus | Allele | Frequency |
|---|---|---|---|
| **Igbo (N = 102)** | TH01 | 10 | 0.005 |
| | D3S1358 | 19 | 0.005 |
| | D19S433 | 9 | 0.005 |
| | SE33 | 13 | 0.010 |
| | D10S1248 | 8 | 0.010 |
| | D10S1248 | 18 | 0.005 |
| | D10S1248 | 19 | 0.005 |
| | D18S51 | 9 | 0.005 |
| **Yoruba (N = 101)** | vWA | 22 | 0.005 |
| | FGA | 17 | 0.005 |
| | D12S391 | 25 | 0.005 |
| | D2S1338 | 15 | 0.005 |
| **Hausa-Fulani (N = 100)** | D18S51 | 23 | 0.005 |
| | D19S433 | 17.2 | 0.005 |
| | FGA | 16.1 | 0.005 |
| | FGA | 31,2 | 0.005 |
| | SE33 | 31.2 | 0.005 |
| | SE33 | 33.2 | 0.005 |
| | D10S1248 | 9 | 0.005 |
| | D12S391 | 24 | 0.010 |

The entire Nigerian dataset recorded 63 different alleles across the 16 autosomal loci, with sizes ranging from 6 to 36. Microvariant alleles were detected at the loci TH01, D18S51, SE33, D2S441, D19S433, D1S1656, D3S1358, FGA, D12S391, and D21S11 among the three ethnic groups, with variations in their respective proportions. The allele frequencies across all 303 unrelated individuals in the Nigerian population ranged from 0.002 to 0.493, with allele 7 at the TH01 marker being detected at the highest frequency (0.493) **(Table 3.1)**. Among the loci surveyed in the total Nigerian population, locus SE33 was the most polymorphic (PIC = 91.7%), the most discriminative (PD = 98.5%), followed by D2S1338 (PIC = 86.8%, PD = 96.9%). The SE33 locus also demonstrated the highest Power of Elimination (PE = 83.8%) and the highest number of allelic variants (v = 34). In comparison, D18S51 (PE = 76.4%) and FGA (v = 19) had the second highest Power of Elimination and number of allelic variants, respectively. Locus SE33 may be considered the most informative marker among the 16 loci **(Table 3.2)**. Conversely, TH01 was the least informative with the lowest Polymorphic Information Content (PIC = 64.0%), Power of Elimination (PE = 42.8%), Power of Discrimination (PD = 85.5%) and the smallest number of allelic variants (v = 6). The Typical Paternity Index ranged from 1.665 (TH01) to 6.313 (SE33) across the 16 loci **(Table 3.2)**. The Combined Power of Elimination (CPE) and the Combined Power of Discrimination (CPD) for the

16 loci were higher than 99.9999%. The Combined Match Probability (CMP) for all 16 loci was 4.71 x $10^{-21}$. This conveys high evidential strength, with a likelihood ratio of 1 in 2.12 × $10^{20}$ for finding two individuals with the same profile within the same population. The high discriminatory power and robust forensic parameters confirm that the QIAGEN™ Investigator ESSplex SE QS Kit is effective for conducting forensic DNA analysis in the Nigerian population.

The analysis of 16 autosomal loci in the Igbo dataset identified 59 distinct alleles, with sizes from 6 to 36. The allele frequencies among 102 unrelated individuals ranged from 0.005 to 0.495, with the highest frequency for Allele 7 at the TH01 marker (0.495) (see **Table 3.3**). Locus SE33 was the most polymorphic (PIC = 91.3%) and discriminative (PD = 98.0%), followed by D2S1338 (PIC = 87.2%, PD = 96.3%). SE33 also had the highest number of allelic variants (v = 24) and Power of Elimination of 76.0% (**Table 3.4**). In contrast, TH01 was the least polymorphic and discriminative (PIC = 63.1%, PD = 84.2%) and had the lowest Power of Elimination (PE = 40.7%) with the smallest number of allelic variants (v = 6). The Typical Paternity Index (TPI) ranged from 1.594 (TH01) to 6.375 (D2S1338) (**Table 3.4**). The Combined Power of Exclusion (CPE) and Discrimination (CPD) for all loci exceeded 99.9999%, and the Combined Match Probability (CMP) was 6.10 x $10^{-20}$, indicating an extremely low chance of two individuals sharing the same profile (L.R = 1 in 1.64 x $10^{19}$).

Considering the Yoruba subpopulation distinctly, the dataset showed 58 distinct alleles, ranging from 6 to 36. The allele frequencies among 101 unrelated individuals ranged from 0.005 to 0.505, with Allele 7 at the TH01 marker being the most frequent (**Table 3.5**). Locus SE33 was the most polymorphic (PIC = 90.2%) and discriminative (PD = 97.6%), followed by D2S1338 (PIC = 86.6%, PD = 96.3%). D2S11 had the highest Power of Elimination (PE = 85.8%), followed by SE33 (PE = 83.8%). Notably, SE33 also had the highest number of allelic variants (v = 27), followed by FGA (v = 17) (**Table 3.6**). Conversely, TH01 was the least polymorphic (PIC = 63.1%) and discriminative (PD = 85.0%) loci and had the lowest Power of Elimination (PE = 40.3%) with the smallest number of allelic variants (v = 5). The Typical Paternity Index ranged from 1.578 (TH01) to 7.214 (D21S11) (**Table 3.6**). The Combined Power of Exclusion (CPE) and Discrimination (CPD) exceeded 99.9999%, and the Combined Match Probability (CMP) was 2.60 x $10^{-20}$, indicating a very low chance of matching profiles in the population (L.R = 1 in 3.84 x $10^{19}$).

The Hausa-Fulani dataset identified 61 distinct alleles, with sizes ranging from 6 to 36. The allele frequencies among 100 unrelated individuals varied from 0.005 to 0.480, with allele 7 at the TH01

marker having the highest frequency (0.480) (**Table 3.7**). Locus SE33 was the most polymorphic (PIC = 92.5%) and discriminative (PD = 98.3%), followed by D2S1338 (PIC = 85.6%) and FGA (PD = 96.3%). SE33 exhibited the highest Power of Elimination (PE = 91.9%) and had the largest number of allelic variants (v = 28), while D18S51 (PE = 83.6%) and D21S11 (v = 17) had the second highest values **(Table 3.8)**. In contrast, TH01 had the lowest polymorphic (PIC = 64.6%), discriminative (PD = 86.0%), and Power of Elimination (PE = 47.6%) values, with the fewest number of allelic variants (v = 5). The Typical Paternity Index varied from 1.852 (TH01) to 12.5 (SE33) across the 16 loci **(Table 3.8)**. The Combined Power of Exclusion (CPE) and Discrimination (CPD) for the 16 loci exceeded 99.9999%, with a Combined Match Probability (CMP) of $3.03 \times 10^{-20}$, indicating a very low chance of two individuals sharing the same profile in this population (L.R = 1 in $3.30 \times 10^{19}$).

**Table 3.9** presents the private alleles (PAs) uniquely identified within specific subpopulations. A total of 20 PAs were documented from a sample of 303 unrelated individuals. Both the Igbo and Hausa subpopulations identified 8 PAs each. In contrast, the Yoruba subpopulation recorded only 4 PAs, the lowest number among the subpopulations studied. It is crucial to recognize that the sample size can greatly influence the distribution of alleles among these Nigerian subpopulations.

### 3.4.4 INTER-SUBPOPULATION RELATIONSHIP

Exact tests of Hardy–Weinberg equilibrium based on Monte Carlo simulations revealed no significant deviations ($p < 0.05$) across loci in the pooled Nigerian population prior to Bonferroni correction **(Table 3.2; Appendix 20)**. Furthermore, no locus remained significant following correction for multiple testing. Within the Igbo subpopulation, no loci exhibited deviations from Hardy–Weinberg equilibrium prior to Bonferroni correction, and all loci conformed to equilibrium expectations following correction **(Table 3.4; Appendix 21)**. In the Yoruba subpopulation, locus D21S11 showed a nominal deviation from the Hardy-Weinberg equilibrium prior to Bonferroni correction **(Table 3.6)**. However, this deviation was not significant after correction for multiple testing, and all other loci remained in Hardy–Weinberg equilibrium **(Appendix 22)**. In the Hausa subpopulation, loci D21S11 and D2S1338 exhibited significant deviations from the Hardy-Weinberg equilibrium prior to Bonferroni correction **(Table 3.8)**. However, none of these loci remained significant following correction for multiple testing **(Appendix 23)**. Overall, these findings indicate that the QIAGEN™ Investigator ESSplex SE QS Kit demonstrates high forensic suitability for DNA profiling within Nigerian subpopulations and the pooled Nigerian population.

**F-statistics Genetic Distance**

Wright's F-statistics were calculated and averaged across all loci to estimate the genetic differentiation and inter-subpopulation relationships among the Igbo, Yoruba, and Hausa-Fulani subpopulations. The $F_{ST}$ values for the entire Nigerian population (as shown in **Table 3.2**) ranged from 0.002 to 0.006 across loci, indicating very low genetic differentiation among the Igbos, Yorubas, and Hausa-Fulanis. Locus D21S11 exhibited the highest $F_{ST}$ value at 0.006, while D3S1358, D19S433, and D2S441 had the lowest $F_{ST}$ values, each at 0.002.

**Table 3.10. $F_{ST}$ value comparison between the Igbo, Yoruba, and Hausa-Fulani subpopulation using the QIAGEN™ Investigator ESSplex SE QS Kit**

| Subpopulation | Igbo | Yoruba | Hausa-Fulani |
|---|---|---|---|
| **Igbo** | 0.000 | | |
| **Yoruba** | 0.002 | 0.000 | |
| **Hausa-Fulani** | 0.003 | 0.002 | 0.000 |

**Table 3.10** presents the pairwise population $F_{ST}$ values, which reflect the level of genetic distinctiveness among the Igbo, Yoruba, and Hausa-Fulani populations. The $F_{ST}$ values indicate very low levels of differentiation between these groups. The Igbo appear slightly more genetically differentiated from the Yoruba, with an $F_{ST}$ value of 0.002. At the same time, they are nearly identical to the Hausa-Fulani, who have an $F_{ST}$ value of 0.003. Similarly, the Yoruba and Hausa-Fulani populations show a minimal $F_{ST}$ value of 0.002. These results signify that the three subpopulations have a high level of genetic similarity.

The $F_{IS}$ values **(Table 3.2)**, also known as the inbreeding coefficient, showed negative values (-0.092 to -0.025) at most loci, indicating an excess of heterozygotes. This suggests no significant impact of inbreeding was detected within the subpopulations. Locus D16S539 exhibited the highest $F_{IS}$ value (0.025), while D19S433 had the lowest (-0.092). Most loci showed negative $F_{IS}$ values, with exceptions including D16S539 (0.025), vWA (0.004), D2S1338 (0.006), and FGA (0.007). The $F_{IS}$ values varied across subpopulations: -0.108 to 0.057 in the Igbo **(Table 3.4)**, -0.119 to 0.094 in the Yoruba **(Table 3.6)**, and -0.118 to 0.090 in the Hausa-Fulani **(Table 3.8)**. Each group displayed only a few positive values, showing overall similar results.

The $F_{IT}$ value was statistically similar to the $F_{IS}$ value based on the expected heterozygosity, as it considers the inbreeding of individuals within an overall population. The $F_{IT}$ values for the Nigerian

population (refer to **Table 3.2**) were generally low, ranging from -0.090 to 0.030 across all loci in the total population. These values indicate no significant inbreeding within or between the populations.

**Population Structure**

*STRUCTURE* was used to examine the inter-subpopulation relationships among the Igbo, Yoruba, and Hausa-Fulani subpopulations. The Mean LnP(K) ± SD analysis (see **Appendix 26**) revealed inconsistencies in estimating the average log probability at $K = 4$, as indicated by the high standard deviation (SD) shown in **Figure 3.5.** The transition from $K = 4$ to $K = 5$ shows a slight increase in log-likelihood but does not correspond to a significant improvement in Delta *K*. The Mean LnP(K) compared the likelihood of different *K* values generated by *STRUCTURE* with the higher value, suggesting a better fit for *K*. The insight on the robustness of the model for a given *K* was provided by the Mean LnP(K)±SD to help assess the stability of the estimate of the population structure across runs.

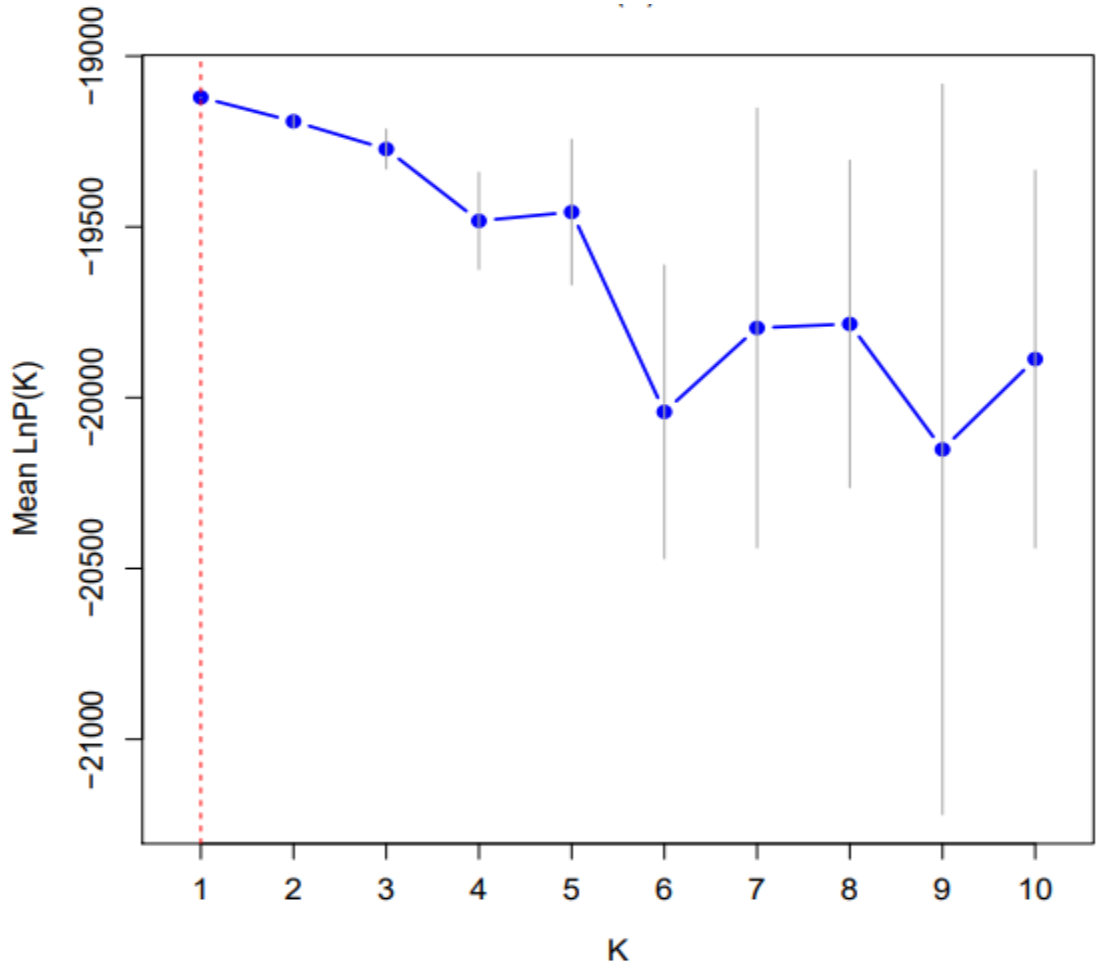



**Figure 3.5. The mean log-likelihood (±SD) of the data for a given number of population (*K*) computed across five repeat runs and iterations of the *STRUCTURE* program.**

According to the method developed by Evano *et al*. (2005), the optimal *K* value identified was 5 (see **Figure 3.6** and **Appendix 26**). This suggests that there are five genetically distinct sub-populations, with the highest Delta *K* recorded for this value. The rate of change in Delta *K* assists in selecting the appropriate population size (*K*) by comparing the likelihood between different *K* values. A higher Delta *K* value indicates the optimal number of populations in the dataset. Although *K*=1 was expected, the Evanno method (Delta *K*) does not assess *K*=1 because Delta *K* is undefined at that level, as it relies on the second order of change, which is also undefined at *K*=1. *STRUCTURE* primarily identifies the most pronounced levels of population structure without using prior information about individual origins (Pritchard *et al*., 2000).

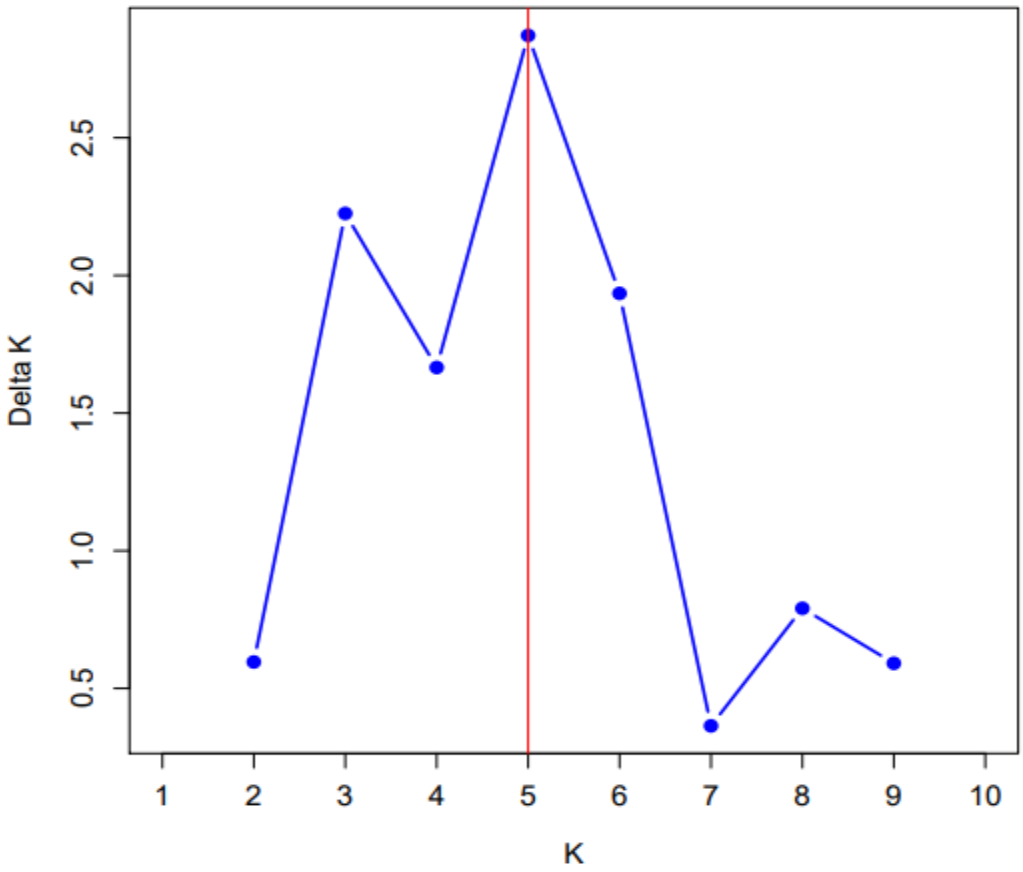



**Figure 3.6. The change in the log-likelihood rate (Delta *K*) for each value of *K* computed across five repeated runs and iterations of the *STRUCTURE* program.**

The analysis of assumed populations across values of *K* (from *K* = 1 to *K* = 10) revealed no population structuring detected among the Igbo, Yoruba, and Hausa-Fulani subpopulations based on the bar plots **(Figure 3.7)**. Although *K* = 5 was identified as the optimal result (Delta *K*), **Figure 3.7** shows no genetic differentiation in the ancestry proportions of individuals within each group for *K* values from *K* = 1 to *K* = 10.

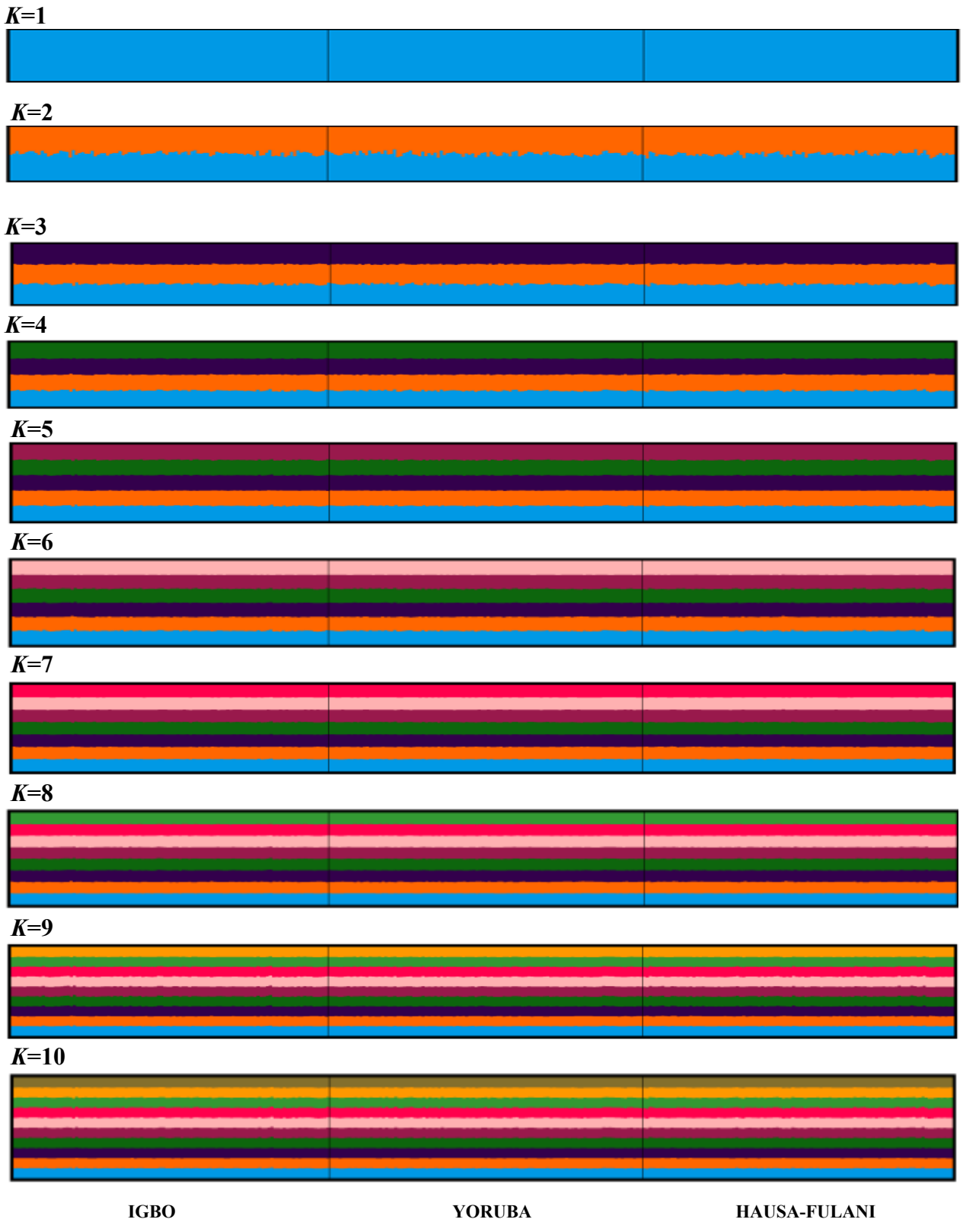



**Figure 3.7. The population stratification of the Igbo, Yoruba, and Hausa-Fulani subpopulations within the Nigerian population. This was computed over five repeated runs and iterations using the *STRUCTURE* program. The assumed populations, ranging from *K*=1 to *K* =10, are represented by a colour bar. Each vertical line corresponds to an individual (total = 303) assessed based on data from 16 autosomal loci using the QIAGEN™ Investigator ESSplex SE QS kit. The amount of colour in each bar represents the individual's genetic ancestry to each *K* population.**

**Principal Component Analysis**

To examine the relationships among different groups in Nigeria, Principal Component Analysis (PCA) was used to investigate population differentiation across the Igbo, Yoruba, and Hausa-Fulani subpopulations. PCA helps uncover hidden relationships among these ethnic groups by analysing genetic data without requiring prior knowledge of their ancestral backgrounds. It simplifies genetic datasets while preserving significant variation. This approach helps find patterns in the Nigerian population genetic structure without bias. Similar samples are expected to cluster together, while dissimilar samples are expected to cluster separately when visualised on a PCA plot. Investigating genetic segregation patterns among the Igbo, Yoruba, and Hausa-Fulani ethnic groups involved analysing the full dataset from 303 unrelated individuals. This analysis utilised loci information from 16 autosomal STR to visualise genetic differentiation. As shown in **Figure 3.8**, the results revealed no differentiation among the 303 unrelated individuals from the three ethnic groups based on the examined samples, as no data clustering was observed from different subpopulations.

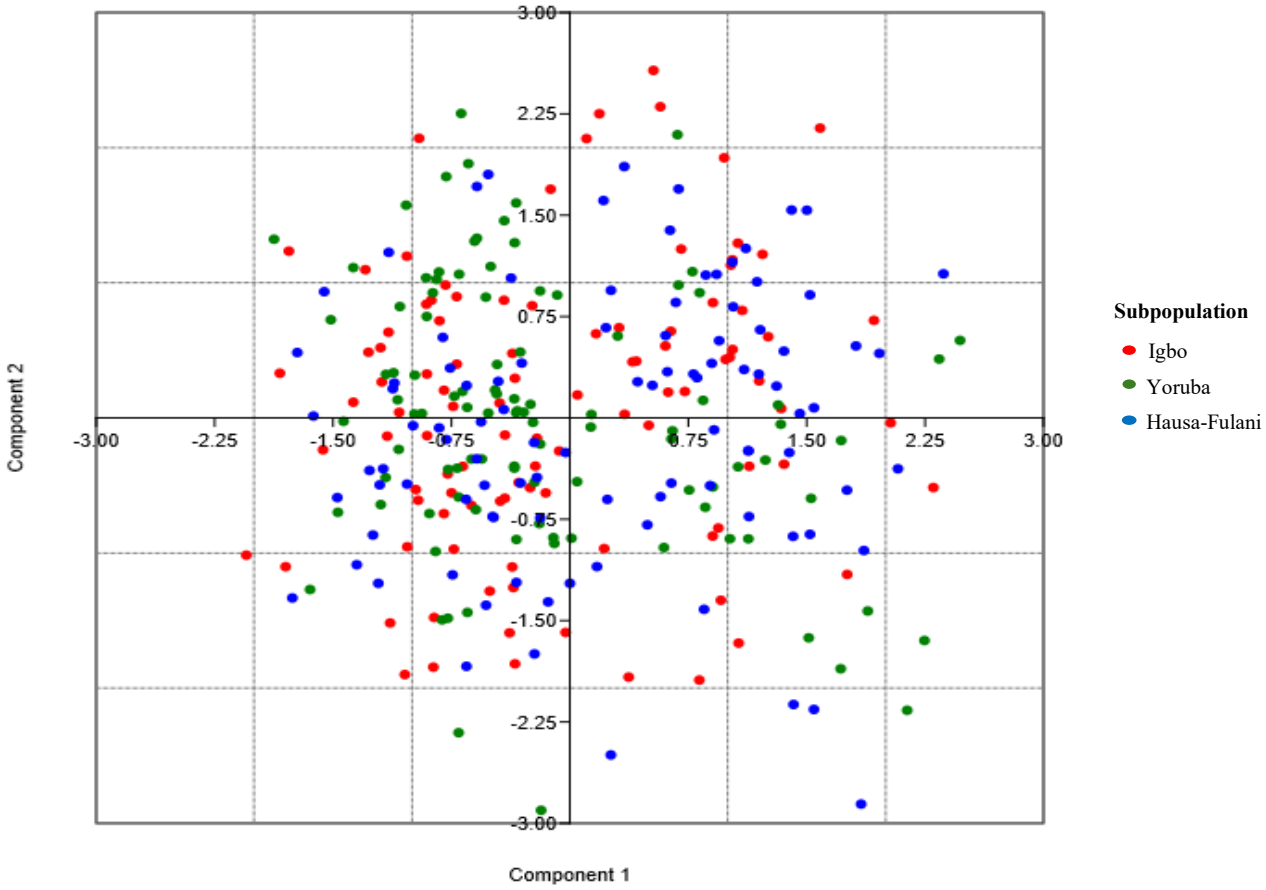



**Figure 3.8. Result of the Principal Component Analysis (PCA) conducted on the Igbo, Yoruba, and Hausa-Fulani subpopulations. This utilised 16 autosomal short tandem repeat (STR) loci from the QIAGEN™ Investigator ESSplex SE QS kit. In the PCA plot, each dot represents one of the 303 samples collected from unrelated individuals within the Nigerian population.**

**Neighbour-Joining (NJ)**

The phylogenetic distance among the three sub-populations was evaluated using neighbour-joining (NJ) method to assess their inter-subpopulation relationships. The tree was constructed using data from the pairwise Nei's genetic distance population matrix (see **Appendix 24**), which was calculated based on 16 autosomal STR loci. The results, shown in **Figure 3.9** displayed a similar pattern of population differentiation to the pairwise $F_{ST}$ values among the Igbo, Yoruba, and Hausa-Fulani ethnic groups. As anticipated, the NJ analysis revealed very low levels of inter-subpopulation variation, suggesting a high degree of genetic similarity between the three populations.

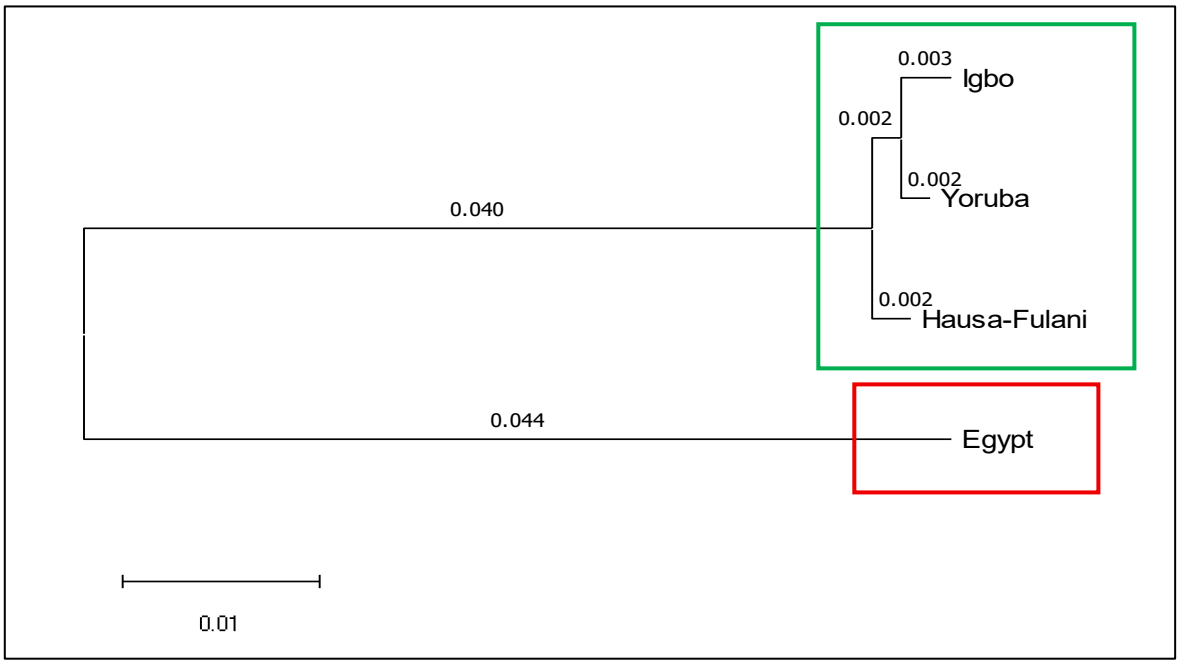



**Figure 3.9. Neighbour-joining (NJ) tree based on pairwise Nei's Genetic Distance illustrating the clustering of the Igbo, Yoruba, and Hausa–Fulani subpopulations, with Egypt included as an outgroup. This analysis utilized 16 autosomal short tandem repeat (STR) loci from the QIAGEN™ Investigator ESSplex SE QS Kit.**

However, the phylogenetic tree indicates that the Igbos are slightly genetically differentiated from the Yorubas than the Hausa-Fulanis. This is likely due to the Yorubas' geographical proximity (located in the South-West) to the Igbos' (located in the South-East). Conversely, the Hausa-Fulanis are slightly genetically closer to the Yorubas than to the Igbos.

## 3.5 DISCUSSION

This research analysed the population genetics of 303 unrelated Nigerians from the Igbo, Yoruba, and Hausa-Fulani ethnic groups using 16 autosomal STR markers from the QIAGEN™ Investigator ESSplex SE QS Kit. All samples produced full STR profiles except for those with off-

ladder alleles. The STR profiles were utilised to calculate allele frequencies and forensic genetic parameters and to assess inter-subpopulation relationships using *STRUCTURE,* PCA, and NJ methods. Previous studies have generated allele frequency data for the Nigerian population. Agbo *et al.* (2017) and Okolie *et al.* (2017) reported this data using different sets of genetic markers: the 21 autosomal STR loci of the Qiagen Investigator 24plexGO! Kit and the 15-loci in the AmpF*l*STR™ Identifiler™ Direct PCR Amplification *Kit* from ThermoFisher Scientific contain core CODIS loci. Hohoff *et al.* (2009) also conducted similar work using the 16 STR loci in the PowerPlex ES Kit from Promega, the AmpF*l*STR Identifiler, Profiler, and SEFiler PCR amplification kits from Applied Biosystems for the autosomal STR alleles. However, no studies have been conducted using the QIAGEN™ Investigator ESSplex SE QS Kit on the Nigerian population. This kit was selected because it includes essential DNA markers (CODIS and ESS loci), has built-in quality controls (QS1 and QS2), and works well with degraded or low amounts of DNA. It helps fill gaps in current data and provides reliable DNA profiles for forensic and population genetics research.

Based on the estimation of the allele frequencies, forensic parameters and test of effectiveness, the overall data showed that all the genetic markers of the QIAGEN™ Investigator ESSplex SE QS Kit are very polymorphic across the Igbo, Yoruba, and Hausa-Fulani subpopulations. It was discovered that locus SE33 was the most polymorphic and consistently discriminative due to possessing the largest number of allelic variants. Although SE33 demonstrates high discriminatory power due to its high level of polymorphism, it is not utilised in every country, such as the USA, the Netherlands, and the UK. This is because it is not included in all STR kits or mandated by all national forensic databases. The reason is usually related to technical issues and regional forensic needs (Guo *et al*., 2014). On the other hand, TH01 ranked the lowest due to having fewer allelic variants, but it was detected as the locus with the highest allele frequency (allele 7 in the TH01 marker) across the sample population and in two sub-populations (Igbo and Yoruba). Other informative loci included D2S1338, D21S11, D2S441, FGA, and D18S51.

The Combined Power of Elimination (CPE) and Discrimination (CPD) calculated for the entire population and subpopulations of the Igbos, Yorubas and Hausa-Fulanis were consistently higher than 99.9%. According to Butler (2006), the Power of Discrimination (PD) considers the probability that two randomly selected unrelated individuals will not have an identical genotype at a specific locus within a population. In contrast, the Power of Exclusion (PE) is the probability of excluding a randomly chosen non-related individual as a contributor of alleles at a given locus in

that population. This further demonstrates the richness of information of the QIAGEN™ Investigator ESSplex SE QS Kit in effectively analysing the Nigerian population and discriminating among individuals.

Exact tests of Hardy–Weinberg equilibrium (HWE) were performed for all autosomal STR loci using Monte Carlo simulations, which are appropriate for highly polymorphic, multi-allelic markers and moderate sample sizes (Guo and Thompson, 1992; Butler, 2015). Statistical significance was assessed before and after Bonferroni correction for multiple testing. In the Nigerian population studied, none of the loci showed significant deviations from the Hardy-Weinberg equilibrium at the 0.05 significance level before Bonferroni correction. Furthermore, all loci conformed to Hardy–Weinberg expectations following correction for multiple testing, indicating overall equilibrium across the combined dataset. Within the Igbo subpopulation, all loci adhered to Hardy–Weinberg equilibrium both before and after Bonferroni correction, with no evidence of deviation detected across the analysed loci. In the Yoruba subpopulation, a nominal deviation from the Hardy-Weinberg equilibrium was seen at locus D21S11, but it was not significant after Bonferroni correction. All other loci met equilibrium expectations. The Hausa-Fulani subpopulation showed nominal deviations at loci D21S11 and D2S1338, but none were significant following Bonferroni correction, indicating that the population generally conformed to Hardy-Weinberg expectations. The few nominal deviations observed before correction could be attributed to sampling effects commonly observed in STR datasets and do not necessarily indicate genotyping error or true biological disequilibrium (Butler, 2015). The absence of significant deviation in the pooled Nigerian population further indicates that pooling individuals from the Igbo, Yoruba, and Hausa subpopulations did not introduce detectable population structure effects. This aligns with the findings of Agbo *et al*. (2017), Okolie *et al*. (2017), and Hohoff *et al*. (2009), who demonstrated conformity with Hardy-Weinberg expectations in the Nigerian population. Allele and genotype frequencies in a population should remain constant from generation to generation unless altered by some evolutionary force, including non-random mating (including inbreeding), genetic drift, natural selection, mutation, migration (gene flow), founder effect, population bottleneck, and meiotic drive (Gillespie, 2004). According to Butler (2015), the problem with a potentially significant HWE test is that this may indicate that evolutionary forces such as natural selection, genetic drift, migration, and mutation are impacting allele/genotype frequencies.

The $F_{ST}$ values of each locus calculated and averaged in this experiment recorded very low levels of differentiation between the Igbo, Yoruba, and Hausa-Fulani subpopulations, indicating a high

level of genetic similarity among the three populations. This outcome aligns with the findings of Akpan *et al*. (2024), who reported a similar pattern of pairwise $F_{ST}$ values indicating genetic similarity among the Igbo, Yoruba, Hausa, Ibibio, and Tiv populations. Their results revealed that the Igbo and Yoruba populations are genetically closer to each other but slightly more distinct from the Hausa. The pairwise $F_{ST}$-based similarity between the Yoruba and Igbo populations was 0.993, indicating the highest genetic similarity among the five studied populations. The Yoruba-Tiv similarity was 0.987, Yoruba-Ibibio was 0.990, and Yoruba-Hausa was 0.992. In contrast, the Igbo-Ibibio similarity was the lowest at 0.895 (Akpan *et al.,* 2024). In a forensic context, the $F_{ST}$ value is often utilised to assess the degree of differentiation and understanding of the structure of a population between different groups or sub-groups (Butler, 2011).

The study, which focused on genetic structure, NJ and PCA analysis, indicated a low genetic differentiation and high level of genetic similarity between the Igbo, Yoruba, and Hausa-Fulani ethnic groups. The analysis of the population structure for the QIAGEN™ Investigator ESSplex SE QS Kit revealed Delta *K* values that suggested the presence of five populations. However, the bar plot of the population stratification showed no genetic differentiation among these subpopulations. This finding shows that the report stating there are 5 subpopulations in Nigeria is not correct. It suggests that the Delta *K* method has limitations in distinguishing between these subpopulations. Therefore, it is recommended to conclude that there is only a single genetic population across the subpopulations in Nigeria, as the bar plot results did not indicate any genetic structure. The PCA results suggest that the Igbo, Yoruba, and Hausa-Fulani may share historical and genetic connections. Over the centuries, ethnic group interactions may have promoted gene flow and reduced stratification between these ethnic groups in Nigeria (Okolie *et al*., 2017; Akpan *et al.*, 2024). The highly overlapping allele frequencies also indicate that the various ethnic groups are genetically close. Although the NJ analysis revealed some subtle genetic clusters indicating that the Igbos are genetically more similar to the Yorubas than to the Hausa-Fulanis, the Hausa-Fulanis show a slight genetic similarity to the Yorubas compared to the Igbos. However, the analysis did not show a well-established population stratification that could differentiate these subpopulations. This lack of distinction may be caused by limited gene flow caused by the geographic distance between southeastern and northern Nigeria, resulting in less migration and intermarriage in these regions than in southern Nigeria (Adeyemo *et al*., 2005). According to Okolie *et al*. (2017), the low genetic differentiation observed in the Nigerian population could be attributed to gene flow, shared ancestry, and the large size of the population with subpopulations continuously mixing through intermarriage, trade, or migration over time. The Igbo in South-East, the Yoruba in the South-West,

and Hausa-Fulani in the North Nigeria the have historically interacted, which may have led to genetic exchange.

## 3.6 CONCLUSION

The research established the forensic utility of the 16 autosomal polymorphic STR markers included in the QIAGEN™ Investigator ESSplex SE QS Kit for analysing the Nigerian population and discriminating among individuals of the Igbo, Yoruba, and Hausa-Fulani tribes in the Nigerian population. The 16 STR markers generated informative allele frequency data and were highly polymorphic and discriminatory, which makes them useful in forensic testing and the demographic and anthropological study of the Nigerian population. The study revealed that locus SE33 was the most informative and remained consistently discriminative due to the locus possessing the most allelic variants in all sub-populations. At the same time, TH01 ranked the lowest due to fewer allelic variants but had the highest detected allele frequency. The Hardy-Weinberg Equilibrium (HWE) displayed no significant deviation from expectations across loci and across subpopulations. The F-statistics, population structure, PCA, NJ analysis conducted in this experiment recorded very low differentiation in the population genetic structure. The study has also revealed that the Igbo ethno-linguistic group is slightly genetically closer to the Yorubas than the Hausa-Fulanis, who are much more immediate to the Yorubas.

Storage and transportation of the QIAGEN™ Investigator ESSplex SE QS kit in warm climates is a limitation if not adequately considered. The kit components are sensitive to storage conditions, requiring regular freezing to maintain the integrity of the reagents effectively. The forensic sector of Nigeria is still developing and underutilised in crime investigation; therefore, the country needs more well-trained personnel with expertise in forensic DNA to utilise the QIAGEN™ Investigator ESSplex SE QS kit successfully and to ensure adequate quality control measures are implemented as there are few experts in forensic genetics in the country. Furthermore, it is also essential to validate the kit's performance with other ethnic groups to ensure reliable and accurate results. Notably, the QIAGEN™ Investigator ESSplex SE QS kit showed good performance across the samples collected from the three prominent ethnic groups, namely the Hausa-Fulani, Igbo and Yoruba subpopulations. However, considering that Nigeria has over 300 ethnic groups with 500 spoken languages and English being the official language, a validation of the kit's performance with other sub-populations is recommended.

# CHAPTER FOUR

# AUTOSOMAL STR ANALYSIS OF THE NIGERIAN POPULATION USING THE GLOBALFILER™ EXPRESS PCR AMPLIFICATION KIT

## 4.1 INRODUCTION

The GlobalFiler™ Express PCR Amplification Kit (Thermo Fisher Scientific, WA, USA) is an STR multiplex PCR kit designed to amplify 21 autosomal STRs, one Y-linked indel, one Y-STR, and the Amelogenin sex chromosome locus from human genomic DNA. In a single reaction, this kit enables amplification and efficient separation of 24 markers using five-colour fluorescent detection. It includes the 7 European Standard Set of Loci (ESSL), 13 original CODIS loci, the highly discriminating SE33 locus, 2 Y-based loci, and a sex chromosome marker. The GlobalFiler™ Express PCR Amplification Kit is optimized for direct DNA amplification from buccal and blood samples collected on a paper substrate such as FTA® Cards (Whatman/GE Healthcare), NucleoCard® (Macherey-Nagel), and 903 Protein Saver Cards (Whatman). It features a robust PCR protocol that can produce results in less than an hour (Flores *et al*., 2014; Wang *et al*., 2015). Highly sensitive and specifically designed for direct amplification from reference samples, such as blood or buccal samples collected on swab substrates or on plain or chemically treated paper, this kit is compatible with ABI 3500 Genetic Analyzers (Thermo Fisher Scientific). Furthermore, the kit has been specifically designed with powerful features for analysing low-template or degraded forensic DNA samples, particularly using short amplicons, robust enzymes, and inhibitor-resistant chemistry. It has been successfully applied in forensic casework, including forensic DNA typing and databasing (Wang *et al*., 2015; Larnane *et al*., 2024).

The number of STR loci used in the United Kingdom increased between 1995 and 2014. This change was driven by advancements in technology that allowed for the development of upgraded STR kits incorporating a larger number of loci. The smaller number of loci had previously resulted in lower discriminatory power. In 2014, England and Wales adopted a new DNA profiling system based on 17 STR loci. Several commercial STR kits were tested and approved for use by different manufacturers. Conversely, Scotland implemented the 24 STR locus system using the GlobalFiler™ kit produced by Thermo Fisher as the standard for DNA profiling, effective from 2015 (Johnson and William, 2004; Hopwood *et al*., 2012). While 17-locus kits are still in use in England and Wales, laboratories have recommended 24-locus kits like GlobalFiler™ for their higher resolution and sensitivity in distinguishing complex mixtures and aiding in mass disaster identification, based on evolving guidance from the UK Forensic Science Regulator and standards set by the National DNA Database (NDNAD) (Hartshorne *et al.*, 2024).

Nigeria has not yet established a national criminal DNA database for forensic purposes, as lawmakers are still debating the project. The Federal Government has shown little commitment to creating this database for several reasons: a lack of political support, a shortage of skilled workers, and inadequate funding (Aborisade *et al*., 2024; Etim-Osa and Etim-Osa, 2019). The existing legal framework in Nigeria poses significant challenges in establishing a forensic database as the legal system currently lacks proper standards for packaging and labelling forensic evidence, as this undermines the admissibility of most legal practitioners in handling criminal cases involving forensic evidence.

## 4.2 AIM AND OBJECTIVES

This study aims to evaluate the population genetics of 303 unrelated Nigerian individuals from three main ethnic groups: Hausa-Fulani, Igbo and Yoruba. The analysis will utilise 21 autosomal STR markers from the 24-locus GlobalFiler™ Express PCR Amplification Kit (Thermo Fisher Scientific), which will be reported for the first time for the Nigerian population. A secondary aim is to assess the genetic information obtained and compare it with data generated using the Qiagen™ Investigator ESSplex SE QS Kit (as reported in **Chapter 3**), thereby contributing to and expanding the existing genetic database of the Nigerian population.

The specific objectives for the work detailed in this chapter are as follows:

1. To generate allele frequency data that will contribute to developing an allele frequency database for each sub-population group and the overall population.
2. To determine forensic parameters for each subpopulation and the overall population and assess the kit's effectiveness and forensic utility.
3. To compare the inter-subpopulation genetic structure and examine the relationships between the sub-populations.
4. To compare the genetic data generated by the GlobalFiler™ Express PCR Amplification Kit with that from the Qiagen™ Investigator ESSplex SE QS Kit, and to assess the concordance of results for loci common to both kits
5. Recommend the most suitable STR kit for establishing a forensic DNA database in Nigeria.

## 4.3 MATERIALS AND METHODS

Blood samples for the Nigerian population study were collected from different locations in Nigeria and the United Kingdom, involving 303 unrelated individuals (167 males and 136 females), as described in **Chapter 2.2**. Donors received participant information sheets and consent forms to sign and return. Participants additionally received verbal explanations regarding the procedures for

collecting the blood samples. STR profiling, data analysis, and statistical calculations were conducted at the University of Strathclyde in Glasgow. Comprehensive procedures for sample collection are outlined in **Chapter 2.2**, and the participation information sheet and consent form are in **Appendices 1** and **2**. The DNA profiling technique, utilising a 1.2mm punch of blood sample from the FTA® card, involved standard PCR and direct PCR with the GlobalFiler™ Express PCR Amplification Kit (Thermo Fisher Scientific) as described in **Chapters 2.5.1** and **2.5.2**, respectively. The capillary electrophoresis method, which detects and separates DNA fragments, was performed using the ABI Genetic Analyzer 3500 with internal lane standard, the GeneScan™ 600 LIZ™ Size Standard v2.0 and GlobalFiler™ Express Allelic Ladder, as detailed in **Chapter 2.5.3**. Data interpretation, including allele calling and genotype determination for the 21 autosomal STR loci, was conducted using GeneMapper® ID-X software version 1.6 (Thermo Fisher Scientific), as described in **Chapter 2.4.4**. Concordance analysis was performed for each sample by comparing the allele calls at shared loci between the QIAGEN™ Investigator ESSplex SE QS Kit and the GlobalFiler™ Express Kit. The procedure for calculating the allele frequencies and various forensic parameters using the STR Analysis for Forensic (*STRAF*-A 2.2.2) online tool and the Genetic Analysis in Excel (GenAIEx 6.5) platform was explained in **Chapter 2.7.1**. The Wright's F-statistics ($F_{ST}$, $F_{IS}$, and $F_{IT}$) were evaluated using the Genetic Analysis in Excel (GenAIEx 6.5) platform to assess genetic variation among the Igbo, Yoruba, and Hausa-Fulani sub-populations compared to the total population, as explained in **Chapter 2.7.1**. To evaluate the inter-subpopulation structure, *STRUCTURE* software version 2.3.4 was employed, as described in **Chapter 2.7.2**. The software PAleontological Statistics (PAST 4.03) conducted Principal Component Analysis (PCA) to assess sub-population differentiation, as explained in **Chapter 2.7.2**. As explained in **Chapter 2.7.2**, the software Molecular Evolutionary Genetics Analysis (MEGA-11) was used to conduct a Neighbour-Joining analysis of the Igbo, Yoruba, and Hausa-Fulani sub-populations.

## 4.4 RESULTS

### 4.4.1 STR GENOTYPING

The GlobalFiler™ Express PCR Amplification Kit successfully produced complete DNA profiles for all 303 samples from blood spots on FTA cards for all 303 samples. This was achieved using two techniques: the solid-phase silica gel column extraction method for 49 samples and the direct PCR amplification approach for 254 samples. Blood samples that were challenging to amplify using direct PCR were successfully amplified using standard PCR, which included DNA extraction and quantification to obtain complete STR profiles. The DNA extraction and quantification

processes overcame PCR inhibition, likely caused by matrix components in the samples that hindered the effectiveness of the direct PCR method. Each DNA profile included 21 polymorphic STR markers, the Y indel, the Y-STR, and the Amelogenin sex marker. Notably, all negative PCR controls included in each assay showed no evidence of contamination. The DNA profile obtained from the positive control sample (**Figure 4.1**) matched the expected profile.

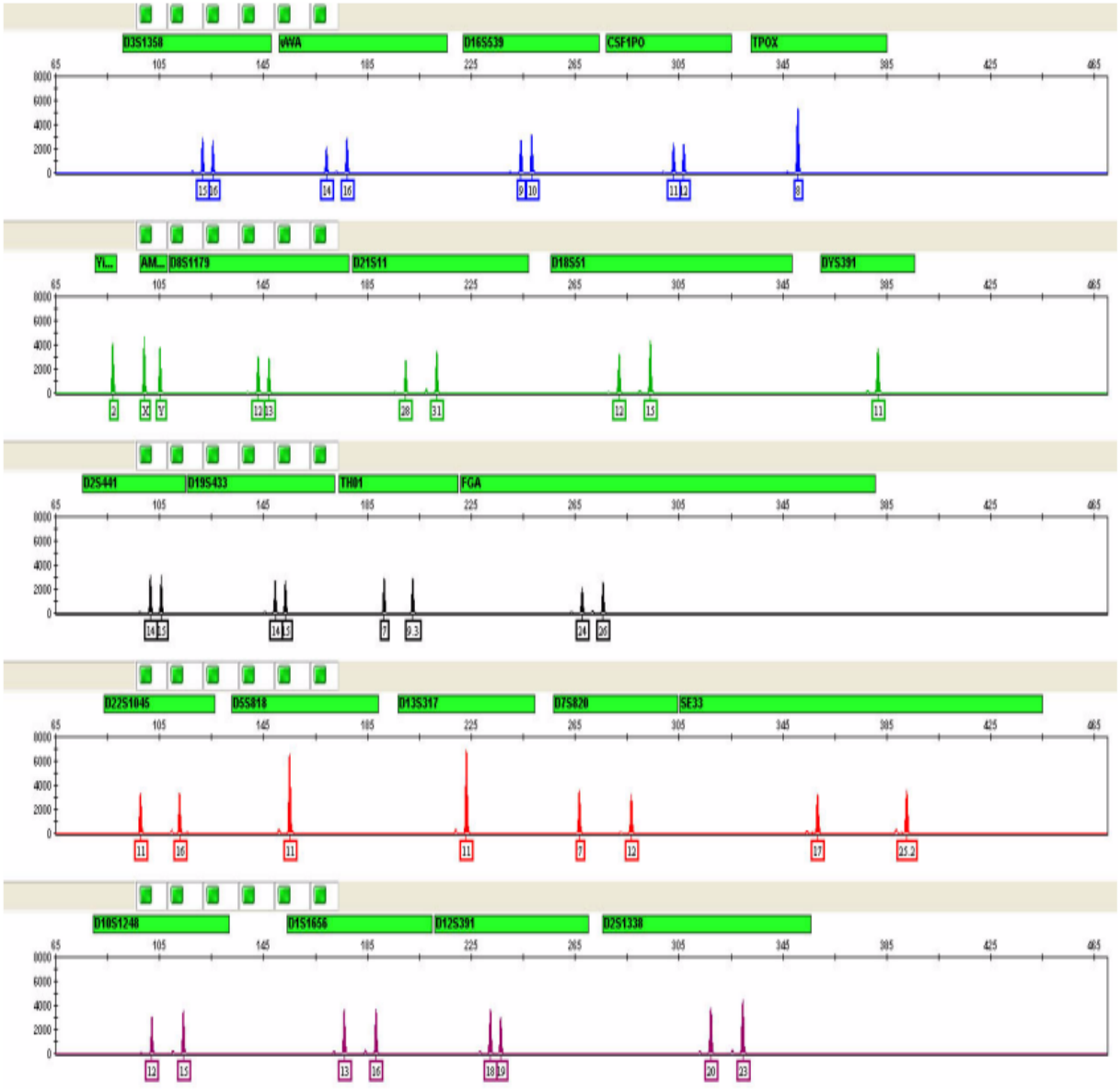



**Figure 4.1. Electropherogram of the GlobalFiler™ Express PCR Amplification DNA Control 007 with the loci labelled with five dye colours distributed across five colour channels.**

### 4.4.2 STR OFF-LADDER ALLELE CALLS AND CONCORDANCE ANALYSIS

The GeneMapper® ID-X v1.6 software was utilised to analyse the data, which included allele calling and genotype determination for 21 autosomal STR loci. During data interpretation, all electropherograms generated with the GlobalFiler™ Express Kit were thoroughly examined for

artefacts, including stutter peaks, pull-ups, and allelic dropouts. No off-ladder alleles or significant artefacts were detected in the analysed samples, and no additional sample clean-up was necessary.

The results from samples processed with the QIAGEN™ Investigator ESSplex SE QS Kit were compared to those obtained from the GlobalFiler™ Express Kit. This comparison was based on allele data generated for the same samples at the same loci for those loci that were surveyed in both kits.

The results generated from the GlobalFiler™ Express Kit were consistent with those from the QIAGEN™ Investigator ESSplex SE QS Kit, except for results affected by off-ladder issues in the Qiagen kit; off-ladder alleles were found in positions lacking bins, leading to their classification as off-ladder.

The alleles that were not identified in the QIAGEN™ Investigator ESSplex SE QS Kit due to the absence of designated bins in the ladder were successfully identified in the GlobalFiler™ Express Kit, which included the necessary bins in its allelic ladder. The resolution of the off-ladder alleles observed in the QIAGEN™ Investigator ESSplex SE QS Kit is illustrated in **Figure 4.2**a and **4.2b** (D12S391 locus), **Figure 4.3** (D3S1358 locus), and **Figure 4.4** (D1S1656 locus), based on the comparison of STR alleles produced by the GlobalFiler™ Express Kit for the same samples.

The alleles identified for the D12S391 locus are 18.1, 20.1, and 17.1 **(Figures 4.2a** and **4.2b)**. For the D3S1358 locus, the allele is 15.2 **(Figure 4.3)**, while the D1S1656 locus shows 13.3 **(Figure 4.4)**.

**GlobalFiler™ Express Kit. - Locus D12S391**

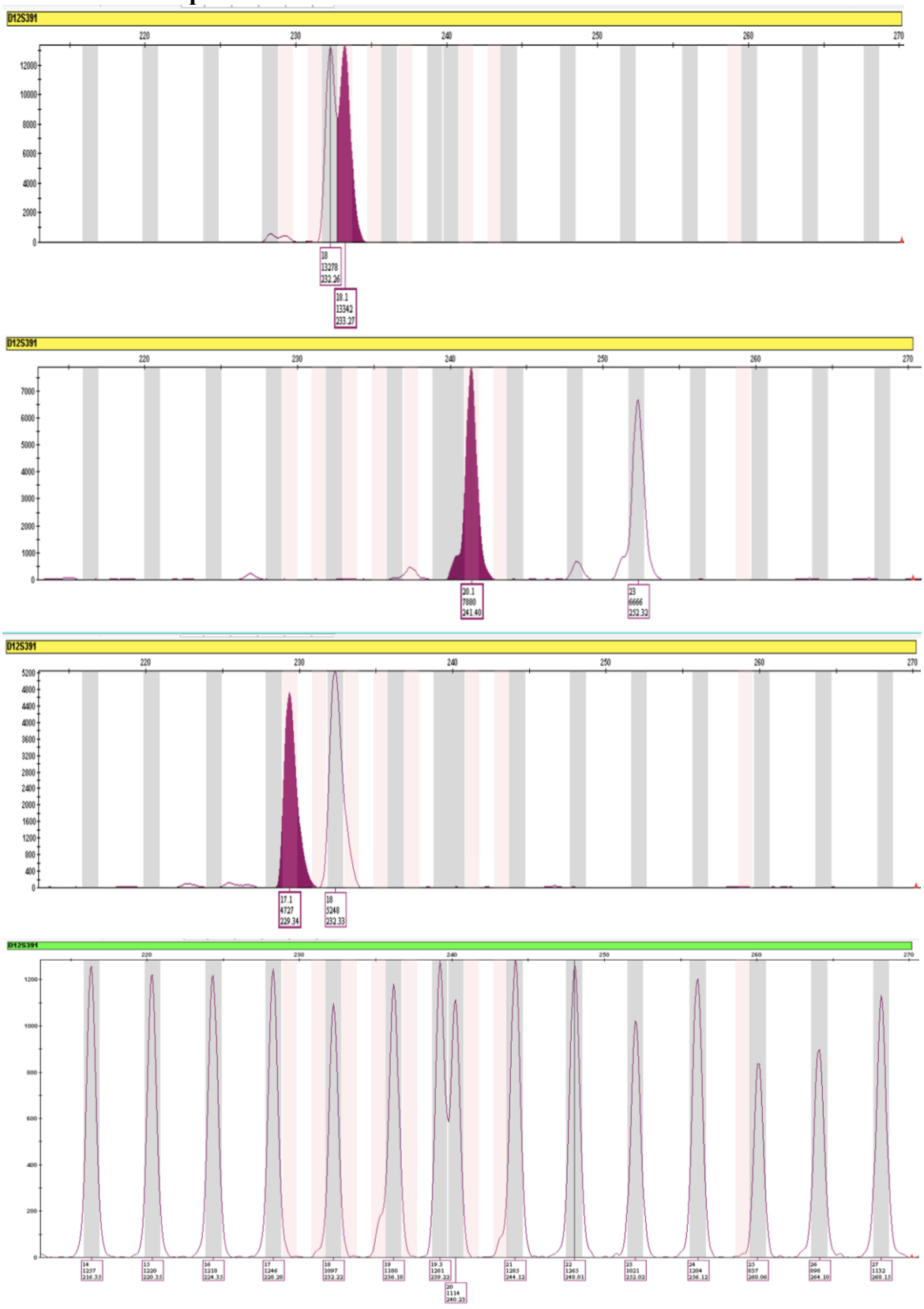



**Figure 4.2a. D12S391 alleles identified on the electropherogram using the GlobalFiler™ Express Kit, included appropriate bins from its allelic ladder with no off-ladder issues. Results were compared to those from the QIAGEN™ Investigator ESSplex SE QS Kit in Figure 4.2b.**

**QIAGEN™ Investigator ESSplex SE QS Kit – Locus D12S391**

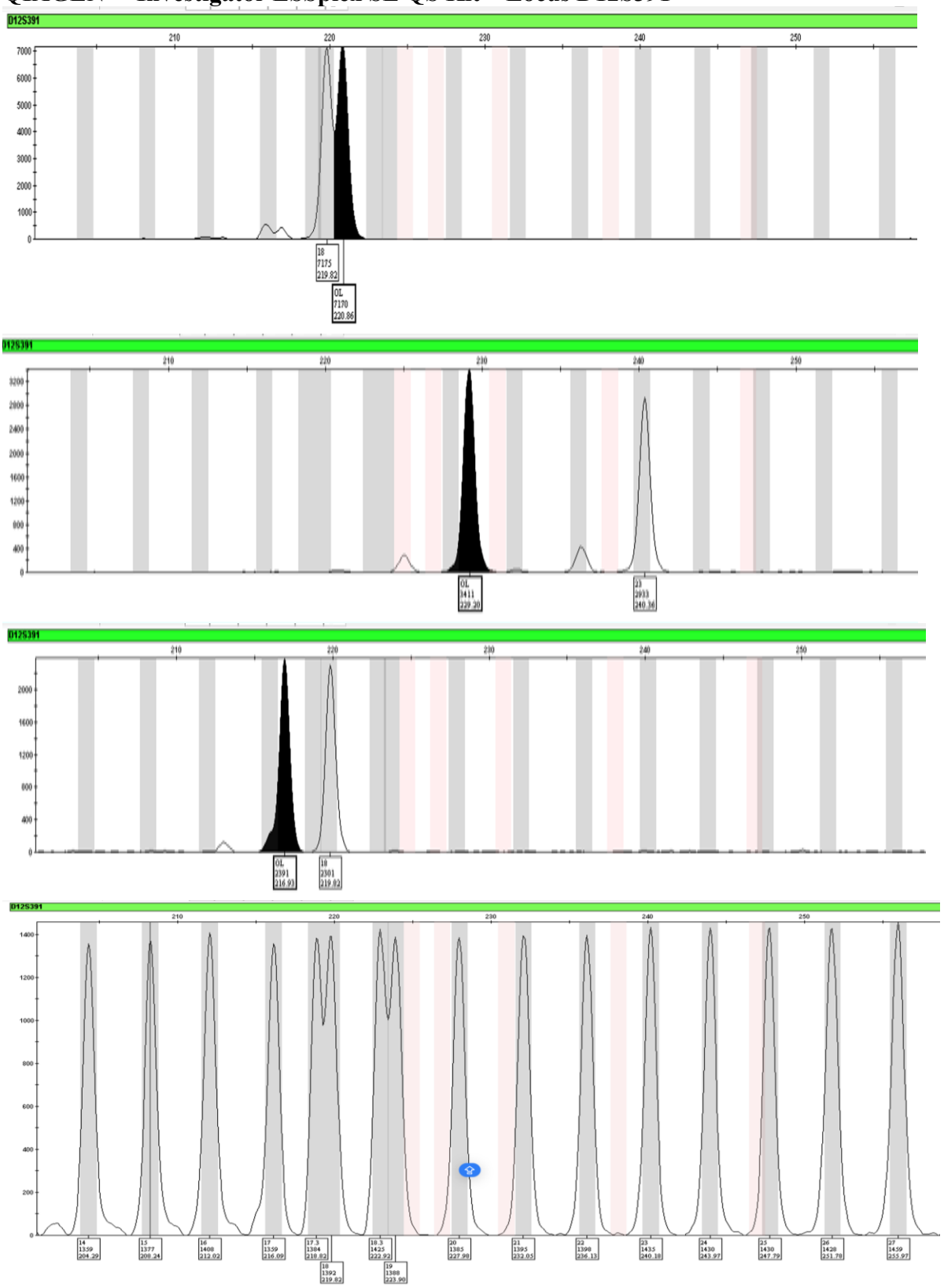
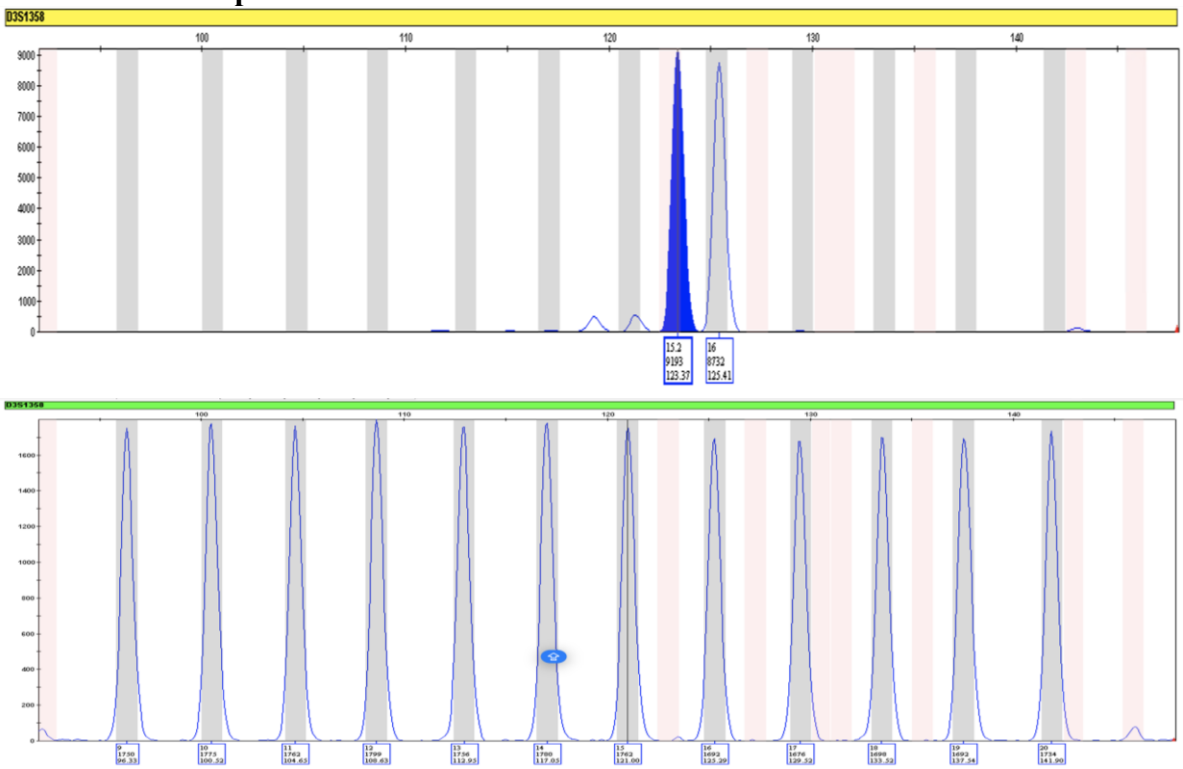



**Figure 4.2b. D12S391 alleles observed on the electropherogram from the QIAGEN™ Investigator ESSplex SE QS Kit. The three allele calls are indicated in purple for the GlobalFiler™ Express Kit (Figure 4.2a) and the off-ladder (OL) alleles and in black for the QIAGEN™ Investigator ESSplex SE QS Kit, in comparison to their respective allelic ladders of the corresponding locus shown in the lower panel. The alleles identified for the D12S391 locus are 18.1, 20.1, and 17.1**

**GlobalFiler™ Express Kit – Locus D3S1358**




**QIAGEN™ Investigator ESSplex SE QS Kit – Locus D3S1358**

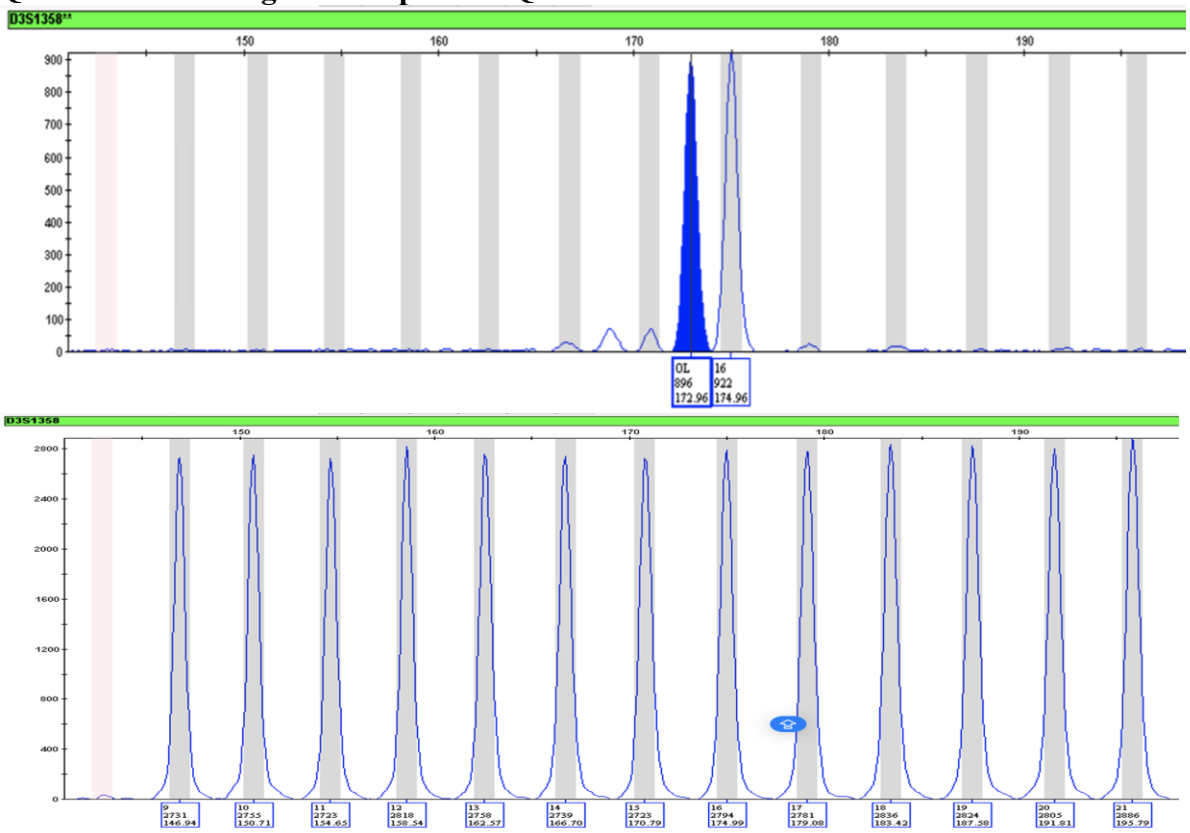
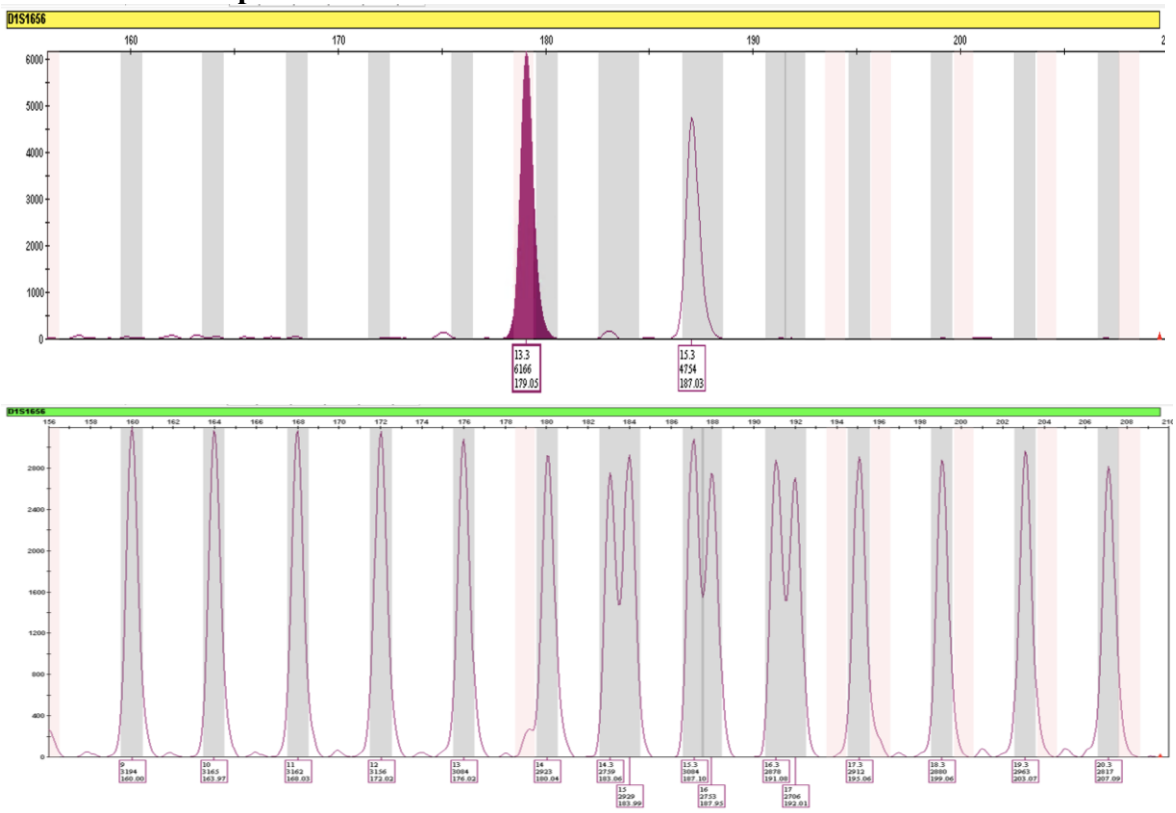



**Figure 4.3. D3S1358 off-ladder alleles observed on the electropherogram from the GlobalFiler™ Express Kit (top) and the QIAGEN™ Investigator ESSplex SE QS Kit (bottom). The off-ladder (OL) alleles are highlighted in blue for both kits, allowing for comparison with their respective allelic ladders for the corresponding locus displayed in the lower panel. The allele identified for the D3S1358 locus is 15.2.**

**GlobalFiler™ Express Kit – Locus D1S1656**




**QIAGEN™ Investigator ESSplex SE QS Kit – Locus D1S1656**

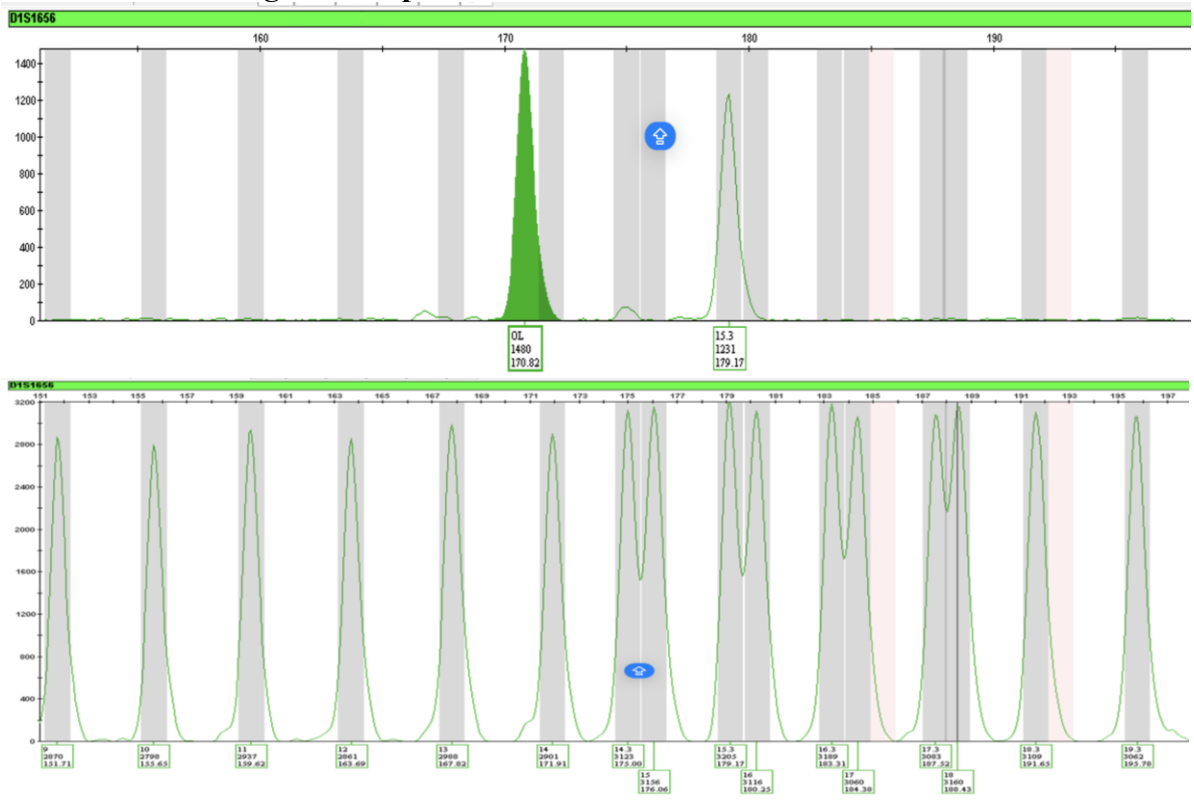



**Figure 4.4. D1S1656 off-ladder alleles observed on the electropherogram from the GlobalFiler™ Express Kit (top) and the QIAGEN™ Investigator ESSplex SE QS Kit (bottom). The off-ladder (OL) alleles are indicated in purple for the GlobalFiler™ Express Kit and in green for the QIAGEN™ Investigator ESSplex SE QS Kit, in comparison to their respective allelic ladders of the corresponding locus shown in the lower panel. The allele identified for the D3S1358 locus is 13.3.**

### 4.4.3 ALLELE FREQUENCIES AND FORENSIC PARAMETERS

The GlobalFiler™ Express PCR Amplification Kit generated complete STR profiles from blood samples on FTA cards collected from 303 unrelated Nigerians. The allele frequencies and forensic parameters of Igbo, Yoruba, and Hausa-Fulani were calculated based on the total number of alleles surveyed in the samples collected from the Nigerian population and separately calculated for the sub-populations. **Tables 4.1** and **4.2** present the allele frequencies and forensic parameters computed for 303 unrelated individuals from the Nigerian population, with all participants from the three subpopulations combined. Among the 303 Nigerian populations were 102, 101 and 100 Igbos, Yorubas and Hausa-Fulanis subpopulations, respectively. **Tables 4.3** and **Table 4.4, Tables 4.5** and **Table 4.6, and Tables 4.7** and **Table 4.8** display the allele frequencies and forensic parameters generated for the Igbo, Yoruba, and Hausa-Fulani sub-populations, respectively. **Table 4.9** presents the private alleles observed among the sub-populations.

The two software programs, STR Analysis for Forensic (*STRAF*-A 2.15), an online tool for analysing STR data, and Genetic Analysis in Excel (GenAIEx 6.5), used for statistical data analysis, facilitated quicker and more practical outputs. These software tools also generated graphical representations (see **Appendices 27 - 47**) after calculating various allele frequencies and forensic parameters.

**Table 4.1. Allele frequencies among 303 Nigerian individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit**

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | | | | 0.002 | | | | | | | | | | | | | | | | | |
| 6 | 0.003 | | | | | | | | | | | | | | 0.002 | | | | 0.127 | 0.073 | |
| 7 | 0.045 | | | 0.005 | | | | | | | | | | | | | | | 0.492 | 0.040 | |
| 8 | 0.083 | 0.003 | | 0.010 | 0.020 | | | | | 0.005 | | | | 0.081 | 0.172 | | | | 0.203 | 0.289 | |
| 9 | 0.045 | 0.002 | | 0.007 | 0.251 | 0.002 | 0.002 | | | | | | 0.003 | 0.015 | 0.153 | | | | 0.135 | 0.259 | |
| 9.3 | | | | | | | | | | | | | | | | | | | 0.041 | | |
| 10 | 0.295 | 0.018 | | 0.028 | 0.124 | | 0.010 | 0.015 | | 0.048 | | 0.066 | | 0.073 | 0.350 | 0.015 | | | 0.002 | 0.127 | |
| 10.2 | 0.002 | | | | | 0.003 | | | | | | | | | | | | 0.003 | | | |
| 10.3 | 0.003 | | | | | | | | | | | | | | | | | | | | |
| 11 | 0.203 | 0.040 | | 0.276 | 0.302 | 0.008 | 0.089 | 0.068 | | 0.158 | | 0.403 | | 0.233 | 0.200 | 0.036 | | | | 0.177 | |
| 11.3 | | | | | | | | | | | | 0.043 | | | | | | | | | |
| 12 | 0.236 | 0.183 | | 0.452 | 0.153 | 0.051 | 0.130 | 0.040 | | 0.053 | | 0.167 | 0.010 | 0.305 | 0.104 | 0.119 | | 0.005 | | 0.033 | |
| 12.2 | | | | | | | 0.048 | | | | | | | | | | | 0.008 | | | |
| 12.3 | 0.002 | | | | | | | | | | | 0.005 | | | | | | | | | |
| 13 | 0.069 | 0.224 | | 0.158 | 0.142 | 0.020 | 0.259 | 0.144 | | | | 0.040 | 0.003 | 0.266 | 0.017 | 0.195 | | 0.003 | | 0.002 | 0.026 |
| 13.2 | | | | | | 0.018 | 0.066 | | | | | | | | | | | | | | |
| 13.3 | | | | 0.002 | | | | 0.002 | | | | 0.002 | | | | | | | | | |
| 14 | 0.012 | 0.254 | | 0.059 | 0.008 | 0.040 | 0.177 | 0.254 | | 0.096 | | 0.239 | 0.111 | 0.013 | 0.003 | 0.366 | | 0.046 | | 0.002 | 0.071 |
| 14.2 | | | | | | 0.002 | 0.079 | | | | | | | | | | | 0.005 | | | |
| 14.3 | | | | | | | | 0.002 | | | | | | | | | | 0.002 | | | |
| 15 | 0.002 | 0.177 | 0.074 | 0.002 | | 0.188 | 0.041 | 0.177 | | 0.213 | 0.002 | 0.036 | 0.274 | 0.013 | | 0.228 | | 0.035 | | | 0.188 |
| 15.2 | | | | | | | 0.053 | | | | | | 0.002 | | | | | 0.005 | | | |
| 15.3 | | | | | | | | 0.008 | | | | | | | | | | | | | |
| 16 | | 0.086 | 0.051 | | | 0.200 | 0.007 | 0.112 | | 0.168 | 0.069 | | 0.318 | 0.002 | | 0.031 | | 0.079 | | | 0.290 |
| 16.1 | | | | | | | | | | | | | | | | | 0.002 | | | | |
| 16.2 | | | | | | | 0.038 | | | | | | | | | | | | | | |
| 16.3 | | | | | | | | 0.104 | | | | | | | | | | | | | |
| 17 | | 0.010 | 0.135 | | | 0.170 | | 0.015 | | 0.221 | 0.089 | | 0.229 | | | 0.010 | 0.002 | 0.099 | | | 0.186 |
| 17.1 | | | 0.002 | | | | | | | | | | | | | | | | | | |
| 17.2 | | | | | | | 0.002 | | | | | | | | | | | | | | |
| 17.3 | | | 0.003 | | | | | 0.033 | | | | | | | | | | | | | |
| 18 | 0.002 | 0.002 | 0.317 | | | 0.107 | | 0.003 | | 0.035 | 0.031 | | 0.050 | | | | 0.012 | 0.102 | | | 0.124 |
| 18.1 | | | 0.002 | | | | | | | | | | | | | | | | | | |
| 18.2 | | | | | | | | | | | | | | | | | 0.021 | | | | |
| 18.3 | | | 0.002 | | | | | 0.017 | | | | | | | | | | | | | |

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | | 0.002 | 0.185 | | | 0.114 | | 0.003 | | 0.003 | 0.191 | | 0.002 | | | | 0.066 | 0.140 | | | 0.074 |
| 19.1 | | | 0.008 | | | | | | | | | | | | | | | | | | |
| 19.2 | | | | | | 0.002 | | | | | | | | | | | 0.010 | 0.003 | | | |
| 19.3 | | | | | | | | 0.007 | | | | | | | | | | | | | |
| 20 | | | 0.134 | | | 0.040 | | | | | 0.063 | | | | | | 0.040 | 0.111 | | | 0.033 |
| 20.1 | | | 0.002 | | | | | | | | | | | | | | | | | | |
| 20.2 | | | | | | | | | | | | | | | | | 0.005 | 0.003 | | | |
| 21 | | | 0.046 | | | 0.026 | | | | | 0.153 | | | | | | 0.112 | 0.084 | | | 0.005 |
| 21.2 | | | | | | | | | | | | | | | | | | 0.003 | | | |
| 22 | | | 0.018 | | | 0.008 | | | | | 0.165 | | | | | | 0.216 | 0.023 | | | 0.002 |
| 22.2 | | | | | | | | | | | | | | | | | 0.002 | 0.008 | | | |
| 23 | | | 0.021 | | | 0.002 | | | | | 0.078 | | | | | | 0.172 | | | | |
| 23.2 | | | | | | | | | | | | | | | | | | 0.008 | | | |
| 24 | | | 0.003 | | | | | | | | 0.069 | | | | | | 0.168 | | | | |
| 24.2 | | | | | | | | | | | | | | | | | | 0.010 | | | |
| 25 | | | 0.002 | | | | | | | | 0.064 | | | | | | 0.073 | 0.002 | | | |
| 25.2 | | | | | | | | | | | | | | | | | | 0.026 | | | |
| 26 | | | | | | | | | 0.005 | | 0.017 | | | | | | 0.045 | 0.002 | | | |
| 26.2 | | | | | | | | | | | | | | | | | | 0.056 | | | |
| 27 | | | | | | | | | 0.079 | | 0.008 | | | | | | 0.030 | 0.002 | | | |
| 27.2 | | | | | | | | | | | | | | | | | | 0.048 | | | |
| 28 | | | | | | | | | 0.297 | | | | | | | | 0.020 | 0.003 | | | |
| 28.2 | | | | | | | | | | | | | | | | | | 0.038 | | | |
| 29 | | | | | | | | | 0.175 | | | | | | | | 0.005 | | | | |
| 29.2 | | | | | | | | | | | | | | | | | | 0.017 | | | |
| 30 | | | | | | | | | 0.117 | | | | | | | | | | | | |
| 30.2 | | | | | | | | | 0.031 | | | | | | | | | 0.017 | | | |
| 31 | | | | | | | | | 0.079 | | | | | | | | | | | | |
| 31.2 | | | | | | | | | 0.050 | | | | | | | | 0.002 | 0.002 | | | |
| 32 | | | | | | | | | 0.015 | | | | | | | | | | | | |
| 32.2 | | | | | | | | | 0.053 | | | | | | | | | | | | |
| 33 | | | | | | | | | 0.007 | | | | | | | | | | | | |
| 33.2 | | | | | | | | | 0.031 | | | | | | | | | 0.002 | | | |
| 34 | | | | | | | | | 0.015 | | | | | | | | | | | | |
| 34.2 | | | | | | | | | 0.002 | | | | | | | | | | | | |
| 35 | | | | | | | | | 0.038 | | | | | | | | | | | | |
| 36 | | | | | | | | | 0.007 | | | | | | | | | | | | |

**Table 4.1. continued**

**Table 4.2. Forensic parameters for 303 Nigerian individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_{exp}$ | $H_o$ | HWE | RMP | PIC | PD | PE | TPI | $F_{ST}$ | $F_{IT}$ | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CSF1PO** | 303 | 606 | 14 | 4.912 | 0.802 | 0.801 | 0.199 | 0.960 | 0.067 | 0.773 | 0.933 | 0.603 | 2.525 | 0.004 | -0.007 | -0.011 |
| **D10S1248** | 303 | 606 | 12 | 5.276 | 0.828 | 0.812 | 0.188 | 0.729 | 0.067 | 0.784 | 0.933 | 0.653 | 2.913 | 0.004 | -0.022 | -0.026 |
| **D12S391** | 303 | 606 | 14 | 5.352 | 0.835 | 0.819 | 0.181 | 0.896 | 0.057 | 0.797 | 0.943 | 0.665 | 3.030 | 0.002 | -0.027 | -0.029 |
| **D13S317** | 303 | 606 | 11 | 3.221 | 0.677 | 0.691 | 0.309 | 0.706 | 0.150 | 0.642 | 0.850 | 0.393 | 1.546 | 0.002 | 0.018 | 0.016 |
| **D16S539** | 303 | 606 | 7 | 4.682 | 0.762 | 0.788 | 0.212 | 0.982 | 0.077 | 0.754 | 0.923 | 0.531 | 2.104 | 0.005 | 0.030 | 0.025 |
| **D18S51** | 303 | 606 | 18 | 7.290 | 0.884 | 0.865 | 0.135 | 0.849 | 0.035 | 0.850 | 0.965 | 0.764 | 4.329 | 0.004 | -0.025 | -0.029 |
| **D19S433** | 303 | 606 | 14 | 4.510 | 0.848 | 0.859 | 0.141 | 0.443 | 0.036 | 0.843 | 0.964 | 0.691 | 3.293 | 0.002 | -0.090 | -0.092 |
| **D1S1656** | 303 | 606 | 16 | 5.680 | 0.858 | 0.854 | 0.146 | 0.331 | 0.041 | 0.836 | 0.959 | 0.711 | 3.523 | 0.004 | -0.042 | -0.045 |
| **D21S11** | 303 | 606 | 16 | 5.830 | 0.848 | 0.847 | 0.153 | 0.072 | 0.046 | 0.831 | 0.954 | 0.691 | 3.293 | 0.006 | -0.024 | -0.029 |
| **D22S1045** | 303 | 606 | 10 | 6.131 | 0.838 | 0.838 | 0.162 | 0.799 | 0.050 | 0.816 | 0.950 | 0.672 | 3.092 | 0.004 | -0.002 | -0.006 |
| **D2S1338** | 303 | 606 | 13 | 8.307 | 0.871 | 0.881 | 0.119 | 0.110 | 0.031 | 0.868 | 0.969 | 0.737 | 3.885 | 0.002 | 0.010 | 0.006 |
| **D2S441** | 303 | 606 | 9 | 3.418 | 0.736 | 0.745 | 0.255 | 0.878 | 0.101 | 0.708 | 0.899 | 0.486 | 1.894 | 0.002 | -0.041 | -0.043 |
| **D3S1358** | 303 | 606 | 9 | 4.100 | 0.759 | 0.757 | 0.243 | 0.994 | 0.105 | 0.715 | 0.895 | 0.525 | 2.075 | 0.004 | -0.004 | -0.007 |
| **D5S818** | 303 | 606 | 9 | 4.342 | 0.733 | 0.771 | 0.229 | 0.499 | 0.094 | 0.733 | 0.906 | 0.481 | 1.870 | 0.007 | 0.048 | 0.041 |
| **D7S820** | 303 | 606 | 8 | 4.418 | 0.739 | 0.775 | 0.225 | 0.310 | 0.084 | 0.741 | 0.916 | 0.492 | 1.918 | 0.005 | 0.044 | 0.040 |
| **D8S1179** | 303 | 606 | 8 | 4.154 | 0.776 | 0.761 | 0.239 | 0.996 | 0.094 | 0.724 | 0.906 | 0.555 | 2.228 | 0.004 | -0.022 | -0.026 |
| **FGA** | 303 | 606 | 19 | 7.395 | 0.855 | 0.869 | 0.131 | 0.498 | 0.034 | 0.854 | 0.966 | 0.704 | 3.443 | 0.005 | 0.011 | 0.007 |
| **SE33** | 303 | 606 | 34 | 12.151 | 0.921 | 0.924 | 0.076 | 0.346 | 0.015 | 0.917 | 0.985 | 0.838 | 6.313 | 0.003 | -0.004 | -0.007 |
| **TH01** | 303 | 606 | 6 | 3.027 | 0.700 | 0.682 | 0.318 | 0.977 | 0.145 | 0.640 | 0.855 | 0.428 | 1.665 | 0.003 | -0.045 | -0.048 |
| **TPOX** | 303 | 606 | 9 | 4.860 | 0.776 | 0.796 | 0.204 | 0.720 | 0.074 | 0.765 | 0.926 | 0.555 | 2.228 | 0.006 | 0.023 | 0.018 |
| **vWA** | 303 | 606 | 10 | 5.489 | 0.812 | 0.819 | 0.181 | 0.902 | 0.059 | 0.794 | 0.941 | 0.621 | 2.658 | 0.003 | 0.007 | 0.004 |

| | | |
|---|---|---|
| N | - | Number of unrelated individuals |
| Allele (n) | - | Number of alleles surveyed |
| Allele (v) | - | Number of different allelic variants |
| Allele (e) | - | Number of effective alleles |
| $H_o$ | - | Observed heterozygosity |
| $H_{exp}$ | - | Expected heterozygosity |
| $H_{om\_exp}$ | - | Expected homozygosity |
| RMP | - | Random Match Probability |
| HWE | - | Hardy-Weinberg Equilibrium probability |
| PIC | - | Polymorphism Information Content |
| PD | - | Power of Discrimination |
| PE | - | Power of Exclusion |
| TPI | - | Typical Paternity Index |
| $F_{ST}$ | - | Fixation Index - **S**ubpopulation within the **T**otal population |
| $F_{IT}$ | - | Fixation Index - **I**ndividual within the **T**otal population |
| $F_{IS}$ | - | Fixation Index - **I**ndividual within the **S**ubpopulation |

**Table 4.3. Allele frequencies among 102 Igbo individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit**

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **5** | | | | 0.005 | | | | | | | | | | | | | | | | | |
| **6** | 0.005 | | | | | | | | | | | | | | | | | | 0.127 | 0.093 | |
| **7** | 0.034 | | | 0.005 | | | | | | | | | | | | | | | 0.495 | 0.049 | |
| **8** | 0.069 | 0.010 | | 0.010 | 0.015 | | | | | 0.005 | | | | 0.078 | 0.157 | | | | 0.225 | 0.260 | |
| **9** | 0.025 | | | 0.005 | 0.206 | 0.005 | 0.005 | | | | | | | 0.015 | 0.152 | | | | 0.113 | 0.255 | |
| **9.3** | | | | | | | | | | | | | | | | | | | 0.034 | | |
| **10** | 0.314 | 0.010 | | 0.034 | 0.132 | | | 0.010 | | 0.044 | | 0.049 | | 0.034 | 0.392 | 0.010 | | | 0.005 | 0.123 | |
| **10.2** | | | | | | 0.005 | | | | | | | | | | | | | | | |
| **11** | 0.255 | 0.020 | | 0.260 | 0.363 | 0.010 | 0.113 | 0.054 | | 0.191 | | 0.402 | | 0.289 | 0.167 | 0.034 | | | | 0.167 | |
| **11.3** | | | | | | | | | | | | 0.029 | | | | | | | | | |
| **12** | 0.211 | 0.157 | | 0.461 | 0.162 | 0.044 | 0.118 | 0.044 | | 0.059 | | 0.181 | | 0.250 | 0.108 | 0.113 | | 0.010 | | 0.049 | |
| **12.2** | | | | | | | 0.054 | | | | | | | | | | | 0.010 | | | |
| **12.3** | 0.005 | | | | | | | | | | | 0.010 | | | | | | | | | |
| **13** | 0.074 | 0.221 | | 0.172 | 0.113 | 0.020 | 0.240 | 0.113 | | | | 0.044 | 0.005 | 0.304 | 0.025 | 0.167 | | 0.010 | | | 0.034 |
| **13.2** | | | | | | 0.039 | 0.059 | | | | | | | | | | | | | | |
| **13.3** | | | | 0.005 | | | | | | | | | | | | | | | | | |
| **14** | 0.010 | 0.294 | | 0.044 | 0.010 | 0.034 | 0.181 | 0.294 | | 0.088 | | 0.240 | 0.118 | | | 0.417 | | 0.054 | | 0.005 | 0.088 |
| **14.2** | | | | | | 0.005 | 0.083 | | | | | | | | | | | 0.005 | | | |
| **15** | | 0.201 | 0.059 | | | 0.186 | 0.020 | 0.162 | | 0.206 | | 0.044 | 0.279 | 0.025 | | 0.221 | | 0.054 | | | 0.157 |
| **15.2** | | | | | | | 0.059 | | | | | | 0.005 | | | | | | | | |
| **15.3** | | | | | | | | 0.010 | | | | | | | | | | | | | |
| **16** | | 0.064 | 0.059 | | | 0.245 | | 0.132 | | 0.157 | 0.069 | | 0.348 | 0.005 | | 0.029 | | 0.059 | | | 0.333 |
| **16.2** | | | | | | | 0.069 | | | | | | | | | | | | | | |
| **16.3** | | | | | | | | 0.108 | | | | | | | | | | | | | |
| **17** | | 0.015 | 0.118 | | | 0.147 | | 0.025 | | 0.206 | 0.098 | | 0.225 | | | 0.010 | | 0.108 | | | 0.181 |
| **17.1** | | | 0.005 | | | | | | | | | | | | | | | | | | |
| **17.3** | | | | | | | | 0.039 | | | | | | | | | | | | | |
| **18** | | 0.005 | 0.324 | | | 0.108 | | | | 0.044 | 0.029 | | 0.020 | | | | | 0.108 | | | 0.103 |
| **18.2** | | | | | | | | | | | | | | | | | 0.025 | | | | |
| **18.3** | | | | | | | | 0.005 | | | | | | | | | | | | | |
| **19** | | 0.005 | 0.201 | | | 0.088 | | | | | 0.181 | | 0.005 | | | | 0.113 | 0.157 | | | 0.069 |
| **19.1** | | | 0.015 | | | | | | | | | | | | | | | | | | |
| **19.2** | | | | | | 0.005 | | | | | | | | | | | 0.005 | 0.005 | | | |

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **19.3** | | | | | | | | 0.005 | | | | | | | | | | | | | |
| **20** | | | 0.147 | | | 0.029 | | | | | 0.059 | | | | | | 0.059 | 0.083 | | | 0.029 |
| **20.2** | | | | | | | | | | | | | | | | | 0.005 | 0.005 | | | |
| **21** | | | 0.044 | | | 0.029 | | | | | 0.167 | | | | | | 0.093 | 0.078 | | | 0.005 |
| **22** | | | 0.025 | | | | | | | | 0.137 | | | | | | 0.230 | 0.020 | | | |
| **22.2** | | | | | | | | | | | | | | | | | | 0.010 | | | |
| **23** | | | 0.010 | | | | | | | | 0.074 | | | | | | 0.176 | | | | |
| **23.2** | | | | | | | | | | | | | | | | | | 0.010 | | | |
| **24** | | | | | | | | | | | 0.098 | | | | | | 0.167 | | | | |
| **24.2** | | | | | | | | | | | | | | | | | | 0.005 | | | |
| **25** | | | | | | | | | | | 0.049 | | | | | | 0.069 | | | | |
| **25.2** | | | | | | | | | | | | | | | | | | 0.029 | | | |
| **26** | | | | | | | | | | | 0.029 | | | | | | 0.020 | | | | |
| **26.2** | | | | | | | | | | | | | | | | | | 0.059 | | | |
| **27** | | | | | | | | | 0.069 | | 0.010 | | | | | | 0.025 | | | | |
| **27.2** | | | | | | | | | | | | | | | | | | 0.069 | | | |
| **28** | | | | | | | | | 0.319 | | | | | | | | 0.010 | | | | |
| **28.2** | | | | | | | | | | | | | | | | | | 0.034 | | | |
| **29** | | | | | | | | | 0.147 | | | | | | | | 0.005 | | | | |
| **29.2** | | | | | | | | | | | | | | | | | | 0.005 | | | |
| **30** | | | | | | | | | 0.137 | | | | | | | | | | | | |
| **30.2** | | | | | | | | | 0.025 | | | | | | | | | 0.015 | | | |
| **31** | | | | | | | | | 0.098 | | | | | | | | | | | | |
| **31.2** | | | | | | | | | 0.044 | | | | | | | | | | | | |
| **32** | | | | | | | | | 0.010 | | | | | | | | | | | | |
| **32.2** | | | | | | | | | 0.039 | | | | | | | | | | | | |
| **33** | | | | | | | | | 0.005 | | | | | | | | | | | | |
| **33.2** | | | | | | | | | 0.025 | | | | | | | | | | | | |
| **34** | | | | | | | | | 0.015 | | | | | | | | | | | | |
| **35** | | | | | | | | | 0.064 | | | | | | | | | | | | |
| **36** | | | | | | | | | 0.005 | | | | | | | | | | | | |

**Table 4.3. continued**

**Table 4.4. Forensic parameters for 102 Igbo individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit.**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_{exp}$ | $H_o$ | HWE | RMP | PIC | PD | PE | TPI | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CSF1PO** | 102 | 204 | 10 | 4.506 | 0.814 | 0.784 | 0.216 | 0.945 | 0.089 | 0.748 | 0.911 | 0.625 | 2.684 | -0.046 |
| **D10S1248** | 102 | 204 | 11 | 4.876 | 0.794 | 0.799 | 0.201 | 0.089 | 0.085 | 0.765 | 0.915 | 0.588 | 2.429 | 0.001 |
| **D12S391** | 102 | 204 | 10 | 5.096 | 0.833 | 0.814 | 0.186 | 0.698 | 0.069 | 0.787 | 0.931 | 0.662 | 3.000 | -0.037 |
| **D13S317** | 102 | 204 | 10 | 3.182 | 0.647 | 0.691 | 0.309 | 0.230 | 0.152 | 0.640 | 0.848 | 0.351 | 1.417 | 0.056 |
| **D16S539** | 102 | 204 | 7 | 4.335 | 0.725 | 0.773 | 0.227 | 0.852 | 0.087 | 0.736 | 0.913 | 0.469 | 1.821 | 0.057 |
| **D18S51** | 102 | 204 | 16 | 6.872 | 0.843 | 0.861 | 0.139 | 0.139 | 0.043 | 0.842 | 0.957 | 0.681 | 3.188 | 0.013 |
| **D19S433** | 102 | 204 | 11 | 4.706 | 0.873 | 0.865 | 0.135 | 0.993 | 0.036 | 0.847 | 0.963 | 0.740 | 3.923 | -0.108 |
| **D1S1656** | 102 | 204 | 13 | 5.118 | 0.824 | 0.842 | 0.158 | 0.164 | 0.052 | 0.821 | 0.948 | 0.643 | 2.833 | -0.024 |
| **D21S11** | 102 | 204 | 14 | 5.500 | 0.833 | 0.839 | 0.161 | 0.054 | 0.052 | 0.818 | 0.948 | 0.662 | 3.000 | -0.019 |
| **D22S1045** | 102 | 204 | 9 | 6.208 | 0.843 | 0.843 | 0.157 | 0.611 | 0.053 | 0.819 | 0.947 | 0.681 | 3.188 | -0.005 |
| **D2S1338** | 102 | 204 | 12 | 8.581 | 0.922 | 0.888 | 0.112 | 0.083 | 0.037 | 0.872 | 0.964 | 0.840 | 6.375 | -0.043 |
| **D2S441** | 102 | 204 | 8 | 3.489 | 0.716 | 0.744 | 0.256 | 0.404 | 0.110 | 0.704 | 0.890 | 0.453 | 1.759 | -0.003 |
| **D3S1358** | 102 | 204 | 7 | 3.783 | 0.745 | 0.739 | 0.261 | 0.876 | 0.121 | 0.689 | 0.879 | 0.501 | 1.962 | -0.013 |
| **D5S818** | 102 | 204 | 8 | 4.054 | 0.735 | 0.757 | 0.243 | 0.868 | 0.102 | 0.712 | 0.898 | 0.485 | 1.889 | 0.024 |
| **D7S820** | 102 | 204 | 6 | 4.141 | 0.706 | 0.762 | 0.238 | 0.276 | 0.094 | 0.726 | 0.906 | 0.437 | 1.700 | 0.069 |
| **D8S1179** | 102 | 204 | 8 | 3.774 | 0.745 | 0.739 | 0.261 | 0.675 | 0.110 | 0.698 | 0.890 | 0.501 | 1.962 | -0.014 |
| **FGA** | 102 | 204 | 14 | 6.897 | 0.814 | 0.861 | 0.139 | 0.169 | 0.045 | 0.841 | 0.955 | 0.625 | 2.684 | 0.048 |
| **SE33** | 102 | 204 | 24 | 11.749 | 0.882 | 0.923 | 0.177 | 0.349 | 0.020 | 0.913 | 0.980 | 0.760 | 4.250 | 0.036 |
| **TH01** | 102 | 204 | 6 | 3.026 | 0.686 | 0.681 | 0.319 | 0.700 | 0.158 | 0.634 | 0.842 | 0.407 | 1.594 | -0.025 |
| **TPOX** | 102 | 204 | 8 | 5.297 | 0.775 | 0.815 | 0.185 | 0.609 | 0.066 | 0.785 | 0.934 | 0.553 | 2.217 | 0.045 |
| **vWA** | 102 | 204 | 9 | 5.161 | 0.843 | 0.810 | 0.19 | 0.292 | 0.070 | 0.783 | 0.930 | 0.681 | 3.188 | -0.046 |

| | | |
|---|---|---|
| N | - | Number of unrelated individuals |
| Allele (n) | - | Number of alleles surveyed |
| Allele (v) | - | Number of different allelic variants |
| Allele (e) | - | Number of effective alleles |
| $H_o$ | - | Observed heterozygosity |
| $H_{exp}$ | - | Expected heterozygosity |
| $H_{om\_exp}$ | - | Expected homozygosity |
| HWE | - | Hardy-Weinberg Equilibrium probability |
| RMP | - | Random Match Probability |
| PIC | - | Polymorphism Information Content |
| PD | - | Power of Discrimination |
| PE | - | Power of Exclusion |
| TPI | - | Typical Paternity Index |
| $F_{IS}$ | - | Fixation Index - **I**ndividual within the **S**ubpopulation |

**Table 4.5. Allele frequencies among 101 Yoruba individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit**

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | | | | | | | | | | | | | | | 0.005 | | | | 0.149 | 0.074 | |
| **7** | 0.059 | | | 0.010 | | | | | | | | | | | | | | | 0.505 | 0.040 | |
| **8** | 0.104 | | | 0.005 | 0.020 | | | | | 0.010 | | | | 0.099 | 0.178 | | | | 0.193 | 0.243 | |
| **9** | 0.045 | | | 0.010 | 0.292 | | | | | | | | 0.005 | 0.010 | 0.188 | | | | 0.109 | 0.267 | |
| **9.3** | | | | | | | | | | | | | | | | | | | 0.045 | | |
| **10** | 0.257 | 0.035 | | 0.025 | 0.109 | | 0.020 | 0.005 | | 0.059 | | 0.059 | | 0.109 | 0.322 | 0.020 | | | | 0.134 | |
| **10.2** | | | | | | | | | | | | | | | | | | 0.005 | | | |
| **10.3** | 0.010 | | | | | | | | | | | | | | | | | | | | |
| **11** | 0.168 | 0.045 | | 0.257 | 0.292 | 0.010 | 0.064 | 0.074 | | 0.144 | | 0.391 | | 0.193 | 0.238 | 0.059 | | | | 0.218 | |
| **11.3** | | | | | | | | | | | | 0.050 | | | | | | | | | |
| **12** | 0.257 | 0.218 | | 0.475 | 0.129 | 0.045 | 0.144 | 0.045 | | 0.054 | | 0.163 | 0.025 | 0.317 | 0.064 | 0.144 | | | | 0.025 | |
| **12.2** | | | | | | | 0.045 | | | | | | | | | | | 0.005 | | | |
| **12.3** | | | | | | | | | | | | 0.005 | | | | | | | | | |
| **13** | 0.069 | 0.243 | | 0.144 | 0.153 | 0.020 | 0.282 | 0.158 | | | | 0.025 | | 0.238 | 0.005 | 0.198 | | | | | 0.035 |
| **13.2** | | | | | | 0.010 | 0.059 | | | | | | | | | | | | | | |
| **14** | 0.025 | 0.203 | | 0.069 | 0.005 | 0.025 | 0.153 | 0.248 | | 0.149 | | 0.272 | 0.129 | 0.030 | | 0.312 | | 0.040 | | | 0.079 |
| **14.2** | | | | | | | 0.084 | | | | | | | | | | | 0.005 | | | |
| **15** | | 0.153 | 0.099 | 0.005 | | 0.198 | 0.059 | 0.193 | | 0.208 | 0.005 | 0.030 | 0.282 | 0.005 | | 0.233 | | 0.035 | | | 0.193 |
| **15.2** | | | | | | | 0.059 | | | | | | | | | | | 0.005 | | | |
| **15.3** | | | | | | | | 0.005 | | | | | | | | | | | | | |
| **16** | | 0.099 | 0.025 | | | 0.158 | 0.005 | 0.094 | | 0.144 | 0.089 | | 0.287 | | | 0.025 | | 0.094 | | | 0.262 |
| **16.2** | | | | | | | 0.025 | | | | | | | | | | | | | | |
| **16.3** | | | | | | | | 0.084 | | | | | | | | | | | | | |
| **17** | | 0.005 | 0.158 | | | 0.183 | | 0.010 | | 0.203 | 0.059 | | 0.223 | | | 0.010 | 0.005 | 0.104 | | | 0.183 |
| **17.3** | | | | | | | | 0.025 | | | | | | | | | | | | | |
| **18** | 0.005 | | 0.302 | | | 0.109 | | 0.005 | | 0.025 | 0.020 | | 0.050 | | | | 0.015 | 0.109 | | | 0.119 |
| **18.1** | | | 0.005 | | | | | | | | | | | | | | | | | | |
| **18.2** | | | | | | | | | | | | | | | | | 0.010 | | | | |
| **18.3** | | | | | | | | 0.040 | | | | | | | | | | | | | |
| **19** | | | 0.178 | | | 0.158 | | 0.005 | | 0.005 | 0.198 | | | | | | 0.035 | 0.149 | | | 0.079 |
| **19.1** | | | 0.010 | | | | | | | | | | | | | | | | | | |
| **19.2** | | | | | | | | | | | | | | | | | 0.015 | 0.005 | | | |
| **19.3** | | | | | | | | 0.010 | | | | | | | | | | | | | |
| **20** | | | 0.139 | | | 0.045 | | | | | 0.094 | | | | | | 0.045 | 0.149 | | | 0.040 |
| **20.1** | | | 0.005 | | | | | | | | | | | | | | | | | | |

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **20.2** | | | | | | | | | | | | | | | | | 0.010 | 0.005 | | | |
| **21** | | | 0.035 | | | 0.030 | | | | | 0.124 | | | | | | 0.084 | 0.084 | | | 0.005 |
| **21.2** | | | | | | | | | | | | | | | | | | 0.005 | | | |
| **22** | | | 0.020 | | | 0.010 | | | | | 0.168 | | | | | | 0.213 | 0.020 | | | 0.005 |
| **22.2** | | | | | | | | | | | | | | | | | | 0.010 | | | |
| **22.3** | | | | | | | | | | | | | | | | | 0.005 | | | | |
| **23** | | | 0.030 | | | | | | | | 0.104 | | | | | | 0.168 | | | | |
| **23.2** | | | | | | | | | | | | | | | | | | 0.010 | | | |
| **24** | | | | | | | | | | | 0.058 | | | | | | 0.203 | | | | |
| **24.2** | | | | | | | | | | | | | | | | | | 0.010 | | | |
| **25** | | | 0.005 | | | | | | | | 0.074 | | | | | | 0.059 | 0.005 | | | |
| **26** | | | | | | | | | 0.010 | | 0.010 | | | | | | 0.069 | 0.005 | | | |
| **26.2** | | | | | | | | | | | | | | | | | | 0.050 | | | |
| **27** | | | | | | | | | 0.109 | | 0.005 | | | | | | 0.035 | 0.005 | | | |
| **27.2** | | | | | | | | | | | | | | | | | | 0.025 | | | |
| **28** | | | | | | | | | 0.267 | | | | | | | | 0.020 | | | | |
| **28.2** | | | | | | | | | | | | | | | | | | 0.020 | | | |
| **29** | | | | | | | | | 0.223 | | | | | | | | 0.010 | | | | |
| **29.2** | | | | | | | | | | | | | | | | | | 0.030 | | | |
| **30** | | | | | | | | | 0.064 | | | | | | | | | | | | |
| **30.2** | | | | | | | | | 0.030 | | | | | | | | | 0.015 | | | |
| **31** | | | | | | | | | 0.059 | | | | | | | | | | | | |
| **31.2** | | | | | | | | | 0.069 | | | | | | | | | | | | |
| **32** | | | | | | | | | 0.010 | | | | | | | | | | | | |
| **32.2** | | | | | | | | | 0.079 | | | | | | | | | | | | |
| **33** | | | | | | | | | 0.005 | | | | | | | | | | | | |
| **33.2** | | | | | | | | | 0.020 | | | | | | | | | | | | |
| **34** | | | | | | | | | 0.015 | | | | | | | | | | | | |
| **35** | | | | | | | | | 0.030 | | | | | | | | | | | | |
| **36** | | | | | | | | | 0.010 | | | | | | | | | | | | |

**Table 4.5. continued**

**Table 4.6. Forensic parameters among 101 Yoruba individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit.**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_{exp}$ | $H_o$ | HWE | RMP | PIC | PD | PE | TPI | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CSF1PO** | 101 | 202 | 10 | 5.324 | 0.812 | 0.821 | 0.179 | 0.188 | 0.066 | 0.794 | 0.934 | 0.621 | 2.658 | 0.001 |
| **D10S1248** | 101 | 202 | 8 | 5.433 | 0.842 | 0.820 | 0.18 | 0.885 | 0.065 | 0.790 | 0.935 | 0.678 | 3.156 | -0.031 |
| **D12S391** | 101 | 202 | 11 | 5.441 | 0.812 | 0.824 | 0.176 | 0.599 | 0.058 | 0.798 | 0.942 | 0.621 | 2.658 | 0.005 |
| **D13S317** | 101 | 202 | 9 | 3.141 | 0.653 | 0.685 | 0.315 | 0.567 | 0.152 | 0.636 | 0.848 | 0.360 | 1.443 | 0.041 |
| **D16S539** | 101 | 202 | 7 | 4.484 | 0.752 | 0.781 | 0.219 | 0.486 | 0.090 | 0.743 | 0.910 | 0.514 | 2.020 | 0.032 |
| **D18S51** | 101 | 202 | 13 | 7.074 | 0.891 | 0.863 | 0.137 | 0.905 | 0.042 | 0.843 | 0.958 | 0.777 | 4.591 | -0.038 |
| **D19S433** | 101 | 202 | 12 | 4.385 | 0.851 | 0.856 | 0.144 | 0.424 | 0.043 | 0.837 | 0.957 | 0.698 | 3.367 | -0.103 |
| **D1S1656** | 101 | 202 | 15 | 5.943 | 0.842 | 0.855 | 0.145 | 0.128 | 0.047 | 0.834 | 0.953 | 0.678 | 3.156 | -0.012 |
| **D21S11** | 101 | 202 | 15 | 5.945 | 0.931 | 0.850 | 0.15 | 0.008 | 0.067 | 0.830 | 0.933 | 0.858 | 7.214 | -0.119 |
| **D22S1045** | 101 | 202 | 10 | 6.454 | 0.881 | 0.849 | 0.151 | 0.102 | 0.059 | 0.826 | 0.941 | 0.757 | 4.208 | -0.043 |
| **D2S1338** | 101 | 202 | 13 | 8.093 | 0.901 | 0.883 | 0.117 | 0.411 | 0.037 | 0.866 | 0.963 | 0.797 | 5.050 | -0.028 |
| **D2S441** | 101 | 202 | 8 | 3.297 | 0.703 | 0.742 | 0.258 | 0.922 | 0.102 | 0.700 | 0.898 | 0.433 | 1.683 | -0.009 |
| **D3S1358** | 101 | 202 | 7 | 4.322 | 0.733 | 0.772 | 0.228 | 0.655 | 0.094 | 0.731 | 0.906 | 0.481 | 1.870 | 0.047 |
| **D5S818** | 101 | 202 | 8 | 4.613 | 0.802 | 0.787 | 0.213 | 0.537 | 0.094 | 0.751 | 0.906 | 0.603 | 2.525 | -0.024 |
| **D7S820** | 101 | 202 | 7 | 4.322 | 0.733 | 0.772 | 0.228 | 0.720 | 0.092 | 0.731 | 0.908 | 0.481 | 1.870 | 0.047 |
| **D8S1179** | 101 | 202 | 8 | 4.633 | 0.802 | 0.788 | 0.212 | 0.928 | 0.081 | 0.752 | 0.919 | 0.603 | 2.525 | -0.023 |
| **FGA** | 101 | 202 | 17 | 7.136 | 0.871 | 0.868 | 0.132 | 0.607 | 0.041 | 0.850 | 0.959 | 0.737 | 3.885 | -0.013 |
| **SE33** | 101 | 202 | 27 | 10.351 | 0.921 | 0.914 | 0.086 | 0.313 | 0.024 | 0.902 | 0.976 | 0.838 | 6.375 | -0.019 |
| **TH01** | 101 | 202 | 5 | 2.960 | 0.683 | 0.675 | 0.325 | 0.877 | 0.150 | 0.631 | 0.850 | 0.403 | 1.578 | -0.032 |
| **TPOX** | 101 | 202 | 7 | 4.919 | 0.792 | 0.801 | 0.199 | 0.543 | 0.080 | 0.767 | 0.920 | 0.584 | 2.405 | 0.006 |
| **vWA** | 101 | 202 | 10 | 5.912 | 0.752 | 0.835 | 0.165 | 0.269 | 0.054 | 0.810 | 0.946 | 0.514 | 2.020 | 0.094 |

N - Number of unrelated individuals
Allele (n) - Number of alleles surveyed
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
$H_o$ - Observed heterozygosity
$H_{exp}$ - Expected heterozygosity
$H_{om\_exp}$ - Expected homozygosity
HWE - Hardy-Weinberg Equilibrium probability
RMP - Random Match Probability
PIC - Polymorphism Information Content
PD - Power of Discrimination
PE - Power of Exclusion
TPI - Typical Paternity Index
$F_{IS}$ - Fixation Index - **I**ndividual within the **S**ubpopulation

**Table 4.7. Allele frequencies among 100 Hausa-Fulani individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit**

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | 0.005 | | | | | | | | | | | | | | | | | | 0.105 | 0.05 | |
| **7** | 0.04 | | | | | | | | | | | | | | | | | | 0.48 | 0.03 | |
| **8** | 0.075 | | | 0.015 | 0.025 | | | | | | | | | 0.065 | 0.18 | | | | 0.19 | 0.365 | |
| **9** | 0.065 | 0.005 | | 0.005 | 0.255 | | | | | | | | 0.005 | 0.02 | 0.12 | | | | 0.18 | 0.255 | |
| **9.3** | | | | | | | | | | | | | | | | | | | 0.045 | | |
| **10** | 0.315 | 0.01 | | 0.025 | 0.125 | | 0.01 | 0.03 | | 0.04 | | 0.09 | | 0.075 | 0.335 | 0.015 | | | | 0.125 | |
| **10.2** | 0.005 | | | | | 0.005 | | | | | | | | | | | | 0.005 | | | |
| **11** | 0.185 | 0.055 | | 0.31 | 0.255 | 0.005 | 0.09 | 0.075 | | 0.14 | | 0.415 | | 0.215 | 0.195 | 0.015 | | | | 0.145 | |
| **11.3** | | | | | | | | | | | | 0.05 | | | | | | | | | |
| **12** | 0.24 | 0.175 | | 0.42 | 0.17 | 0.065 | 0.13 | 0.03 | | 0.045 | | 0.15 | 0.005 | 0.35 | 0.14 | 0.1 | | 0.005 | | 0.025 | |
| **12.2** | | | | | | | 0.045 | | | | | | | | | | | 0.01 | | | |
| **13** | 0.065 | 0.21 | | 0.16 | 0.16 | 0.02 | 0.255 | 0.16 | | | | 0.05 | 0.005 | 0.255 | 0.02 | 0.22 | | | | 0.005 | 0.01 |
| **13.2** | | | | | | 0.005 | 0.08 | | | | | | | | | | | | | | |
| **13.3** | | | | | | | | 0.005 | | | | 0.005 | | | | | | | | | |
| **14** | | 0.265 | | 0.065 | 0.01 | 0.06 | 0.195 | 0.22 | | 0.05 | | 0.205 | 0.085 | 0.01 | 0.01 | 0.37 | | 0.045 | | | 0.045 |
| **14.2** | | | | | | | 0.07 | | | | | | | | | | | 0.005 | | | |
| **14.3** | | | | | | | | 0.005 | | | | | | | | | | 0.005 | | | |
| **15** | 0.005 | 0.175 | 0.065 | | | 0.18 | 0.045 | 0.175 | | 0.225 | | 0.035 | 0.26 | 0.01 | | 0.23 | | 0.015 | | | 0.215 |
| **15.2** | | | | | | | 0.04 | | | | | | | | | | | 0.01 | | | |
| **15.3** | | | | | | | | 0.01 | | | | | | | | | | | | | |
| **16** | | 0.095 | 0.07 | | | 0.195 | 0.015 | 0.11 | | 0.205 | 0.055 | | 0.32 | | | 0.04 | | 0.085 | | | 0.275 |
| **16.1** | | | | | | | | | | | | | | | | | 0.005 | | | | |
| **16.2** | | | | | | | 0.02 | | | | | | | | | | | | | | |
| **16.3** | | | | | | | | 0.12 | | | | | | | | | | | | | |
| **17** | | 0.01 | 0.13 | | | 0.18 | | 0.01 | | 0.255 | 0.115 | | 0.24 | | | 0.01 | | 0.085 | | | 0.195 |
| **17.2** | | | | | | | 0.005 | | | | | | | | | | | | | | |
| **17.3** | | | 0.01 | | | | | 0.035 | | | | | | | | | | | | | |
| **18** | | | 0.325 | | | 0.105 | | 0.005 | | 0.035 | 0.045 | | 0.08 | | | | 0.02 | 0.09 | | | 0.15 |
| **18.2** | | | | | | | | | | | | | | | | | 0.03 | | | | |
| **18.3** | | | 0.005 | | | | | 0.005 | | | | | | | | | | | | | |
| **19** | | | 0.175 | | | 0.095 | | 0.005 | | 0.005 | 0.19 | | | | | | 0.05 | 0.115 | | | 0.075 |
| **19.2** | | | | | | | | | | | | | | | | | 0.01 | | | | |
| **19.3** | | | | | | | | 0.005 | | | | | | | | | | | | | |
| **20** | | | 0.115 | | | 0.045 | | | | | 0.035 | | | | | | 0.015 | 0.1 | | | 0.03 |
| **21** | | | 0.06 | | | 0.02 | | | | | 0.17 | | | | | | 0.16 | 0.09 | | | 0.005 |

| Allele | CSF1PO | D10S1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D2S441 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | SE33 | TH01 | TPOX | vWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **22** | | | 0.01 | | | 0.015 | | | | | 0.19 | | | | | | 0.205 | 0.035 | | | |
| **22.2** | | | | | | | | | | | | | | | | | | 0.005 | | | |
| **23** | | | 0.025 | | | 0.005 | | | | | 0.055 | | | | | | 0.17 | | | | |
| **23.2** | | | | | | | | | | | | | | | | | | 0.005 | | | |
| **24** | | | 0.01 | | | | | | | | 0.055 | | | | | | 0.135 | | | | |
| **24.2** | | | | | | | | | | | | | | | | | | 0.015 | | | |
| **25** | | | | | | | | | | | 0.07 | | | | | | 0.09 | 0.005 | | | |
| **25.2** | | | | | | | | | | | | | | | | | | 0.045 | | | |
| **26** | | | | | | | | | 0.005 | | 0.01 | | | | | | 0.045 | | | | |
| **26.2** | | | | | | | | | | | | | | | | | | 0.06 | | | |
| **27** | | | | | | | | | 0.06 | | 0.01 | | | | | | 0.03 | | | | |
| **27.2** | | | | | | | | | | | | | | | | | | 0.05 | | | |
| **28** | | | | | | | | | 0.305 | | | | | | | | 0.03 | 0.01 | | | |
| **28.2** | | | | | | | | | | | | | | | | | | 0.06 | | | |
| **29** | | | | | | | | | 0.155 | | | | | | | | | | | | |
| **29.2** | | | | | | | | | | | | | | | | | | 0.015 | | | |
| **30** | | | | | | | | | 0.15 | | | | | | | | | | | | |
| **30.2** | | | | | | | | | 0.04 | | | | | | | | | 0.02 | | | |
| **31** | | | | | | | | | 0.08 | | | | | | | | | | | | |
| **31.2** | | | | | | | | | 0.035 | | | | | | | | 0.005 | 0.005 | | | |
| **32** | | | | | | | | | 0.025 | | | | | | | | | | | | |
| **32.2** | | | | | | | | | 0.04 | | | | | | | | | | | | |
| **33** | | | | | | | | | 0.01 | | | | | | | | | | | | |
| **33.2** | | | | | | | | | 0.05 | | | | | | | | | 0.005 | | | |
| **34** | | | | | | | | | 0.015 | | | | | | | | | | | | |
| **34.2** | | | | | | | | | 0.005 | | | | | | | | | | | | |
| **35** | | | | | | | | | 0.02 | | | | | | | | | | | | |
| **36** | | | | | | | | | 0.005 | | | | | | | | | | | | |

**Table 4.7. continued**

**Table 4.8. Forensic parameters among 100 Hausa-Fulani individuals typed at 21 autosomal STR loci using the GlobalFiler™ Express PCR Amplification Kit.**

| Locus | N | Allele (n) | Allele (v) | Allele (e) | $H_{obs}$ | $H_{exp}$ | $H_o$ | HWE | RMP | PIC | PD | PE | TPI | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CSF1PO** | 100 | 200 | 10 | 4.763 | 0.780 | 0.797 | 0.203 | 0.581 | 0.075 | 0.765 | 0.925 | 0.562 | 2.273 | 0.013 |
| **D10S1248** | 100 | 200 | 9 | 5.323 | 0.850 | 0.816 | 0.184 | 0.884 | 0.069 | 0.786 | 0.931 | 0.695 | 3.333 | -0.047 |
| **D12S391** | 100 | 200 | 12 | 5.379 | 0.860 | 0.824 | 0.176 | 0.869 | 0.058 | 0.800 | 0.942 | 0.715 | 3.571 | -0.056 |
| **D13S317** | 100 | 200 | 7 | 3.298 | 0.730 | 0.700 | 0.3 | 0.995 | 0.154 | 0.647 | 0.846 | 0.476 | 1.852 | -0.048 |
| **D16S539** | 100 | 200 | 7 | 5.009 | 0.810 | 0.806 | 0.194 | 0.929 | 0.073 | 0.772 | 0.927 | 0.618 | 2.632 | -0.012 |
| **D18S51** | 100 | 200 | 15 | 7.460 | 0.920 | 0.871 | 0.129 | 0.728 | 0.042 | 0.852 | 0.958 | 0.836 | 6.250 | -0.062 |
| **D19S433** | 100 | 200 | 13 | 4.353 | 0.820 | 0.859 | 0.141 | 0.177 | 0.045 | 0.839 | 0.955 | 0.637 | 2.778 | -0.065 |
| **D1S1656** | 100 | 200 | 16 | 5.780 | 0.910 | 0.864 | 0.136 | 0.303 | 0.047 | 0.845 | 0.953 | 0.816 | 5.556 | -0.100 |
| **D21S11** | 100 | 200 | 17 | 5.598 | 0.780 | 0.847 | 0.153 | 0.009 | 0.051 | 0.828 | 0.949 | 0.562 | 2.273 | 0.050 |
| **D22S1045** | 100 | 200 | 9 | 5.416 | 0.790 | 0.819 | 0.181 | 0.538 | 0.063 | 0.790 | 0.937 | 0.581 | 2.381 | 0.031 |
| **D2S1338** | 100 | 200 | 12 | 7.590 | 0.790 | 0.873 | 0.127 | 0.034 | 0.044 | 0.856 | 0.956 | 0.581 | 2.381 | 0.090 |
| **D2S441** | 100 | 200 | 8 | 3.412 | 0.790 | 0.753 | 0.247 | 0.935 | 0.098 | 0.718 | 0.902 | 0.581 | 2.381 | -0.118 |
| **D3S1358** | 100 | 200 | 8 | 4.144 | 0.800 | 0.763 | 0.237 | 0.560 | 0.115 | 0.719 | 0.885 | 0.599 | 2.500 | -0.054 |
| **D5S818** | 100 | 200 | 8 | 4.095 | 0.660 | 0.760 | 0.24 | 0.063 | 0.102 | 0.718 | 0.898 | 0.369 | 1.471 | 0.127 |
| **D7S820** | 100 | 200 | 7 | 4.605 | 0.780 | 0.787 | 0.213 | 0.030 | 0.089 | 0.751 | 0.911 | 0.562 | 2.273 | 0.004 |
| **D8S1179** | 100 | 200 | 8 | 3.994 | 0.780 | 0.753 | 0.247 | 0.983 | 0.104 | 0.711 | 0.896 | 0.562 | 2.273 | -0.040 |
| **FGA** | 100 | 200 | 15 | 7.516 | 0.880 | 0.874 | 0.126 | 0.661 | 0.037 | 0.855 | 0.963 | 0.755 | 4.167 | -0.015 |
| **SE33** | 100 | 200 | 28 | 13.450 | 0.960 | 0.934 | 0.066 | 0.955 | 0.017 | 0.925 | 0.983 | 0.919 | 12.500 | -0.037 |
| **TH01** | 100 | 200 | 5 | 3.047 | 0.730 | 0.692 | 0.308 | 0.995 | 0.140 | 0.646 | 0.860 | 0.476 | 1.852 | -0.087 |
| **TPOX** | 100 | 200 | 8 | 4.185 | 0.760 | 0.765 | 0.235 | 0.183 | 0.102 | 0.727 | 0.898 | 0.527 | 2.083 | 0.001 |
| **vWA** | 100 | 200 | 9 | 5.234 | 0.840 | 0.813 | 0.187 | 0.352 | 0.080 | 0.782 | 0.920 | 0.675 | 3.125 | -0.038 |

| | | |
|---|---|---|
| N | - | Number of unrelated individuals |
| Allele (n) | - | Number of alleles surveyed |
| Allele (v) | - | Number of different allelic variants |
| Allele (e) | - | Number of effective alleles |
| $H_o$ | - | Observed heterozygosity |
| $H_{exp}$ | - | Expected heterozygosity |
| $H_{om\_exp}$ | - | Expected homozygosity |
| HWE | - | Hardy-Weinberg Equilibrium probability |
| RMP | - | Random Match Probability |
| PIC | - | Polymorphism Information Content |
| PD | - | Power of Discrimination |
| PE | - | Power of Exclusion |
| TPI | - | Typical Paternity Index |
| $F_{IS}$ | - | Fixation Index - **I**ndividual within the **S**ubpopulation |

**Table 4.9. Summary of private alleles by subpopulation using the GlobalFiler™ Express PCR Amplification Kit**

| Sub-population/ Ethnic group | Locus | Allele | Frequency |
|---|---|---|---|
| Igbo (N = 102) | D3S1358 | 19 | 0.005 |
| | TPOX | 14 | 0.005 |
| | D18S51 | 9 | 0.005 |
| | D19S433 | 9 | 0.005 |
| | TH01 | 10 | 0.005 |
| | D5S818 | 16 | 0.005 |
| | D13S317 | 5 | 0.005 |
| | SE33 | 13 | 0.010 |
| | D10S1248 | 8 | 0.010 |
| | D10S1248 | 18 | 0.005 |
| | D10S1248 | 19 | 0.005 |
| Yoruba (N = 101) | vWA | 22 | 0.005 |
| | CSF1PO | 18 | 0.005 |
| | FGA | 17 | 0.005 |
| | D13S317 | 15 | 0.005 |
| | D7S820 | 6 | 0.005 |
| | D12S391 | 25 | 0.005 |
| | D2S1338 | 15 | 0.005 |
| Hausa (N = 100) | CSF1PO | 15 | 0.005 |
| | TPOX | 13 | 0.005 |
| | D18S51 | 23 | 0.005 |
| | D19S433 | 17.2 | 0.005 |
| | FGA | 16.1 | 0.005 |
| | FGA | 31.2 | 0.005 |
| | D7S820 | 14 | 0.010 |
| | SE33 | 31.2 | 0.005 |
| | SE33 | 33.2 | 0.005 |
| | D10S1248 | 9 | 0.005 |
| | D12S391 | 24 | 0.010 |

Across the 21 autosomal loci, the entire Nigerian dataset recorded 72 different alleles, with sizes ranging from 5 to 36. Microvariant alleles were found at TH01, CSF1PO, D18S51, SE33, D2S441, D19S433, D13S317, D1S1656, D3S1358, FGA, D12S391, and D21S11 among the three ethnic groups in different proportions. The allele frequencies among 303 unrelated individuals in the Nigerian population ranged from 0.002 to 0.492, with allele 7 at the TH01 locus being detected at the highest frequency, followed by allele 12 in the D13S317 locus at 0.452 **(Table 4.1)**. Among the loci surveyed in the overall Nigerian population, the SE33 locus was the most polymorphic (PIC = 91.7%), the most discriminative (PD = 98.5%), followed by the D2S1338 (PIC = 86.8, PD = 96.9%). The SE33 locus also demonstrated the highest Power of Elimination (PE) at 83.8% and

had the most allelic variants, totalling 34. In comparison, D18S51 (PE = 76.4%) and FGA (v = 19 alleles) had the second-highest power of elimination and allelic variants, respectively. Locus SE33 is the most informative marker among the 21 loci analysed **(Table 4.2)**. In contrast, TH01 exhibited the lowest polymorphic information content (PIC = 64.0%), followed closely by D13S317 (PIC = 64.2%). D13S317 also recorded the lowest Power of Discrimination (PD = 85.0%) and Power of Elimination (PE = 39.3%), while TH01 closed behind with PD = 85.5% and PE = 42.8%. TH01 had the smallest number of allelic variants (v = 6), followed by D13S317 (v = 7). The Typical Paternity Index (PI) varied from 1.546 (D13S317) to 6.313 (SE33) across 21 loci **(Table 4.2)**. The Combined Power of Elimination (CPE) and Discrimination (CPD) for these loci exceeded 99.9999%. The Combined Match Probability (CMP) for all 21 loci was calculated at $2.77 \times 10^{-26}$, indicating a high evidential strength. The likelihood of encountering two individuals with the same profile within the same population was estimated at 1 in $3.61 \times 10^{25}$. The comprehensive data from these forensic parameters underscores the effectiveness of the GlobalFiler™ Express PCR Amplification Kit in analysing the Nigerian population.

The analysis of 21 autosomal loci in the Igbo dataset identified 64 distinct alleles, ranging in size from 5 to 36. The allele frequencies among 102 individuals ranged from 0.005 to 0.495. The highest frequency was for Allele 7 at the TH01 locus (0.495), followed by Allele 12 at the D13S317 locus (0.461) (**Table 4.3**). Locus SE33 was the most polymorphic (PIC = 91.3%) and discriminative (PD = 98.0%), with the highest number of allelic variants (v = 24). D2S1338 followed with a PIC of 87.2%, PD of 96.4%, and the highest power of elimination (PE) at 84.0%, while SE33 had a PE of 76.0% (**Table 4.4**). Conversely, TH01 was the least polymorphic and discriminative (PIC = 63.1%, PD = 84.2%), followed by the D13S317 (PIC = 64.0%; PD = 84.8%). Both loci had the lowest Power of Elimination (PE = 40.7%). Additionally, TH01 and D7S820 exhibited the smallest number of allelic variants (v =6). The Typical Paternity Index (TPI) ranged from 1.417 (D13S317) to 6.375 (D2S1338) (**Table 4.4**). The Combined Power of Exclusion (CPE) and Discrimination (CPD) for all loci exceeded 99.9999%, and the Combined Match Probability (CMP) was $5.23 \times 10^{-25}$, indicating an extremely low chance of two individuals sharing the same profile (L.R = 1 in 1 in $1.91 \times 10^{24}$).

Considering the Yoruba subpopulation distinctly, the dataset showed 65 distinct alleles, ranging from 6 to 36. The allele frequencies among 101 unrelated individuals ranged from 0.005 to 0.505, with Allele 7 at the TH01 locus (0.505) being the most frequent, followed closely by Allele 12 at the D13S317 locus (0.475) (**Table 4.5**). Locus SE33 was the most polymorphic (PIC = 90.2%) and

discriminative (PD = 97.6%), followed by D2S1338 (PIC = 86.6%, PD = 96.3%). D2S11 had the highest Power of Elimination (PE = 85.8%), followed by SE33 (PE = 76.0%). Notably, SE33 also had the highest number of allelic variants (v = 27), followed by FGA (v = 17) (**Table 4.6**). Conversely, the TH01 locus had the lowest Polymorphic Information Content (PIC = 63.1%), followed closely by D13S317 (PIC = 63.6%). Additionally, D13S317 had the lowest Power of Discrimination (PD = 84.8%), with TH01 close behind (PD = 85.0%). The D13S317 locus had the lowest Power of Elimination (PE = 36.6%), followed by TH01 (PE = 40.3%), which also had the fewest number of allelic variants (v = 5). The Typical Paternity Index ranged from 1.443 (D13S3171) to 7.214 (D21S11) (**Table 4.6**). The Combined Power of Exclusion (CPE) and Discrimination (CPD) exceeded 99.9999%, and the Combined Match Probability (CMP) was 1.81 x $10^{-25}$, indicating a very low chance of matching profiles in the population (L.R = 1 in 5.52 x $10^{24}$).

The Hausa-Fulani dataset identified 63 distinct alleles, with sizes ranging from 6 to 36. Among 100 unrelated individuals, allele frequencies ranged from 0.005 to 0.480, with allele 7 at the TH01 locus being the most frequent (0.480), followed by Allele 12 at the D13S317 locus (0.420) (**Table 4.7**). Locus SE33 was the most polymorphic (PIC = 92.5%) and discriminative (PD = 98.3%), followed by D2S1338 (PIC = 85.6%) and FGA (PD = 96.3%). SE33 exhibited the highest Power of Elimination (PE = 91.9%) and had the largest number of allelic variants (v = 28), while D18S51 (PE = 83.6%) and D21S11 (v = 17) had the second highest values, respectively **(Table 4.8).** In contrast, TH01 was the least polymorphic (PIC = 64.6%) and discriminative (PD = 86.0%), closely followed by D13S317 (PIC = 64.0%, PD = 84.8%). Both had the lowest Power of Elimination (PE = 47.6%), with TH01 having the least number of allelic variants (v =5). The Typical Paternity Index varied from 1.852 (TH01 and D13S317) to 12.5 (SE33) across the 21 loci **(Table 4.8)**. The Combined Power of Exclusion (CPE) and Discrimination (CPD) for these loci exceeded 99.9999%, with a Combined Match Probability (CMP) of 3.24 x $10^{-25}$, indicating a very low chance of two individuals sharing the same profile in this population (L.R = 1 in 3.09 x $10^{24}$).

**Table 4.9** presents the private alleles (PAs) exclusively identified within specific subpopulations. A total of 20 PAs were documented from a sample of 303 unrelated individuals. Both the Igbo and Hausa subpopulations identified 11 PAs each. In contrast, the Yoruba subpopulation recorded only 7 PAs, the lowest number among the subpopulations studied. It is crucial to recognize that the sample size can greatly influence the distribution of alleles among these Nigerian subpopulations.

### 4.4.4 INTER-SUBPOPULATION RELATIONSHIP

The study found no significant deviations from Hardy–Weinberg equilibrium in the overall Nigerian population (see **Table 4.2** and **Appendix 48**), as determined by Exact tests based on Monte Carlo simulations, with all loci exhibiting p-values greater than 0.05 ($p > 0.05$) before Bonferroni correction. Furthermore, no loci remained significant after correcting for multiple testing. In the Igbo subpopulation, all loci conformed to the Hardy–Weinberg equilibrium both before and after Bonferroni correction **(Table 4.4; Appendix 49)**. The Yoruba subpopulation showed a nominal deviation at locus D21S11 before correction **(Table 4.6)**, but this was not significant after correction, with all other loci in equilibrium **(Appendix 50)**. In the Hausa subpopulation, loci D21S11, D2S1338, and D7S820 initially deviated from Hardy-Weinberg equilibrium **(Table 4.8)**, but none maintained significance after correction **(Appendix 51)**. These findings indicate that the GlobalFiler™ Express PCR Amplification Kit is highly effective for DNA profiling in forensic investigations, both within these subpopulations and in the broader Nigerian population.

**F-statistics Genetic Distance**

Wright's F-statistics were calculated and averaged across all loci to estimate the genetic differentiation and inter-subpopulation relationships among the Igbo, Yoruba, and Hausa-Fulani subpopulations. The $F_{ST}$ values for the Nigerian populations (see **Table 4.2**) ranged from 0.002 to 0.007 across loci, indicating very little genetic differentiation among the Igbos, Yorubas, and Hausa-Fulanis. Locus D5S818 exhibited the highest $F_{ST}$ value at 0.007, while loci D12S391, D13S317, D19S433, D2S1338, and D2S441 had the lowest $F_{ST}$ values at 0.002.

**Table 4.10. $F_{ST}$ value comparison between the Igbo, Yoruba, and Hausa-Fulani subpopulation using the GlobalFiler™ Express PCR Amplification Kit**

| Sub-Population | Igbo | Yoruba | Hausa-Fulani |
|---|---|---|---|
| **Igbo** | 0.000 | | |
| **Yoruba** | 0.003 | 0.000 | |
| **Hausa-Fulani** | 0.003 | 0.003 | 0.000 |

**Table 4.10** compares the pairwise population $F_{ST}$ values among the Igbo, Yoruba, and Hausa-Fulani subpopulations. The $F_{ST}$ values were very low (0.003), indicating minimal genetic differentiation and high genetic similarity among the three ethnic groups based on data generated

using the GlobalFiler™ Express PCR Amplification Kit. The results suggest that these subpopulations are largely genetically homogeneous.

The $F_{IS}$ values **(Table 4.2)**, also known as the inbreeding coefficient, showed negative values (-0.092 to -0.040) at most loci, indicating an excess of heterozygotes. This suggests no significant impact of inbreeding was detected within the subpopulations. Locus D7S820 exhibited the highest $F_{IS}$ value (0.025), while D19S433 had the lowest (-0.092). Most loci showed negative $F_{IS}$ values, with exceptions including D16S539 (0.025), vWA (0.004), D2S1338 (0.006), and FGA (0.007). The $F_{IS}$ values varied across subpopulations: -0.108 to 0.069 in the Igbo **(Table 4.4)**, -0.119 to 0.094 in the Yoruba **(Table 4.6)**, and -0.118 to 0.090 in the Hausa-Fulani **(Table 4.8)**. Each group displayed only a few positive values, showing overall similar results.

The $F_{IT}$ value was statistically similar to the $F_{IS}$ value as both measure inbreeding but at different hierarchical levels—$F_{IT}$ reflects inbreeding of individuals relative to the total population. The $F_{IT}$ values for the Nigerian population (**Table 4.2**) were generally low, ranging from -0.090 to 0.048 across all loci in the total population. These values indicate no significant inbreeding within or between the populations, indicating that the overall population is genetically diverse.

**Population Structure**

To further access the inter-subpopulation relationships among the Igbo, Yoruba, and Hausa-Fulani subpopulations using the 21 autosomal STR loci generated by the GlobalFiler™ Express PCR Amplification Kit, *STRUCTURE* was employed. The Mean LnP(K) ± SD analysis (see **Appendix 54**) demonstrated inconsistency in estimating the average log probability, as evidenced by the standard deviation (SD) illustrated in **Figure 4.5**. The standard deviation increased significantly from $K$=3 to $K$=10, and the log-likelihood values did not rise with $K$.

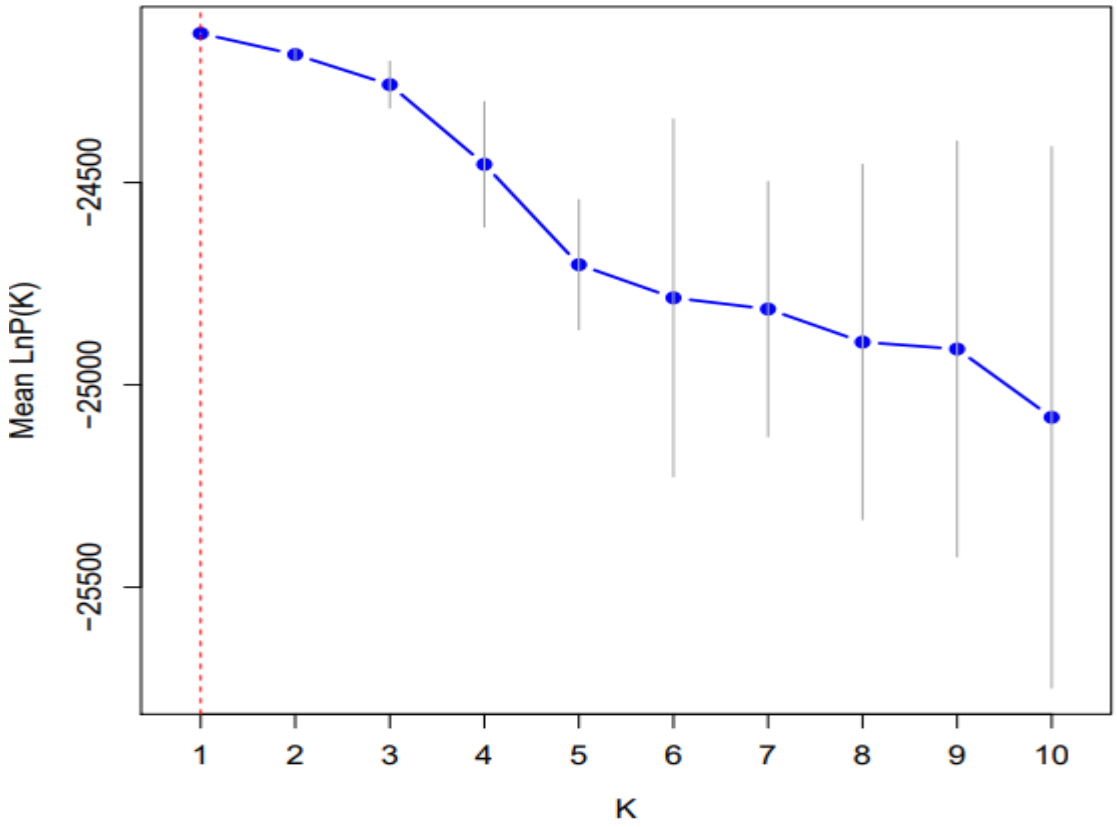



**Figure 4.5. The mean log-likelihood (±SD) of the data for a given number of population (*K*) computed across five repeat runs and iterations of the *STRUCTURE* program.**

The mean LnP(K) compared the likelihood of different *K* values generated by *STRUCTURE* with the higher value, suggesting a better fit for *K*. The insight on the robustness of the model for a given *K* was provided by the Mean LnP(K)±SD to help assess the stability of the estimate of the population structure.

In accordance with the method established by Evano *et al*. (2005), the optimal *K* value was determined to be 3 (see **Figure 4.6** and **Appendix 54**). This suggests three genetically distinct sub-populations, with the highest Delta *K* recorded for this value. The rate of change in Delta *K* assists in selecting the appropriate population size (*K*) by comparing the likelihood between different *K* values. A higher Delta *K* value indicates the optimal number of populations in the dataset. Although *K*=1 was expected, the Evanno method (Delta *K*) does not assess *K*=1 because Delta *K* is undefined at that level, as it relies on the second order of change, which is also undefined at *K*=1. *STRUCTURE* primarily identifies the most pronounced levels of population structure without using prior information about individual origins (Pritchard *et al*., 2000).

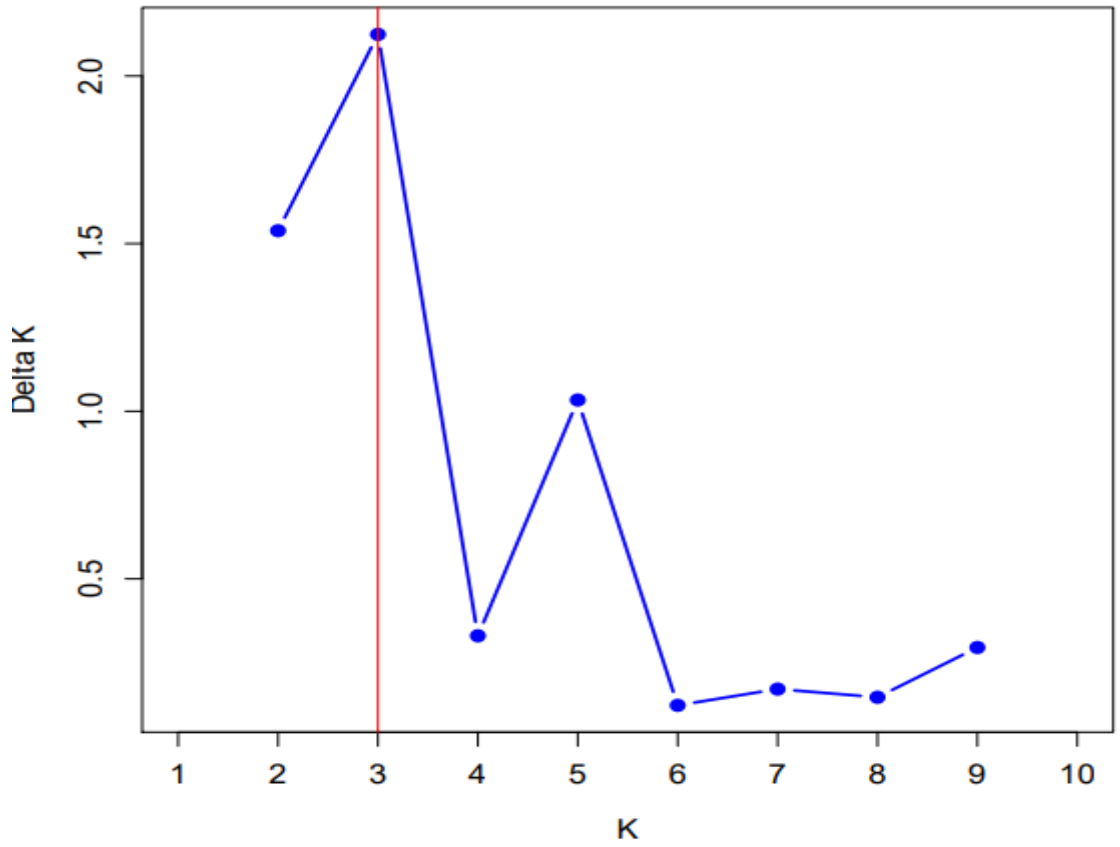
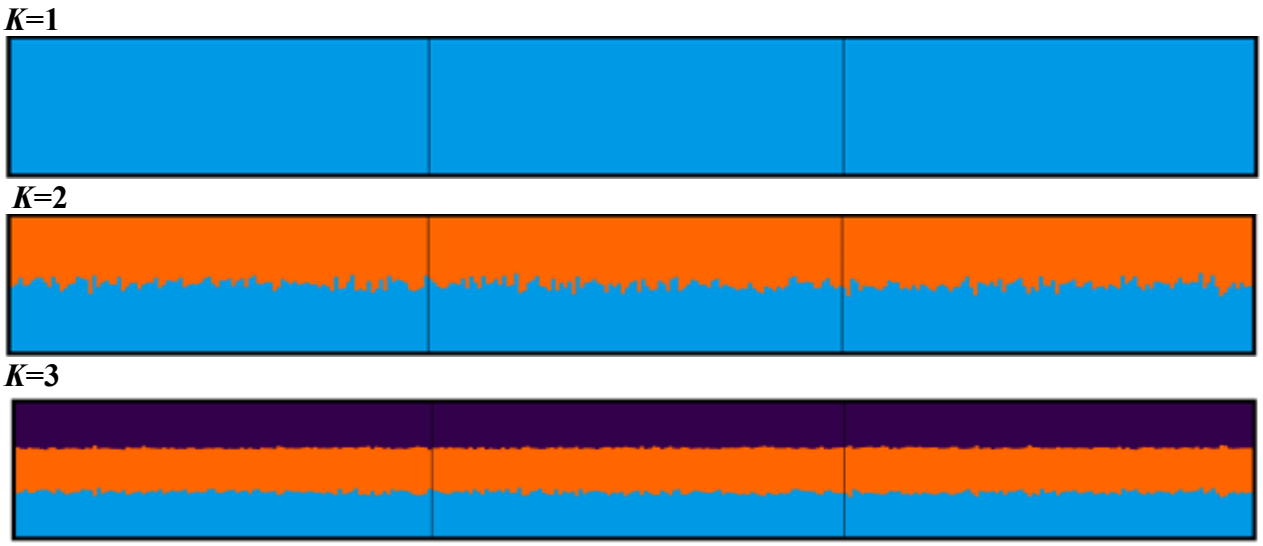



**Figure 4.6. The change in the log-likelihood rate (Delta *K*) for each value of *K* computed across five repeated runs and iterations of the *STRUCTURE* program.**

The analysis of inferred populations across *K* values ranging from 1 to 10 indicated a lack of population structuring among the Igbo, Yoruba, and Hausa-Fulani subpopulations. While *K* = 3 emerged as the optimal value, aligning with the total number of subpopulations examined in this study, **Figure 4.7** illustrates that there was no genetic differentiation in the ancestry proportions of individuals within each group for *K* values spanning from 1 to 10.



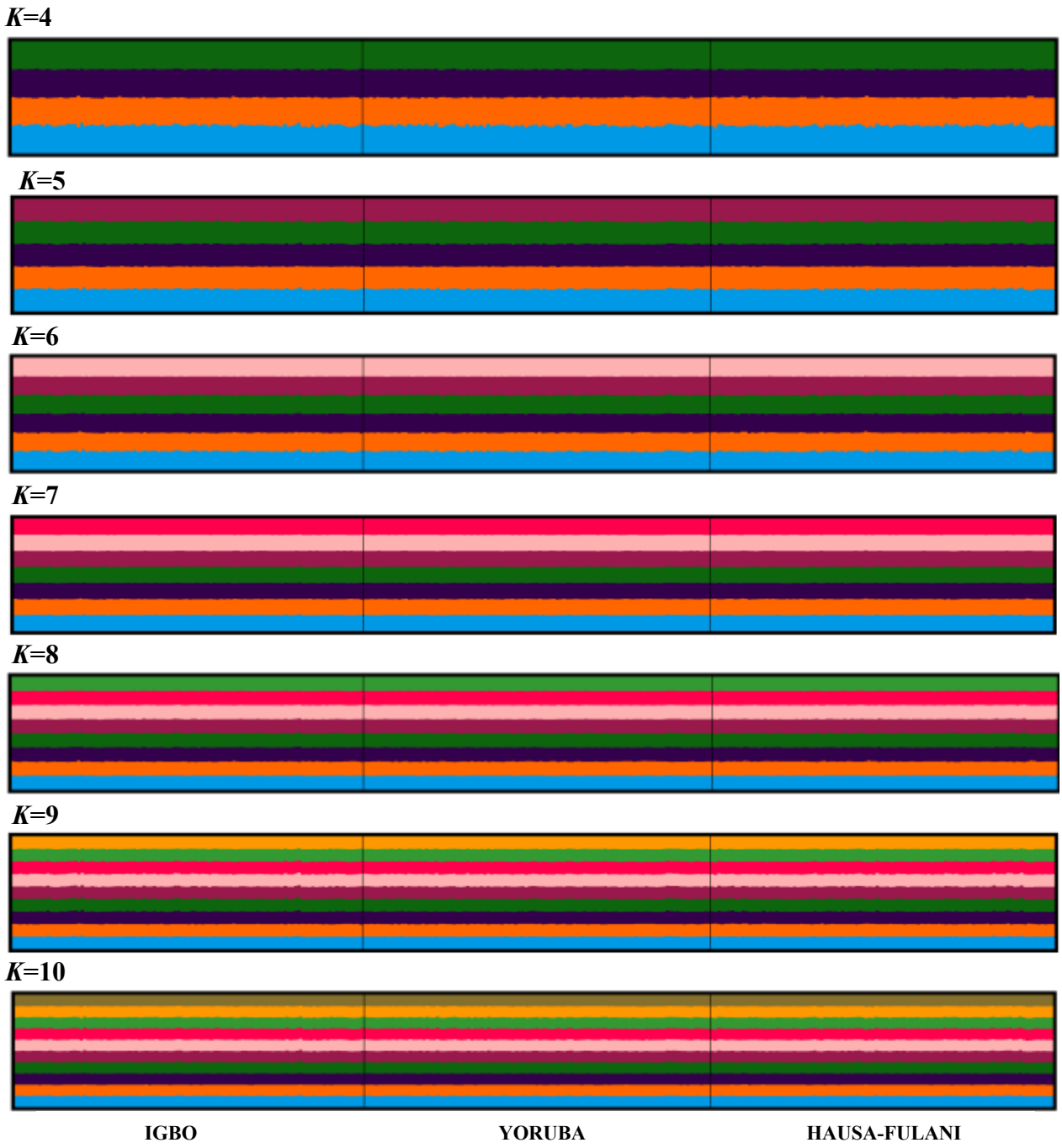



**Figure 4.7. Population stratification of the Nigerian population, specifically focusing on the Igbo, Yoruba, and Hausa-Fulani subpopulations. This analysis was conducted through five repeated runs and iterations utilizing the *STRUCTURE* program. The inferred populations, ranging from *K*=3 to *K*=10, are represented by a corresponding colour bar. Each vertical line denotes an individual (totalling 303) assessed based on data obtained from 21 autosomal loci, using the GlobalFiler™ Express PCR Amplification Kit. The proportion of colour in each bar indicates the individual's genetic ancestry related to each defined population.**

**Principal Component Analysis**

To further investigate the relationships among subpopulations in Nigeria, Principal Component Analysis (PCA) was utilised to assess population differentiation across the Hausa-Fulani, Yoruba, and Igbo subpopulations.

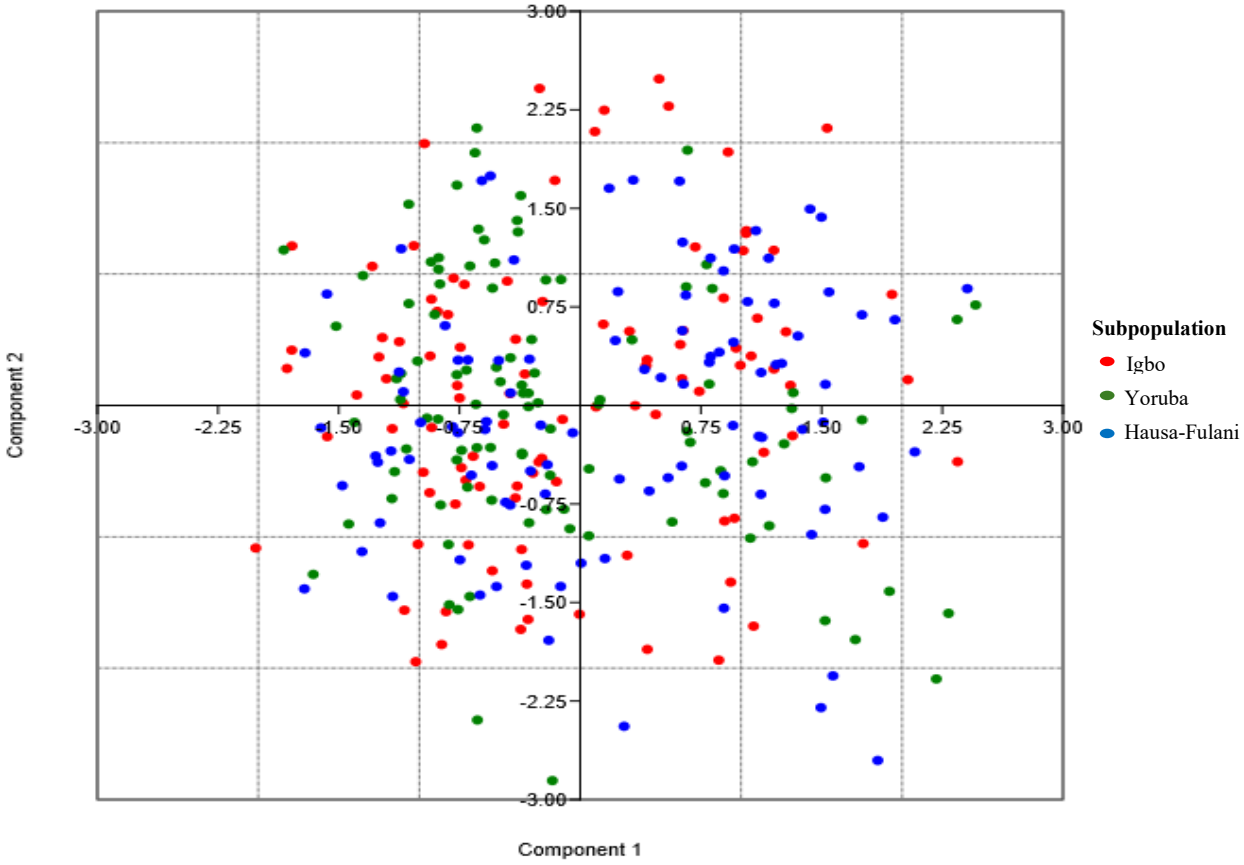



**Figure 4.8. Result of the Principal Component Analysis (PCA) performed on the subpopulations of Igbo, Yoruba, and Hausa-Fulani. This analysis was conducted using 21 autosomal short tandem repeat (STR) loci from the GlobalFiler™ Express PCR Amplification Kit. In the PCA plot, each point represents one of the 303 samples obtained from unrelated individuals within the Nigerian population.**

PCA helps uncover hidden relationships among these ethnic groups by analysing genetic data without requiring prior knowledge of their ancestral backgrounds. It simplifies genetic datasets while preserving significant variations. This approach helps find patterns in the Nigerian population genetic structure without bias. Similar samples are expected to cluster together, while dissimilar samples are visualised to cluster separately when visualised on a PCA plot. Investigating genetic segregation patterns among the Igbo, Yoruba, and Hausa-Fulani ethnic groups involved analysing the full dataset from 303 unrelated individuals. This analysis utilised loci information from the 21 autosomal STR to visualise genetic differentiation. As shown in **Figure 4.8**, the results revealed no differentiation among the 303 unrelated individuals from the three ethnic groups based on the examined samples, as no data clustering was observed from different subpopulations.

**Neighbour-Joining (NJ)**

The phylogenetic distance among the three sub-populations was evaluated using the Neighbour-Joining (NJ) method to assess their inter-subpopulation relationships. The phylogenetic tree was constructed utilizing data derived from the pairwise Nei's Genetic Distance population matrix (**Appendix 52**), which was calculated based on 21 autosomal short tandem repeat (STR) loci. The results in **Figure 4.9** showed a similar pattern of population differentiation to the pairwise $F_{ST}$ values among the Igbo, Yoruba, and Hausa-Fulani ethnic groups. As anticipated, the NJ analysis revealed very low levels of inter-subpopulation variation, indicating a high degree of genetic similarity among the three populations.

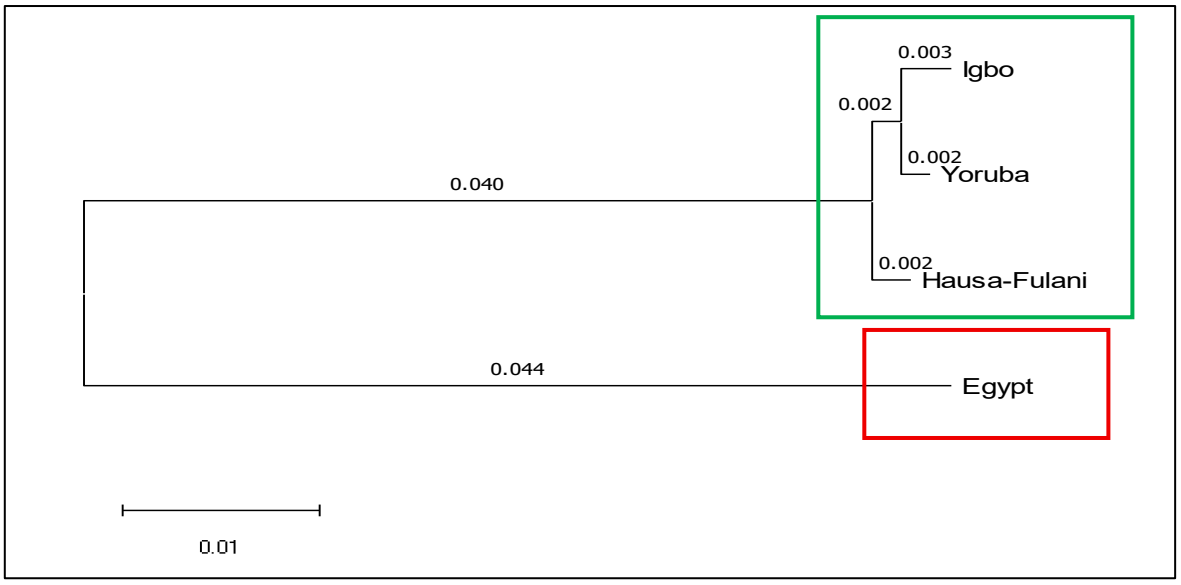



**Figure 4.9. Neighbour-joining (NJ) tree based on pairwise Nei's genetic distance, illustrating the clustering of the Igbo, Yoruba, and Hausa–Fulani subpopulations, with Egypt included as an outgroup. This analysis employed 21 autosomal short tandem repeats (STR) loci obtained from the GlobalFiler™ Express PCR Amplification Kit.**

The genetic analysis suggests that the Igbos exhibit a marginally closer genetic relationship to the Yorubas than the Hausa-Fulanis. This phenomenon may be attributed to the geographical proximity of the Igbos, which are situated in the southeastern region, to the Yorubas, which are located in the southwestern region. In contrast, the Hausa-Fulanis demonstrate a slightly closer genetic affinity to the Yorubas than the Igbos.

## 4.5 DISCUSSION

This study evaluated the population genetics of 303 unrelated Nigerians from the Igbo, Yoruba, and Hausa-Fulani ethnic groups using 21 autosomal STR markers from the GlobalFiler™ Express Kit. All samples yielded full STR profiles, which were utilised to calculate allele frequency data

and forensic genetic parameters and assess inter-subpopulation relationships using *STRUCTURE*, PCA, and NJ methods. Additionally, the results obtained from the GlobalFiler™ Express Kit were compared with those generated using the Qiagen™ Investigator ESSplex SE QS Kit to determine the more suitable STR system for creating a national forensic DNA database in Nigeria. Considerations in this comparison included the presence or absence of off-ladder alleles.

Different STR kits have various alleles in their ladders based on their design, target users, and validation methods (Kutranov, 2011). The GlobalFiler™ Express Kit includes a wider range of alleles, making it suitable for many areas worldwide. This broader selection lowers the chance of getting off-ladder results (Martín *et al*., 2014). In contrast, the QIAGEN™ Investigator ESSplex SE QS Kit is made explicitly for Southern and Central Europe. Its allele selection is based on the European Standard Set (ESS), which is used in forensic databases and casework in European countries (Barbaro *et al*., 2024). This makes the 24 STR-locus GlobalFiler™ Kit a suitable choice for investigating population genetics in Nigeria. Many countries worldwide, including Scotland adopted the 24 STR-locus system using the GlobalFiler™ kit produced by Thermo Fisher as the standard for DNA profiling as it one of the most informative and discriminatory kits available. (Johnson and William, 2004; Hopwood *et al*., 2012).

It is crucial to recognise that DNA sequencing is the most precise method for identifying off-ladder (OL) alleles, as size-based electrophoretic methods fall short in confirming their repeat structure (Fujii *et al*., 2016). However, this study did not use sequencing since the GlobalFiler™ Express Kit generated complete genotypes for all samples, including those initially classified as OL by the QIAGEN™ Investigator ESSplex SE QS Kit. The genotypes from the GlobalFiler™ Express Kit were compared to those from the QIAGEN™ Investigator ESSplex SE QS Kit, revealing consistency in all matching alleles except the OL ones. The OL loci identified by the QIAGEN™ Investigator ESSplex SE QS Kit were genotyped using allele calls from the GlobalFiler™ Express Kit, which offers a more comprehensive allelic ladder that covers additional bins.

The analysis of forensic parameters and allele frequencies indicated that majority of the genetic markers in the GlobalFiler™ Express Kit are highly polymorphic across the Igbo, Yoruba, and Hausa-Fulani subpopulations. Notably, locus SE33 was the most significant and consistently discriminative marker due to its many allelic variants. Other important loci included D2S1338, D21S11, D2S441, FGA, and D18S51. In contrast, TH01 and D13S317 showed the lowest variability, with fewer allelic variants. The combined power of elimination (CPE) and the combined

power of discrimination (CPD) for the overall population and the sub-populations of the Igbos, Yorubas, and Hausa-Fulanis were found to be significantly higher than 99.9%.

Furthermore, the allele frequencies from the GlobalFiler™ Express Kit were consistent with those of the QIAGEN™ Investigator ESSplex SE QS Kit. STR kits demonstrated high discriminative power. In addition, Exact tests of Hardy–Weinberg equilibrium (HWE) were conducted for all autosomal STR loci using Monte Carlo simulations, suitable for highly polymorphic markers and moderate sample sizes (Guo and Thompson, 1992; Butler, 2015). Statistical significance was assessed before and after Bonferroni correction for multiple testing. In the Nigerian population, no loci showed significant deviations from the Hardy-Weinberg equilibrium at the 0.05 level before or after correction, indicating overall equilibrium. The Igbo subpopulation adhered to the Hardy-Weinberg equilibrium, whereas the Yoruba showed a nominal deviation at locus D21S11, which was not significant after correction. The Hausa-Fulani subpopulation showed nominal deviations at D21S11, D19S433, and D2S1338, but none were significant. These deviations could be attributed to sampling effects rather than actual biological disequilibrium. Overall, pooling individuals from the Igbo, Yoruba, and Hausa subpopulations did not introduce detectable population structure effects. This finding aligns with the results from Agbo *et al*. (2017), Okolie *et al.* (2017), and Hohoff *et al*. (2009), which also demonstrated conformity with Hardy-Weinberg Equilibrium expectations in the Nigerian population regarding multi-allelic loci.

From a technical standpoint, the GlobalFiler™ Express PCR Amplification Kit is optimized for direct PCR amplification from buccal and blood samples collected on a paper substrate without requiring a punch buffer. A punch buffer is used to prepare discs from dried biological samples on FTA® paper for PCR in direct amplification workflows. It breaks down cells and neutralises PCR inhibitors, increasing time and reagent costs. The GlobalFiler™ Express PCR Amplification Kit also features a robust PCR protocol that produces results in under an hour, compared to the QIAGEN™ Investigator ESSplex SE QS Kit, which requires slightly more than one hour. While the QIAGEN kit achieves excellent outputs with direct PCR, it necessitates an additional punch buffer for sample preparation, potentially incurring extra financial costs.

The assessment of inter-subpopulation relationships among the three major ethnic groups in Nigeria—Hausa-Fulani, Igbo, and Yoruba—using the GlobalFiler™ Express Kit yielded results similar to those obtained with the QIAGEN™ Investigator ESSplex SE QS Kit. This research, which focused on F-statistics, genetic structure, Principal Component Analysis, and Neighbour-

Joining analysis, found no inter-subpopulation differentiation or structures among these ethnic groups. The outcome of the parameters also suggests that the Igbo, Yoruba, and Hausa-Fulani may share historical and genetic connections. The study demonstrates that, irrespective the different languages spoken and divergent cultures exhibited by these ethnic groups, the differences show no strong correlation with their genetic differences. This is likely due to overlapping geographical boundaries facilitating intermixing through trade, marriage, and migration (Okolie *et al.*, 2017; Gomes *et al*., 2019).

## 4.6 CONCLUSION

The research reported in this chapter highlighted the extensive information provided by 21 polymorphic autosomal STR markers using the GlobalFiler™ Express Kit, which effectively analysed the Nigerian population and distinguished between individuals from the Igbos, Yorubas, and Hausa-Fulanis. However, this analysis did not extend to the sub-population level, as the Nigerian subpopulations are not isolated enough to become genetically differentiated to an extent that this can be detected using the STR kits employed in this study The investigation of Nigeria's three main ethnic groups—Hausa-Fulani, Igbo and Yoruba—with the GlobalFiler™ Express Kit yielded results similar to those of the QIAGEN™ Investigator ESSplex SE QS Kit. Through F-statistics, Principal Component Analysis, and Neighbour-Joining analysis, the study found no differentiation or genetic structures among these groups.

The 24-locus GlobalFiler™ Express Kit demonstrated a clear advantage for establishing a database in Nigeria compared to the 17-locus QIAGEN™ Investigator ESSplex SE QS Kit. The 21 autosomal loci offered more markers for analysis, resulting in higher discriminatory power, better likelihood ratios, and lower random match probabilities than the 16 loci available in the QIAGEN™ Investigator ESSplex SE QS Kit. No off-ladder issues were identified when using the GlobalFiler™ Express Kit. This kit had more alleles in the allelic ladder, allowing for better identification of all alleles across the sampled Nigerian population. Thus, the GlobalFiler™ Express Kit effectively identified off-ladder alleles that could not be identified in the QIAGEN™ Investigator ESSplex SE QS Kit.

A comparative analysis of genotyping results from the GlobalFiler™ Express Kit and the QIAGEN™ Investigator ESSplex SE QS Kit showed high concordance across shared autosomal STR loci. Electropherogram review confirmed no significant artefacts, such as allelic dropout or off-ladder alleles, in profiles from the GlobalFiler™ Express Kit. Although off-ladder alleles were

observed in profiles generated using the QIAGEN™ Investigator ESSplex SE QS Kit, these differences were due to the limitations of the allelic ladder, not because of mistakes in the genotyping process. The affected alleles were successfully resolved by comparing them with corresponding GlobalFiler™ Express Kit profiles, which incorporate a more extensive allelic ladder that accommodates additional microvariant alleles. While DNA sequencing remains the definitive method for characterising off-ladder alleles, the use of GlobalFiler™ Express Kit data provided reliable genotype resolution in this study. These findings show that STR typing systems are reliable for forensic DNA analysis. They also emphasise the importance of having complete allelic ladders. This helps reduce mistakes when classifying unusual alleles in routine forensic testing.

One significant advantage of the GlobalFiler™ Express PCR Amplification Kit is its lower overall cost than the QIAGEN™ Investigator ESSplex SE QS Kit. For a country like Nigeria, which has not yet established a national forensic DNA database, the cost of setting up such a database is an essential factor to consider. The GlobalFiler™ Express Kit costs approximately £7,290 for 400 reactions, while the QIAGEN™ Investigator ESSplex SE QS Kit costs around £9,230 for the same number of reactions. If budget concerns are essential, a more affordable and effective option is the GlobalFiler™ Express PCR Amplification Kit. The study shows this kit is a good choice for creating a national DNA database in Nigeria.

The GlobalFiler™ Express Kit has shown promising results with samples from Nigeria's three prominent ethnic groups: Hausa-Fulani, Igbo, and Yoruba. This kit includes a broader range of alleles, making it suitable for various regions worldwide. As a result, the GlobalFiler™ Kit is a good choice for investigating population genetics in Nigeria. However, since Nigeria is a country of over 300 ethnic groups and 500 spoken languages, testing the kit's performance with other sub-populations is essential.

# CHAPTER FIVE

# ANALYSIS OF Y-STRS IN THE NIGERIAN POPULATION USING THE POWERPLEX® Y23 SYSTEM KIT

## 5.1 INRODUCTION

Y-Short Tandem Repeats (Y-STRs) are repetitive sequences found on the Y chromosome. These Y-STR markers are passed down solely through the male line, making them essential for forensic, anthropological, and genealogical research (Butler 2015). Y-STRs are specifically used to identify male lineages. Y chromosomes are present only in males, and these markers remain stable within patrilineal lineages, with exceptions for mutations. This stability is due to the fact that the Y-chromosome does not go through recombination across generations, except to a limited extent in the pseudoautosomal regions. After all, it is inherited in a haploid manner (Roewer, 2013). Due to the absence of recombination in Y-STR loci, they have proven valuable in tracking paternal ancestry over many generations, as a result of the stability of the lineage. Y-STR markers are passed down only from father to son, making them ideal for studying male-line ancestry, unlike autosomal STRs which are inherited from both parents (Jobling and Tyler-Smith, 2003; Kayser, 2017).

The Promega PowerPlex® Y23 System Kit (Promega Corporation, Madison, Wisconsin, United States) is a Y-STR multiplex PCR kit designed to amplify 23 Y-STRs from human genomic DNA. This kit enables the amplification and separation of 23 markers simultaneously, using four-colour fluorescent detection (including Fluorescein, JOE, TMR-ET, and CXR-ET) in a single reaction (Thompson *et al*., 2013). It is suitable for automatically analysing DNA fragments. The PowerPlex® Y23 System includes the core loci of the European Minimal Haplotype (EMH)—DYS393, DYS392, DYS391, DYS390, DYS389II, DYS389I, DYS385a/b, and DYS19 (Thompson *et al.*, 2013), the two additional loci recommended by the Scientific Working Group - DNA Analysis Methods (SWGDAM) (DYS439 and DYS438) (Pascali *et al*., 1998; Lee, 2004), the system contains loci used in the U.S. core Y-STR set (Y-GATA-H4, DYS635, DYS458, DYS456, DYS448, and DYS437) as well as six additional loci (DYS481, DYS533, DYS549, DYS570, DYS576, DYS643). Furthermore, six highly discriminating Y-STR loci that are included in the Promega PowerPlex® Y23 System Kit (DYS643, DYS576, DYS570, DYS549, DYS533, and DYS481) have been mapped, and the rates of mutation, studied with DYS570 and DYS576 discovered to be rapidly mutating with rates >1% (Kayser *et al.*, 2004; D'Amato *et al*., 2010; Geppert *et al*., 2009). The Promega PowerPlex® Y23 System Kit is optimized for direct DNA amplification and standard PCR amplification from blood and buccal samples collected on FTA cards. Additionally, it incorporates an internal lane standard to ensure accuracy and reliability in

results. The following loci are detected in the blue dye channel, labelled with Fluorescein: DYS576, DYS448, DYS389II, DYS389I, and DYS19. The green dye channel includes the JOE-labelled STR amplification products for DYS549, DYS533, DYS481, DYS438, DYS437, and DYS391. In the yellow dye channel, TMR-ET labelled STR amplicons are present DYS643, DYS635, DYS570, DYS439, DYS392, and DYS390. The red dye channel contains DNA fragments labelled with CXR-ET for Y-GATA-H4, DYS458, DYS456, DYS393, and DYS385a/b. Lastly, the internal lane standard is detected in the orange dye channel, labelled with WEN dye as the Internal Lane Standard 500 Y23 (WEN ILS 500 Y23). It is compatible with the ABI 3500 Genetic Analyzers (Thermo Fisher Scientific) (Thompson *et al*., 2013).

Some researchers have studied the genetic diversity of Y-chromosome STR loci among Indigenous ethnic groups in Nigeria however, none have analysed or documented the Y-STRs in Nigeria using the Promega PowerPlex® Y23. Cole-Showers (2014) used 11 Y-STR loci in 463 samples from different ethnolinguistic groups, showing stronger genetic ties to sub-Saharan Africa than North Africa or the Middle East. Chiedozie and Isaac (2015) examined two Igbo males from Owerri with 16 Y-STR loci and mtDNA HVS1, revealing a common paternal lineage approximately 90 generations ago. Fakorede *et al*. (2019) studied 110 unrelated Yoruba males in Lagos using five Y-STR markers and confirmed their effectiveness in forensic and paternity testing. Martinez *et al*. (2017) evaluated 27 Y-STR loci from 142 samples across Nigeria's major ethnic groups, noting lower haplotype diversity than East African populations. Fakorede *et al.* (2024) analysed 461 saliva samples from unrelated males, including 96 Igbos, 139 Hausas, and 226 Yorubas, using 10 Y-STR loci in the UniQTyper™ Y-10 system. They found that DYS626 had the highest allelic richness while DYS504 had the lowest, identifying 403 unique singletons (haplotypes observed only once) and 430 haplotypes.

## 5.2 AIM AND OBJETIVES

This research seeks to assess the population genetics of 167 unrelated male individuals across three Nigerian prominent ethnic groups: Hausa-Fulani, Igbo and Yoruba. The study employed 23 Y-STR markers utilising the PowerPlex® Y23 System Kit from Promega Corporation. It is worth noting that the use of the Promega PowerPlex Y23 System Kit has not been previously documented in the Nigerian population.

The specific objectives for the work detailed in this chapter are as follows:

1. Generate Y-STR profiles and allele frequency data to create a database for each ethnic group and the overall population.

2. Ascertain forensic parameters for each ethnic group and the overall population while evaluating the kit's effectiveness.
3. Evaluate the genetic structure within each population and examine the relationships between the different ethnic groups.

## 5.3 MATERIALS AND METHODS

The study involving the Nigerian population collected blood samples at multiple locations in Nigeria and the United Kingdom, encompassing a cohort of 167 unrelated males, as detailed in **Chapter 2.2**. Donors received participant information sheets and consent forms to sign and return. Furthermore, the participants were offered verbal explanations of the blood sampling procedures. The Y-STR profiling, associated data, and statistical analyses were performed at the University of Strathclyde in Glasgow. Detailed procedures for sample collection can be found in **Chapter 2.2**, and the participation information sheet and consent form are in **Appendices 1** and **2**. The DNA profiling methodology involved a 1.2 mm punch from the FTA® card. It incorporated both standard PCR and direct PCR techniques, employing the Promega PowerPlex® Y23 System Kit (Promega Corporation, Madison, Wisconsin, United States), as described in **Chapters 2.6.1** and **2.6.2**, respectively. The capillary electrophoresis method, which identifies and separates DNA fragments, was conducted using the Genetic Analyzer 3500 alongside the WEN Internal Lane Standard (WEN ILS) 500 Y23 and PowerPlex® Y23 System Allelic Ladder, as outlined in **Chapter 2.6.3.** Data interpretation, encompassing the 23 Y-STR loci allele calling, was executed using GeneMapper® ID-X software version 1.6 (Applied Biosystems - Life Technologies, USA), as detailed in **Chapter 2.6.4**. Allele frequencies and various forensic parameters [Gene Diversity (GD), Haplotype Match Probability (HMP), Polymorphic Information Content (PIC), and Discrimination Capacity (DC)] were generated using the STR Analysis for Forensic (*STRAF*-A 2.2.2) online tool and the Genetic Analysis in Excel (GenAIEx 6.5) platform as explained in **Chapter 2.7.1**. Haplotype Diversity (HD) was calculated using the formula: $HD = n(1-\sum pi2)/(n-1)$, where $n$ = total number of individuals in the sample, and $pi$ = the frequency of the $i$th haplotype. This was generated using a Python script leveraging the Pandas and NumPy libraries (Python Software Foundation, 2024; McKinney, 2010; Harris *et al*., 2020). Python was also used with the Matplotlib and Seaborn libraries (Python Software Foundation, 2024; Hunter, 2007; Waskom, 2021) to calculate allelic richness. In the calculations for allelic richness, the multi-copy locus DYS385a/b was analysed by designating the smaller allele as DYS385a and the larger allele as DYS385b within each haplotype. Separating DYS385a/b by assigning the smaller repeat value to DYS385a and the larger value to DYS385b was done to ensure consistency across the dataset and to provide a reproducible, systematic method of allele designation. This follows the standard practices in population genetics

(Kittler *et al*., 2003; Mann *et al*., 2024). DYS389 I and DYS389 II were treated as separate loci during Y-STR data analysis and were analysed in the same manner as other Y-STR markers, consistent with the allele reporting format of commercial Y-STR typing systems (Kayser *et al.,* 1997; Butler, 2015). No decomposition of DYS389 II into DYS389b (DYS389 II – DYS389 I) was performed. The PAleontological Statistics (PAST 4.03) software was utilised to conduct a Principal Component Analysis (PCA) to assess sub-population differentiation within the Y-STR dataset, as explained in **Chapter 2.7.2**. As explained in **Chapter 2.7.2**, the Molecular Evolutionary Genetics Analysis (MEGA-11) software was used to perform a Neighbour-Joining analysis of the Igbo, Yoruba, and Hausa-Fulani sub-populations.

## 5.4 RESULTS

### 5.4.1 Y-STR HAPLOTYPING

The Promega PowerPlex® Y23 System Kit effectively generated complete DNA profiles from blood spots on FTA cards for all 167 samples. This was accomplished through two methodologies: the solid-phase silica gel column extraction method, which was applied to 33 samples, and the direct PCR amplification approach, utilized for 134 samples.


**Figure 5.1. Electropherogram produced by the Promega PowerPlex® Y23 System 2800M Control DNA. The loci are designated using four distinct dye colours, which are distributed across four separate channels.**

Blood samples that were challenging to amplify using direct PCR were successfully amplified using standard PCR, which included DNA extraction and quantification to obtain complete STR profiles. The DNA extraction and quantification processes overcame PCR inhibition, likely caused by matrix components in the samples that hindered the effectiveness of the direct PCR method. Each DNA profile generated comprised data at all 23 Y-STR markers. Importantly, all negative controls incorporated in each assay exhibited no evidence of contamination. Furthermore, the DNA profile obtained from the positive control sample (**Figure 5.1**) was consistent with the expected profile.

### 5.4.2 ALLELE FREQUENCIES AND FORENSIC PARAMETERS

The Promega PowerPlex® Y23 System Kit successfully generated full Y-STR profiles from blood samples obtained from 167 unrelated Nigerian males. Allele frequencies and forensic parameters were calculated for the Igbo, Yoruba, and Hausa-Fulani ethnic groups based on the surveyed samples. **Tables 5.1, 5.2,** and **5.3** present detailed results for the Nigerian male population**,** comprising participants from all three subpopulations. Among the 167 Nigerian males were 60, 55, and 52 Igbos, Yorubas, and Hausa-Fulanis. **Tables 5.4, 5.5,** and **5.6, 5.7, 5.8,** and **5.9,** and **5.10, 5.11,** and **5.12** showcase the allele frequencies and forensic parameters generated for the Igbo, Yoruba, and Hausa-Fulani subpopulations, respectively. The forensic parameters evaluated from the allele frequencies of unrelated Nigerian males encompass several key metrics, including Gene Diversity (GD), Haplotype Match Probability (HMP), Polymorphic Information Content (PIC), and Discrimination Capacity (DC) for each locus.

**Table 5.13** presents a detailed overview of the private alleles observed within the subpopulations. These private alleles and the related forensic parameters are essential for assessing how informative and specific the Y-STR data is to each population. **Tables 5.14**, **5.15**, and **5.16** present the Y-STR haplotype frequencies observed in the Igbo, Yoruba, and Hausa-Fulani ethnic groups, respectively, as generated using the PowerPlex® Y23 System Kit. These tables summarise the distribution and frequencies of distinct Y-STR haplotypes within each population and form the basis for subsequent forensic and population genetic interpretation, as Y-chromosomal loci are inherited as a single non-recombining haplotype.

**Table 5.17** summarizes the number of unique and repeated haplotypes, Discrimination Capacity (DC) and Haplotype Diversity (HD) for each ethnolinguistic group as well as for the combined dataset.

## Table 5.1. Allele frequency distribution results among 167 Nigerian males using the Promega PowerPlex® Y23 System Kit

| Allele | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | DYS437 | DYS438 | DYS439 | DYS448 | DYS456 | DYS458 | DYS481 | DYS533 | DYS549 | DYS570 | DYS576 | DYS635 | DYS643 | YGATAH4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **8** | | | | | 0.036 | | | | | | | | | | 0.006 | | | | | | 0.036 |
| **9** | | | | | 0.024 | | | | 0.036 | | | | | | 0.036 | 0.024 | | | | 0.030 | 0.048 |
| **10** | | | | | 0.838 | | | | 0.305 | 0.024 | | | | | 0.156 | 0.030 | | | | | 0.222 |
| **11** | | 0.012 | | | 0.096 | 0.725 | | | 0.545 | 0.317 | | | | | 0.521 | 0.509 | | | | 0.204 | 0.407 |
| **11.3** | | | | | | | | | | | | | | | 0.006 | | | | | | |
| **12** | | 0.144 | | | 0.006 | 0.030 | 0.030 | | 0.048 | 0.389 | | | | | 0.228 | 0.347 | | 0.006 | | 0.054 | 0.192 |
| **12.1** | | | | | | | | | | | | | | | 0.006 | | | | | | |
| **12.3** | | 0.006 | | | | | | | | | | | | | | | | | | | |
| **13** | | 0.611 | | | | 0.246 | 0.473 | 0.132 | 0.030 | 0.162 | | 0.030 | 0.006 | | 0.030 | 0.060 | 0.072 | 0.030 | | 0.383 | 0.042 |
| **13.2** | | | | | | | | | | | | | | | | | 0.012 | | | | |
| **14** | 0.018 | 0.222 | | | | | 0.329 | 0.743 | 0.036 | 0.096 | 0.006 | 0.132 | 0.036 | | 0.012 | 0.018 | 0.012 | 0.048 | | 0.329 | 0.024 |
| **14.2** | | | | | | | | | | | | | 0.006 | | | | | | | | |
| **15** | 0.497 | 0.006 | | | | | 0.162 | 0.090 | | 0.012 | | 0.395 | 0.156 | | | 0.006 | 0.030 | 0.222 | | | 0.024 |
| **15.3** | | | | | | | | | | | | 0.006 | | | | | | | | | |
| **16** | 0.317 | | | | | | 0.006 | 0.024 | | | | 0.228 | 0.329 | | | 0.006 | 0.138 | 0.311 | | | 0.006 |
| **17** | 0.168 | | | | | | | 0.012 | | | | 0.108 | 0.341 | | | | 0.335 | 0.263 | 0.036 | | |
| **18** | | | | | | | | | | | 0.006 | 0.054 | 0.090 | | | | 0.150 | 0.078 | | | |
| **19** | | | | | | | | | | | 0.114 | | 0.036 | | | | 0.150 | 0.024 | 0.030 | | |
| **20** | | | | 0.012 | | | | | | | 0.180 | 0.018 | | | | | 0.090 | 0.018 | 0.090 | | |
| **20.3** | | | | | | | | | | | | | | | | | | | 0.006 | | |
| **21** | | | | 0.766 | | | | | | | 0.563 | 0.030 | | 0.006 | | | 0.012 | | 0.539 | | |
| **22** | | | | 0.108 | | | | | | | 0.114 | | | 0.054 | | | | | 0.204 | | |
| **23** | | | | 0.042 | | | | | | | 0.018 | | | 0.042 | | | | | 0.078 | | |
| **24** | | | | 0.054 | | | | | | | | | | 0.102 | | | | | 0.012 | | |
| **25** | | | | 0.018 | | | | | | | | | | 0.293 | | | | | 0.006 | | |
| **26** | | | | | | | | | | | | | | 0.204 | | | | | | | |
| **27** | | | | | | | | | | | | | | 0.066 | | | | | | | |
| **28** | | | | | | | | | | | | | | 0.210 | | | | | | | |
| **29** | | | 0.186 | | | | | | | | | | | 0.024 | | | | | | | |
| **30** | | | 0.275 | | | | | | | | | | | | | | | | | | |
| **31** | | | 0.323 | | | | | | | | | | | | | | | | | | |
| **32** | | | 0.198 | | | | | | | | | | | | | | | | | | |
| **33** | | | 0.018 | | | | | | | | | | | | | | | | | | |

**Table 5.2. Frequency and count of allelic combinations for the DYS385ab locus among 167 Nigerian males using the Promega PowerPlex® Y23 System Kit.**

| DYS385a/b Alleles | Frequency | Count |
|---|---|---|
| **10,10** | 0.006 | 1 |
| **10,11** | 0.006 | 1 |
| **11,11** | 0.018 | 3 |
| **11,12** | 0.024 | 4 |
| **11,13** | 0.018 | 3 |
| **12,12** | 0.006 | 1 |
| **12,13** | 0.012 | 2 |
| **12,14** | 0.006 | 1 |
| **12,15** | 0.012 | 2 |
| **13,13** | 0.018 | 3 |
| **13,14** | 0.012 | 2 |
| **13,14.2** | 0.006 | 1 |
| **13,15** | 0.071 | 12 |
| **13,16** | 0.006 | 1 |
| **13,17** | 0.018 | 3 |
| **13,18** | 0.012 | 2 |
| **14,14** | 0.030 | 5 |
| **14.2,14.2** | 0.006 | 1 |
| **14,15** | 0.012 | 2 |
| **14,16** | 0.012 | 2 |
| **14,17** | 0.030 | 5 |
| **14,18** | 0.006 | 1 |
| **14,19** | 0.024 | 4 |
| **14,20** | 0.006 | 1 |
| **14,21** | 0.006 | 1 |
| **15,15** | 0.018 | 3 |
| **15,16** | 0.042 | 7 |
| **15,16.2** | 0.006 | 1 |
| **15,17** | 0.030 | 5 |
| **15,18** | 0.006 | 1 |
| **15,19** | 0.012 | 2 |
| **15,20** | 0.018 | 3 |
| **16,16** | 0.036 | 6 |
| **16,17** | 0.054 | 9 |
| **16,18** | 0.030 | 5 |
| **16,19** | 0.024 | 4 |
| **16,20** | 0.006 | 1 |
| **17,17** | 0.054 | 9 |
| **17,18** | 0.090 | 15 |
| **17,19** | 0.036 | 6 |
| **17,20** | 0.018 | 3 |
| **17,21** | 0.012 | 2 |
| **18,18** | 0.054 | 9 |
| **18,19** | 0.006 | 1 |
| **18,20** | 0.012 | 2 |
| **19,19** | 0.030 | 5 |
| **19,20** | 0.006 | 1 |

**Note:** Locus DYS385a/b contains 47 different allelic combinations of duplicate alleles found among the 167 unrelated Nigerian males.

**Table 5.3. Forensic parameters for 167 Nigerian males typed at 23 Y-STR loci using the Promega PowerPlex® Y23 System Kit.**

| Locus | N | Allele (v) | Allele (e) | GD | PIC | HMP | DC |
|---|---|---|---|---|---|---|---|
| **DYS19** | 167 | 4 | 2.658 | 0.628 | 0.554 | 0.376 | 0.624 |
| **DYS389I** | 167 | 6 | 2.249 | 0.560 | 0.503 | 0.443 | 0.557 |
| **DYS389II** | 167 | 5 | 3.933 | 0.750 | 0.700 | 0.254 | 0.746 |
| **DYS390** | 167 | 6 | 1.655 | 0.398 | 0.376 | 0.604 | 0.396 |
| **DYS391** | 167 | 5 | 1.401 | 0.288 | 0.271 | 0.714 | 0.286 |
| **DYS392** | 167 | 3 | 1.706 | 0.416 | 0.350 | 0.586 | 0.414 |
| **DYS393** | 167 | 5 | 2.783 | 0.645 | 0.574 | 0.359 | 0.641 |
| **DYS437** | 167 | 5 | 1.732 | 0.425 | 0.393 | 0.577 | 0.423 |
| **DYS438** | 167 | 6 | 2.525 | 0.608 | 0.544 | 0.396 | 0.604 |
| **DYS439** | 167 | 6 | 3.469 | 0.716 | 0.663 | 0.288 | 0.712 |
| **DYS448** | 167 | 7 | 2.664 | 0.628 | 0.585 | 0.375 | 0.625 |
| **DYS456** | 167 | 9 | 4.053 | 0.763 | 0.727 | 0.242 | 0.758 |
| **DYS458** | 167 | 8 | 3.841 | 0.745 | 0.699 | 0.260 | 0.740 |
| **DYS481** | 167 | 9 | 5.224 | 0.813 | 0.783 | 0.191 | 0.809 |
| **DYS533** | 167 | 9 | 2.787 | 0.654 | 0.605 | 0.350 | 0.650 |
| **DYS549** | 167 | 8 | 2.596 | 0.619 | 0.548 | 0.385 | 0.615 |
| **DYS570** | 167 | 10 | 5.194 | 0.814 | 0.787 | 0.191 | 0.809 |
| **DYS576** | 167 | 9 | 4.432 | 0.779 | 0.740 | 0.226 | 0.774 |
| **DYS635** | 167 | 9 | 2.861 | 0.656 | 0.616 | 0.348 | 0.652 |
| **DYS643** | 167 | 5 | 3.327 | 0.704 | 0.644 | 0.301 | 0.699 |
| **Y-GATA-H4** | 167 | 9 | 3.874 | 0.746 | 0.706 | 0.258 | 0.742 |
| **DYS385a/b** | 167 | 14 | 7.893 | 0.969 | 0.969 | 0.331 | 0.103 |

N - Number of unrelated males
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
GD - Gene Diversity
PIC - Polymorphic Information Content
HMP - Haplotype Match Probability
DC - Discrimination Capacity

## Table 5.4. Allele frequency distribution results among 60 Igbo males using the Promega PowerPlex® Y23 System Kit

| Allele | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | DYS437 | DYS438 | DYS439 | DYS448 | DYS456 | DYS458 | DYS481 | DYS533 | DYS549 | DYS570 | DYS576 | DYS635 | DYS643 | YGATAH4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **8** | | | | | | | | | | | | | | | | | | | | | 0.017 |
| **9** | | | | | 0.050 | | | | 0.083 | | | | | | 0.033 | 0.050 | | | | 0.033 | |
| **10** | | | | | 0.833 | | | | 0.233 | 0.017 | | | | | 0.117 | 0.017 | | | | | 0.250 |
| **11** | | | | | 0.117 | 0.733 | | | 0.483 | 0.317 | | | | | 0.533 | 0.517 | | | | 0.167 | 0.383 |
| **11.3** | | | | | | | | | | | | | | | 0.017 | | | | | | |
| **12** | | 0.183 | | | | | 0.033 | | 0.117 | 0.400 | | | | | 0.217 | 0.267 | | | | 0.100 | 0.233 |
| **12.1** | | | | | | | | | | | | | | | 0.017 | | | | | | |
| **13** | | 0.650 | | | | 0.267 | 0.500 | 0.133 | 0.017 | 0.167 | | | 0.017 | | 0.033 | 0.100 | | | | 0.350 | 0.083 |
| **13.2** | | | | | | | | | | | | | | | | | 0.033 | | | | |
| **14** | 0.050 | 0.167 | | | | | 0.317 | 0.733 | 0.067 | 0.067 | | 0.150 | 0.017 | | 0.033 | 0.050 | 0.017 | 0.050 | | 0.350 | 0.033 |
| **15** | 0.483 | | | | | | 0.133 | 0.100 | | 0.033 | | 0.517 | 0.117 | | | | 0.033 | 0.133 | | | |
| **16** | 0.317 | | | | | | 0.017 | 0.033 | | | | 0.200 | 0.367 | | | | 0.217 | 0.417 | | | |
| **17** | 0.150 | | | | | | | | | | | 0.117 | 0.283 | | | | 0.433 | 0.283 | 0.050 | | |
| **18** | | | | | | | | | | | | 0.017 | 0.133 | | | | 0.017 | 0.083 | | | |
| **19** | | | | | | | | | | | 0.217 | | 0.067 | | | | 0.167 | 0.017 | 0.067 | | |
| **20** | | | | 0.017 | | | | | | | 0.133 | | | | | | 0.067 | 0.017 | 0.017 | | |
| **20.3** | | | | | | | | | | | | | | | | | | | 0.017 | | |
| **21** | | | | 0.600 | | | | | | | 0.533 | | | 0.017 | | | 0.017 | | 0.500 | | |
| **22** | | | | 0.183 | | | | | | | 0.117 | | | 0.083 | | | | | 0.217 | | |
| **23** | | | | 0.067 | | | | | | | | | | 0.050 | | | | | 0.100 | | |
| **24** | | | | 0.100 | | | | | | | | | | 0.100 | | | | | 0.033 | | |
| **25** | | | | 0.033 | | | | | | | | | | 0.250 | | | | | | | |
| **26** | | | | | | | | | | | | | | 0.267 | | | | | | | |
| **27** | | | | | | | | | | | | | | 0.083 | | | | | | | |
| **28** | | | | | | | | | | | | | | 0.133 | | | | | | | |
| **29** | | | 0.167 | | | | | | | | | | | 0.017 | | | | | | | |
| **30** | | | 0.250 | | | | | | | | | | | | | | | | | | |
| **31** | | | 0.333 | | | | | | | | | | | | | | | | | | |
| **32** | | | 0.250 | | | | | | | | | | | | | | | | | | |

**Table 5.5. Frequency and count of allelic combinations for the DYS385ab locus among 60 Igbo males using the Promega PowerPlex® Y23 System Kit.**

| DYS385a/b Alleles | Frequency | Count |
|---|---|---|
| **12,13** | 0.017 | 1 |
| **12,14** | 0,017 | 1 |
| **13,13** | 0.033 | 2 |
| **13,14.2** | 0.017 | 1 |
| **13,15** | 0.133 | 8 |
| **14,15** | 0.017 | 1 |
| **14,16** | 0.017 | 1 |
| **14,17** | 0.017 | 1 |
| **14,19** | 0.050 | 3 |
| **15,15** | 0.033 | 2 |
| **15,16** | 0.067 | 4 |
| **15,16.2** | 0.017 | 1 |
| **15,17** | 0.017 | 1 |
| **15,19** | 0.017 | 1 |
| **15,20** | 0.033 | 2 |
| **16,16** | 0.067 | 4 |
| **16,17** | 0.050 | 3 |
| **16,18** | 0.033 | 2 |
| **16,19** | 0.033 | 2 |
| **17,17** | 0.100 | 6 |
| **17,18** | 0.083 | 5 |
| **17,19** | 0.017 | 1 |
| **18,18** | 0.067 | 4 |
| **18,19** | 0.017 | 1 |
| **19, 19** | 0.017 | 1 |

**Note:** Locus DYS385a/b contains 25 different allelic combinations of duplicate alleles found among the 60 unrelated Igbo males.

**Table 5.6. Forensic parameters for 60 Igbo males typed at 23 Y-STR loci using the Promega PowerPlex® Y23 System Kit.**

| Locus | N | Allele (v) | Allele (e) | GD | PIC | HMP | DC |
|---|---|---|---|---|---|---|---|
| **DYS19** | 60 | 4 | 2.786 | 0.652 | 0.577 | 0.359 | 0.641 |
| **DYS389I** | 60 | 3 | 2.067 | 0.525 | 0.462 | 0.484 | 0.516 |
| **DYS389II** | 60 | 4 | 3.789 | 0.749 | 0.687 | 0.264 | 0.736 |
| **DYS390** | 60 | 6 | 2.442 | 0.601 | 0.554 | 0.409 | 0.591 |
| **DYS391** | 60 | 3 | 1.407 | 0.294 | 0.267 | 0.711 | 0.289 |
| **DYS392** | 60 | 2 | 1.642 | 0.398 | 0.315 | 0.609 | 0.391 |
| **DYS393** | 60 | 5 | 2.707 | 0.641 | 0.567 | 0.369 | 0.631 |
| **DYS437** | 60 | 4 | 1.765 | 0.441 | 0.402 | 0.567 | 0.433 |
| **DYS438** | 60 | 6 | 3.191 | 0.698 | 0.646 | 0.313 | 0.687 |
| **DYS439** | 60 | 6 | 3.403 | 0.718 | 0.656 | 0.294 | 0.706 |
| **DYS448** | 60 | 4 | 2.757 | 0.648 | 0.589 | 0.363 | 0.637 |
| **DYS456** | 60 | 5 | 2.913 | 0.668 | 0.612 | 0.343 | 0.657 |
| **DYS458** | 60 | 7 | 3.982 | 0.762 | 0.711 | 0.251 | 0.749 |
| **DYS481** | 60 | 9 | 5.607 | 0.830 | 0.799 | 0.178 | 0.822 |
| **DYS533** | 60 | 8 | 2.675 | 0.662 | 0.613 | 0.349 | 0.651 |
| **DYS549** | 60 | 6 | 2.830 | 0.658 | 0.598 | 0.353 | 0.647 |
| **DYS570** | 60 | 9 | 3.704 | 0.742 | 0.695 | 0.270 | 0.730 |
| **DYS576** | 60 | 7 | 3.550 | 0.731 | 0.676 | 0.282 | 0.718 |
| **DYS635** | 60 | 8 | 3.163 | 0.696 | 0.650 | 0.316 | 0.684 |
| **DYS643** | 60 | 5 | 3.523 | 0.728 | 0.666 | 0.284 | 0.716 |
| **Y-GATA-H4** | 60 | 6 | 3.673 | 0.740 | 0.682 | 0.272 | 0.728 |
| **DYS385a/b** | 60 | 11 | 6.075 | 0.835 | 0.821 | 0.165 | 0.400 |

N - Number of unrelated males
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
GD - Gene Diversity
PIC - Polymorphic Information Content
HMP - Haplotype Match Probability
DC - Discrimination Capacity

## Table 5.7. Allele frequency distribution results among 55 Yoruba males using the Promega PowerPlex® Y23 System Kit

| Allele | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | DYS437 | DYS438 | DYS439 | DYS448 | DYS456 | DYS458 | DYS481 | DYS533 | DYS549 | DYS570 | DYS576 | DYS635 | DYS643 | YGATAH4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **8** | | | | | | | | | | | | | | | 0.018 | | | | | | 0.018 |
| **9** | | | | | | | | | | | | | | | 0.018 | | | | | 0.018 | 0.127 |
| **10** | | | | | 0.927 | | | | 0.345 | | | | | | 0.145 | | | | | | 0.145 |
| **11** | | 0.018 | | | 0.073 | 0.818 | | | 0.618 | 0.345 | | | | | 0.545 | 0.545 | | | | 0.218 | 0.455 |
| **12** | | 0.073 | | | | | 0.036 | | 0.018 | 0.418 | | | | | 0.218 | 0.400 | | 0.018 | | | 0.200 |
| **13** | | 0.564 | | | | 0.182 | 0.364 | 0.055 | 0.018 | 0.164 | | 0.055 | | | 0.055 | 0.036 | 0.073 | | | 0.400 | 0.018 |
| **14** | | 0.327 | | | | | 0.418 | 0.836 | | 0.073 | 0.018 | 0.127 | 0.018 | | | | | 0.036 | | 0.364 | 0.018 |
| **15** | 0.473 | 0.018 | | | | | 0.182 | 0.073 | | | | 0.418 | 0.182 | | | 0.018 | | 0.255 | | | 0.018 |
| **16** | 0.309 | | | | | | | 0.018 | | | | 0.164 | 0.309 | | | | 0.055 | 0.327 | | | |
| **17** | 0.218 | | | | | | | 0.018 | | | | 0.164 | 0.436 | | | | 0.255 | 0.273 | 0.018 | | |
| **18** | | | | | | | | | | | | 0.018 | 0.036 | | | | 0.291 | 0.055 | | | |
| **19** | | | | | | | | | | | 0.036 | | 0.018 | | | | 0.182 | 0.036 | 0.018 | | |
| **20** | | | | | | | | | | | 0.200 | 0.036 | | | | | 0.127 | | 0.127 | | |
| **21** | | | | 0.964 | | | | | | | 0.618 | 0.018 | | | | | 0.018 | | 0.673 | | |
| **22** | | | | 0.018 | | | | | | | 0.109 | | | 0.036 | | | | | 0.127 | | |
| **23** | | | | | | | | | | | 0.018 | | | | | | | | 0.036 | | |
| **24** | | | | | | | | | | | | | | 0.145 | | | | | | | |
| **25** | | | | 0.018 | | | | | | | | | | 0.327 | | | | | | | |
| **26** | | | | | | | | | | | | | | 0.145 | | | | | | | |
| **27** | | | | | | | | | | | | | | 0.055 | | | | | | | |
| **28** | | | | | | | | | | | | | | 0.255 | | | | | | | |
| **29** | | | 0.200 | | | | | | | | | | | 0.036 | | | | | | | |
| **30** | | | 0.345 | | | | | | | | | | | | | | | | | | |
| **31** | | | 0.291 | | | | | | | | | | | | | | | | | | |
| **32** | | | 0.127 | | | | | | | | | | | | | | | | | | |
| **33** | | | 0.036 | | | | | | | | | | | | | | | | | | |

**Table 5.8. Frequency and count of allelic combinations for the DYS385ab locus among 55 Yoruba males using the Promega PowerPlex® Y23 System Kit.**

| DYS385a/b Alleles | Frequency | Count |
|---|---|---|
| **11,11** | 0.018 | 1 |
| **11,12** | 0.018 | 1 |
| **11,13** | 0.036 | 2 |
| **13,13** | 0.018 | 1 |
| **13,15** | 0.055 | 3 |
| **13,16** | 0.018 | 1 |
| **13,17** | 0.055 | 3 |
| **13,18** | 0.018 | 1 |
| **14,14** | 0.036 | 2 |
| **14.2,14.2** | 0.018 | 1 |
| **14,15** | 0.018 | 1 |
| **14,17** | 0.036 | 2 |
| **14,18** | 0.018 | 1 |
| **14,19** | 0.018 | 1 |
| **14,20** | 0.018 | 1 |
| **14,21** | 0.018 | 1 |
| **15,17** | 0.036 | 2 |
| **15,18** | 0.018 | 1 |
| **15,20** | 0.018 | 1 |
| **16,17** | 0.164 | 9 |
| **16,18** | 0.036 | 2 |
| **16,19** | 0.018 | 1 |
| **17,17** | 0.036 | 2 |
| **17,18** | 0.073 | 4 |
| **17,19** | 0.018 | 1 |
| **17,20** | 0.018 | 1 |
| **17,21** | 0.036 | 2 |
| **18,18** | 0.036 | 2 |
| **18,20** | 0.036 | 2 |
| **19,19** | 0.018 | 1 |

**Note:** Locus DYS385a/b contains 25 different allelic combinations of duplicate alleles found among the 55 unrelated Yoruba males.

**Table 5.9. Forensic parameters for 55 Yoruba males typed at 23 Y-STR loci using the Promega PowerPlex® Y23 System Kit.**

| Locus | N | Allele (v) | Allele (e) | GD | PIC | HMP | DC |
|---|---|---|---|---|---|---|---|
| **DYS19** | 55 | 3 | 2.728 | 0.645 | 0.560 | 0.367 | 0.633 |
| **DYS389I** | 55 | 5 | 2.322 | 0.580 | 0.496 | 0.431 | 0.569 |
| **DYS389II** | 55 | 5 | 3.824 | 0.752 | 0.693 | 0.261 | 0.739 |
| **DYS390** | 55 | 3 | 1.076 | 0.072 | 0.070 | 0.929 | 0.071 |
| **DYS391** | 55 | 2 | 1.156 | 0.137 | 0.126 | 0.865 | 0.135 |
| **DYS392** | 55 | 2 | 1.424 | 0.303 | 0.253 | 0.702 | 0.298 |
| **DYS393** | 55 | 4 | 2.928 | 0.671 | 0.591 | 0.341 | 0.659 |
| **DYS437** | 55 | 5 | 1.412 | 0.297 | 0.279 | 0.708 | 0.292 |
| **DYS438** | 55 | 4 | 1.991 | 0.507 | 0.406 | 0.502 | 0.498 |
| **DYS439** | 55 | 4 | 3.065 | 0.686 | 0.613 | 0.326 | 0.674 |
| **DYS448** | 55 | 6 | 2.293 | 0.574 | 0.522 | 0.436 | 0.564 |
| **DYS456** | 55 | 8 | 4.007 | 0.764 | 0.720 | 0.250 | 0.750 |
| **DYS458** | 55 | 6 | 3.115 | 0.692 | 0.622 | 0.321 | 0.679 |
| **DYS481** | 55 | 7 | 4.549 | 0.795 | 0.748 | 0.220 | 0.780 |
| **DYS533** | 55 | 6 | 2.703 | 0.642 | 0.584 | 0.370 | 0.630 |
| **DYS549** | 55 | 4 | 2.178 | 0.551 | 0.444 | 0.459 | 0.541 |
| **DYS570** | 55 | 7 | 4.825 | 0.807 | 0.763 | 0.207 | 0.793 |
| **DYS576** | 55 | 7 | 3.965 | 0.762 | 0.705 | 0.252 | 0.748 |
| **DYS635** | 55 | 6 | 2.054 | 0.523 | 0.481 | 0.487 | 0.513 |
| **DYS643** | 55 | 4 | 2.940 | 0.672 | 0.589 | 0.340 | 0.660 |
| **Y-GATA-H4** | 55 | 8 | 3.505 | 0.728 | 0.678 | 0.285 | 0.715 |
| **DYS385a/b** | 55 | 12 | 7.370 | 0.951 | 0.946 | 0.049 | 0.545 |

N - Number of unrelated males
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
GD - Gene Diversity
PIC - Polymorphic Information Content
HMP - Haplotype Match Probability
DC - Discrimination Capacity

## Table 5.10. Allele frequency distribution results among 52 Hausa males using the Promega PowerPlex® Y23 System Kit

| Allele | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | DYS437 | DYS438 | DYS439 | DYS448 | DYS456 | DYS458 | DYS481 | DYS533 | DYS549 | DYS570 | DYS576 | DYS635 | DYS643 | YGATAH4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **8** | | | | | 0.115 | | | | | | | | | | | | | | | | 0.077 |
| **9** | | | | | 0.019 | | | | 0.019 | | | | | | 0.058 | 0.019 | | | | 0.038 | 0.019 |
| **10** | | | | | 0.750 | | | | 0.346 | 0.058 | | | | | 0.212 | 0.077 | | | | | 0.269 |
| **11** | | 0.019 | | | 0.096 | 0.615 | | | 0.538 | 0.288 | | | | | 0.481 | 0.462 | | | | 0.231 | 0.385 |
| **12** | | 0.173 | | | 0.019 | 0.096 | 0.019 | | | 0.346 | | | | | 0.250 | 0.385 | | | | 0.058 | 0.135 |
| **12.3** | | 0.019 | | | | | | | | | | | | | | | | | | | |
| **13** | | 0.615 | | | | 0.288 | 0.558 | 0.212 | 0.058 | 0.154 | | 0.038 | | | | 0.038 | 0.154 | 0.096 | | 0.404 | 0.019 |
| **14** | | 0.173 | | | | | 0.250 | 0.654 | 0.038 | 0.154 | | 0.115 | 0.077 | | | | 0.019 | 0.058 | | 0.269 | 0.019 |
| **14.2** | | | | | | | | | | | | | 0.019 | | | | | | | | |
| **15** | 0.538 | | | | | | 0.173 | 0.096 | | | | 0.231 | 0.173 | | | | 0.058 | 0.288 | | | 0.058 |
| **15.3** | | | | | | | | | | | | 0.019 | | | | | | | | | |
| **16** | 0.327 | | | | | | | 0.019 | | | | 0.327 | 0.308 | | | 0.019 | 0.135 | 0.173 | | | 0.019 |
| **17** | 0.135 | | | | | | | 0.019 | | | | 0.038 | 0.308 | | | | 0.308 | 0.231 | 0.038 | | |
| **18** | | | | | | | | | | | 0.019 | 0.135 | 0.096 | | | | 0.154 | 0.096 | | | |
| **19** | | | | | | | | | | | 0.077 | | 0.019 | | | | 0.096 | 0.019 | | | |
| **20** | | | | 0.019 | | | | | | | 0.212 | 0.019 | | | | | 0.077 | 0.038 | 0.135 | | |
| **21** | | | | 0.750 | | | | | | | 0.538 | 0.077 | | | | | | | 0.442 | | |
| **22** | | | | 0.115 | | | | | | | 0.115 | | | 0.038 | | | | | 0.269 | | |
| **23** | | | | 0.058 | | | | | | | 0.038 | | | 0.077 | | | | | 0.096 | | |
| **24** | | | | 0.058 | | | | | | | | | | 0.058 | | | | | | | |
| **25** | | | | | | | | | | | | | | 0.308 | | | | | 0.019 | | |
| **26** | | | | | | | | | | | | | | 0.192 | | | | | | | |
| **27** | | | | | | | | | | | | | | 0.058 | | | | | | | |
| **28** | | | | | | | | | | | | | | 0.250 | | | | | | | |
| **29** | | | 0.192 | | | | | | | | | | | 0.019 | | | | | | | |
| **30** | | | 0.231 | | | | | | | | | | | | | | | | | | |
| **31** | | | 0.346 | | | | | | | | | | | | | | | | | | |
| **32** | | | 0.212 | | | | | | | | | | | | | | | | | | |
| **33** | | | 0.019 | | | | | | | | | | | | | | | | | | |

**Table 5.11. Frequency and count of allelic combinations for the DYS385ab locus among 52 Hausa-Fulani males using the Promega PowerPlex® Y23 System Kit.**

| DYS385a/b Alleles | Frequency | Count |
|---|---|---|
| **10,10** | 0.019 | 1 |
| **10,11** | 0.019 | 1 |
| **11,11** | 0.038 | 2 |
| **11,12** | 0.058 | 3 |
| **11,13** | 0.019 | 1 |
| **12,12** | 0.019 | 1 |
| **12,13** | 0.019 | 1 |
| **12,15** | 0.038 | 2 |
| **13,14** | 0.038 | 2 |
| **13,15** | 0.019 | 1 |
| **13,18** | 0.019 | 1 |
| **14,14** | 0.058 | 3 |
| **14,16** | 0.019 | 1 |
| **14,17** | 0.038 | 2 |
| **15,15** | 0.019 | 1 |
| **15,16** | 0.038 | 2 |
| **15,17** | 0.038 | 2 |
| **15,19** | 0.019 | 1 |
| **16,16** | 0.038 | 2 |
| **16,18** | 0.019 | 1 |
| **16,19** | 0.019 | 1 |
| **16,20** | 0.019 | 1 |
| **17,17** | 0.019 | 1 |
| **17,18** | 0.115 | 6 |
| **17,19** | 0.076 | 4 |
| **17,20** | 0.038 | 2 |
| **18,18** | 0.058 | 3 |
| **19,19** | 0.058 | 3 |
| **19,20** | 0.019 | 1 |

**Note:** Locus DYS385a/b contains 29 different allelic combinations of duplicate alleles found among the 52 unrelated Hausa-Fulani males.

**Table 5.12. Forensic parameters for 52 Hausa-Fulani males typed at 23 Y-STR loci using the Promega PowerPlex® Y23 System Kit.**

| Locus | N | Allele (v) | Allele (e) | GD | PIC | HMP | DC |
|---|---|---|---|---|---|---|---|
| **DYS19** | 52 | 3 | 2.410 | 0.597 | 0.509 | 0.415 | 0.585 |
| **DYS389I** | 52 | 5 | 2.242 | 0.572 | 0.513 | 0.439 | 0.561 |
| **DYS389II** | 52 | 5 | 3.919 | 0.759 | 0.700 | 0.255 | 0.745 |
| **DYS390** | 52 | 5 | 1.716 | 0.425 | 0.394 | 0.583 | 0.417 |
| **DYS391** | 52 | 5 | 1.707 | 0.422 | 0.388 | 0.586 | 0.414 |
| **DYS392** | 52 | 3 | 2.122 | 0.539 | 0.457 | 0.471 | 0.529 |
| **DYS393** | 52 | 4 | 2.476 | 0.608 | 0.535 | 0.404 | 0.596 |
| **DYS437** | 52 | 5 | 2.074 | 0.528 | 0.470 | 0.482 | 0.518 |
| **DYS438** | 52 | 5 | 2.410 | 0.597 | 0.511 | 0.415 | 0.585 |
| **DYS439** | 52 | 5 | 3.942 | 0.761 | 0.704 | 0.254 | 0.746 |
| **DYS448** | 52 | 6 | 2.811 | 0.657 | 0.604 | 0.356 | 0.644 |
| **DYS456** | 52 | 9 | 4.761 | 0.814 | 0.773 | 0.201 | 0.799 |
| **DYS458** | 52 | 7 | 4.199 | 0.780 | 0.728 | 0.235 | 0.765 |
| **DYS481** | 52 | 8 | 4.794 | 0.807 | 0.762 | 0.209 | 0.791 |
| **DYS533** | 52 | 4 | 2.926 | 0.671 | 0.601 | 0.342 | 0.658 |
| **DYS549** | 52 | 6 | 2.709 | 0.643 | 0.562 | 0.369 | 0.631 |
| **DYS570** | 52 | 8 | 5.587 | 0.837 | 0.800 | 0.179 | 0.821 |
| **DYS576** | 52 | 8 | 5.261 | 0.826 | 0.785 | 0.190 | 0.810 |
| **DYS635** | 52 | 6 | 3.363 | 0.716 | 0.658 | 0.297 | 0.703 |
| **DYS643** | 52 | 5 | 3.406 | 0.720 | 0.655 | 0.294 | 0.706 |
| **Y-GATA-H4** | 52 | 9 | 4.012 | 0.765 | 0.716 | 0.249 | 0.751 |
| **DYS385a/b** | 52 | 11 | 9.620 | 0.973 | 0.950 | 0.047 | 0.560 |

N - Number of unrelated males
Allele (v) - Number of different allelic variants
Allele (e) - Number of effective alleles
GD - Gene Diversity
PIC - Polymorphic Information Content
HMP - Haplotype Match Probability
DC - Discrimination Capacity

**Table 5.13. Summary of private alleles by sub-population using the Promega PowerPlex® Y23 System Kit**

| Subpopulation/ Ethnic group | Locus | Allele | Frequency |
|---|---|---|---|
| Igbo (N = 60) | DYS19 | 14 | 0.050 |
| | DYS481 | 21 | 0.017 |
| | DYS549 | 14 | 0.050 |
| | DYS533 | 14 | 0.033 |
| | DYS635 | 24 | 0.033 |
| | DYS439 | 15 | 0.033 |
| | DYS393 | 16 | 0.017 |
| | DYS458 | 13 | 0.017 |
| Yoruba (N = 55) | DYS576 | 12 | 0.018 |
| | DYS389I | 15 | 0.018 |
| | DYS448 | 14 | 0.018 |
| | DYS549 | 15 | 0.018 |
| | DYS533 | 8 | 0.018 |
| | DYS385b | 21 | 0.073 |
| Hausa (N = 52) | DYS576 | 13 | 0.096 |
| | DYS448 | 18 | 0.019 |
| | DYS391 | 8 | 0.115 |
| | DYS391 | 12 | 0.019 |
| | DYS549 | 16 | 0.019 |
| | DYS635 | 25 | 0.019 |
| | DYS392 | 12 | 0.096 |
| | Y-GATA-H4 | 16 | 0.019 |
| | DYS385a | 10 | 0.038 |
| | DYS385b | 10 | 0.019 |

## Table 5.14. Haplotype Frequency Specific to the Igbo Ethnic Group for the PowerPlex® Y23 System Kit

| POPULATION | HAPLOTYPE | FREQUENCY | COUNT |
|---|---|---|---|
| Igbo1 | 16-12.0-21-31-15-10-24-12-11.0-10-14-16.0-22.0-21-13-11-13-15-16.0-16.0-11-14.0-19.0 | 1 | 1 |
| Igbo10 | 16-13.0-22-31-15-10-26-12-11.0-11-14-20.0-21.0-21-14-11-13-13-17.0-16.0-10-15.0-15.0 | 1 | 1 |
| Igbo11 | 16-13.0-21-32-15-10-26-11-11.0-11-13-13.2-21.0-22-14-11-14-13-16.0-15.0-10-16.0-17.0 | 1 | 1 |
| Igbo12 | 16-13.0-21-31-16-9-25-11-10.0-11-14-17.0-22.0-22-11-11-13-13-17.0-15.0-12-17.0-17.0 | 1 | 1 |
| Igbo13 | 17-13.0-21-31-15-10-25-11-11.0-11-14-16.0-21.0-21-11-11-14-15-15.0-15.0-11-14.0-17.0 | 1 | 1 |
| Igbo14 | 17-13.0-21-30-16-11-25-11-12.0-11-13-17.0-21.0-21-12-11-14-14-16.0-15.0-11-15.0-15.0 | 1 | 1 |
| Igbo15 | 15-13.0-19-31-15-10-25-13-13.0-11-14-17.0-24.0-21-12-11-14-14-16.0-16.0-11-16.0-17.0 | 1 | 1 |
| Igbo16 | 17-12.0-21-31-16-11-27-13-14.0-10-14-17.0-21.0-21-12-13-13-15-16.0-15.0-12-13.0-13.0 | 1 | 1 |
| Igbo17 | 16-13.0-20-30-17-10-26-11-11.0-10-13-17.0-21.0-21-13-11-14-14-16.0-15.0-11-15.0-16.0 | 1 | 1 |
| Igbo18 | 17-12.0-19-30-16-11-26-12-11.0-10-15-15.0-19.0-22-14-13-14-13-16.0-16.0-12-16.0-19.0 | 1 | 1 |
| Igbo19 | 16-14.0-21-32-15-10-26-12-12.0-11-14-17.0-19.0-21-12-11-13-14-15.0-16.0-10-17.0-17.0 | 1 | 1 |
| Igbo2 | 16-13.0-21-30-15-10-24-11-11.0-14-14-16.0-22.0-21-12-11-13-15-17.0-15.0-10-17.0-17.0 | 1 | 1 |
| Igbo20 | 16-13.0-21-30-17-10-24-11-11.0-11-14-17.0-22.0-21-15-11-13-14-17.0-15.0-10-18.0-19.0 | 1 | 1 |
| Igbo21 | 17-14.0-21-31-16-10-26-11-11.0-11-14-16.0-20.0-21-12-11-14-15-18.0-17.0-10-14.0-19.0 | 1 | 1 |
| Igbo22 | 16-13.0-19-30-15-10-25-12-11.0-14-14-16.0-21.0-22-11-11-14-12-18.0-15.0-11-18.0-18.0 | 1 | 1 |
| Igbo23 | 17-12.0-21-31-15-10-25-11-11.0-11-13-17.0-21.0-21-13-11-13-15-17.0-15.0-12-19.0-19.0 | 1 | 1 |
| Igbo24 | 16-13.0-21-30-16-10-26-12-11.0-11-13-17.0-22.0-21-12-11-14-13-16.0-16.0-12-16.0-17.0 | 1 | 1 |
| Igbo25 | 16-14.0-22-32-15-10-28-11-11.3-10-14-16.0-21.0-21-12-11-14-13-19.0-15.0-11-17.0-17.0 | 1 | 1 |
| Igbo26 | 15-13.0-21-32-15-10-25-12-12.0-11-14-16.0-21.0-21-12-11-13-14-16.0-15.0-11-16.0-16.0 | 1 | 1 |
| Igbo27 | 17-13.0-21-29-16-10-25-11-12.0-11-13-16.0-21.0-21-12-11-14-14-17.0-15.0-10-15.0-16.0 | 1 | 1 |
| Igbo28 | 17-12.0-21-30-15-10-28-11-11.0-11-14-19.0-21.0-22-12-11-14-13-17.0-15.0-12-15.0-16.2 | 1 | 1 |
| Igbo29 | 18-13.0-22-29-16-10-25-11-11.0-10-14-17.0-21.0-21-12-11-14-14-17.0-15.0-8-12.0-13.0 | 1 | 1 |
| Igbo3 | 17-13.0-21-31-17-10-25-10-11.0-11-14-17.0-21.0-21-11-11-13-13-16.0-15.0-11-18.0-18.0 | 1 | 1 |
| Igbo30 | 16-13.0-21-32-15-10-22-11-12.1-10-14-17.0-20.3-21-13-11-13-14-15.0-16.0-10-16.0-16.0 | 1 | 1 |
| Igbo31 | 16-12.0-21-29-17-10-26-11-11.0-12-14-17.0-21.0-21-11-13-11-14-18.0-16.0-10-15.0-19.0 | 1 | 1 |
| Igbo32 | 17-13.0-21-32-15-10-26-11-10.0-9-15-17.0-20.3-23-11-11-12-13-16.0-15.0-11-17.0-17.0 | 1 | 1 |
| Igbo33 | 16-14.0-21-31-15-10-27-12-11.0-9-14-19.0-22.0-21-13-11-11-14-16.0-14.0-11-16.0-17.0 | 1 | 1 |
| Igbo34 | 17-14.0-19-31-16-10-24-11-9.0-11-14-19.0-21.0-24-11-11-12-13-13.0-14.0-12-17.0-18.0 | 1 | 1 |
| Igbo35 | 15-13.0-20-31-16-10-25-11-11.0-12-14-17.0-21.0-21-12-13-12-14-17.0-17.0-12-15.0-17.0 | 1 | 1 |
| Igbo36 | 16-12.0-22-29-15-9-28-12-11.0-11-14-17.0-21.0-22-12-13-11-13-15.0-15.0-8-15.0-20.0 | 1 | 1 |
| Igbo37 | 17-14.0-22-29-16-10-25-14-11.3-12-14-16.0-19.0-21-11-11-14-14-17.0-16.0-8-16.0-18.0 | 1 | 1 |
| Igbo38 | 14-13.0-21-32-17-10-25-11-10.0-11-14-16.0-21.0-24-12-11-11-15-16.0-14.0-12-17.0-18.0 | 1 | 1 |
| Igbo39 | 14-13.0-19-31-15-10-22-12-10.0-11-14-20.0-23.0-23-13-13-11-14-16.0-16.0-10-16.0-17.0 | 1 | 1 |
| Igbo4 | 14-13.0-20-32-15-10-25-13-11.0-11-14-15.0-24.0-21-13-11-13-13-15.0-17.0-11-15.0-20.0 | 1 | 1 |
| Igbo40 | 17-13.0-20-32-16-10-28-11-11.0-12-14-17.0-21.0-21-11-11-13-13-17.0-16.0-13-17.0-18.0 | 1 | 1 |
| Igbo41 | 17-13.0-22-32-15-10-26-11-12.0-12-15-16.0-17.0-24-11-13-13-13-18.0-15.0-10-14.0-16.0 | 1 | 1 |
| Igbo42 | 16-13.0-22-31-17-10-26-12-10.0-9-15-17.0-17.0-24-10-11-11-13-19.0-14.0-12-16.0-17.0 | 1 | 1 |
| Igbo43 | 16-13.0-19-31-16-10-27-11-11.0-10-14-19.0-23.0-22-12-11-12-13-16.0-14.0-10-16.0-16.0 | 1 | 1 |
| Igbo44 | 19-14.0-19-32-14-10-21-12-11.0-12-14-19.0-23.0-25-11-13-12-13-18.0-17.0-13-13.0-15.0 | 1 | 1 |
| Igbo45 | 18-12.0-19-31-16-9-26-14-12.0-10-16-19.0-21.0-22-13-11-11-13-16.0-17.0-11-13.0-15.0 | 1 | 1 |
| Igbo46 | 16-14.0-19-32-15-10-22-14-12.0-9-14-17.0-23.0-25-13-13-13-13-17.0-15.0-11-13.0-13.0 | 1 | 1 |
| Igbo47 | 16-13.0-21-31-16-11-24-12-11.0-10-14-21.0-21.0-21-11-11-14-13-16.0-14.0-14-17.0-17.0 | 1 | 1 |
| Igbo48 | 18-13.0-20-31-17-10-26-11-12.0-11-14-20.0-21.0-21-12-13-11-14-19.0-15.0-11-15.0-16.0 | 1 | 1 |
| Igbo49 | 16-12.0-21-30-16-10-26-11-11.0-10-14-17.0-22.0-22-12-13-13-13-15.0-15.0-13-17.0-19.0 | 1 | 1 |
| Igbo5 | 17-13.0-19-30-16-10-25-11-13.0-11-14-19.0-22.0-21-12-11-13-14-17.0-15.0-12-14.0-19.0 | 1 | 1 |
| Igbo50 | 16-13.0-21-32-15-11-24-13-11.0-11-14-20.0-21.0-21-10-11-14-13-17.0-17.0-10-16.0-19.0 | 1 | 1 |
| Igbo51 | 16-13.0-21-32-15-10-28-11-12.0-11-14-18.0-22.0-20-11-11-14-13-16.0-15.0-12-18.0-18.0 | 1 | 1 |
| Igbo52 | 15-12.0-19-31-14-10-23-13-11.0-9-14-17.0-21.0-24-11-11-14-14-17.0-15.0-11-16.0-16.0 | 1 | 1 |
| Igbo53 | 16-13.0-21-29-17-10-28-9-10.0-11-14-17.0-22.0-22-11-13-11-14-15.0-15.0-11-16.0-18.0 | 1 | 1 |
| Igbo54 | 15-13.0-20-32-15-11-27-11-11.0-11-14-19.0-23.0-21-11-11-12-13-17.0-15.0-12-15.0-16.0 | 1 | 1 |
| Igbo55 | 16-13.0-19-29-15-11-22-13-12.0-12-14-17.0-23.0-24-12-13-14-13-19.0-15.0-13-16.0-17.0 | 1 | 1 |
| Igbo56 | 15-14.0-21-29-14-10-23-12-10.0-11-14-17.0-21.0-21-12-11-13-14-16.0-15.0-13-18.0-18.0 | 1 | 1 |
| Igbo57 | 16-12.0-19-29-15-10-26-9-12.0-9-14-16.0-21.0-21-13-13-14-15-18.0-17.0-12-13.0-14.2 | 1 | 1 |
| Igbo58 | 14-13.0-21-29-16-9-28-9-9.0-10-16-16.0-21.0-23-12-13-14-12-17.0-14.0-11-17.0-18.0 | 1 | 1 |
| Igbo59 | 20-13.0-22-31-16-10-26-12-12.0-10-15-17.0-17.0-23-11-11-13-13-18.0-14.0-11-13.0-15.0 | 1 | 1 |
| Igbo6 | 17-13.0-21-30-15-10-27-12-14.0-14-13-19.0-22.0-21-14-11-14-13-17.0-15.0-11-14.0-15.0 | 1 | 1 |
| Igbo60 | 14-13.0-21-30-15-10-28-11-11.0-11-14-19.0-22.0-21-12-11-11-13-16.0-16.0-10-17.0-18.0 | 1 | 1 |
| Igbo7 | 16-13.0-21-30-15-11-25-11-11.0-10-13-17.0-21.0-21-12-11-14-13-16.0-17.0-10-15.0-16.0 | 1 | 1 |
| Igbo8 | 16-13.0-21-30-15-10-23-11-11.3-11-14-14.0-21.0-21-11-11-13-15-18.0-15.0-11-16.0-16.0 | 1 | 1 |
| Igbo9 | 17-14.0-21-30-17-10-29-11-12.1-14-13-13.2-19.0-22-12-13-13-13-16.0-15.0-11-16.0-17.0 | 1 | 1 |

**Haplotypes are reported as allele strings in a fixed marker order corresponding to the PowerPlex® Y23 System Kit. The marker order used is: DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS456, YGATAH4, DYS385a, and DYS385b.**

## Table 5.15. Haplotype Frequency Specific to the Yoruba Ethnic Group for the PowerPlex® Y23 System Kit

| POPULATION | HAPLOTYPE | FREQUENCY | COUNT |
|---|---|---|---|
| Yoruba1 | 12-11.0-19-31-17-10-22-11-8.0-10-17-20.0-20.0-21-11-11-14-13-15.0-16.0-11-13.0-15.0 | 1 | 1 |
| Yoruba10 | 19-14.0-14-29-17-10-26-11-13.0-10-14-18.0-17.0-22-11-11-13-13-18.0-15.0-10-11.0-12.0 | 1 | 1 |
| Yoruba11 | 15-13.0-21-30-17-10-29-11-11.0-11-15-20.0-21.0-21-13-11-14-14-17.0-15.0-12-14.2-14.2 | 1 | 1 |
| Yoruba12 | 17-13.0-20-29-15-10-28-12-10.0-10-15-18.0-20.0-21-12-11-14-13-17.0-21.0-11-17.0-19.0 | 1 | 1 |
| Yoruba13 | 15-13.0-21-29-16-10-25-11-11.0-11-14-18.0-21.0-21-12-11-14-14-16.0-15.0-11-14.0-15.0 | 1 | 1 |
| Yoruba14 | 16-14.0-21-31-16-10-24-11-11.0-10-14-17.0-22.0-21-12-11-13-14-16.0-14.0-12-11.0-13.0 | 1 | 1 |
| Yoruba15 | 15-12.0-21-31-15-10-28-12-12.0-10-14-19.0-21.0-21-11-11-13-13-16.0-15.0-9-15.0-17.0 | 1 | 1 |
| Yoruba16 | 15-13.0-20-30-16-11-25-12-10.0-11-14-16.0-22.0-21-12-11-11-14-17.0-16.0-9-17.0-21.0 | 1 | 1 |
| Yoruba17 | 17-13.0-21-30-16-10-25-11-11.0-11-14-17.0-21.0-21-12-11-13-14-15.0-17.0-11-18.0-20.0 | 1 | 1 |
| Yoruba18 | 15-13.0-21-30-15-10-28-12-11.0-11-14-19.0-21.0-21-11-13-13-14-15.0-16.0-12-17.0-20.0 | 1 | 1 |
| Yoruba19 | 15-13.0-21-30-17-10-28-11-11.0-10-14-19.0-21.0-21-13-11-14-13-15.0-15.0-11-14.0-19.0 | 1 | 1 |
| Yoruba2 | 17-14.0-19-30-16-10-22-13-11.0-12-14-18.0-23.0-25-12-11-11-13-17.0-15.0-12-16.0-17.0 | 1 | 1 |
| Yoruba20 | 15-13.0-21-29-16-10-29-11-11.0-11-15-20.0-21.0-21-13-11-14-14-17.0-15.0-9-11.0-13.0 | 1 | 1 |
| Yoruba21 | 17-12.0-21-29-15-10-28-12-11.0-11-14-21.0-20.0-21-12-11-13-14-17.0-15.0-10-15.0-17.0 | 1 | 1 |
| Yoruba22 | 15-13.0-21-29-16-10-28-11-12.0-10-14-17.0-22.0-21-11-11-11-14-16.0-17.0-11-17.0-21.0 | 1 | 1 |
| Yoruba23 | 16-12.0-22-30-15-11-28-12-11.0-11-14-20.0-22.0-21-12-11-11-14-17.0-15.0-12-18.0-18.0 | 1 | 1 |
| Yoruba24 | 16-13.0-22-31-15-10-28-12-11.0-11-14-19.0-21.0-21-11-11-14-13-17.0-15.0-11-17.0-21.0 | 1 | 1 |
| Yoruba25 | 16-14.0-21-31-15-10-27-11-10.0-11-14-17.0-21.0-21-11-11-11-14-16.0-16.0-10-14.0-17.0 | 1 | 1 |
| Yoruba26 | 16-14.0-21-31-17-10-27-11-10.0-10-14-17.0-21.0-21-12-11-13-14-16.0-17.0-11-13.0-15.0 | 1 | 1 |
| Yoruba27 | 16-14.0-21-30-15-10-25-11-12.0-11-14-17.0-21.0-21-12-12-14-15-15.0-15.0-9-13.0-13.0 | 1 | 1 |
| Yoruba28 | 15-14.0-21-29-16-10-26-11-13.0-11-14-19.0-21.0-21-12-11-9-12-16.0-15.0-11-15.0-18.0 | 1 | 1 |
| Yoruba29 | 19-13.0-21-31-15-10-24-11-12.0-11-14-18.0-22.0-21-14-11-9-14-17.0-14.0-11-17.0-18.0 | 1 | 1 |
| Yoruba3 | 18-14.0-22-31-15-10-24-12-12.0-10-16-16.0-19.0-21-12-11-9-13-19.0-14.0-12-17.0-18.0 | 1 | 1 |
| Yoruba30 | 15-14.0-21-30-15-10-26-12-13.0-10-14-19.0-21.0-21-12-11-14-12-17.0-14.0-12-13.0-16.0 | 1 | 1 |
| Yoruba31 | 15-13.0-21-31-16-10-28-11-8.0-10-14-13.0-21.0-21-12-11-13-13-17.0-13.0-11-14.0-14.0 | 1 | 1 |
| Yoruba32 | 15-13.0-21-30-16-10-28-11-11.0-11-14-20.0-21.0-21-12-11-13-13-17.0-15.0-11-19.0-19.0 | 1 | 1 |
| Yoruba33 | 16-13.0-21-30-17-10-25-12-11.0-11-14-17.0-20.0-21-14-11-13-14-16.0-15.0-11-14.0-20.0 | 1 | 1 |
| Yoruba34 | 16-14.0-21-29-17-10-25-12-10.0-10-14-17.0-21.0-21-12-11-13-15-15.0-13.0-12-17.0-18.0 | 1 | 1 |
| Yoruba35 | 17-13.0-21-30-15-10-26-11-11.0-11-14-18.0-21.0-21-12-11-11-15-16.0-16.0-11-14.0-18.0 | 1 | 1 |
| Yoruba36 | 17-13.0-22-31-15-10-25-11-12.0-10-14-17.0-21.0-21-12-11-13-13-16.0-17.0-9-17.0-18.0 | 1 | 1 |
| Yoruba37 | 16-14.0-20-30-16-10-25-12-11.0-11-14-17.0-21.0-21-11-11-13-14-17.0-14.0-10-18.0-18.0 | 1 | 1 |
| Yoruba38 | 17-15.0-21-33-16-10-25-11-11.0-11-14-19.0-22.0-21-11-11-13-15-17.0-15.0-10-13.0-17.0 | 1 | 1 |
| Yoruba39 | 16-13.0-22-33-15-10-25-11-11.0-10-13-16.0-21.0-21-12-11-11-15-15.0-16.0-15-14.0-21.0 | 1 | 1 |
| Yoruba4 | 17-14.0-22-32-16-11-26-11-10.0-10-15-18.0-20.0-21-12-11-13-13-18.0-18.0-14-16.0-18.0 | 1 | 1 |
| Yoruba40 | 17-14.0-20-31-16-10-24-12-11.0-10-14-18.0-21.0-21-13-11-13-14-16.0-17.0-15-11.0-11.0 | 1 | 1 |
| Yoruba41 | 17-14.0-20-30-17-10-24-12-11.0-11-14-18.0-21.0-21-13-11-14-14-16.0-17.0-12-18.0-18.0 | 1 | 1 |
| Yoruba42 | 16-13.0-20-31-17-10-25-11-12.0-10-13-13.0-21.0-21-14-13-13-14-17.0-17.0-11-13.0-17.0 | 1 | 1 |
| Yoruba43 | 18-14.0-21-32-17-10-27-11-12.0-11-14-20.0-21.0-21-13-11-11-13-16.0-15.0-9-13.0-17.0 | 1 | 1 |
| Yoruba44 | 16-13.0-23-32-15-10-25-12-11.0-11-14-19.0-21.0-21-11-11-11-14-16.0-16.0-10-17.0-17.0 | 1 | 1 |
| Yoruba45 | 16-13.0-21-31-16-10-25-12-10.0-10-14-17.0-20.0-21-11-11-13-15-15.0-20.0-11-13.0-18.0 | 1 | 1 |
| Yoruba46 | 17-14.0-20-31-15-10-24-11-11.0-11-14-18.0-21.0-21-11-11-14-14-16.0-14.0-10-17.0-17.0 | 1 | 1 |
| Yoruba47 | 17-13.0-21-29-16-10-25-11-12.0-11-14-17.0-21.0-21-12-11-11-15-14.0-17.0-11-16.0-18.0 | 1 | 1 |
| Yoruba48 | 17-14.0-20-32-17-10-24-12-11.0-11-14-18.0-21.0-21-13-11-11-14-16.0-17.0-11-15.0-20.0 | 1 | 1 |
| Yoruba49 | 14-13.0-21-30-15-10-28-11-11.0-10-14-13.0-21.0-21-14-11-14-13-17.0-20.0-9-16.0-17.0 | 1 | 1 |
| Yoruba5 | 16-13.0-21-32-15-11-24-12-11.0-11-14-20.0-21.0-21-11-11-14-13-17.0-14.0-11-17.0-18.0 | 1 | 1 |
| Yoruba50 | 16-14.0-21-29-15-10-25-12-11.0-11-14-18.0-21.0-21-11-12-14-13-17.0-15.0-12-16.0-17.0 | 1 | 1 |
| Yoruba51 | 15-13.0-21-31-15-10-26-12-11.0-11-14-19.0-22.0-21-12-11-14-13-17.0-15.0-12-14.0-17.0 | 1 | 1 |
| Yoruba52 | 17-13.0-21-30-15-10-25-11-11.0-11-14-18.0-21.0-21-11-11-13-14-16.0-15.0-11-13.0-15.0 | 1 | 1 |
| Yoruba53 | 16-13.0-21-30-16-10-25-11-11.0-11-13-17.0-21.0-21-13-11-13-15-17.0-16.0-11-16.0-17.0 | 1 | 1 |
| Yoruba54 | 14-13.0-21-29-15-10-28-12-10.0-11-14-19.0-21.0-21-11-11-14-13-17.0-15.0-9-16.0-17.0 | 1 | 1 |
| Yoruba55 | 16-13.0-21-30-15-10-25-12-11.0-10-14-17.0-21.0-21-14-11-13-15-17.0-16.0-10-16.0-17.0 | 1 | 1 |
| Yoruba6 | 18-13.0-20-31-15-10-26-13-11.0-11-14-18.0-21.0-21-11-11-11-15-15.0-13.0-15-16.0-19.0 | 1 | 1 |
| Yoruba7 | 16-13.0-20-30-17-10-28-12-12.0-11-14-13.0-20.0-21-11-11-14-13-17.0-15.0-11-14.0-14.0 | 1 | 1 |
| Yoruba8 | 15-12.0-20-30-15-10-28-11-12.0-11-14-18.0-21.0-21-12-11-14-13-15.0-15.0-12-17.0-21.0 | 1 | 1 |
| Yoruba9 | 17-13.0-21-30-15-10-26-11-12.0-11-14-18.0-23.0-21-13-11-13-14-17.0-15.0-11-18.0-20.0 | 1 | 1 |

**Haplotypes are reported as allele strings in a fixed marker order corresponding to the PowerPlex® Y23 System Kit. The marker order used is: DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS456, YGATAH4, DYS385a, and DYS385b.**

## Table 5.16. Haplotype Frequency Specific to the Hausa-Fulani Ethnic Group for the PowerPlex® Y23 System Kit

| POPULATION | HAPLOTYPE | FREQUENCY | COUNT |
|---|---|---|---|
| **Hausa1** | 16-13.0-21-33-16-12-28-12-11.0-11-14-15.0-21.0-21-11-12-14-14-17.0-15.0-16-16.0-16.0 | 1 | 1 |
| **Hausa10** | 16-14.0-20-33-15-10-22-11-11.0-13-14-19.0-25.0-24-12-11-13-13-15.0-21.0-16-11.0-11.0 | 1 | 1 |
| **Hausa11** | 16-13.0-22-33-16-10-27-12-11.0-11-14-13.0-21.0-21-13-12-11-13-17.0-14.0-16-11.0-12.0 | 1 | 1 |
| **Hausa12** | 15-13.0-20-33-16-11-28-11-11.0-13-13-19.0-21.0-22-12-11-13-15-15.0-16.0-11-19.0-19.0 | 1 | 1 |
| **Hausa13** | 13-13.0-20-33-15-10-24-11-12.0-13-13-16.0-21.0-21-14-11-14-14-17.0-21.0-11-16.0-20.0 | 1 | 1 |
| **Hausa14** | 15-12.3-21-33-15-8-26-12-11.0-11-13-16.0-21.0-21-13-12-14-13-15.0-16.0-16-14.0-16.0 | 1 | 1 |
| **Hausa15** | 17-12.0-21-33-15-8-24-11-10.0-11-15-16.0-20.0-22-12-12-13-14-16.0-21.0-11-16.0-16.0 | 1 | 1 |
| **Hausa16** | 17-14.0-22-33-16-10-26-12-10.0-11-15-15.0-17.0-23-13-11-13-13-17.0-13.0-16-15.0-16.0 | 1 | 1 |
| **Hausa17** | 17-13.0-21-31-15-10-26-11-12.0-13-13-17.0-21.0-21-12-12-11-13-14.0-15.0-11-18.0-18.0 | 1 | 1 |
| **Hausa18** | 14-13.0-20-33-15-11-28-12-11.0-13-13-13.0-21.0-21-11-11-14-14-16.0-16.0-16-14.0-14.0 | 1 | 1 |
| **Hausa19** | 15-13.0-20-30-16-10-28-11-12.0-11-14-19.0-22.0-21-10-12-13-13-16.0-16.0-11-16.0-16.0 | 1 | 1 |
| **Hausa2** | 15-13.0-21-33-15-10-26-11-12.0-11-14-17.0-20.0-21-12-11-14-14-15.0-16.0-16-11.0-12.0 | 1 | 1 |
| **Hausa20** | 17-13.0-21-33-16-10-25-12-12.0-13-14-18.0-23.0-21-11-11-11-15-17.0-13.0-11-14.0-14.0 | 1 | 1 |
| **Hausa21** | 15-13.0-21-33-15-10-26-12-11.0-11-14-17.0-22.0-21-11-11-11-14-15.0-15.0-16-17.0-18.0 | 1 | 1 |
| **Hausa22** | 13-12.0-21-33-16-8-25-11-11.0-13-13-17.0-20.0-21-10-12-13-13-14.0-16.0-11-14.0-14.0 | 1 | 1 |
| **Hausa23** | 16-13.0-21-33-15-10-25-11-10.0-13-14-17.0-21.0-21-11-11-14-13-15.0-15.0-11-18.0-18.0 | 1 | 1 |
| **Hausa24** | 16-13.0-22-32-15-10-26-12-11.0-13-14-20.0-22.0-21-12-11-13-13-17.0-15.0-11-15.0-17.0 | 1 | 1 |
| **Hausa25** | 17-13.0-21-33-16-10-26-12-10.0-11-13-17.0-22.0-21-12-11-14-13-16.0-18.0-16-10.0-10.0 | 1 | 1 |
| **Hausa26** | 16-13.0-21-31-15-10-28-11-11.0-11-14-17.0-21.0-21-13-13-14-14-18.0-17.0-11-16.0-16.0 | 1 | 1 |
| **Hausa27** | 17-13.0-22-33-15-10-26-12-12.0-11-14-16.0-21.0-21-13-11-13-13-16.0-16.0-16-15.0-15.0 | 1 | 1 |
| **Hausa28** | 17-13.0-21-33-15-10-26-16-12.0-13-14-17.0-21.0-21-13-11-11-15-15.0-15.0-11-12.0-15.0 | 1 | 1 |
| **Hausa29** | 19-13.0-21-33-16-10-26-11-11.0-11-15-17.0-22.0-21-11-12-13-13-16.0-13.0-11-11.0-12.0 | 1 | 1 |
| **Hausa3** | 15-11.0-21-33-16-10-22-12-11.0-11-14-17.0-17.0-22-13-11-13-13-14.0-16.0-16-11.0-11.0 | 1 | 1 |
| **Hausa30** | 15-12.3-20-33-15-10-28-11-12.0-11-14-18.0-21.0-21-12-11-14-13-15.0-15.0-11-13.0-14.0 | 1 | 1 |
| **Hausa31** | 17-13.0-21-33-15-10-24-10-11.0-11-14-18.0-22.0-21-11-11-11-14-16.0-14.0-11-14.0-17.0 | 1 | 1 |
| **Hausa32** | 15-14.0-19-32-15-10-28-12-10.0-11-14-18.0-17.0-21-12-11-11-13-17.0-18.0-11-10.0-11.0 | 1 | 1 |
| **Hausa33** | 18-14.0-22-33-15-10-28-10-12.0-11-14-17.0-22.0-21-12-12-11-13-17.0-16.0-16-12.0-13.0 | 1 | 1 |
| **Hausa34** | 20-13.0-21-33-16-10-25-11-11.0-11-14-17.0-20.0-21-11-11-14-15-16.0-15.0-11-12.0-15.0 | 1 | 1 |
| **Hausa35** | 14-13.0-20-31-15-10-25-12-10.0-13-15-20.0-21.0-22-13-12-13-14-17.0-16.0-16-12.0-12.0 | 1 | 1 |
| **Hausa36** | 14-13.0-19-31-17-10-27-11-11.0-11-14-13.0-21.0-24-12-11-14-15-17.0-18.0-16-18.0-18.0 | 1 | 1 |
| **Hausa37** | 15-13.0-22-33-17-8-28-11-11.0-11-14-18.0-21.0-21-11-12-13-13-18.0-16.0-11-19.0-19.0 | 1 | 1 |
| **Hausa38** | 18-13.0-20-32-16-10-23-11-10.0-11-15-13.0-22.0-23-12-11-13-13-14.2-14.0-16-17.0-18.0 | 1 | 1 |
| **Hausa39** | 15-12.0-21-33-17-10-28-10-11.0-13-14-18.0-25.0-21-13-11-14-13-15.0-16.0-11-13.0-18.0 | 1 | 1 |
| **Hausa4** | 17-13.0-21-33-16-10-25-12-12.0-11-13-17.0-22.0-22-12-11-13-13-17.0-16.0-16-13.0-14.0 | 1 | 1 |
| **Hausa40** | 13-14.0-19-33-15-11-23-12-12.0-13-15-14.0-23.0-24-11-11-14-13-18.0-13.0-16-16.0-18.0 | 1 | 1 |
| **Hausa41** | 13-12.0-20-33-15-8-23-11-12.0-11-14-13.0-21.0-21-13-13-13-14-17.0-13.0-11-18.0-18.0 | 1 | 1 |
| **Hausa42** | 13-12.0-20-31-15-8-25-12-11.0-11-14-16.0-20.0-21-12-11-14-15-17.0-13.0-11-14.0-17.0 | 1 | 1 |
| **Hausa43** | 18-13.0-21-33-15-10-25-12-11.0-13-13-18.0-22.0-21-12-12-13-13-16.0-16.0-16-17.0-19.0 | 1 | 1 |
| **Hausa44** | 13-12.0-21-33-15-8-25-11-11.0-11-13-13.0-20.0-21-12-11-14-13-14.0-13.0-16-17.0-20.0 | 1 | 1 |
| **Hausa45** | 17-13.0-21-33-16-10-25-11-11.0-13-14-13.0-21.0-21-11-11-13-15-16.0-15.0-16-15.0-17.0 | 1 | 1 |
| **Hausa46** | 15-14.0-19-32-15-11-23-16-12.0-11-14-19.0-23.0-22-10-11-13-13-18.0-16.0-11-15.0-19.0 | 1 | 1 |
| **Hausa47** | 15-13.0-20-33-17-11-28-11-11.0-11-14-17.0-23.0-21-12-12-14-12-16.0-21.0-11-16.0-19.0 | 1 | 1 |
| **Hausa48** | 15-13.0-21-33-17-10-28-11-11.0-13-14-20.0-21.0-21-11-11-13-13-16.0-18.0-16-17.0-17.0 | 1 | 1 |
| **Hausa49** | 15-14.0-23-32-17-10-27-11-11.0-11-13-18.0-21.0-21-11-12-13-14-16.0-13.0-11-13.0-15.0 | 1 | 1 |
| **Hausa5** | 17-12.0-21-32-16-10-25-10-12.0-11-14-15.0-22.0-21-12-11-11-14-19.0-16.0-11-17.0-19.0 | 1 | 1 |
| **Hausa50** | 16-13.0-21-31-16-10-25-16-11.0-11-14-16.0-20.0-21-13-12-13-13-17.0-16.0-11-15.0-16.0 | 1 | 1 |
| **Hausa51** | 16-13.0-23-33-15-10-25-12-10.0-11-14-13.0-23.0-21-12-11-13-13-17.0-15.3-16-11.0-12.0 | 1 | 1 |
| **Hausa52** | 15-13.0-21-33-15-10-28-11-11.0-13-14-19.0-21.0-21-11-12-13-13-17.0-13.0-10-11.0-13.0 | 1 | 1 |
| **Hausa6** | 17-13.0-21-33-15-10-28-12-11.0-11-14-19.0-21.0-21-12-11-14-13-18.0-15.0-11-19.0-20.0 | 1 | 1 |
| **Hausa7** | 18-12.0-19-33-15-10-25-16-11.0-13-14-16.0-21.0-22-11-11-13-13-16.0-16.0-16-18.0-18.0 | 1 | 1 |
| **Hausa8** | 18-14.0-21-33-17-10-25-11-11.0-11-14-17.0-22.0-21-11-11-13-14-16.0-15.0-11-17.0-19.0 | 1 | 1 |
| **Hausa9** | 16-14.0-21-30-16-10-27-12-11.0-11-14-17.0-22.0-21-11-12-14-14-16.0-21.0-11-17.0-20.0 | 1 | 1 |

**Haplotypes are reported as allele strings in a fixed marker order corresponding to the PowerPlex® Y23 System Kit. The marker order used is: DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS456, YGATAH4, DYS385a, and DYS385b.**

**Table 5.17. Summary of unique and recurrent haplotypes, Haplotype Diversity (HD), and Discrimination Capacity (DC) for each ethnolinguistic group and the combined dataset.**

| Ethnolinguistic group | Number of unique haplotypes | Number of recurrent haplotypes within a group | Haplotypes Diversity (%) | Discrimination Capacity (%) |
|---|---|---|---|---|
| **Igbo** | 60 | 0 | 98.3 | 99.9 |
| **Yoruba** | 55 | 0 | 98.2 | 99.9 |
| **Hausa-Fulani** | 52 | 0 | 98.1 | 99.9 |
| **Combined dataset** | 167 | 0 | 99.4 | 99.9 |

The findings related to allele frequencies, forensic parameters, and private alleles obtained from the Genetic Analysis in Excel (GenAIEx 6.5) software and the STR Analysis for Forensics (STRAF-A 2.15), were shown in the tables above. In addition, graphical representations of this data are provided in the **Appendices 55** to **77** for further reference.

The Nigerian dataset comprised 33 alleles across Y-STR loci, with repeat numbers ranging from 8 to 33. Microvariant alleles were identified at the loci DYS458, DYS533, DYS576, and DYS635 among the three ethnic groups, showing varying proportions. Among 167 unrelated males, allele frequencies for single-copy loci ranged from 0.006 to 0.838, with the highest frequency for allele 10 in the DYS391 marker at 0.838, followed by allele 21 in the DYS390 marker at 0.766 (see **Table 5.1**). The allele frequency of the multicopy DYS385 locus, which includes DYS385a and DYS385b, was analysed as a single marker (DYS385a/b) encapsulating two allelic combinations (refer to **Table 5.2**). A total of 47 allelic combinations were identified, featuring 14 different alleles. Microvariant alleles (14.2 and 16.2) were observed in the DYS385a/b locus. Allele frequencies ranged from 0.006 to 0.090, with the highest frequency for allele 17,18 at 0.090 (15 counts) and allele 13,15 at 0.071 (12 counts) (**Table 5.2**). The DYS385a/b locus was the most polymorphic (PIC = 96.9%), diverse (GD = 96.9%), and had the highest number of allelic variants (v = 14). It was followed by DYS570 (PIC = 78.7%, GD = 81.4%, v = 10)**.** The DYS391 locus demonstrated the highest haplotype match probability (HMP = 71.0%), followed by DYS390 (HMP = 60.4%). The most discriminative loci were DYS481 and DYS570 (DC = 80.9%), with DYS576 next (DC = 77.4%). Overall, locus DYS385a/b is considered the most informative among the 23 loci studied **(Table 5.3)**. Conversely, DYS391 was the least polymorphic (PIC = 27.1%), diverse (GD = 28.8%), and discriminatory (DC = 28.6%). DYS385a/b had the lowest Haplotype Match Probability (HMP = 3.1%), while DYS19 had the fewest allelic variants (v= 4) **(Table 5.3)**. The Haplotype Diversity (HD) for the entire Nigerian population was 99.4%, with a high Discrimination Capacity of 99.9999% across the 23 loci **(Table 5.17)**. The forensic parameter values obtained from the Promega PowerPlex® Y23 System Kit indicate a highly diverse dataset. Every individual in the population possessed a unique haplotype, with no recurrent haplotypes (see **Tables 5.14** to **5.16** for haplotype frequencies).

The Igbo dataset identified 31 distinct alleles, with sizes ranging from 8 to 32. Microvariant alleles were found at loci DYS533, DYS570, and DYS635. Among the 60 unrelated males, allele frequencies varied from 0.017 to 0.833, with allele 10 in marker DYS391 having the highest frequency (0.833), followed by alleles 11 and 14 in markers DYS392 and DYS437 with frequencies

of 0.733 (see **Table 5.4**). The multicopy DYS385a/b marker showed 25 allelic combinations consisting of 11 different alleles (refer to **Table 5.5**), including micro variants 14.2 and 16.2. Allele frequencies ranged from 0.017 to 0.133, with the most frequent being 13,15 at 0.133 (8 counts), followed by 17,17 at 0.100 (6 counts) (**Table 5.5**). The DYS385a/b locus was the most polymorphic (PIC = 82.1%), diverse (GD = 83.5%), and had the highest number of allelic variants (v = 11). It was followed by DYS481 (PIC = 79.9%, GD = 83.0%, v = 9). The DYS391 locus exhibited the highest Haplotype Match Probability (HMP = 71.1%), followed by DYS392 (HMP = 60.9%). The DYS481 locus had the highest Discrimination Capacity (DC = 82.2%), with DYS458 next (DC = 74.9%). Among the 23 Y-STR loci in the Igbo sub-population, DYS385a/b was the most informative **(Table 5.6)**. In contrast, the DYS391 locus was the least polymorphic (PIC = 26.7%), discriminatory (DC = 28.9%), and diverse (GD = 29.4%). The DYS392 locus recorded the lowest number of allelic variants (v = 2). Additionally, the DYS385a/b locus showed the lowest Haplotype Match Probability (HMP = 3.1%) (**Table 5.6**). The Igbo subpopulation's Haplotype Diversity (HD) was 98.3%, and the Discrimination Capacity (DC) across the 23 loci was extremely high, measured at 99.9999% **(Table 5.17)**. Every individual possessed a unique haplotype, with no recurrent haplotypes (see **Table 5.14** for haplotype frequency).

The Yoruba subpopulation exhibited 26 distinct alleles, ranging from 8 to 33 in size, with no microvariant alleles identified. Among 55 unrelated males, allele frequencies varied from 0.018 to 0.927. Allele 10 in the DYS391 locus had the highest frequency (0.927), followed by allele 14 in the DYS437 locus (0.733) (see **Table 5.7**). The multicopy DYS385a/b locus exhibited 25 allelic combinations consisting of 12 different alleles (refer to **Table 5.8**). Allele frequencies ranged from 0.018 to 0.164, with allele combinations 16,17 being the most frequent at 0.164 (9 counts), followed by allele combinations 17,18 at 0.073 (4 counts) (**Table 5.8**). The DYS385a/b locus was identified as the most polymorphic (PIC = 94.6%) and diverse (GD = 95.1%), followed by DYS570 (PIC of 76.3%, GD of 80.7%). The DYS385a/b locus also had the highest number of allelic variants (v = 12), followed by DYS456 and Y-GATA-H4 loci (v = 8). The DYS390 locus showed the highest Haplotype Match Probability (HMP = 92.9%), followed by the DYS391 locus (HMP = 86.2%). The highest Discrimination Capacity was exhibited by DYS570 (DC = 79.3%), followed by DYS481 (DC = 78.0%). The DYS385a/b locus was the most informative among the 23 Y-STR loci studied in the Yoruba subpopulation. (**Table 5.9**). Conversely, the DYS390 locus was the least polymorphic (PIC = 7.0%), discriminatory (DC = 7.1%), and diverse (GD = 7.2%). The DYS391 and DYS392 loci recorded the lowest number of allelic variants (v =2). Additionally, the DYS385a/b locus had the lowest Haplotype Match Probability (HMP = 4.9%) (**Table 5.9**). The

Yoruba subpopulation's Haplotype Diversity (HD) was 98.2%, and the Discrimination Capacity (DC) across the 23 loci was extremely high, measuring 99.9999% (**Table 5.17**). Every individual in the population possessed a unique haplotype with no recurrent haplotypes (see **Table 5.15** for haplotype frequencies).

The analysis of 23 Y-STR loci within the Hausa-Fulani dataset identified 29 distinct alleles, with sizes ranging from 8 to 33. Microvariant alleles were found at the loci DYS389I, DYS456, and DYS458. In a survey of 52 unrelated males, allele frequencies varied from 0.019 to 0.750, with alleles 10 and 21 at loci DYS391 and DYS390 having the highest frequency (0.750), followed by allele 14 at locus DYS437 (0.654) (see **Table 5.10**). The multicopy DYS385a/b locus displayed 29 allelic combinations consisting of 11 different alleles (refer to **Table 5.11**). Allele frequencies ranged from 0.019 to 0.115. The most frequent allele combination was 17,18 at 0.115 (6 counts), followed by 17,19 at 0.076 (4 counts) (**Table 5.11**). Among the loci examined, the DYS385a/b locus was found to be the most polymorphic (PIC = 95.0%) and diverse (GD = 97.3%), followed by the DYS570 (PIC = 80.0%, GD = 83.7%). The DYS385a/b locus also had the highest number of allelic variants (v = 11), while the DYS456 and Y-GATA-H4 loci followed (v = 9). The DYS391 locus had the highest Haplotype Match Probability (HMP = 58.6%), closely followed by DYS390 (HMP = 58.3%). The highest Discrimination Capacity (DC) was seen in the DYS570 locus (DC = 82.1%), followed by DYS576 (81.0%). Among the 23 Y-STR loci in the Hausa-Fulani subpopulation, DYS385a/b was the most informative **(Table 5.12)**. In contrast, the DYS391 locus had the lowest polymorphic (PIC = 38.8%), discriminatory (DC = 41.4%), and gene diversity (GD = 42.2%) values. The DYS19 and DYS392 loci had the fewest number of allelic variants (v = 3), while the DYS385a/b locus recorded the lowest HMP (4.7%) **(Table 5.12)**. The Hausa-Fulani subpopulation's Haplotype Diversity (HD) was 98.1%, and the overall discrimination capacity across the 23 loci was extremely high at 99.9999% **(Table 5.17)**. Every individual in the population possessed a unique haplotype with no recurrent haplotypes (see **Table 5.16** for haplotype frequency).

**Table 5.13** displays the private alleles (PAs) exclusively identified within distinct subpopulations. A total of 24 PAs were documented from a sample of 167 unrelated males. The Hausa subpopulation, which consisted of 50 unrelated males, recorded the highest number, with 10 PAs. The Yoruba and Igbo subpopulations, comprising 60 and 52 unrelated males, identified 8 and 6 PAs, respectively. The Igbo subpopulation recorded the lowest number of private alleles.

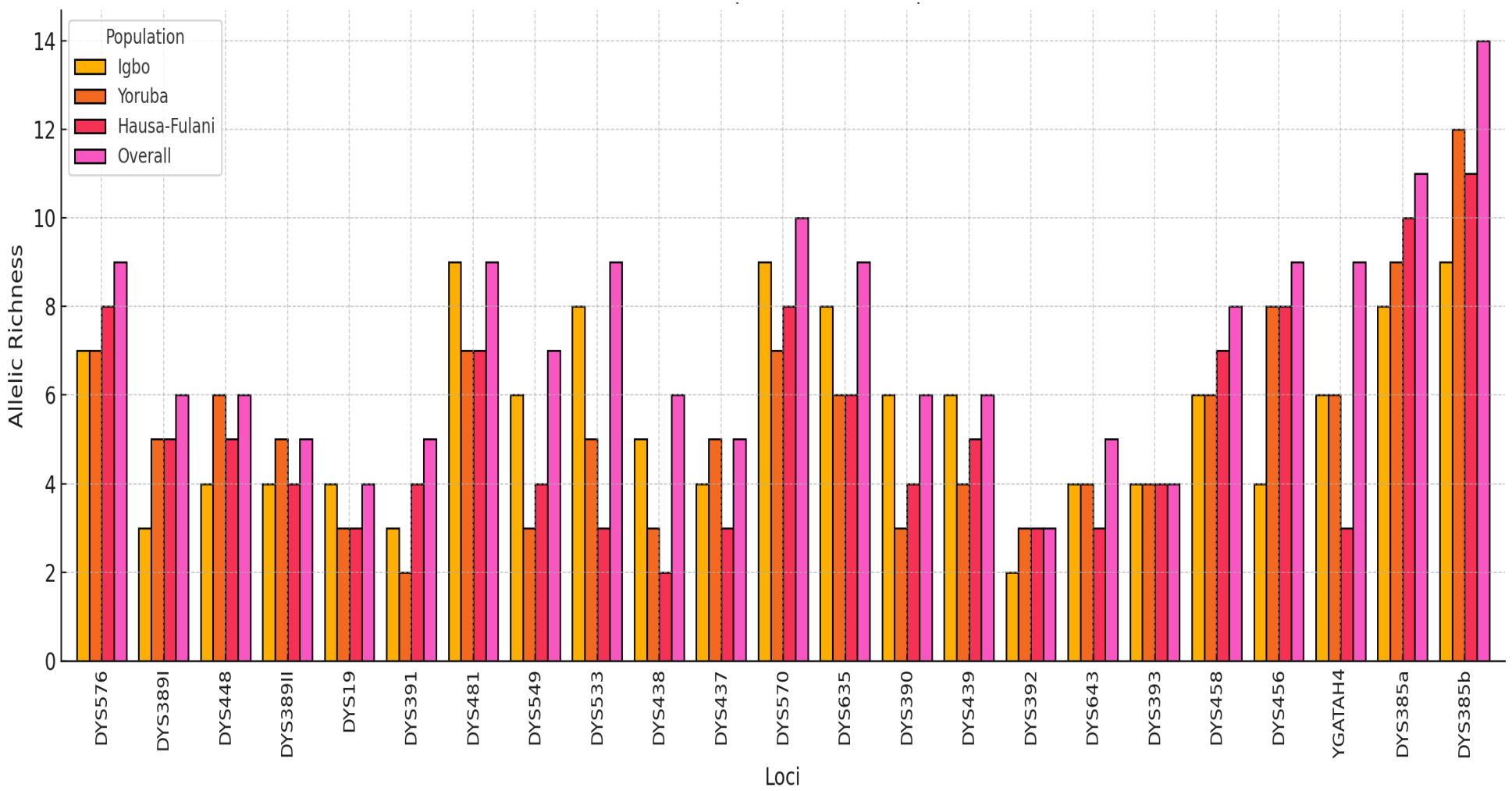



**Figure 5.2. Allelic richness per locus and population for the Igbo (yellow bars), Yoruba (orange bars), Hausa-Fulani (red bars), and the overall population (purple bars).**

**Figure 5.2** shows the allelic richness for each locus and population. Allelic richness refers to how many distinct alleles are observed at a locus in a population and is used as a descriptive indicator of genetic diversity rather than for statistical inference. It revealed that the DYS385a/b locus has the highest allelic richness in the overall population, with counts of 11 and 14 allelic richness. It is followed by the DYS570 locus, which has a count of 10. This trend is consistent across the three ethnic groups: the Igbo population exhibits the highest allelic richness at DYS385a/b with counts of 18 and 19, the Yoruba have counts of 9 and 12, and the Hausa-Fulani display counts of 10 and 11. In contrast, the DYS392 locus demonstrated the lowest allelic richness, with only 3 unique alleles present in the overall Nigerian population (See **Appendix 80** for the raw data).

### 5.4.3 INTER-SUBPOPULATION RELATIONSHIP

**Principal Component Analysis**

To investigate the genetic variation among sub-populations in Nigeria based on Y-STR marker values, PCA was used to assess population differentiation. The dataset includes Y-STR values from 167 Nigerian males, categorized into three ethnic groups: 60 Igbo individuals, 55 Yoruba individuals, and 52 Hausa-Fulani individuals. **Figure 5.3** represents the PCA plot generated in which each individual is represented by a dot: the Igbo sub-population is depicted with red dots, the Yoruba with green dots, and the Hausa-Fulani with blue dots. PCA, an unsupervised clustering technique, effectively reduces the high-dimensional Y-STR data (23 loci per person) into a lower-dimensional space (PC1 and PC2) while retaining most genetic variation. This analytical process minimizes the loss of information without requiring prior knowledge about the ancestral origins of the Nigerian population. Similar samples are expected to cluster in a PCA plot, while dissimilar samples should be distinctly separated. Y-STRs are transmitted paternally from father to son in a non-recombinant form. This characteristic provides valuable insights into population structure and ancestral or historical lineage patterns. The PCA plot illustrates overlapping groups, indicating shared ancestry or genetic similarity among the Igbo, Yoruba, and Hausa ethnic groups. Conversely, distinct clusters or individuals indicate variations in chromosomal lineage distributions.

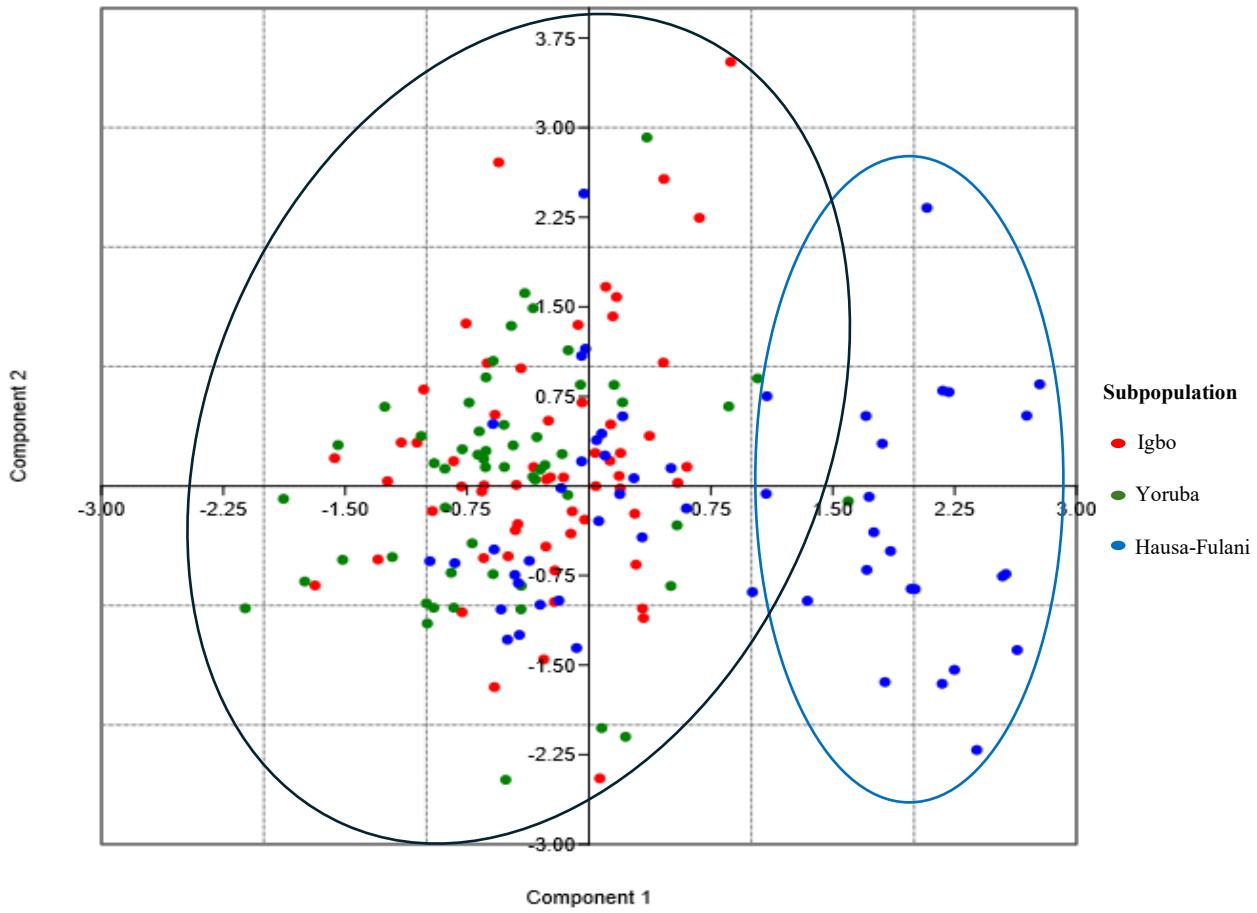



**Figure 5.3. Result of the Principal Component Analysis (PCA) performed on the subpopulations of Igbo, Yoruba, and Hausa-Fulani. This analysis was conducted using 23 Y-STR loci from the Promega PowerPlex® Y23 System Kit. In the PCA plot, each point represents one of the 167 samples obtained from unrelated males within the Nigerian population.**

The PCA plot **(Figure 5.3)** shows that the Igbo and Yoruba ethnic groups show some overlap, suggesting a shared ancestry. Additionally, some individuals from the Hausa-Fulani ethnic group exhibit overlapping paternal ancestry with the Igbos and Yorubas, suggesting historical gene flow. However, a subset of the Hausa-Fulani individuals is positioned away from the cluster, likely due to mixed paternal ancestry from genetically distinct lineages within the Hausa and Fulani populations. While mutations and allelic variation are inherent sources of genetic diversity and may contribute to minor dispersion, admixture is the primary driver of this separation. This observation further suggests that the Hausa-Fulani population likely has a history of genetic admixture from intermarriage between the Hausa and Fulani ethnic groups. Over generations, this blending of lineages has increased genetic diversity, as seen in the PCA plot, where some individuals stand out or are grouped differently. Overall, the PCA analysis did not clearly separate the three ethnic groups based on the Y-STR dataset but revealed patterns in paternal lineage. The overlap between Igbo

and Yoruba suggests shared ancestry or gene flow. At the same time, the partial separation of some Hausa-Fulani individuals indicates possible admixture reflecting intermarriage between the Hausa and Fulani populations.

**Neighbour-Joining (NJ)**

The Neighbour-Joining (NJ) method was used to visualise the phylogenetic distances between the three subpopulations, allowing for an assessment of their inter-subpopulation relationships. Genetically close populations will have closer branches. Genetically distant populations will have farther branches. The phylogenetic tree was created using data from the pairwise Nei's genetic distance population matrix of the 23 Y-STR loci from the subpopulations (see **Appendix 78**). The result illustrates the genetic distances between the Igbo, Yoruba, and Hausa-Fulani ethnic groups, as shown in **Figure 5.4.** Y-STRs were used to examine population-level differences in paternal genetic structure, highlighting the genetic relatedness and divergence among these groups.

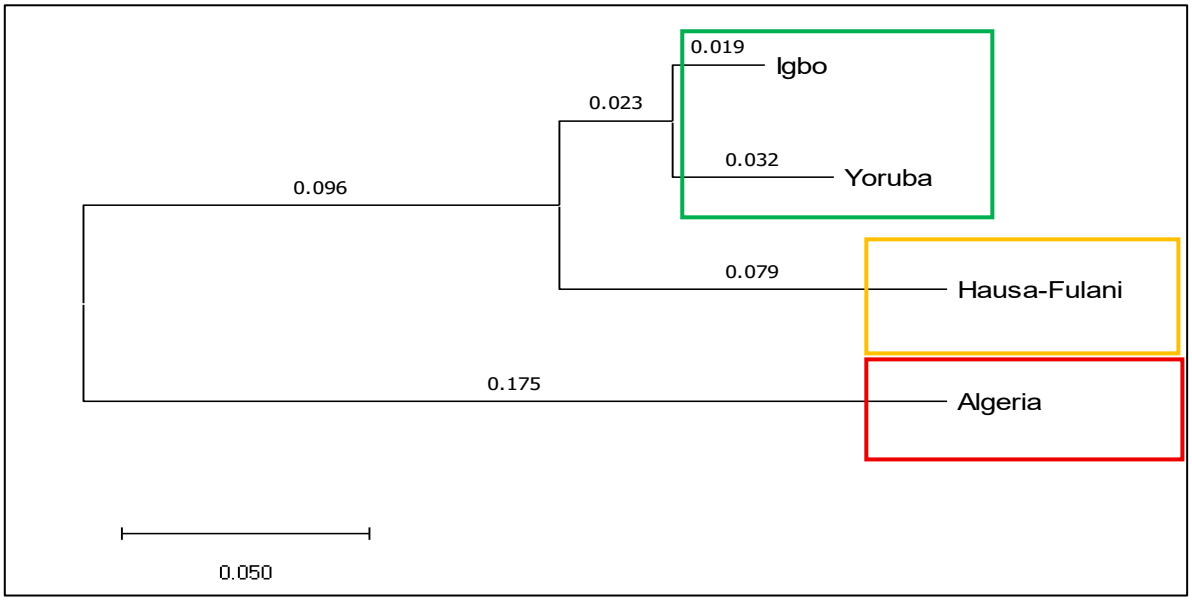



**Figure 5.4. Neighbour Joining (NJ) tree constructed from the pairwise Nei Genetic Distance population matrix, which illustrates the clustering of the Igbo, Yoruba, and Hausa-Fulani subpopulations with Algeria included as an outgroup. This analysis employed the 23 Y-STR loci obtained from the Promega PowerPlex® Y23 System Kit.**

As expected, the Neighbour-Joining (NJ) tree **(Figure 5.4)** shows that the Yoruba and Igbo ethnic groups have a low genetic distance, reflected by their short branch lengths, suggesting they are closely related at the paternal lineage level. In contrast, both the Igbo and Yoruba groups exhibit a moderate genetic distance from the Hausa-Fulani, indicating more distinct paternal genetic backgrounds. The genetic distance between the Yoruba and the Hausa-Fulani is slightly lower than

between the Igbo and the Hausa-Fulani, suggesting differences in their historical paternal connections.

## 5.4 DISCUSSION

Nigeria has a diverse population comprising more than 300 ethnic groups and over 500 languages, each with its unique customs and traditions (Agbaire and Dunne, 2024). This study assessed the genetic diversity in Nigeria's multi-ethnic and multicultural population. Some researchers have studied the genetic diversity of Y-chromosome STR loci among Indigenous ethnic groups in Nigeria, but none have utilised the Promega PowerPlex® Y23. In ancestry and population genetics, Y chromosome haplogroups map human migration patterns and establish paternal lineage. They are also utilised in biogeography to study population movements, as Y-STR analyses provide insights into human populations. Y-STR analysis has also been employed in genealogical studies, including surname research and historical investigations, to trace family lineages since all biological males in a lineage share identical Y-STR haplotype (Forouzesh *et al*., 2022; Roewer, 2019; Syndercombe, 2021).

To develop a robust and reliable reference database, it is essential to determine how Y-chromosomal haplotypes are distributed within each population. Generating high-resolution haplotype data and studying genetic diversity in well-studied populations are essential for accurately estimating the chances of finding a match when examining Y-STR profiles from questioned samples in forensic investigations (Gill, 2002; Schneider, 2007). The DYS385a/b locus exhibited the highest levels of polymorphic information content (PIC) and gene diversity (GD) among the markers studied. However, it also had the lowest haplotype match probability (HMP) within the Nigerian population and across the three ethnic groups: Hausa-Fulani, Igbo, and Yoruba. This low HMP is likely due to it being a multicopy marker with two alleles per individual. This increases genetic diversity and lowers the probability that unrelated individuals will share the same haplotype at this locus. The results showing that the DYS385a/b locus has the highest GD are consistent with findings by Fakorede *et al*. (2024), who analysed 461 saliva samples from unrelated Nigerian males, which included 96 Igbos, 139 Hausas, and 226 Yorubas, using the UniQTyper™ Y-10 system that also incorporates the DYS385a/b locus. Furthermore, the DYS481 and DYS570 loci exhibited strong discrimination capacity (DC) throughout the Nigerian population and within each of the three ethnic groups. The DYS481 locus showed the highest DC for the Igbo subpopulation, while the DYS570 locus had the highest DC for the Yoruba and Hausa-Fulani subpopulations. In contrast, the DYS391 locus exhibited the lowest levels of polymorphic

information content (PIC), gene diversity (GD), discrimination capacity (DC) across the Nigerian population and in the Igbo and Hausa-Fulani subpopulations. The Yoruba subpopulation, however, showed the lowest DC for the DYS390 locus. Despite this, the DYS390 and DYS391 loci demonstrated the highest haplotype match probability (HMP) within the Nigerian population and the various ethnic groups examined. In the Nigerian dataset, 167 individuals exhibited 167 unique haplotypes with no shared haplotypes. The overall haplotype diversity (HD) was 0.994, indicating strong individual discrimination potential. Subpopulations showed similar patterns: the Igbo group had an HD of 0.983, the Yoruba group 0.982, and the Hausa-Fulani group 0.981, all with zero unique haplotypes. The low genetic diversity values of DYS390 and DYS391 aligns with the findings of Martinez *et al*. (2017), who identified low diversity among the loci DYS390, DYS391, DYS438, DYS392, DYS437, and DYS533. Martinez *et al*. (2017) also reported a similar finding of haplotype diversity of 0.9998, where 140 haplotypes were identified, including two haplotypes shared by two individuals, indicating that Y-STR markers are very effective for distinguishing between different individuals when looking at multi-locus haplotypes.

Allelic richness is the number of different alleles found at a genetic locus in a population. It provides an idea of how variable that location is (Nei, 1987; Hartl and Clark, 2007; Butler, 2015). In this study, allelic richness differed among various loci and populations, reflecting the known characteristics of Y-STR markers. The duplicated locus DYS385a/b had the highest allelic richness, which makes sense because it has multiple copies. On the other hand, the nested loci DYS389 I and DYS389 II had lower diversity due to sharing a similar repeat structure. Rapidly mutating Y-STRs like DYS570 and DYS576 showed higher allelic richness in all populations, highlighting their usefulness in distinguishing Y-STR haplotypes.

According to Shinagawa *et al*. (2022), rare haplotypes can appear in Y-STR data due to several factors, such as recent or unique mutations, high genetic diversity stemming from multiple ancestral lineages or extensive historical migrations, sample size, significant genetic variation, the founder effect, and population structure. Every participant in this study had a distinct Y-STR profile, indicating that no two individuals possessed the same Y-STR haplotype. This indicates a very high level of genetic diversity among the unrelated Nigerian males analysed, with most haplotypes being rare and found in only one individual (i.e., singletons). Differences in sample composition may influence the variation in Y-STR haplotype frequencies between Nigerian males sampled in the United Kingdom and those in Nigeria, as UK-based individuals represent a limited subset of the broader Nigerian genetic diversity. When specific haplotypes are common in a population, they

may indicate shared paternal ancestry or genetic relatedness (Jobling and Tyler-Smith, 2003). However, this study's absence of shared haplotypes shows how effective Y-STR markers are for identifying individuals. This information can be helpful in forensic and population genetics to analyse individual identification, conduct kinship analyses, or study genetic relationships or structures within and between populations (Kayser, 2017).

The NJ and PCA analysis indicate that the Igbo and Yoruba ethnic groups are closely related and overlap, reflecting some degree of shared paternal ancestry. Conversely, the Hausa-Fulani ethnic group is more distantly related to the Igbo and Yoruba, suggesting more distinct paternal genetic backgrounds. The findings of the NJ and PCA are consistent with historical and anthropological data. The Igbo and Yoruba are speakers of the Niger Congo – Atlantic Congo language, which may account for some shared paternal ancestry (Hyman *et al*., 2019). In contrast, the Hausa-Fulani are speakers of the Chadic language, which is part of the Afro-Asiatic language family, suggesting a different migration and paternal ancestry pattern (Newman, 2009). Historically, the Hausa-Fulani have mixed with Sahelian and North African groups, whereas the Yoruba and Igbo have shared ancestry specifically within West Africa. It is not surprising to find unique traits among individuals within sub-populations and ethnic groups. Nigeria's known demographic history shows that certain ethnic groups are distinct but have overlapping ancestral lineage due to the country's heterogeneous nature, as this study's findings agree. The NJ and PCA provided valuable forensic insight into the genetic structure of Igbo, Yoruba, and Hausa-Fulani ethnic groups using the Y-STR markers generated from the Promega PowerPlex® Y23 System Kit. It demonstrates that the Y-STR markers can differentiate the Nigerian ethnic groups based on the paternal ancestral lineage. Examining a male suspect's Y-STR haplotype in forensic investigations can help identify the ancestry by comparing it to reference haplotypes in global databases, such as the Y Chromosome Haplotype Reference Database (YHRD). This method helps trace a male's geographic origin linked to crime scene evidence.

## 5.5 CONCLUSION

This research highlights the rich genetic information provided by the 23 Y-STR markers in the Promega PowerPlex® Y23 System Kit. It demonstrates their effectiveness in analysing the Nigerian population and distinguishing individuals from the Igbo, Yoruba, and Hausa-Fulani ethnic groups. The 23 Y-STR markers produced informative haplotype frequency data and recorded highly polymorphic, diverse, and discriminatory alleles. These outcomes can be applied in criminal investigations, paternity tests, and studies on the demographics and anthropology of the Nigerian

population. Y-STR analysis is indispensable for determining paternal lineages in forensic investigations, particularly in cases involving unidentified male individuals, kinship analysis or paternity testing. By creating a Y-STR haplotype database for specific subpopulations, haplotype frequencies can be estimated more accurately.

Importantly, Y-STR analysis is particularly valuable in sexual offence investigations, where male DNA is often present in low quantities or is mixed with an excess of female DNA. In such cases, autosomal STR analysis may be limited. In contrast, Y-STR typing enables the selective amplification and identification of male-specific genetic material, facilitating the detection and interpretation of male profiles in complex DNA mixtures. This makes Y-STR analysis indispensable for the identification of male contributors in cases involving sexual assault, unidentified male remains, or disputed paternity. The forensic data generated by this study are therefore of practical relevance. They should be actively utilised in forensic investigations, especially given the current challenges related to insecurity and criminal activity in Nigeria. Establishing subpopulation-specific databases would allow for more accurate estimation of haplotype frequencies and improve the calculation of match probabilities in forensic casework, thereby enhancing the reliability and evidential value of Y-STR profiling.

The analysis of an ancestry lineage is essential in applying an investigation lead. Suppose an unknown Y-STR profile is found during an analysis at a crime scene and is linked to a specific Y-STR cluster or group. In that case, forensic scientists can infer the suspect's potential ethnic background by examining the haplotype or utilising the YHRD database. Furthermore, Y-STR databases and reference populations are essential in forensic comparisons of sexual assault cases, paternity disputes, and haplogroup analysis. A crucial aspect of the importance of Y-STR databases and reference populations is the utility of population-specific Y-STR haplotype frequencies. The outcome of this study reinforces the need for ethnic-specific forensic databases, as the three Nigerian populations display genetic differentiation, so using a generalised African database might lead to some match probability errors. Statistical reliability is ensured in a well-curated Nigerian Y-STR database as the likelihood of an individual or suspect's ancestry can be determined. However, Y-STR testing alone cannot confirm an individual's identity, as males from the same paternal lineage will exhibit identical Y-STR haplotypes. It is essential to note that the Y-STR profile data generated from the Promega PowerPlex Y23 System Kit effectively differentiated the Nigerian ethnic groups. This outcome will further aid ancestry inference in forensic cases.

# CHAPTER SIX

# GLOBAL META-ANALYSIS OF AUTOSOMAL STR AND Y-STR DATASETS

## 6.1 INTRODUCTION

Understanding genetic structure reveals how genetic variation is distributed among different populations, how these populations are evolutionarily related, and the degree to which they interact with each other through gene flow or isolation. (Bossart and Prowell, 1998). Genetic structure within a population (or "within-population structure") describes the genetic differences among individuals in a single group. The differences within a population can be affected by several factors. These include mating patterns, population size, inbreeding, and genetic drift (Hoelzel *et al*., 2002). In contrast, the genetic structure among populations (also known as "between-population structure") expounds on the genetic distinctiveness of different groups. Natural selection, cultural or language differences, geographic barriers, migration, and gene flow can influence this genetic distinctiveness (Destro-Bisol *et al*., 2004; Kohler, 2023).

Analysing the genetic structure within and between populations requires statistical tools such as Principal Component Analysis (PCA), Wright's F-statistics ($F_{ST}$), *STRUCTURE*, and Analysis of Molecular Variance (AMOVA). Additional statistical tools used in this field are Neighbour Joining (NJ), Nei's Genetic Distance, Multi-Dimensional Scaling (MDS), and Slatkin's R-statistic ($R_{ST}$). **Analysis of Molecular Variance (AMOVA)** assesses genetic variation across multiple hierarchical levels of population structure (within populations, among populations, and among groups of populations). It produces Phi statistics (Φ-statistics) that help to quantify genetic differences. AMOVA determines if there is significant genetic differentiation among populations, indicating potential evolutionary divergence (Li *et al*., 2022). **Principal Component Analysis (PCA)** is a method employed for dimensionality reduction that recognises linear arrangements of variables and encapsulates the most data variance, making it more straightforward to visualise clustering and relationships among individuals or populations and infer population structure and admixture (mixing) between populations (Elhaik, 2022). ***STRUCTURE*** is a model-guided statistical algorithm that estimates the number of genetic populations ($K$) and assigns individual memberships to each of the $K$ populations probabilistically. It determines the genetic composition of groups by assessing the number of individuals that belong to various genetic groups or subpopulations. Additionally, *STRUCTURE* helps detect admixture and ancestral components within these populations (Pritchard *et al*., 2000). **Wright's F-statistics ($F_{ST}$)** measures the genetic differentiation among populations. It calculates the proportion of genetic differences attributed to variations between populations compared to the total genetic variation resulting from differences

in allele frequencies. This helps compare pairwise population distances or assess overall population structure (Wright, 1949). **Neighbour-Joining (NJ)** is a method that constructs phylogenetic relationships between populations based on genetic distance metrics, often visualised as an accurately scaled tree (Tamura *et al.*, 2021). **Multi-Dimensional Scaling (MDS)** is a technique that transforms high-dimensional datasets such as genetic distance matrices into low-dimensional space for visualisation of inter-population relationships, often highlighting clusters and gradients of genetic similarity while preserving the distances between data points, allowing for visualisation of population structure (Delicado and Pachón-García, 2024). **Nei's Genetic Distance** is a statistical technique that measures genetic divergence between populations based on allele frequency differences (Kanaka *et al*., 2023). **Slatkin's R-statistic** (***$R_{ST}$***) is a stepwise mutation model-based equivalent of $F_{ST}$ that measures of genetic differentiation between populations for microsatellite (STR) data, particularly useful for Y-STRs and other microsatellites where mutations accumulate in a stepwise manner (Holsinger and Weir, 2009).

According to extensive anatomical, archaeological, and genetic studies, modern humans (*Homo sapiens*) emerged in Africa approximately 200,000 to 300,000 years ago (Schlebusch and Jakobsson, 2018). Following their emergence, early modern humans gradually dispersed out of Africa in one or more migration waves, with the most widely supported dispersal event occurring around 50,000 to 70,000 years ago. This *Out of Africa* migration led to the colonisation of other continents (Stringer, 2016; Wang *et al*., 2021; Guanglin *et al.,* 2023). As these human population ancestral groups settled in different environments across Europe, Asia, the Americas, and beyond, population divergence occurred due to genetic drift, local adaptation, and geographic isolation—resulting in the genetic structure observed today. The signatures of these ancient demographic events and population movements can still be detected using various molecular markers, including mitochondrial DNA (mtDNA), single nucleotide polymorphisms (SNPs), Y-chromosomal STRs (Y-STRs), and autosomal short tandem repeats (STRs) (Excoffier *et al.*, 1992; Cavalli-Sforza *et al*., 1994; Pritchard *et al*., 2000).

Africa has a rich and complex history of human migration and intermixing, which has shaped the unique genetic patterns observed in modern populations. The continent is remarkably diverse—genetically, culturally, and linguistically—and is home to a vast array of ethnic and indigenous groups (Campbell and Tishkoff, 2008; Lipson *et al*., 2022). These primary indigenous groups are divided geographically into two regions: Saharan Africa and Sub-Saharan Africa. **Saharan Africa** is a vast desert region spanning much of North Africa, encompassing parts of Algeria, Libya, Egypt, Mauritania, Mali, Niger, Chad, and Sudan. The five North African countries—Morocco, Algeria,

Tunisia, Libya, and Egypt—border the Mediterranean Sea, but only their inland areas are part of the Sahara. **Sub-Saharan Africa** includes 49 countries south of the Sahara Desert and the transitional Sahel zone, spanning West, East, Central, and Southern Africa. The Sahara separates North Africa from Sub-Saharan Africa, highlighting differences in language, religion, history, and genetics. (Henn *et al*., 2012). The Berbers, Nubians, Toubou, Sahrawis, and Zaghawa people are among the indigenous populations of Saharan Africa who existed before the Arab-Islamic expansion. Genetic studies of these groups reveal a mixture of ancient North African lineages, back-migrations from Eurasia, and limited interbreeding with sub-Saharan Africans, particularly in the southern regions of the Sahara. Although these populations are historically and genetically distinct from sub-Saharan groups, they are not entirely isolated. Gene flow across the Sahara has occurred due to trade interactions in the Sahel and the trans-Saharan region (Henn *et al*., 2012; Fadhlaoui-Zid *et al*., 2013; Salem *et al*., 2025). Sub-Saharan Africa (SSA) Indigenous groups include the Bantu, Cushitic, Khoisan, Benue-Congolese, Ethio-Semitic, Gur, Mende, Nilotic, and Pygmy peoples. Sub-Saharan Africa is characterised by its vast and diverse regions, languages, cultures, ethnic groups, and genetic backgrounds. Historically, this area has been less influenced by Arabic-Islamic culture compared to Saharan Africa, although there is some overlap, particularly in the Sahel region (Campbell and Tishkoff, 2008; Tishkoff *et al*., 2009). Sub-Saharan Africa is recognised as the cradle of modern humans due to its high genetic diversity. This diversity comes from its ancient family lines and the varied histories of its people (Tishkoff *et al*., 2009). In North Africa, the genetic landscape is shaped by a mixture of indigenous Berber ancestry, ancient Eurasian back-migrations, and later Arab expansions following the Islamic conquests starting in the 7th century CE. These historical events introduced Middle Eastern genetic components into North African gene pools (Henn *et al*., 2012; Arauna *et al*., 2017). Due to admixture and migration, North African populations developed unique features that distinguish them from sub-Saharan Africans in terms of phenotype, culture, and language (Henn *et al*., 2012). Globally, the Transatlantic Slave Trade, which occurred from the 15th to the 19th centuries, had a significant effect on genetic patterns in the Americas. Millions of enslaved Africans, mainly from Central and West Africa, were taken to the New World against their will. Today, the genetic influence of these individuals is still evident in Afro-descendant populations across the Caribbean, South America, and North America, and the Studies of Y-STR and autosomal STR show that these populations mainly have African ancestry, along with varying amounts of European and Native American heritage (Salas *et al*., 2004; Bryc *et al*., 2010; Schlebusch, and Jakobsson, 2018). These historical movements have contributed to the complex mosaic of human diversity observable through modern genomic tools. A breakdown of the Indigenous groups, their features, locations, skin tones, and examples is presented in **Table 6.1.**

**Table 6.1. Summary of the features, locations, skin tones, language family, and ethnic groups of the major indigenous groups in Africa (Ehret, 1995; Wood *et al*., 2005; Campbell and Tishkoff, 2008; Tishkoff *et al*., 2009; Becker *et al*., 2011; Henn *et al*., 2012; Pagani *et al*., 2012; Perry *et al*., 2014; Hollfelder *et al*., 2017; Fregel *et al*., 2018; Deubel, 2020)**

| Indigenous Group | Physical Feature | Region/ Country | Skin Tone | Language Family | Notable Ethnic Group |
|---|---|---|---|---|---|
| Bantu | Medium to tall statue. Slender build. Varying height with some tending to be shorter. | **Central, East, and Southern Africa:** Angola, Botswana, Cameroon, DR Congo, Kenya, Mozambique, Tanzania, Zambia, Zimbabwe, and South Africa | Medium to dark skin. Some may be lighter skin. | Niger-Congo | Kimbundu, Bakongo, Tswana, Kalanga, Beti, Bulu, Konga, Luba, Lubya, Meru, Makua, Sena, Chaga, Haya, Bemba, Tonga, Shona, Ndebele, Zulu, Xhosa |
| Berber | Medium to strong facial features. Wavy to curly dark hair. Brown or light eyes. | **North Africa:** Algeria, Morocco, Libya, and Tunisia | Generally light brown to olive. Often dark to medium to light skin. | Afroasiatic | Mozabite, Siwa Berber, Tunisian Berber, Kabyle, Riffian, Shilha, Zayanes, Tuareg, Chaoui. |
| Benue-Congo (Non-Bantu) | Medium to tall stature, Broad nose. full lips, kinky hair. Round or oval faces, almond-shaped or round eyes. Average to athletic body build. | **West Africa:** Nigeria, Benin, and Cameroon | Brown to dark skin | Niger-Congo | Igbo, Yoruba, Ebira, Edo, Gwari, Ibibio, Idoma, Igala, Jukun, Nupe, Tiv |
| Cushite | Tall height, slender stature. Narrow faces, long noses, and high cheekbones. Tightly or loosely coiled hair texture. Brown to dark brown eyes. Moderately full lips. | **East Africa:** Ethiopia, Somalia, Djibouti, Eritrea, and Kenya, Sudan | Brown to dark skin (but not deep black) | Afroasiatic | Oroma, Somali, Afar, Beja, Sidama, Agaw, Saho, Burji, Hadiya |
| Gur | Medium to tall stature. Strong and muscular frame. Tightly coiled or kinky hair. Broad nose, full lips, oval or rounded face. | **West Africa:** Benin, Burkina Faso, Ghana, Ivory Coast, Mali, Niger, and Togo, | Medium to dark brown skin | Niger-Congo | Mossi, Bobo, Lobi, Dagara, Gurunsi, Bissa, Dagomba, Mamprusi, Frafra, Dagaaba, Talensi, Builsa, Gurma, Kotokoli, Lamba, Biali, Senufo, Koulango, Djimini, Kurumba, |
| Khoisan | Short to medium height. Slender and wiry frame build. Tightly coiled hair, flat broad nose and almond-shaped eyes. Steatopygia, especially women. | **Southern Africa:** Botswana, Namibia, South Africa, and Zambia | Light brown to golden brown skin | Khoisan | San, Khoikhoi |
| Mende | Tightly coiled or kinky hair. Broad nose, full lips, oval or round faces. Average height, athletic or medium build. Brown to dark brown eyes. | **West Africa:** Guinea, Liberia, Senegal, and Sierra Leone | Medium to dark brown skin | Niger-Congo | Mende |

| Indigenous Group | Physical Feature | Region/ Country | Skin Tone | Language Family | Notable Ethnic Group |
|---|---|---|---|---|---|
| Nilotes | Exceptionally tall (often 6+feet for men). Lean, long limb, narrow torso, (slim body fat percentage). High cheekbones, narrow nose, elongated face. Tightly coiled, short to medium length hair. Brown to dark brown, almond shaped eye. | **Central and East Africa:** DR Congo, Ethiopia, Kenya, Tanzania, South Sudan, and Uganda | Very dark to deep brown skin | Nilo-Saharan | Dinka, Nuer, Shilluk, Luo. Maasai, Turkana, Kalenjin, Teso |
| Nubian | Generally tall and lean. High cheekbones, long narrow noses, oval faces. Curly to coiled, medium to dark hair. Almond-shaped, dark brown eyes. | **North Africa:** Southern Egypt, and North Sudan | Medium to dark brown skin | Nilo-Saharan | Nobiin, Dongolawi, Halfawein, Mahas, Midob, Kenzi, Fadicca |
| Pygmy | Short stature (adults average 4'11 to 5'1 in height) – biologically adaptation for hunting in rainforest environment. Lean and compact body build. Broad noses, full lips and round faces. | **Central Africa:** Burundi, Cameroon, Central African Republic, Democratic Republic of Congo, Congo-Brazzaville, Gabon, Rwanda, and Uganda | Dark brown to deep black skin | Niger-Congo Nilo-Saharan | Asua, Mbuti, Twa, Baka, Bakola, Bedzam, Babongo, Batwa |
| Sahrawis | Wavy and curly, dark brown or black hair. Oval to angular faces, high cheek bones, and long noses. Medium to tall statue. Brown to dark brown eyes. | **North Africa:** Western Sahara, Southern Morocco, Algeria, Mauritania, Northern Mali | Light brown to dark brown skin | Afroasiatic | Sahrawis |
| Semite | Medium height, slender to average stature. Oval or narrow face, high cheekbones, pointed nose. Curly to coiled hairs with looser textures. Brown to dark eyes. Moderate to full lips. | **Horn of Africa:** Ethiopia, and Somalia, | Light brown to medium brown skin (some with olive undertone). | Afroasiatic | Amhara, Tigrayans, Tigrinya, Tigre, Gurage, Harari, Argobba |

**Table 6.1. continued**

The language families in Africa correspond to significant Indigenous population groupings. However, factors such as historical migrations, cultural changes, assimilation, and language shifts have added complexity to the distribution of languages. As a result, there are exceptions and overlaps, making languages practical but not definitive indicators for identifying indigenous groups in Africa (Campbell and Tishkoff, 2008; Tishkoff *et al*., 2009).

Africa has four prominent language families, each with various subgroups: 1. **Afroasiatic:** This family includes Semitic, Cushitic, Chadic, Berber, Egyptian (extinct), and Omotic languages (Frajzyngier and Shay, 2019). 2. **Khoisan:** Comprising Khoe, Tuu, Kx'a, Hadza, and Sandawe languages (Güldemann, 2014). 3. **Niger-Congo:** The largest and most diverse family, including Atlantic, Mende, Gur (Voltaic), Kwa, Volta-Niger (East Kwa), Benue-Congo, Bantu, Adamawa-Ubangi, Kordofanian, Ijoid, and Dogon languages (Williamson and Blench, 2000). 4. **Nilo-Saharan:** Encompasses the Nilotic, eastern Sudanic, central Sudanic, Saharan, Fur, and Songhay languages (Bender, 2000).

Among these language families, the Niger-Congo is the most diverse, with over 1,500 languages spoken by more than 500 million people across West, Central, East, and Southern Africa (Good, 2017; Heine and Nurse, 2000). The Afroasiatic language family follows next, comprising over 300 languages spoken by more than 350 million people in the Horn of Africa and North Africa (Hayward, 2000). The Nilo-Saharan family comes next, with over 200 languages spoken by more than 60 million people (Dimmendaal, 2008). Finally, the Khoisan family is the least diverse, featuring fewer than 40 languages spoken by over 300,000 people. It is also Africa's smallest and most endangered language family (Vossen, 2013).

## 6.2 AIM AND OBJETIVES

This investigation aims to analyse the genetic connections between the Nigerian population and diverse populations worldwide, utilising autosomal STR (Short Tandem Repeat) and Y-STR data. It is noteworthy that a comprehensive assessment comparing the Nigerian population with other groups from South America, North America, Europe, Asia, and Africa has not been undertaken to this extent previously.

The specific objectives for the work detailed in this chapter are as follows:

1. To examine how the Nigerian population differs from other populations worldwide using autosomal STR data with *STRUCTURE* analysis.

2. To calculate the genetic divergence and differentiation of the autosomal STR using Nei's genetic distance and $F_{ST}$ based genetic distance method, respectively.
3. To assess the genetic differentiation of Y-STR data among populations using Slatkin's $R_{ST}$ based genetic distance method.
4. To conduct a comparison of the Nigerian population with other populations using Multi-Dimensional Scaling (MDS) analysis based on autosomal STR and Y-STR data.
5. To analyse and visualise the relationships between the Nigerian population and other global populations using Neighbour-Joining (NJ) analysis based on autosomal STR and Y-STR data.
6. To quantify and partition genetic variation within and among populations in the autosomal STR and Y-STR datasets using AMOVA, and to assess their genetic structure and differentiation.

## 6.3 METHODOLOGY

Chapters 3-5 reported autosomal STR data obtained from 303 unrelated Nigerian individuals and Y-STR data from 167 unrelated Nigerian males. These findings were compared with corresponding autosomal STR data from 7,361 unrelated individuals (males and females), bringing the total number of unrelated individuals to 7,664. Additionally, haplotype data from 7,181 males contributed to a total of 7,348 participants with Y chromosomes in unrelated population samples from various regions around the world, including South America, North America, Europe, Asia, and Africa

### 6.3.1 DATASET COLLECTION

**Autosomal STR Datasets**

The autosomal STR datasets were sourced from the supplementary materials of previously published studies. The sampled population from Nigeria was compared with 22 reference populations, resulting in 23 population groups from 20 countries across five continents. Each country was represented as a single population group, except for the USA, which had four population groups. **Table 6.2** summarizes this information, including the population, assigned code, size, continent, and source.

**Y-STR Datasets**

The Y-STR datasets utilized in this research were obtained from the supplementary materials of previously published research. The sampled population from Nigeria, including the Igbo, Yoruba, and Hausa-Fulani ethnic groups, was compared with 47 reference populations, resulting in 50 population groups from 36 countries across 12 geopolitical regions and five continents. Each

population group was categorized based on its association with specific ethnolinguistic groups. **Table 6.3** summarises this information, including the population, language family, assigned code, size, ethnicity, continent, geopolitical region, and source.

**Table 6.2. Size and continent, assigned code and source of each population sample for the autosomal STR datasets**

| | Population | Assigned Code | Size | Continent | Source |
|---|---|---|---|---|---|
| **1** | Nigeria | NIG | 303 | Africa | Present study |
| **2** | Egypt | EGY | 265 | | AbdEl-Hafez *et al*., 2019 |
| **3** | Ghana | GHA | 109 | | Kofi *et al*., 2020 |
| **4** | Morocco | MOR | 239 | | Mertens *et al*., 2011 |
| **5** | Mozambique | MOZ | 160 | | Semo *et al.,* 2017 |
| **6** | Somalia | SOM | 404 | | Tillmar *et al*., 2009 |
| **7** | Bahrain | BAH | 543 | Asia | Al-Snan *et al*., 2019 |
| **8** | China | CHI | 200 | | Rashid *et al*., 2020 |
| **9** | India | IND | 200 | | |
| **10** | Laos | LAO | 451 | | Than *et al.,* 2022 |
| **11** | Malaysia | MAL | 200 | | Rashid *et al*., 2020 |
| **12** | Saudi Arabia | S-ARA | 500 | | Alsafiah *et al*., 2017 |
| **13** | Thailand | THA | 334 | | Than *et al.,* 2022 |
| **14** | Italy | ITA | 441 | Europe | Brisighelli *et al*., 2009 |
| **15** | Spain | SPA | 496 | | Barrio *et al*., 2019 |
| **16** | Honduras | HON | 174 | North America | Herrera-Paz *et al*., 2008 |
| **17** | Mexico | MEX | 500 | | Rubi-Castellanos *et al*., 2009 |
| **18** | USA-African American | U-AFM | 340 | | Hill *et al*., 2013 |
| **19** | USA-Asian | U-ASI | 97 | | |
| **20** | USA-Caucasian | U-CAU | 361 | | |
| **21** | USA-Hispanic | U-HIS | 236 | | |
| **22** | Brazil | BRA | 494 | South America | Rodrigues *et al*., 2009 |
| **23** | Colombia | COL | 617 | | Sánchez-Diz *et al.,* 2009 |
| **-** | **TOTAL** | **-** | **7,664** | **-** | **-** |

**Table 6.3. Size and continent, assigned code and source of each population sample for the Y-STR datasets**

| | Population/ Country | Language Family | Assigned Code | Ethnic Group | Size | Geopolitical Region | Continent | Source |
|---|---|---|---|---|---|---|---|---|
| 1 | Nigeria | Niger Congo-Igbo | NIG-IB | Igbo | 60 | West Africa | Africa | Present study |
| 2 | | Niger Congo-Yoruba | NIG-YO | Yoruba | 55 | | | |
| 3 | | Afroasiatic | NIG-HF | Hausa-Fulani | 52 | | | |
| 4 | Benin | Niger Congo-Atlantic | BEN-AC | Fon, Yoruba.B | 138 | | | Fortes-Lima *et al*. 2015 |
| 5 | | Niger Congo-Mende | BEN-M | Bariba | 58 | | | |
| 6 | Burkina Faso | | BF-M | Bissa, Marka, Samo North, Samo South, | 153 | | | De Filippo *et al*., 2011 |
| 7 | | Niger Congo-Gur | BF-G | Kassana, Lyela, Mosi, Nuna, Pana, Samoya | 183 | | | |
| 8 | Ivory Coast | Niger Congo-Atlantic | IVC-AC | Ahizi | 49 | | | Fortes-Lima *et al*., 2015 |
| 9 | | Niger Congo-Mende | IVC-M | Yacouba | 41 | | | |
| 10 | Senegal | | SEN-M | Mandenka | 15 | | | De Filippo *et al*., 2011 |
| 11 | Angola | Bantu West | ANG-B | AngOtherBantu, Ganguela, NyanekaNkhumbi, Umbunda | 230 | Central Africa | | |
| 12 | Cameroon | Niger Congo-Pygmy | CAM-P | Baka CAM, Bakola | 27 | | | |
| 13 | | Bantu West | CAM-B | Fang, Ngumba | 28 | | | |
| 14 | Central African Republic | Pygmy | CAR-P | Biaka | 23 | | | |

| | Population/ Country | Language Family | Assigned Code | Ethnic Group | Size | Geopolitical Region | Continent | Source |
|---|---|---|---|---|---|---|---|---|
| 15 | Democratic Republic Congo | Bantu West | DRC-B | Mbala, Mbuun, Pende, Yansi | 52 | Central Africa | Africa | De Filippo *et al*., 2011 |
| 16 | | Nilotic Pygmy | DRC-P | Mbutu | 11 | | | |
| 17 | Gabon | Bantu West | GAB-B | Akele, Ateke, Bekwill, Benga, Duma, Eshiva, Eviya, Fang, Galoa, Kota, Makina, Mbaouin, Ndumu, Nzebi, Obamba, Okande, Orungu, Punu, Shake, Tsogo | 795 | | | |
| 18 | | Pygmy West | GAB-P | Baka-GAB | 33 | | | |
| 19 | Ethiopia | Afroasiatic | ETH-AA | OmoValley | 70 | East Africa | | Haddish *et al*., 2022 |
| 20 | Kenya | Bantu West | KEN-B | BantuKenya | 10 | | | De Filippo *et al*., 2011 |
| 21 | | Nilotic | KEN-NS | Maasai | 51 | | | |
| 22 | Tanzania | Afroasiatic | TZ-AA | Burunge, WaFiome | 23 | | | |
| 23 | | Nilotic | TZ-NS | Datogo | 31 | | | |
| 24 | | Khoisan | TZ-K | Hadza, Sandawe | 121 | | | |
| 25 | | Bantu East | TZ-B | Mbugwe, Sukuma,Turu | 64 | | | |
| 26 | Uganda | Nilotic | UGA-NS | KarimojongJieDodos | 118 | | | |
| 27 | Algeria | Afroasiatic | ALG-AA | Mozabite | 20 | North Africa | | Triki-Fendri *et al.*, 2013 |
| 28 | Libya | | LIB-AA | Arabs, Berbers | 142 | | | |

**Table 6.3. continued**

| | Population/ Country | Language Family | Assigned Code | Ethnic Group | Size | Geopolitical Region | Continent | Source |
|---|---|---|---|---|---|---|---|---|
| 29 | Botswana | Bantu East | BOT-B | Kalanga, Tswana | 38 | Southern Africa | Africa | De Filippo *et al*., 2011 |
| 30 | Namibia | Khoisan | NAM-K | San | 6 | | | |
| 31 | South Africa | Bantu East | SA-B | BantuSouth | 269 | | | Leat *et al*., 2007 |
| 32 | Zambia | | ZAE-B | Bisa, Kunda | 69 | | | De Filippo *et al*., 2011 |
| 33 | | Bantu West | ZAW-B | Aushi, Bemba, Chewa, Chokwe, Few, Kangala, Kwamashi, Lala, Lenje, Lozi, Luchazi, Lunda, Luvale, Luyana, Kaonde, Mambwe, Mbukushu, Mbunda, Ngoni, Nkoya, Nsenga, Nyengo, Shanjo, Subiya, Tebele, Tonga, Totela, Tumbuka, Tswana, Umbundu, Yeyi | 478 | | | |
| 34 | China | Sino-Tibetan | CHI | Han Chinese | 246 | East Asia | Asia | Purps *et al*., 2014 |
| 35 | Iraq | Afroasiatic-Semitic | IRQ | Iraqi Arab | 124 | Middle East Asia | | Taqi *et al*., 2015 |
| 36 | Israel | | ISR | Israeli Arab | 154 | | | Fernandes *et al*., 2011 |
| 37 | Kuwait | | KUW | Kuwaiti Arab | 249 | | | Taqi *et al*., 2015 |
| 38 | Lebanon | | LEB | Lebanese Arab | 505 | | | |
| 39 | United Arab Emirates | | UAE | Emirati Arab | 278 | | | |

**Table 6.3. continued**

| | Population/ Country | Language Family | Assigned Code | Ethnic Group | Size | Geopolitical Region | Continent | Source |
|---|---|---|---|---|---|---|---|---|
| 40 | India | Dravidian | IND | Tamil-South India | 126 | South Asia | Asia | Purps *et al*., 2014 |
| 41 | Philippine | Austronesian | PHP | Filipino | 629 | Southeast Asia | | Taqi *et al*., 2015 |
| 42 | Italy | Indo-European (Italian) | ITA | Italian | 157 | Southern Europe | Europe | Purps *et al*., 2014 |
| 43 | Spain | Indo-European (Spanish) | SPA | Spanish | 126 | | | |
| 44 | Germany | Indo-European (German) | GER | German | 131 | Western Europe | | |
| 45 | United Kingdom | Indo-European (English) | UK-BW | British-White | 161 | | | |
| 46 | United Kingdom | | UK-BA | British-Black | 171 | | | |
| 47 | United States of America | | US-AA | African American | 509 | United States | North America | |
| 48 | Jamaica | Jamaican Patois | JAM | Afro-Jamaican | 66 | Caribbean | | |
| 49 | Brazil | Indo-European (Portuguese) | BRA | Mixed Brazilian | 120 | Latin America | South America | |
| 50 | Peru | Indo-European (Spanish) | PER | Mixed Peruvian | 83 | | | |
| - | **TOTAL** | - | - | - | **7,348** | - | - | - |

**Table 6.3. continued**

### 6.3.2 INTER-POPULATION GENETIC ANALYSIS

The statistical and inter-population genetic analysis for the autosomal STR datasets was conducted using 13 autosomal STR loci that were common to the compiled datasets: TH01, D3S1358, vWA, D21S11, D16S539, D8S1179, D18S51, FGA, CSF1PO, TPOX, D5S818, D13S317, and D7S820. These were analysed across 23 globally distributed populations, as listed in **Table 6.2**. Regarding the Y-STR datasets, the statistical and inter-population genetic analysis utilized 12 Y-STR loci: DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439. This analysis was performed across 50 globally distributed populations, as shown in **Table 6.3**.

**Wright's F-statistics ($F_{ST}$)**

The Genetic Analysis in Excel (GenAIEx 6.5) platform, developed by Peakall and Smouse (2006, 2012), was used to calculate Wright's $F_{ST}$ to assess the genetic structure on a global level, showing how much genetic variation exists between different populations compared to within populations. $F_{ST}$ aims to investigate inter-population relationships between the Nigerian population and various population groups worldwide based on the formula below (Weir and Cockerham, 1984; Weir, 1996):

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Where $H_T$ represents the variation within the total population (the total expected heterozygosity if all subpopulations were combined). In contrast, $H_S$ represents the variation within subpopulations (the average expected heterozygosity across all subpopulations). In this context, the different population groups from various regions served as the subpopulations alongside the Nigerian population. $F_{ST}$ was used to measure the degree of distinctiveness between these population groups based solely on autosomal STR data. It examines how heterozygosity decreases due to population structure, with values ranging from 0 to 1; 0 signifies no genetic divergence between populations, and 1 represents complete genetic separation.

In a global context, $F_{ST}$ is interpreted as follows:

| | | |
|---|---|---|
| < 0.05 | = | Low genetic divergence |
| 0.05 – 0.15 | = | Moderate genetic divergence |
| 0.15 – 0.25 | = | High genetic divergence |
| > 0.25 | = | Very high genetic divergence |

The Python program with Matplotilib and Seaborn libraries (Python Software Foundation, 2024; Hunter, 2007; Waskom, 2021) was also utilized to visualize $F_{ST}$ in the form of heat maps.

**Nei's Genetic Distance**

The Genetic Analysis in Excel (GenAIEx 6.5) platform (Peakall and Smouse, 2006; 2012), was also utilized to generate pairwise matrices for both Nei's genetic distance and Nei's genetic identity. Nei's genetic distance was used to measure genetic divergence between the Nigerian population and the various population groups around the world, following the formula provided by Nei (1972; 1987):

$$D = -\ln(I)$$

Where D represents Nei's genetic distance, while *I* represents Nei's genetic identity between two populations calculated as:

$$I = \frac{\sum p_{xi} \cdot p_{yi}}{\sqrt{\sum p_{xi}^2 \cdot \sum p_{yi}^2}}$$

Where *Pxi* is the frequency of allele *i* in population X, while *Pyi* is the frequency allele *i* in population Y. Nei's genetic distance was used to assess the genetic level among these different population groups based solely on the autosomal STR datasets. It measures population structure, with values (*D*) ranging from 0 to 1; 0 indicates no shared alleles between populations, while 1 indicates populations are genetically identical.

In a global context, Nei's genetic distance is interpreted as follows:

| | | |
|---|---|---|
| < 0.01 | = | Populations are genetically identical |
| 0.01 – 0.05 | = | Very closely related |
| 0.05 – 0.15 | = | Moderately related |
| 0.15 – 0.25 | = | Distantly related |
| > 0.25 | = | Very divergent |

**Slatkin's R-statistic (*$R_{ST}$*)**

Arlequin version 3.5 software (Excoffier and Lischer, 2010) was employed to calculate Slatkin's R-statistic ($R_{ST}$) to evaluate the global genetic divergence between the Nigerian population and various population groups, as well as to assess the amount of genetic variation between populations versus within them. $R_{ST}$ is comparable to $F_{ST}$ but is specifically designed to account for the stepwise mutation model (SMM), which is appropriate for STRs. The calculation was performed based on the formula below (Slatkin, 1995):

$$R_{ST} = \frac{V_a}{V_t}$$

Where $V_a$ represents the variance among populations (i.e., the between-population component of allele size variance), and $V_t$ represents the total variance (i.e., within and between population variance). This study used $R_{ST}$ to analyse solely the Y-STR datasets because Y-STR alleles are assumed to mutate by gaining or losing repeat units' step by step. $R_{ST}$ evaluates population structure, with values ranging from 0 to 1; 0 indicates no genetic divergence between populations, while 1 indicates complete genetic divergence.

In a global context, $R_{ST}$ is interpreted as follows:

| | | |
|---|---|---|
| < 0.05 | = | No genetic divergence |
| 0.05 – 0.15 | = | Moderate genetic divergence |
| > 0.15 | = | High genetic divergence |

**Multi-Dimensional Scaling (MDS)**

This is a statistical tool for visualising genetic structure (similarity or dissimilarity between datasets). It reveals clusters of related populations, patterns of gene flow and identifies outliers or hybrid populations by projecting populations or individuals into a low-dimensional space, allowing for clearer representation of genetic distances between them (Saeed *et al*., 2018). To determine the differentiation and structure among populations for the Multi-Dimensional Scaling (MDS) scatter plot, the Python program leveraging the Matplotilib, Pandas, and NumPy libraries was utilised (Python Software Foundation, 2024; Hunter, 2007; McKinney, 2010; Harris *et al*., 2020). MDS uses a matrix of pairwise distances (e.g., genetic distances like Nei's or $R_{ST}$) to represent each population as a point on a plot. The closer the two points are to one another, the more genetically similar the populations are. Conversely, populations farther apart on the plot are more genetically distinct. To perform multivariate analysis for the inter-population relationships, the autosomal STR dataset was evaluated using the pairwise unbiased Nei's genetic distance matrix for the 24 population groups included in this study. In contrast, the Y-STR datasets were assessed using the pairwise unbiased $R_{ST}$ genetic distance matrix for the 50 population groups compiled for this research. The parameters involved comparing the MDS Dimension 1, which captures the highest variance on the X-axis and the MDS Dimension 2, which reports the second highest variance on the Y-axis. This format provides the best low-dimensional plot of the data.

**Neighbour-Joining (NJ) Analysis**

Molecular Evolutionary Genetics Analysis (MEGA-11) software (Tamura *et al*., 2021) was employed to conduct a Neighbour-Joining analysis of the population groups evaluated in this study.

This analysis evaluated the phylogenetic relatedness among these population groups based on the assembled 23 population groups of the autosomal STR dataset and the compiled 50 population groups of the Y-STR datasets. The Neighbour-Joining tree for the autosomal STR datasets was constructed based on the pairwise matrix of unbiased Nei's genetic distance for the 24 population groups. In contrast, the Neighbour-Joining tree for the Y-STR datasets was constructed utilising the pairwise matrix of unbiased $R_{ST}$ unbiased genetic distance for the 50 population groups. The Neighbour-Joining tree illustrated the population clusters and evolutionary inter-population relationships among the groups.

**Analysis of Molecular Variance (AMOVA)**

Arlequin version 3.5 software (Excoffier and Lischer, 2010) was employed to perform an Analysis of Molecular Variance (AMOVA). This statistical approach partitions genetic differences within and among populations by analysing molecular data based on genotypes from the autosomal STR datasets and the haplotype data from the Y-STR datasets. The Phi statistics (Φ-statistics) produced in the analysis quantified genetic structure and differences and assessed potential evolutionary divergence. Statistical significance was also evaluated through P-value calculations.

**Population Structure Analysis**

*STRUCTURE* software version 2.3.4 was employed to analyse the structure of different populations. This program utilises a Bayesian inference-based clustering technique to impute individuals to genetic groups based on their genotypes, using multi-locus genotype data from our sample set (Pritchard *et al*., 2000). The data was organised into groups representing the 23 global population groups in the autosomal STR datasets. To effectively assess the *STRUCTURE* output using the Structure Selector (Evanno *et al*., 2005) and CLUMPAK (Clustering Markov Packager Across K) (Kopelman *et al*., 2015) online tools, the total population of 7,664 unrelated individuals was reduced proportionately to 5,000. This adjustment was necessary because 5,000 is the maximum number of individuals the software can analyse simultaneously. The population sizes for the various groups are as follows: Nigeria: 252, Ghana: 109, Morocco: 239, Egypt: 241, Mozambique: 160, Somalia: 241, US African American: 241, US Caucasian: 241, USA Hispanic: 241, USA Asian: 97, Brazil: 241, Mexico: 241, Honduras: 174, Colombia: 241, Malaysia: 200, China: 200, India: 200, Bahrain: 241, Laos: 241, Saudi Arabia: 241, Thailand: 241, Italy: 241, and Spain: 241. This totals 5,000 unrelated individuals. *STRUCTURE* software evaluated various potential population sizes (*K*) to estimate the proportion of each individual's genome derived from the defined populations. The analysis tested *K* values ranging from 1 to 25 and was set to repeat

ten times for each *K*. The parameters for the study were: number of individuals = 5,000, ploidy = 2, and number of loci = 13, with missing values coded as -9. The burn-in period was set to 50,000 iterations, followed by a Markov Chain Monte Carlo (MCMC) iteration of 200,000. To evaluate the likelihood of the observed data (X) for different *K* values, posterior probabilities were calculated for each *K* using the Structure Selector software and the results with bar plots generated by CLUMPAK. The highest Ln P(X | K) value indicated the appropriate number of clusters (*K*). Additionally, the maximum rate of change (Delta *K*) in Ln P(X | K) values between consecutive *K* estimates assisted in determining the optimal number of genetic clusters within the dataset.

## 6.4 RESULTS

### 6.4.1 AUTOSOMAL STR INTER-POPULATION GENETICS

To demonstrate the genetic relationship between Nigeria and various previously reported global populations, pairwise $F_{ST}$ genetic analyses, Multi-Dimensional Scaling (MDS), Neighbour-Joining (NJ), and Analysis of Molecular Variance (AMOVA) were conducted. These analyses were based on the 13 autosomal STR loci the Nigerian population dataset shares with the 22 global population group datasets listed in **Table 6.2**.

#### F-statistics ($F_{ST}$) Genetic Distance

The heatmap in **Figure 6.1** shows the genetic differentiation between Nigeria and other global populations based on pairwise population $F_{ST}$ values. The colour intensity corresponds to the $F_{ST}$ values detailed in **Appendix 81**. Darker blue shades indicate higher $F_{ST}$ values, implying more genetic differences. Lighter blue shades indicate lower $F_{ST}$ values, suggesting more genetic similarity. Nigeria and Ghana exhibit very low genetic differentiation ($F_{ST}$ = 0.002), which is understandable given their close geographical closeness and shared cultural and religious connection in West Africa. Similarly, Nigeria and African Americans in the US show a comparable low level of genetic divergence ($F_{ST}$ = 0.002). This aligns with many African Americans' ancestral roots in West Africa, including Nigeria and Ghana. This also suggests an ancestral link to Africa, likely from the Transatlantic Slave Trade that began in West Africa. Nigeria exhibited moderate genetic differentiation from Mozambique ($F_{ST}$ = 0.005) and Somalia ($F_{ST}$ = 0.008), which are sub-Saharan African populations located in Southern and East Africa, respectively. Egypt ($F_{ST}$ = 0.012) and Morocco ($F_{ST}$ = 0.013) in North Africa exhibited moderate genetic differentiation from Nigeria, suggesting some genetic distinctiveness. This pattern shows how North Africa and Sub-Saharan Africa have been separated historically and geographically.

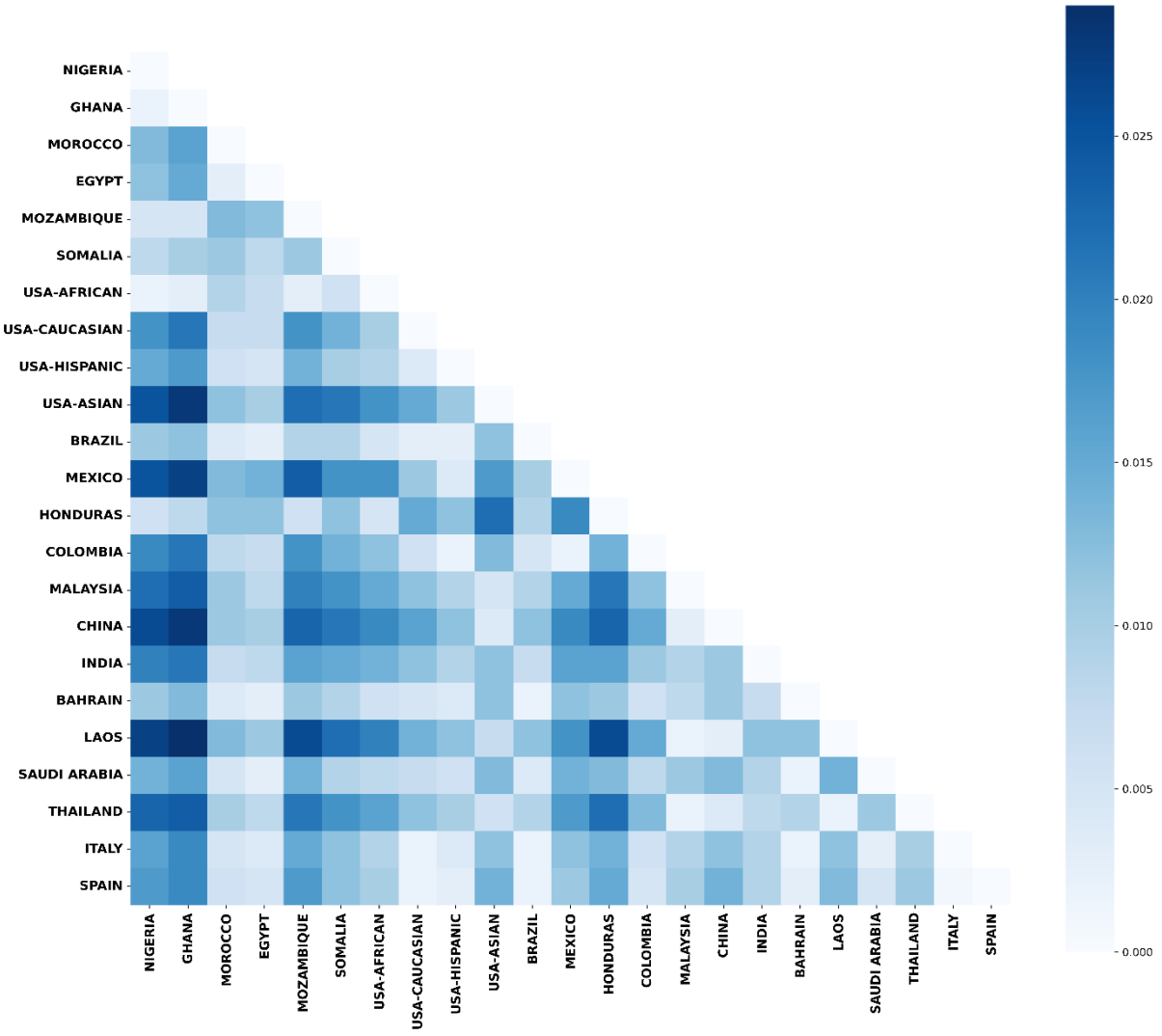



**Figure 6.1. Heatmap displaying the pairwise $F_{ST}$-based genetic distance matrix among global population groups based on the genotypes of 13 autosomal STR markers collected from 7,664 individuals across 23 population groups. The colour intensity corresponds to the $F_{ST}$ values, which can be found in Appendix 81.**

Beyond Africa, the Nigerian population exhibited moderate to very high genetic differentiation compared to other populations. Nigerians are genetically closer to the Honduran population ($F_{ST}$ = 0.006) in Central America (subregion in North American continent) than to North African countries like Morocco, Egypt, and Libya. This shows a genetic link between some North American people and populations in West Africa, especially Ghana and Nigeria. This connection demonstrates evidence of historical migration from Africa to America. In South America, Brazil's genetic distance from Nigeria is moderate ($F_{ST}$ = 0.011), while Colombia ($F_{ST}$ = 0.019) and Mexico ($F_{ST}$ = 0.025) show high genetic differentiation. In Asia, Middle Eastern countries such as Bahrain ($F_{ST}$ = 0.011) and Saudi Arabia ($F_{ST}$ = 0.014) display moderate genetic differentiation from Nigeria. In contrast, populations from other regions in Asia, including China ($F_{ST}$ = 0.026), India ($F_{ST}$ = 0.025),

Laos ($F_{ST}$ = 0.027), Malaysia ($F_{ST}$ = 0.022), and Thailand ($F_{ST}$ = 0.023), demonstrate high to very high genetic divergence from Nigeria. In Europe, Italy and Spain show moderate genetic differentiation from Nigeria, with $F_{ST}$ values of 0.016 and 0.017, respectively. These values indicate that most genetic variation (~98%) occurs within populations, while only a small fraction (~1.6–1.7%) distinguishes the populations. In the United States, Caucasians show a moderate genetic distance from Nigeria, with an $F_{ST}$ value of 0.018, comparable to their European counterparts. Meanwhile, Hispanics have an $F_{ST}$ value of 0.015. The Asian populations in the U.S. ($F_{ST}$ = 0.025) display high genetic differences, corresponding closely to those of various populations in Asia.

**Multi-Dimensional Scaling (MDS)**

The Multi-Dimensional Scaling (MDS) analysis was conducted using pairwise unbiased Nei's genetic distance matrix values (see **Appendix 88**), derived from 13 autosomal STR loci across 24 population groups representing South America, North America, Europe, Asia, and Africa. The two-dimensional MDS plot **(Figure 6.2)** showed apparent genetic differences and regional group patterns.

The populations of Africa are shown as red dots in the MDS plot. Nigeria (NIG) and Ghana (GHA) have a close genetic similarity. This similarity shows their geographic and historical connections in West Africa. Although geographically distant in East and Southern Africa, Somalia (SOM) and Mozambique (MOZ) are somewhat separated from West African groups, suggesting subtle genetic differentiation. Morocco (MOR) and Egypt (EGY) are more peripheral among Africans, which aligns with their North African origin and intermediate genetic profiles between Sub-Saharan Africa and Europe/Asia. Egypt (EGY) is positioned between African and Middle Eastern/European groups. It is not closely clustered with Sub-Saharan African populations like Nigeria, Ghana, or Mozambique. Egypt is genetically closer to Morocco and Middle Eastern populations than to populations like Nigeria or Somalia.

The South American populations were represented as blue dots. Brazil (BRA), and Colombia (COL), are closer to the US Hispanics (U-HIS). Mexico (MEX), which is geographically located in North America is closer Colombia and US-Hispanics indicating genetic similarity between Latin American and Hispanic populations.

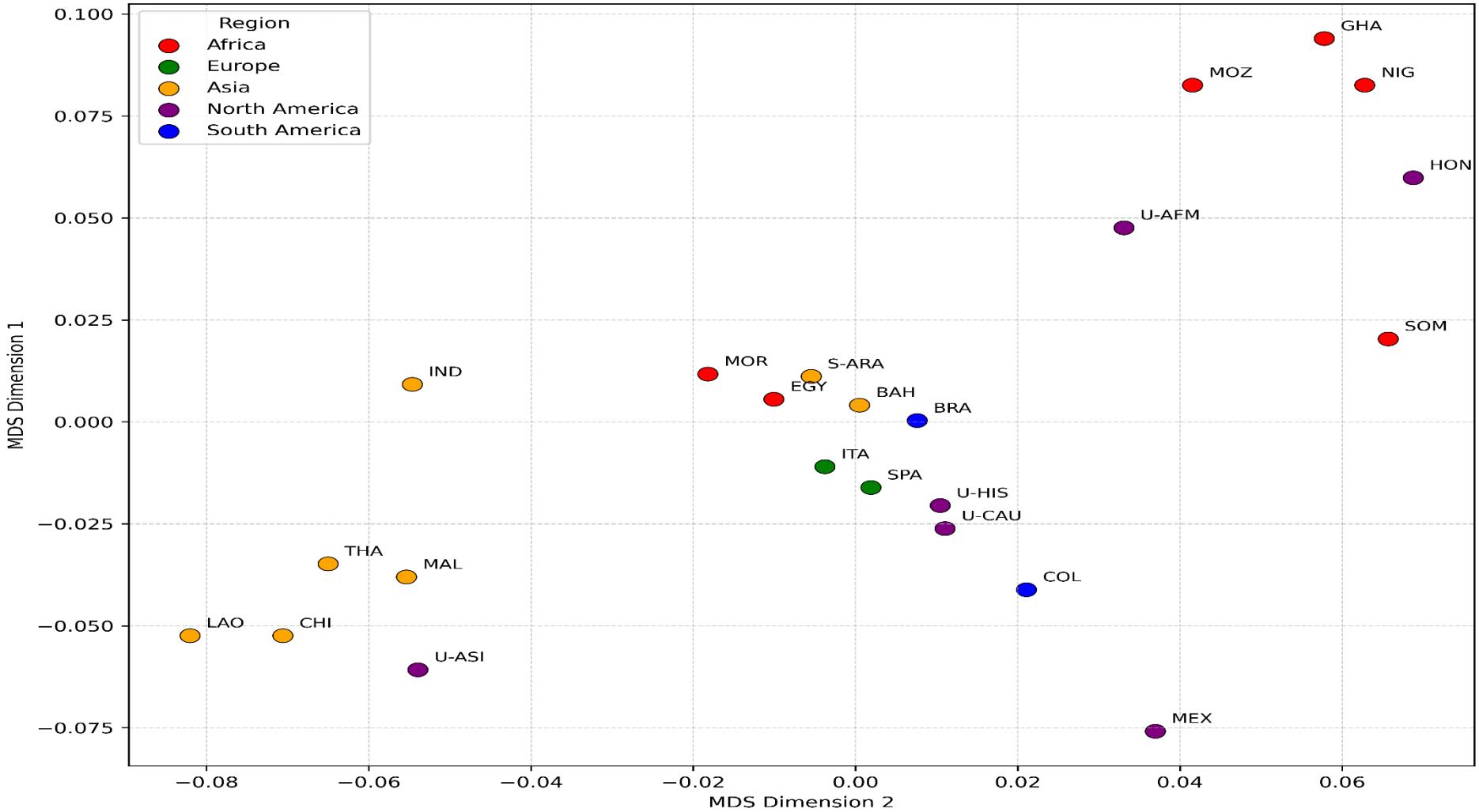



**Figure 6.2. A two-dimensional MDS plot visualising genetic distances among 23 global populations based on 13 autosomal STR loci using the Nei's genetic distance matrix. The populations are colour-coded by their respective continental regions. Dimension 1 (y-axis) likely reflects the broad continental structure and significant genetic divide, like those between Africa, Asia, and Europe. Dimension 2 (x-axis) captures admixture patterns, and within-continent variation.**

Honduras (HON), located in Central America, a subregion of the North American continent, represents the Garifuna population living along the Caribbean coast in this study. Although the Garifuna may share some genetic features with other Latin American populations, they show a closer genetic relationship to African populations, particularly West African groups such as Nigeria (NIG) and Ghana (GHA). This pattern reflects their predominantly African ancestry, originating in the transatlantic African diaspora and subsequent historical admixture.

The Asian populations, represented by orange dots, form a broad cluster indicating intra-regional diversity yet general separation from the African (sub-Saharan) cluster. Bahrain (BAH) and Saudi Arabia (S-ARA) in the Middle East cluster close to European and North African populations. China (CHI), Malaysia (MAL), Laos (LAO), and Thailand (THA) form a distinct cluster. India (IND) is slightly offset but remains closely grouped to the Asian cluster. Italy (ITA) and Spain (SPA) are part of the European populations, shown in green. As expected, they form a close group of European populations with similar demographic histories. Italy and Spain's populations also cluster near Middle Eastern and Latin American populations, reflecting geographic continuity with the Middle East and historical gene flow across the Mediterranean. The closeness to Latin American populations reflects colonial migrations from Europe to South America. The population groups in North America are represented by purple dots and include US African Americans (U-AFM), US Caucasians (U-CAU), US Asians (U-ASI), and US Hispanics (U-HIS). As expected, the US population is diverse, with different ethnic groups that show various backgrounds and genetic traits. African Americans (U-AFM) tend to cluster closer to the African population group. Asians in the US are genetically similar to the larger Asian population. At the same time, the US Caucasians (U-CAU) and the US Hispanics (U-HIS) are closer to the European groups and the Latin American groups.

**Neighbour-Joining (NJ)**

The Neighbour-Joining (NJ) method was employed to visualise the phylogenetic distances among the 23 population groups. Genetically close populations will have closer branches. Genetically distant populations will have farther branches. The construction of the phylogenetic tree was performed using data from the pairwise unbiased Nei's genetic distance matrix values (see **Appendix 82**), which was calculated based on 13 autosomal STR loci. The results demonstrate the genetic distances and regional group patterns, as displayed in **Figure 6.3**.

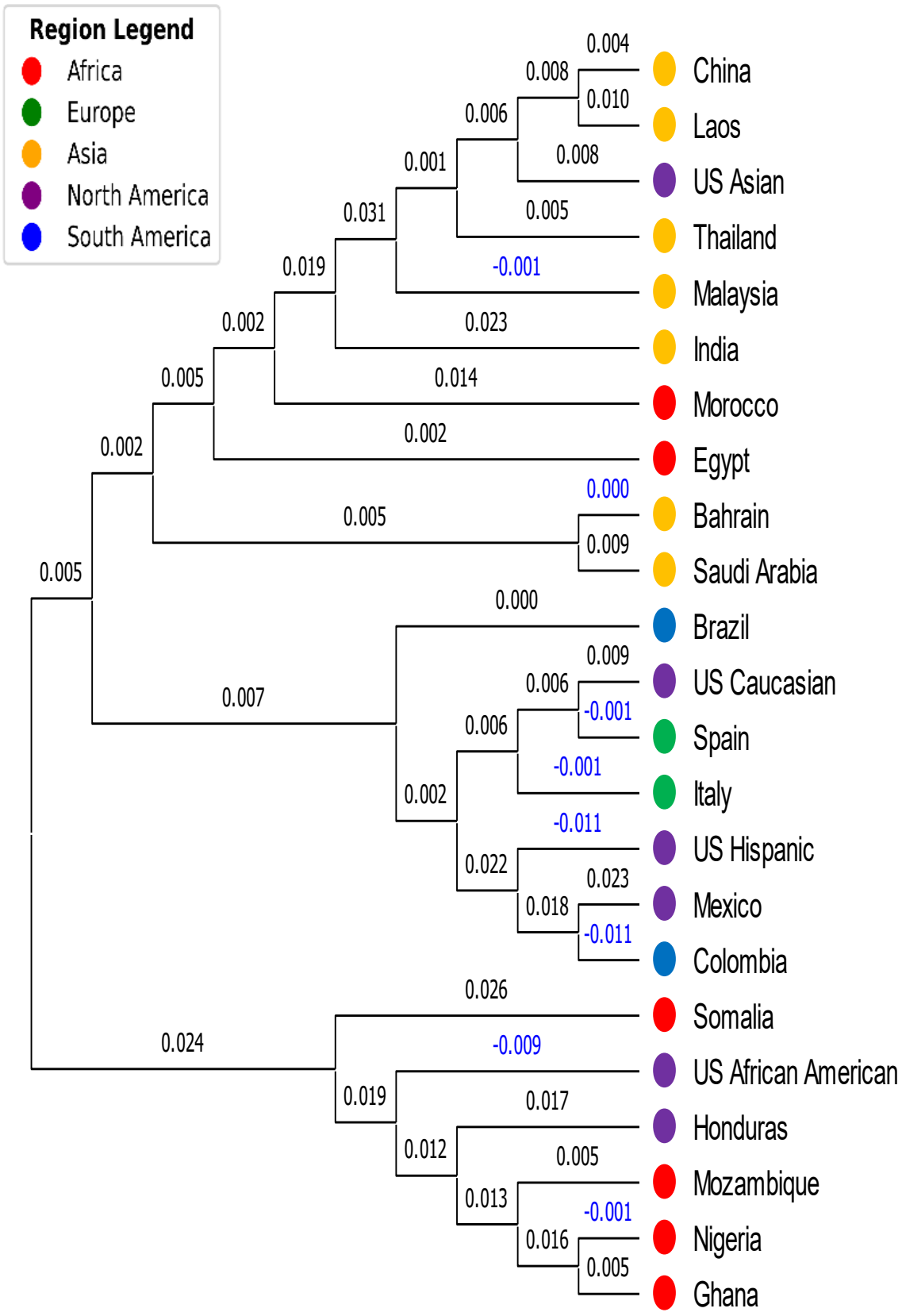



**Figure 6.3. Neighbour Joining (NJ) tree constructed from pairwise unbiased Nei's unbiased genetic distance population matrix, illustrating the clustering of 23 global population groups. The populations are colour-coded by their respective continental regions, with branch lengths representing genetic distances.**

As anticipated, Nigeria and Ghana are closely clustered, showing significant genetic similarity between these West African populations. Mozambique is in East Africa but is also linked to Nigeria and Ghana, which shows a genetic connection among the people in sub-Saharan Africa. Honduras is close to this group because of African ancestry from historical migration, which greatly influenced the genetic makeup of Afro-Caribbean and Latin American populations.

Similarly, US African American populations are closely clustered with the African region, indicating significant African ancestry. Sub-Saharan African countries such as Nigeria, Ghana, Mozambique, and Somalia form a clear cluster due to their shared evolutionary history. Honduras, which is mainly comprised of the Garifuna people who live on the Caribbean coast, also fall within this African group, reflecting their African roots and historical background. In contrast, North African populations, such as those from Egypt and Morocco, group more closely with Middle Eastern populations like Saudi Arabia and Bahrain, forming a Middle East–North Africa (MENA) grouping. This pattern suggests shared ancestry and historical connections between North Africa and the Arabian Peninsula.

Saudi Arabia and Bahrain are closely clustered, showing they share significant genetic similarities as part of the Middle Eastern Asia block. China, India, Malaysia, Laos, and Thailand formed distinct Asian subclusters with US Asian individuals grouping with East and Southeast Asian reference populations. Additionally, China and Laos are closely associated, which aligns with their shared East Asian autosomal ancestry.

Spain, Italy, Brazil, and US Caucasian form a close European cluster, reflecting shared Western European ancestry. Brazil is grouped closely with Italy, likely due to the predominance of Southern European male ancestry from colonisers and immigrants. US Hispanics, Mexico, and Colombia form a compact cluster with European groups, reflecting their tri-hybrid ancestry (European, Native American, and African) but driven mainly by European autosomal components.

**Analysis of Molecular Variance (AMOVA)**

AMOVA was performed to partition genetic variation across three hierarchical levels: first, among the five continental regions, which included South America, North America, Europe, Asia, and Africa; second, among populations within those regions, represented by 23 distinct global population groups (see **Table 6.2**); and third, within populations, comprising a total of 7,664

unrelated individuals distributed across each of the population groups. **Table 6.4** summarises the result of the AMOVA analysis, while **Table 6.5** displays the Phi statistic (Φ-statistics) and p-values.

**Table 6.4. Summary of AMOVA for 13 autosomal STR loci across 23 global populations involving 7,664 unrelated individuals grouped into five continental regions**

| Source of Variation | Degrees of freedom (df) | Sum of Squares | Variance Components | % Variation | Interpretation |
|---|---|---|---|---|---|
| **Among Regions** | 4 | 531.18 | 0.0693 | 0.667 | **Low genetic differentiation** among continental regions (e.g., Africa vs. Asia). |
| **Among Populations within regions** | 18 | 1019.95 | 0.1331 | 1.281 | **Low genetic differentiation** among populations within continental regions but statistically significant (e.g., Nigeria vs. Ghana). |
| **Within Populations** | 7,641 | 78063.5 | 10.1871 | 98.052 | **High genetic differentiation** among individuals within populations. (e.g. Nigerian individuals) |
| **Total** | **7,663** | **79614.7** | **10.3895** | **100** | |

The result **(Table 6.4)** showed that most genetic variation (98.05%) is due to individual differences. This high level of variability is attributed to the highly polymorphic and informative characteristics of autosomal STR markers. In contrast, only 0.67% of the genetic variation is accounted for by differences between continental regions, which result from shared ancestry and admixture within these regions despite their geographic distances (for example, between Africa, North America, and South America). However, the significant AMOVA result shows that continental regions are still genetically distinguishable. Only 1.28% of the differences among populations occur within the same region. Factors like gene flow, past migrations, cultural practices, and geographical closeness likely cause these differences. These findings demonstrate strong population structure at the individual level and show minimal divergence among populations within the same region.

**Table 6.5. Summary of the Phi (Φ)statistics and p-values based on the AMOVA analysis of the autosomal STR dataset**

| Statistic | Φ Value | P-value | Interpretation |
|---|---|---|---|
| **$\Phi_{CT}$ (Among Regions)** | 0.0067 | < 0.0001 | Significant differentiation among continents (Africa, Asia, Europe, North and South America). This means that 0.67% of the genetic differentiation is due to variation between regions, indicating low structure across continents. |
| **$\Phi_{SC}$ (Among Populations within Regions)** | 0.0129 | < 0.0001 | Low but significant differentiation among populations within regions (e.g., African populations), contributing 1.29% to total variance. This shows that even populations in the same region are genetically differentiated. |
| **$\Phi_{ST}$ (Among Populations Overall)** | 0.0195 | < 0.0001 | The total genetic differentiation across all populations and regions is 1.95%. This suggests an overall low genetic structure among all 23 global populations. |

The Φ-statistics **(Table 6.5)** helped to measure genetic differences. The $\Phi_{CT}$ value of 0.0067 shows a small but statistically significant genetic differences between continents ($P < 0.0001$). This means there is genetic variation across different continents. The overall $\Phi_{ST}$ value of 0.0665 reflected little divergence among all populations, which was also statistically significant ($P < 0.0001$), highlighting the low partitioning of autosomal STR variation at the global level. In contrast, the $\Phi_{SC}$ value is 0.0129. Although this number is lower and shows variation among populations within regions, it is still statistically significant ($P < 0.0001$). This suggests that specific populations have had historical separations, migration patterns, and specific evolutionary changes.

### *STRUCTURE* Analysis

*STRUCTURE* was used to evaluate genetic affinities among 23 global population groups using 13 autosomal STR markers. The outputs were uploaded to the Structure Selector for analysis, which calculated the Mean LnP(K) ± SD (see **Appendix 85**) and Delta *K* values to determine the most probable number of genetic clusters (*K*), tested from 1 to 25. Results were visualized using the CLUMPAK package within the Structure Selector software.

**Figure 6.4.** shows the average log-likelihood values [LnP(K)] corresponding to each assumed number of genetic clusters ($K = 1$ to 25) calculated across ten independent *STRUCTURE* runs per *K*. Each point on the plot represents the average LnP(K). In contrast, the vertical error bars represent the standard deviation (SD), indicating the variability of results across replicates. The mean LnP(K) compares the likelihood of different *K* values generated by *STRUCTURE* with the higher value, suggesting a better fit for *K*. The insight on the robustness of the model for a given *K* was provided

by the Mean LnP(K)±SD to help assess the stability of the estimate of the population structure. The LnP(K) values generally increase with higher *K*, reflecting an improved fit of the model as more clusters are added. However, this increase often reaches a plateau, where adding more clusters results in diminishing gains in likelihood, or the estimates become unstable. The plot helps to visually assess model fit, with the plateau suggesting the point beyond which additional clusters may not capture meaningful population structure and introduce instability, suggestive of overfitting. This provides the foundational values used in the Evanno method (Delta *K*) to infer the most probable number of genetic clusters (optimal *K*). *K* = 3 is emphasized because it represents where the Evanno method identified the most significant change, indicated by the highest Delta *K*. The value of *K* = 3 yields a strong, stable dataset partition. While *K* values greater than 3 may capture additional structure, they can also reflect noise, admixture, or overfitting.

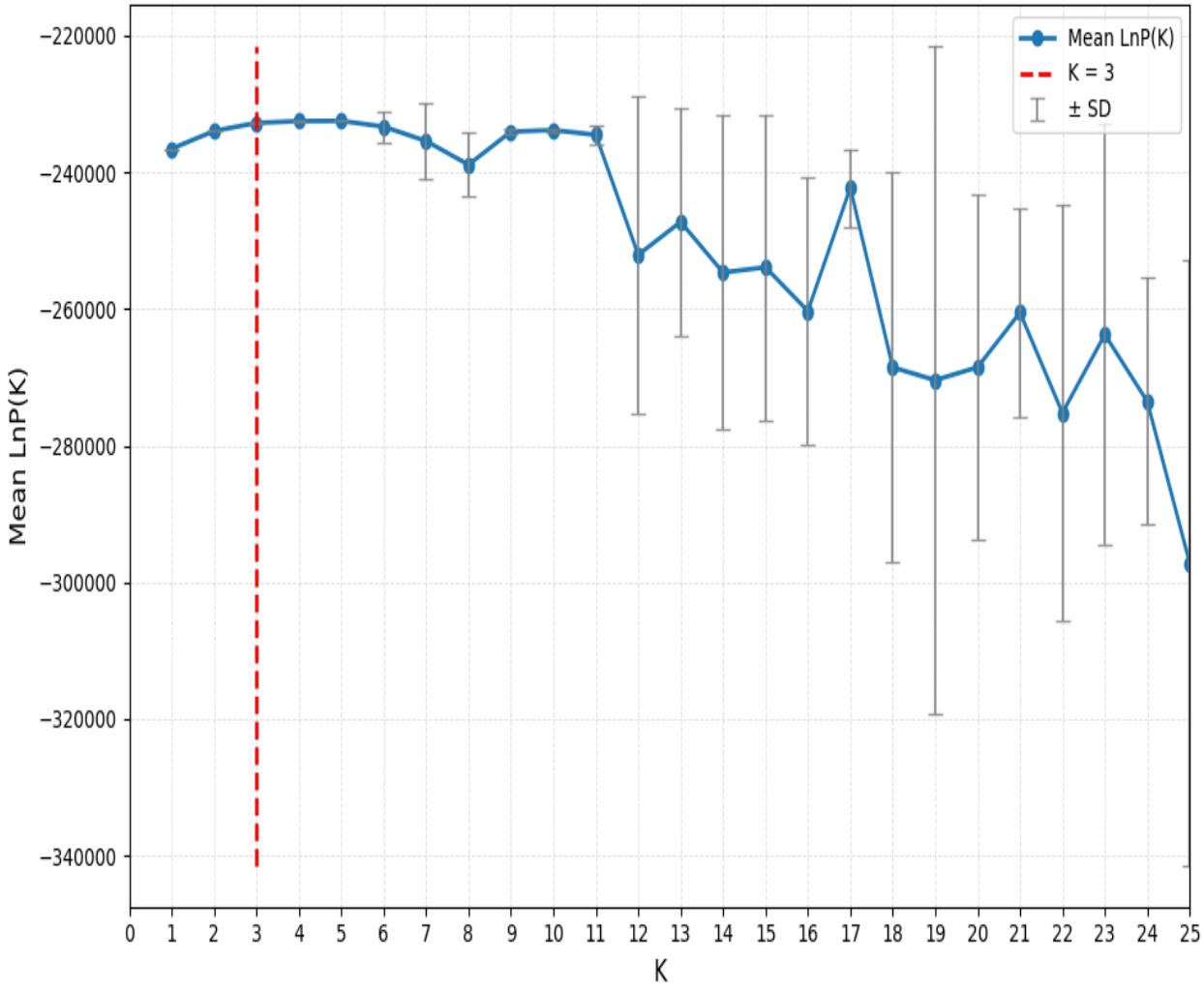



**Figure 6.4. Mean log-likelihood of the data [LnP(*K*)] ± standard deviation (SD) plotted against the number of inferred populations (*K*), calculated across ten independent runs of the *STRUCTURE* program. The red dashed line indicates *K* = 3, corresponding to the optimal number of clusters inferred from the ΔK method.**

**Figure 6.5.** presents the Delta *K* values (see **Appendix 85**) calculated using the Evanno method, which estimates the LnP(K) change rate between successive *K* values. A clear peak in Delta *K* is observed at *K* = 3, identifying it as the most probable number of genetic clusters. This suggests four genetically distinct populations, with the highest Delta *K* recorded for this value. The rate of change in Delta *K* assists in selecting the appropriate population size (*K*) by comparing the likelihood between different *K* values. A higher Delta *K* value indicates the optimal number of populations in the dataset. Delta *K* is particularly informative when the LnP(K) curve lacks a distinct peak, as it highlights the point of most significant improvement in model fit before the curve stabilizes.

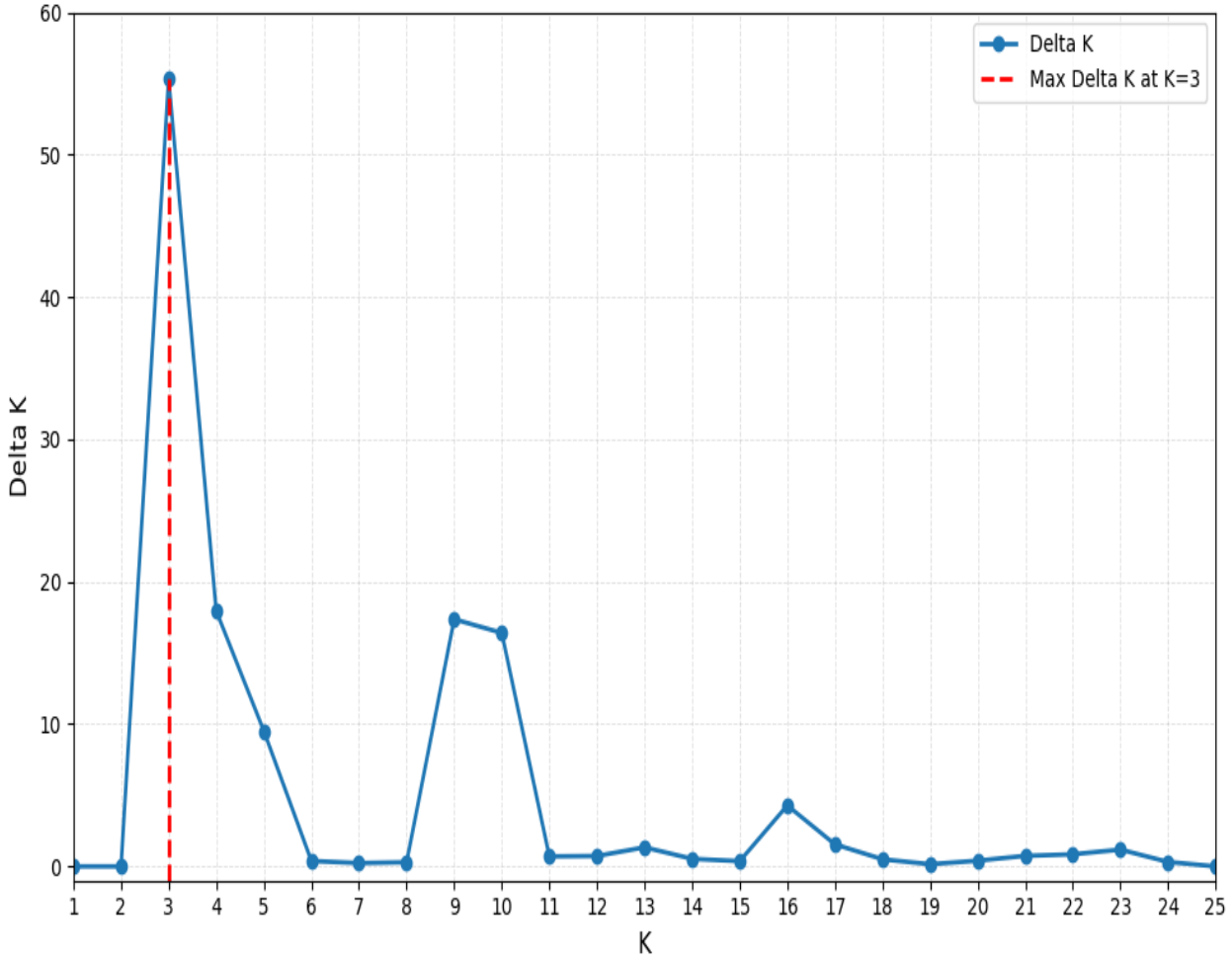



**Figure 6.5. The change in the log-likelihood rate (Delta *K*) for each value of *K* computed across ten repeated runs and iterations of the *STRUCTURE* program. The plot suggests four genetically distinct populations, with the highest Delta *K* recorded for this value.**

The analysis of inferred populations across values of *K* (from *K* = 1 to *K* = 25) revealed a clear population structuring among the 23 global populations. Although *K* = 3 was identified as the optimal result, **Figure 6.6** illustrates genetic differentiation in ancestry proportions among individuals within each group for selected *K* values (*K* = 3, *K* = 6, *K* = 23, and *K* = 25).

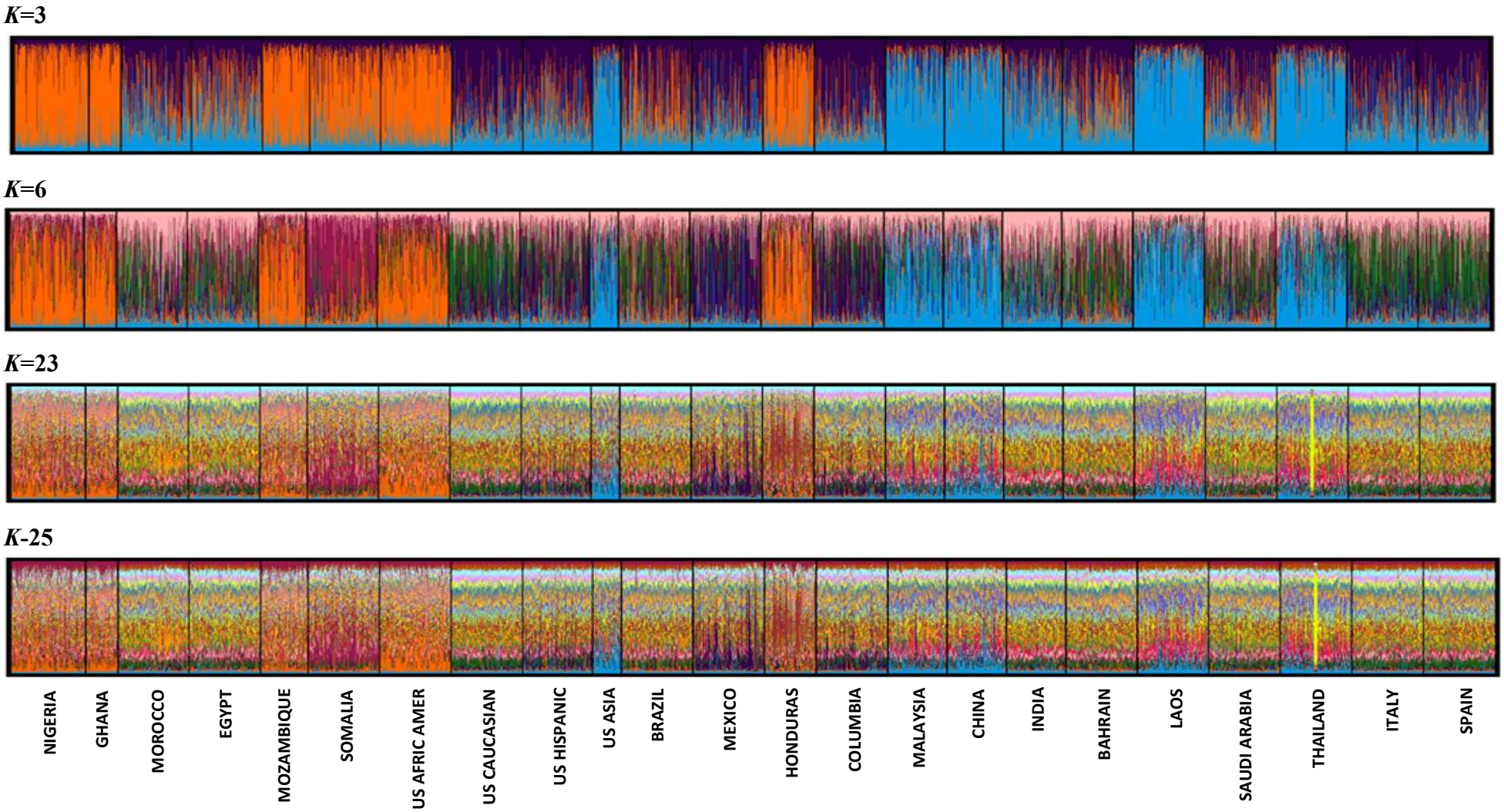



**Figure 6.6. The population stratification of 23 global populations. This analysis was conducted over ten repeated runs and iterations using the *STRUCTURE* program. The assumed populations (*K*) were intentionally selected based on biological relevance and are represented as bar plots for *K*=3, *K*=6, *K*=23, and *K*=25. Each vertical line corresponds to an individual (total = 5,000) assessed based on data from 13 autosomal loci. The amount of colour in each bar indicates the individual's genetic ancestry to each *K* population.**

The selection of *K* values 3, 6, 23, and 25 was intentional, based on biological relevance from *STRUCTURE* and CLUMPAK analysis.

*K* = 3 is the best point to consider when studying Delta *K* (statistical peak of Delta *K*) and gives a broad ancestral structure. Delta *K* shows how likely it is to see changes between different values of *K*. It helps find old population splits and provides a clear view of the global structure. *K*=3 shows the broad and consistent clusters at a continental level: 1. the Sub-Saharan African cluster (e.g. Nigeria, Ghana, Mozambique, and Somalia); 2. the East and Southern Asian cluster (e.g. China, Malaysia, Laos, Thailand and US Asian); and 3. Europe, North African, and Latino ancestry cluster (e.g. Italy, Spain, Egypt, Morocco, US Caucasians, US Hispanics, Brazil, Colombia and Mexico) with admixed profiles observed in populations from the Americas and parts of the Middle East and South Asia. The clustering at this level reflects deep ancestral divergence, likely capturing early human population splits across Africa, Asia, and Europe.

*K* = 6 shows clearer patterns of regional substructure emerge within continents, allowing finer resolution of population differentiation and admixture. African populations remain distinct, while additional substructure becomes apparent, highlighting historical gene flow and admixture events. Latin American populations—Brazil, Colombia, Honduras, and Mexico—showed contributions from both European and Native American-like clusters and African influence. Honduras (Garifuna) exhibits a more substantial genetic alignment with African ancestry, while Mexico, Colombia, and Brazil tend to align more closely with European ancestry and the US Hispanics. Somalia exhibits a blend of genetic sources with patterns reflecting North African and Eurasian influences. African Americans in the US closely align with populations from other African countries, such as Ghana and Nigeria. The separation between South Asian (India) and East Asian populations (China, Malaysia) also became more distinct, reflecting regional differentiation within Asia. Overall, this level captures more recent historical admixture processes, including transatlantic gene flow, while maintaining interpretable population structure.

*K* = 23 corresponds to the labelled populations and highlights specific signals related to each population. This analysis examined 23 known populations to determine whether the *STRUCTURE* program could accurately identify genetic groupings. With *K* = 23, most individuals were assigned to a single dominant population cluster, allowing them to be recognised by their origins. Admixed populations such as US Hispanic, Brazil, and US African American continued to show contributions from multiple ancestral clusters, highlighting complex ancestry profiles. Morocco

and Egypt showed genetic contributions that may have come from other North African populations, or Middle Eastern sources. The exclusive clustering suggests that the *STRUCTURE* program effectively captures known population boundaries.

$K$ = 25 is the maximum $K$ run in the analysis. It highlights the risk of overfitting because it introduces additional complexity, but it can be valuable for exploring substructure or rare ancestry components. It shows fine-scale cluster fragmentation, and some populations split or show overlapping signals, suggesting instability or noise across replicate runs. Therefore, $K$=25 should be interpreted cautiously when validated with additional evidence such as $F_{ST}$, MDS, or NJ.

### 6.4.2 Y-STR INTER-POPULATION GENETIC DISTANCE

To illustrate the genetic relationship between Nigeria and various previously reported global populations, several analyses were conducted, including pairwise comparisons, Multi-Dimensional Scaling (MDS), Neighbour-Joining (NJ), and Analysis of Molecular Variance (AMOVA). These analyses were based on the 12 Y-STR loci shared by the datasets collected from the Nigerian sub-populations (Igbo, Yoruba, and Hausa-Fulani) and the 47 global population groups listed in **Table 6.3**.

#### Multi-Dimensional Scaling (MDS)

The Multi-Dimensional Scaling (MDS) plot **(Figure 6.7)** was constructed from the values of the pairwise unbiased $R_{ST}$ genetic distance matrix **(Appendix 84)**, which was derived from 12 Y-STR loci across 50 population groups. Each population group was categorized by its language family to enhance the understanding of their ancestral connections with other groups. The two-dimensional MDS plot illustrates significant regional clustering and differentiation based on male-lineage genetic variation. The populations are categorized by region: Africa, Europe, South America, North America, and Asia. Population labels encompass the country along with linguistic or ethnic affiliations, reflecting genetic, cultural, and historical patterns.

The African populations exhibited the highest level of internal diversity, characterised by several sub-clusters represented as blue dots. West African populations, including the three Nigerian subpopulations—Igbo (NIG-IB), Yoruba (NIG-YO), and Hausa-Fulani (NIG-HF)—along with Benin-*Niger Congo-Atlantic* (BEN-AC), Benin-*Niger Congo-Mende* (BEN-M), Burkina Faso-*Niger Congo-Mur* (BF-G), Ivory Coast-*Niger Congo-Atlantic* (IVC-AC), Ivory Coast-*Niger*

*Congo-Mende* (IVC-M), and Senegal-*Niger Congo-Mende* (SEN-M) clustered relatively close together, reflecting shared ancestry in Upper Guinea and coastal regions.

Central and East African groups such as Angola-*Bantu West* (ANG-B), Cameroon-*Niger Congo-Pygmy* (CAM-P), Central African Republic-*Pygmy* (CAR-P), Democratic Republic Congo-*Bantu West* (DRC-B), Democratic Republic Congo-*Nilotic-Pygmy* (DRC-P), Gabon-*Pygmy West* (GAB-P), Kenya-*Bantu Wes*t (KEN-B), Kenya-*Nilotic* (KEN-NS), Tanzania-*Afroasiatic* (TZ-AA), Tanzania-*Khoisan* (TZ-K), Tanzania-*Nilotic* (TZ-NS), and Uganda-*Nilotic* (UGA-NS) occupied more peripheral or spread-out positions around the West African cluster.

Notably, South Africa-*Bantu West* (SA-B), Ethiopia-*Afroasiatic* (ETH-AA), and Namibia-*Khoisan* (NAM-K) stood far from the central African cluster, indicating a highly distinct Y-STR profile, possibly reflecting ancient divergence, demographic drift, or limited gene flow with northern or western African populations. Botswana-*Bantu East* (BOT-B) in Southern Africa, Tanzania-*Bantu East* (TZ-B) in East Africa, and the Bantu populations in Zambia (ZAE-B and ZAW-B) in Southern Africa exhibit moderate divergence from other African clusters. However, they cluster closely as Bantu-speaking languages despite being located in different geographic regions of Africa. This suggests that Bantu-speaking populations share a common ancestry.

Orange dots represented the Asian populations. China-*Sino-Tibetan* (CHI), India-*Dravidian* (IND), and the Philippines-*Austronesian* (PHP) were clustered together, positioned away from African and European populations. This tight grouping reflects genetic continuity across East and South Asia and the relatively low Y-STR differentiation within this region. Middle Eastern populations, including Iraq-*Afroasiatic-Semitic* (IRQ), Israel-*Afroasiatic-Semitic* (ISR), Kuwait-*Afroasiatic-Semitic* (KUW), Lebanon-*Afroasiatic-Semitic* (LEB), and the United Arab Emirates-*Afroasiatic-Semitic* (UAE), form a closely related regional cluster. They are also near some North African groups, such as Algeria-*Afroasiatic* (ALG-AA) and Libya-*Afroasiatic* (LIB-AA). This suggests that these populations not only share a common language family but also have a shared ancestry lineage.

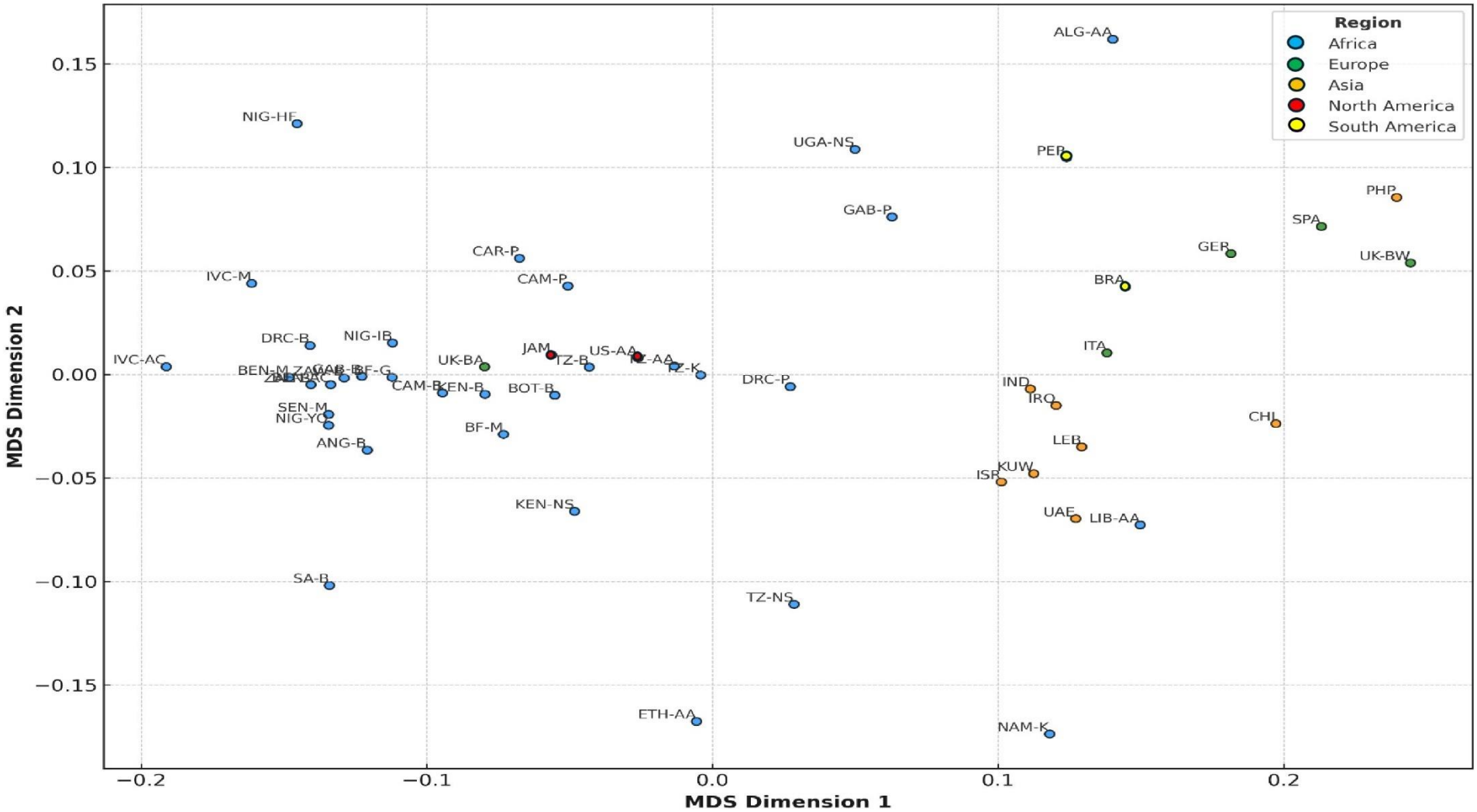



**Figure 6.7. A two-dimensional MDS plot visualising genetic distances among 50 global populations based on 12 Y-TR loci using the $R_{ST}$ genetic distance matrix. The populations are colour-coded by their respective continental regions. Dimension 1 (x-axis) captured intra-African structure and divergence among Asia and European populations. Dimension 2 (y-axis) captures the intra-African structure and divergence among Asia, Europe, and the admixed North and South American populations.**

The populations from North America were represented by red dots on the MDS plot. The US-based African Americans (US-AA) and Jamaica-*Patois* (JAM) were found to be closely aligned with several African populations, particularly those from Nigeria (NIG-IB and NIG-YO), Botswana (BOT-B), Burkina Faso (BF-M), Cameroon (CAM-B), Democratic Republic of Congo (DRC-P), Kenya (KEN-B and KEN-NS), and Tanzania (TZ-AA and TZ-K). This alignment indicates a stronger continuity with their ancestral origins. However, the Nigerian-*Hausa-Fulani* subpopulation does not demonstrate as close a genetic relationship with the African American group compared to the Nigerian Igbos and Yorubas. This suggests that the Nigerian-*Hausa-Fulani* people migrated to the United States in smaller numbers than the Igbos and Yorubas, leading to a more distant connection in male ancestry.

In the MDS plot, the South American populations are represented by yellow dots. The Brazil-*Indo-European-Portuguese* (BRA) and Peru-*Indo-European-Spanish* (PER) populations are positioned between the European and African populations, indicating a known admixture that involves enslaved African males, European colonists, and contributions from Indigenous Americans.

The European populations are shown as green dots in the MDS plot. Germany-*Indo-European-German* (GER), Italy-*Indo-European-Italian* (ITA), Spain-*Indo-European-Spanish* (SPA), and United Kingdom-*Indo-European-English* (UK-BW) are grouped closely together, forming a well-defined cluster. The European population displays proximity to the Asian populations, possibly indicating shared Eurasian male lineages and deep ancestral connection. The British African (UK-BA) group is genetically similar to several other African populations, particularly those from Nigeria (NIG-IB and NIG-YO), Benin (BEN-AC), Botswana (BOT-B), Burkina Faso (BF-G), Cameroon (CAM-B), and Kenya (KEN-B). Overall, the British African group is genetically more closely related to African populations than to European populations. This indicates their genetic profile shares significant similarities with various African groups, suggesting a strong ancestral connection between British Africans and their African roots. This aligns with historical patterns of migration and ancestry. Moreover, the Nigerian-*Hausa-Fulani* subpopulation does not show as close a genetic relationship with the British African group as the Nigerian Igbos and Yorubas. This suggests that the Nigerian-Hausa-Fulani people migrated to the United Kingdom in smaller numbers than the Igbos and Yorubas, resulting in a more distant male ancestry connection.

**Neighbour-Joining (NJ)**

The Neighbour-Joining (NJ) method was utilised to visualise the phylogenetic distances among the 50 population groups. This approach enabled an assessment of their ancestral relationships. The optimal tree is based on the accuracy of the values of the pairwise unbiased $R_{ST}$ genetic distance matrix (see **Appendix 84**), which is derived from 12 Y-STR loci. The tree is scaled, with branch lengths reflecting the pairwise values used to estimate genetic distances. Genetically similar populations will have closer branches, while those that are genetically distant will have branches that are farther apart. These results illustrate the paternal ancestral genetic distances and regional group patterns, as shown in **Figure 6.8.**

The three Nigerian ethnic groups (Hausa-Fulani, Igbo, and Yoruba) form a distinct West African cluster with clear branch separations for the Nigerian populations. The Yoruba and Igbo populations are closely related and show strong genetic connections to the Benin (Niger-Congo-Mende) and Ivory Coast (Niger-Congo-Atlantic) groups. The Yoruba population also exhibited a close genetic relationship with the Benin (Niger-Congo-Atlantic) population. The close relationship between the Yoruba people is evident as they historically spread across Nigeria and Benin before the colonial separation of these West African countries, indicating they share a common genetic ancestry with the Yoruba communities in this region. The Yorubas are recognised as a minority in the Benin Republic. In contrast, the Hausa-Fulani group appears to be genetically close to Senegal and Ivory Coast populations, specifically within the Niger-Congo-Mende language group. The Hausa-Fulani population is located between the Igbo and Senegal populations and still exhibits some shared ancestry with the Yorubas and Igbos; however, it is classified within the Niger-Congo-Mende cluster. Overall, Nigerian populations, including the Igbo, Yoruba, and Hausa-Fulani ethnic groups, are genetically distinct from Afroasiatic-speaking North Africans from countries such as Libya and Algeria, as well as the Middle Eastern populations, including Iraq, Israel, Kuwait, and Lebanon. Additionally, they share distant connections with Asians from China, India, and the Philippines and Europeans from Italy, the United Kingdom, and Spain. The Nigerian ethnic groups are part of a larger clade that includes various West and Central African communities, such as those in Benin, Ivory Coast, Gabon, Burkina Faso, Cameroon, and Angola, which predominantly speak Bantu or Niger-Congo languages.

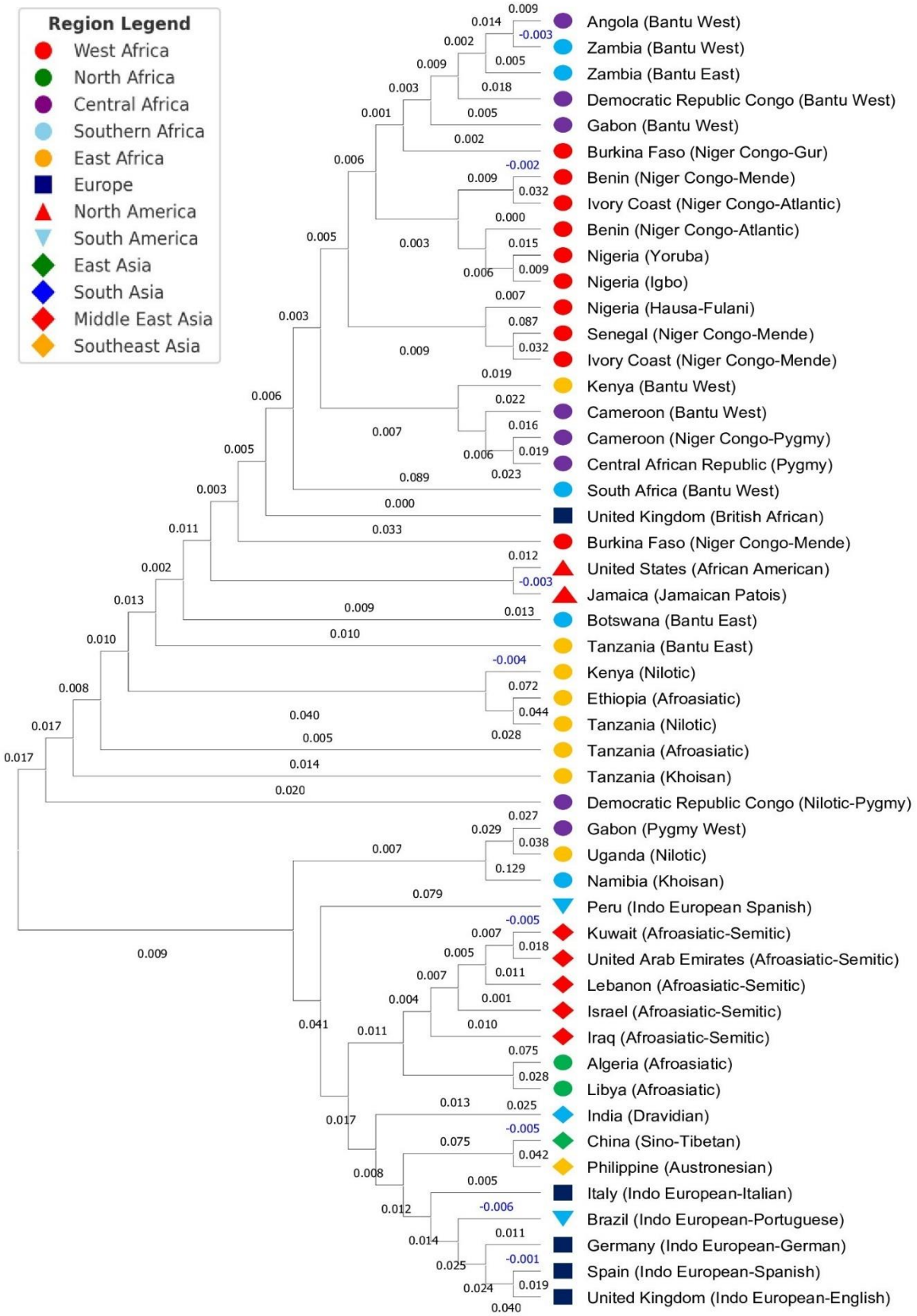



**Figure 6.8. Neighbour-Joining (NJ) tree constructed from the values of the pairwise unbiased $R_{ST}$ genetic distance matrix, illustrating the clustering of 50 global population groups, using 12 Y-STR loci. The populations are colour-coded according to their respective continental geopolitical regions, with branch lengths representing paternal ancestral genetic distances.**

Populations such as those in Angola (Bantu West), Zambia (Bantu East and West), the Democratic Republic of Congo (Bantu West), Gabon (Bantu West), Benin (Niger-Congo Atlantic and Mende), Ivory Coast (Niger-Congo Atlantic and Mende), Senegal (Niger-Congo Mende), and Nigeria (Yoruba, Igbo, Hausa-Fulani) form a closely-knit cluster. This group primarily consists of speakers of the Niger-Congo language family (both Atlantic and Mende) and various Bantu-speaking groups (both East and West). The genetic similarities and shared ancestral lineages within this population are likely a result of the Bantu expansion and deep West African ancestry.

Cameroon (Niger-Congo Pygmy) and the Central African Republic (Vitious Pygmy) are closely clustered together, while Cameroon (Bantu West) and Kenya (Bantu West) are nearby but distinct. This indicates the divergence between Pygmy and Bantu populations within Central Africa. Additionally, Kenya (Nilotic), Ethiopia (Afroasiatic), and Tanzania (Afroasiatic, Nilotic, Khoisan, and Bantu East) form a separate East African branch, highlighting their different ancestries and languages (Nilotic, Afroasiatic, and Khoisan). Furthermore, the Tanzania (Bantu East) cluster aligns closely with Botswana (Bantu East).

The African populations in the United States (African Americans), the United Kingdom (British Africans), and Jamaica (Jamaican Patois speakers) are closely grouped within the African genetic clade. However, Jamaica clusters closely with African Americans, which reflects the historical legacy of the transatlantic slave trade and their common African ancestry.

Populations such as Kuwait, the United Arab Emirates, Lebanon, Israel, and Iraq form a cohesive Afroasiatic-Semitic cluster. In contrast, the Dravidian lineage from India and the Eastern Asian lineages are somewhat more distinct, indicating moderate divergence. The populations of China (Sino-Tibetan) and the Philippines (Austronesian) cluster with moderate branch lengths, suggesting a shared Asian ancestry. Furthermore, Asian populations are generally more distant from African groups, as expected from their geographic and evolutionary history.

Populations from Italy, Germany, Spain, and the United Kingdom (all Indo-European groups) form a close-knit and shallow cluster, indicating high genetic similarity. Brazil is located nearby in this cluster, which reflects its substantial European genetic influence, particularly from Portuguese ancestry. Peru, whose population has Indo-European and Spanish ties, is genetically situated near groups influenced by Middle Eastern and African ancestry, likely due to historical admixture.

**Analysis of Molecular Variance (AMOVA)**

AMOVA was performed using data from 12 Y-STR loci collected from 7,348 unrelated males across 50 globally distributed populations. This analysis classified genetic variation into three nested levels: first, among continental regions, which included South America, North America, Europe, Asia, and Africa; second, among populations within those regions, represented by 50 distinct global population groups (refer to **Table 6.3**); and third, within populations, comprising the total of 7,348 unrelated individuals distributed across each of these groups. A summary of the AMOVA findings can be found in **Table 6.6**, while **Table 6.7** displays the Phi statistics (Φ-statistics) along with their corresponding p-values.

**Table 6.6. Summary of AMOVA for 12 Y-STR loci across 50 global populations involving 7,348 unrelated individuals grouped into five continental regions**

| Source of Variation | Degrees of freedom (df) | Sum of Squares | Variance Components | % Variation | Interpretation |
|---|---|---|---|---|---|
| **Among Regions** | 4 | $9.5 \times 10^3$ | $2.3 \times 10^3$ | 88.11 | **High genetic variation** between continental regions (e.g., Africa vs. Asia). |
| **Among Populations within regions** | 45 | $1.4 \times 10^4$ | $3.2 \times 10^2$ | 11.57 | **Moderate genetic variation** among populations within continental regions but statistically significant (e.g., Igbo vs. Yoruba). |
| **Within Populations** | 7,297 | $6.4 \times 10^4$ | 8.82 | 0.32 | **Low genetic variation** among individuals within populations. (e.g. Igbo individuals) |
| **Total** | **7,348** | **$8.8 \times 10^4$** | **$2.7 \times 10^3$** | **100%** | |

**Table 6.6.** shows the outcomes of the AMOVA analysis and illustrates partitions of the genetic differentiation. Most genetic variation, accounting for 88.11%, is attributed to differences among major geographic regions: South America, North America, Europe, Asia, and Africa. A moderate proportion of genetic variation, 11.57%, occurs between populations within the same region, indicating some regional sub-structure. For instance, this variation can be observed when comparing the Nigerian (Igbo) population to the Nigerian (Yoruba) population, as well as between the Angola (Bantu West) population and the Botswana (Bantu West) population in Africa, and among populations in China and India in Asia. Only a tiny fraction, just 0.32%, of the genetic

differentiation is found among individuals within a population, which reflects the relatively low diversity among Y-STRs within these groups.

**Table 6.7. Summary of the Phi (Φ)statistics and p-values based on the AMOVA analysis of the Y-STR dataset**

| Statistic | Φ Value | P-value | Interpretation |
|---|---|---|---|
| **$\Phi_{RT}$ (Among Regions)** | 0.881 | < 0.001 | Strong genetic structure among continents (Africa, Asia, Europe, North and South America). This means that 88.1% of the genetic differentiation is due to variation between regions, indicating high differentiation across continents. |
| **$\Phi_{PR}$ (Among Populations within Regions)** | 0.973 | < 0.001 | Very strong genetic structure among populations within regions (e.g., African populations), contributing 97.3% to total variance. This shows that even populations in the same region are very highly differentiated. |
| **$\Phi_{PT}$ (Among Populations Overall)** | 0.997 | < 0.001 | Nealy complete genetic structure globally. The total genetic differentiation across all populations and regions is 99.7%. This implies that overall, 50 global populations are extremely structured. |

The Φ-statistics **(Table 6.7)** were used to measure genetic differences among populations. The $\Phi_{PT}$ value of 0.997 indicates that the populations are highly structured genetically ($P < 0.001$), signifying significant genetic variation across different continents. The overall $\Phi_{RT}$ value of 0.881 demonstrates considerable differentiation among all populations, which is also statistically significant ($P < 0.001$). This suggests a strong regional structure in global Y-STR variation. The $\Phi_{PR}$ value of 0.973 shows a substantial difference among populations within regions, which is statistically significant ($P < 0.001$). This indicates that certain groups of people have gone through substantial historical separations, migration patterns, and changes over time.

## 6.5 DISCUSSION

There are generally considered to be seven continents on Earth: Africa, Europe, South America, North America, Asia, and Antarctica. Among these continents, Africa is considered to be the most genetically diverse (Campbell and Tishkoff, 2008). Africa has over 2,000 ethnolinguistic populations and exhibits remarkable genetic variations (Tishkoff *et al*., 2009). It retains the highest level of genetic diversity and has been recognised as the cradle of modern human civilisation (Guanglin *et al.,* 2023). As humans migrated out of Africa, two factors—the founder effect and genetic drift—led to a reduction in genetic diversity in populations outside Africa. The migration carried with it only a small part of the genetic diversity that existed in their African homeland. This is why non-African populations have lower genetic diversity. The loss of this diversity accounts for the genetic differences observed in this study between African populations and those from other

regions (Stringer, 2016; Wang *et al*., 2021). Modern human origins can be traced to Sub-Saharan Africa, particularly in East Africa and certain regions of Southern Africa. These areas have the oldest genetic roots of humanity. Within the continental regions, populations exhibit less noticeable differences and more complex genetic patterns. This observation aligns with the research conducted by Cavalli-Sforza *et al*. (1994), Tishkoff *et al*. (2009), Henn *et al*. (2011), and Jobling and Tyler-Smith (2017), which emphasised that human history—marked by ancient separations between continents, founder effects, and gene flow between areas—has significantly influenced genetic variation.

**The autosomal STR dataset inference**

The analysis of autosomal STR datasets using Nei's and $F_{ST}$ genetic distances and techniques such as MDS (Multi-Dimensional Scaling), NJ (Neighbour-Joining), *STRUCTURE* and AMOVA (Analysis of Molecular Variance) revealed a strong genetic differentiation within and between the studied global populations across the continents. The $F_{ST}$ and Nei genetic distance analyses yielded genetic variation patterns among populations. The MDS and NJ outputs were plotted using pairwise Nei's unbiased genetic distance data, showing relatively similar patterns. The bar plots generated from the *STRUCTURE* analysis illustrated the populations' genetic structure and historical admixture. The AMOVA results showed how variance is distributed among different continents, populations within those continents, and individuals within each population. The magnitude of both pairwise $F_{ST}$ values and AMOVA Φ-statistics observed in this study is consistent with previously reported global estimates for autosomal STR loci. Steele *et al*. (2014), in work by Syndercombe-Court and Balding, demonstrated that worldwide population differentiation at forensic STR markers is generally low, with typical $F_{ST}$ /Φ values ranging from approximately 0.01 to 0.03, reflecting that the vast majority of human genetic variation occurs within populations. The results obtained here, including low differentiation between Nigeria and populations with African ancestry, higher differentiation with East and Southeast Asian populations, and an overall $\Phi_{ST}$ value of approximately 0.019, fit well within this expected global range. These statistical tools effectively highlighted significant differences between continents and more detailed population structures based on autosomal STR diversity, and they support the robustness and forensic relevance of the present dataset.

**Africa**

West African groups formed a close cluster, particularly those from Nigeria and Ghana. This is expected because Ghana and Nigeria share similar cultures, geographies, and religions. A study by

Adeyemo *et al*. (2005) analysed 372 autosomal microsatellite (STR) loci across four West African ethnic groups: the Ga-Adangbe and Akan from Ghana and the Igbo and Yoruba from Nigeria. The study found that, while subtle differences allowed for the distinction between Ghanaian and Nigerian groups, the overall genetic differentiation was minimal. Mozambique and Somalia have more divergent populations, likely due to the influence of East and Southern African ancestors. Somalia bridges the Middle East and sub-Saharan Africa, highlighting its genetic connections to both regions. Similarly, Morocco and Egypt are genetically similar and may serve as a North African bridge connecting Africa and the Middle East. Egypt's genetic distance is between Africa and Middle Eastern countries, such as Bahrain and Saudi Arabia, reflecting a blend of West Asian and North African genetic backgrounds. North African and Middle Eastern populations have Afroasiatic ancestry and experienced historical gene flow across the Red Sea and Mediterranean (Hodgson *et al*., 2014). The genetic background of North Africans comes from a mix of local Berber ancestry, ancient Eurasian migrations, and Arab expansions starting in the 7th century CE. These events introduced Middle Eastern traits to North African populations (Arauna *et al*., 2017). This blending resulted in unique characteristics distinguishing North Africans from sub-Saharan Africans in appearance, culture, and language (Henn *et al*., 2012).

**Asia**

Populations from the Middle East, including Bahrain and Saudi Arabia, have genetic similarities to populations in Europe and North Africa. This shows that the Mediterranean corridor was critical in allowing geneflow between these regions throughout history (Henn *et al*., 2012; Arauna *et al*., 2017). North Africa's population has a genetic mix influenced by several historical events. These include native Berber ancestry, the movement of people from Eurasia, and Arab migrations after the 7th century CE Islamic conquests. These events added Middle Eastern genetic traits to North African populations (Arauna *et al*., 2017). Due to admixture, North Africans developed unique features that distinguish them from sub-Saharan Africans in areas like appearance, culture, and language (Henn *et al*., 2012). In contrast, Asian populations from China, Laos, Malaysia, and Thailand are closely grouped with the Asian population in the United States. The images of the *STRUCTURE* data output suggest that India appears genetically more aligned with Middle Eastern populations, particularly Bahrain, than those from China, Laos, Malaysia, and Thailand. However, the Neighbour-Joining output indicates that India is categorised with other populations within the Asian clade. This demonstrates that the Asian populations share a genetic background and remain connected to the broader region (Li *et al*., 2008; Reich *et al*., 2009).

**Europe**

The two European populations examined in this study—Italy and Spain—share similar genetic backgrounds. This resemblance comes from their close geographical proximity and shared widespread historical experiences as populations from Southern Europe. While not genetically identical, Italy and Spain show significant similarity. People from Southern Europe share a close ancestry with Caucasians in the United States. This shows their shared European roots and genetic ties. A study by Seldin *et al*. (2006) and Sarno *et al.* (2014) found that individuals of southern European ancestry, specifically Italians and Spaniards, largely clustered together, indicating significant genetic similarities because of their geographical closeness and intertwined histories.

**Latin Americans in North and South America**

The Latin American populations, including Brazil, Mexico, Colombia, and Honduras, are intermediate between Europe, Africa, and the Americas. Mexico and Colombia share a closer genetic relatedness with the Hispanic population in the United States. At the same time, Brazil shows genetic similarities with Italians and Spaniards in Europe, and Caucasians in the United States. Although Honduras is located in Central America, the population analysed here comprises the Garifuna of the Caribbean coast of Central America and shows closer genetic affinity to African populations, reflecting their predominantly African ancestry. During the Spanish colonial period, enslaved Africans were brought to Honduras primarily for labour in mining and agriculture. Enslaved Africans in Honduras experienced social and geographic mobility, particularly along the Atlantic coast, where they could settle and work as labourers (Lokken, 2013). Latin Americans' genetic history and connection reflect their unique tri-hybrid background, encompassing European, African, and Native American roots.

**North America**

The populations of North America, including African Americans, Asians, Caucasians, and Hispanics in the United States, as well as those from Honduras and Mexico, demonstrate a diverse genetic mixture of ancestry and historical backgrounds. African Americans and people from Honduras share genetic ties with African populations. This shows the historical links to the trans-Atlantic slave trade (Salas *et al*., 2005). In contrast, US Asians genetically belong to the broader Asian group, clustering especially with China, Malaysia, Thailand and Laos, which aligns with expectations (Tian *et al*., 2008; Bryc *et al*., 2015). US Caucasians genetically cluster near Southern European populations (Italy and Spain), confirming their European ancestry. Meanwhile, US Hispanics and the population from Mexico and Colombia are genetically positioned between

European (Italy and Spain) and Latin American groups (Brazil), reflecting their mixed ancestry, which includes Indigenous American and European roots. The genetic structure in North America aligns with the study conducted by Bryc *et al*. (2015), who examined the genetic backgrounds of African Americans, European Americans, and Latin Americans in the United States. The DNA samples were typed using SNP markers, and their findings gave clear insights into the genetic backgrounds of these groups. The outcome revealed the diverse mix of ancestries in North America, shaped by years of migration, colonisation, and cultural exchange.

The AMOVA result revealed that individuals within each population (within-population) accounted for most of the diversity in the sample. The high genetic differentiation observed among individuals within populations (98.05%) aligns with findings from Rosenberg *et al*. (2002) and Adeyemo *et al*. (2005), who noted that the genetic variation among individuals within a population account for the most significant variation in analyses of autosomal STR datasets. Low genetic variation was found among continental regions and among populations within continents. All Φ-statistics were significant ($P < 0.0001$), confirming that the observed genetic structure is statistically meaningful. These variations may be linked to a lengthy trade history and other interactions between continents. Factors such as Transatlantic slavery, migration, intermarriage, and a common language contribute to gene flow and genetic admixture, which help reduce genetic differences among regions and populations within a region (Gouveia *et al*, 2020).

**The Y-STR dataset inference**

The Y-STR datasets were analysed using $R_{ST}$ genetic distances along with techniques such as Multi-Dimensional Scaling (MDS), Neighbour-Joining (NJ), and Analysis of Molecular Variance (AMOVA). The pairwise unbiased $R_{ST}$ genetic distance matrix value was used to plot MDS and NJ for the Y-STR datasets. $R_{ST}$ is preferred for the analysis of Y-STRs because it accounts for differences in allele sizes and follows the stepwise mutation model, which accurately represents the way Y-STRs mutate (Balloux *et al*., 2002; Excoffier *et al*., 2005). However, autosomal STR alleles recombine each generation, mixing alleles from both parents and making the stepwise mutation pattern less clear over generations (Weir, 1996). The Y-STR markers effectively revealed the ancestral relationships among population groups in Nigeria, Africa, and other continental regions. The MDS and NJ methods demonstrated similar output patterns at the intercontinental level. However, the NJ method was more effective in clustering populations based on geographical regions at the continental region level. For instance, it distinguished the West African population from the East African population, identified admixtures between Central and Southern African

populations, and displayed distinctions between the North African and the rest of the African populations. The AMOVA results illustrated the variance distribution across continents, populations within those continents, and individuals within each population. The statistical tools employed in this study highlighted significant differences in paternal inheritance across continents. These tools also allowed for tracking genetic lineage based on paternal inheritance using Y-STR diversity. The Y-STR markers also showed how genetic structures vary across different continents.

**African Extensive Male Lineage Structure**

African populations show apparent differences in their male lineage structures. Distinct differences and complex connections among male family lines characterise significant genetic diversity among these groups. Groups from West Africa, such as the Igbo, Yoruba, and Hausa-Fulani in Nigeria, along with populations from Burkina Faso, Senegal, Ivory Coast, and Benin, form closely related genetic clusters. Similarly, populations from Central Africa, including those in Cameroon, Angola, the Democratic Republic of Congo, the Central African Republic, and Gabon, exhibit strong genetic connections. The West African group clusters correspond to the Niger-Congo language family, which includes Atlantic, Gur, and Mende languages. In contrast, the Central African group clusters correspond with the Bantu-West (sub-group of the Niger-Congo language family) speaking populations. Cameroon serves as a bridge between West Africa and Central Africa, featuring a diverse population that includes speakers of Niger-Congo, Pygmy, and Bantu languages. Located in Central Africa, Cameroon shares borders with Gabon, the Democratic Republic of Congo, and Equatorial Guinea to the south, the Central African Republic to the east, Nigeria to the west, and Chad to the north. This geographic positioning facilitates gene flow and cultural exchange among the Niger-Congo-speaking communities from West Africa and the Bantu and Pygmy language groups predominantly found in Central Africa. This pattern also reflects a significant impact of Bantu expansion by spreading common Y-chromosomal haplotypes across much of sub-Saharan Africa (Wood *et al*., 2005; De Filippo *et al*., 2011). Populations in East Africa, including Nilotic, Afroasiatic, and Khoisan speakers from Kenya, Tanzania, and Uganda, as well as Southern African populations like Bantu speakers from South Africa and Botswana and Khoisan speakers from Namibia, show a clear genetic divergence from West African lineages. Notably, Nilotic, Khoisan, and Pygmy groups are genetically distinct from Niger-Congo speakers, indicating that they possess ancient, divergent paternal lineages. These findings support earlier research by Wood *et al*. (2005) and De Filippo *et al*. (2011). Additionally, Nilotic and Pygmy language-speaking groups from the Democratic Republic of Congo (DRC) and Pygmy-speaking groups from Gabon in Central Africa cluster near the East African clade, reflecting their shared male ancestry. This demonstrates the

genetic relationship between Pygmies and Nilotes. A similar connection has also been noted by Bitani *et al*. (2011).

**The Retention of African Y-Chromosome Lineages in the African Diaspora**

The MDS and NJ data output clearly illustrated that populations of African ancestry in the United States, United Kingdom, and Jamaica closely cluster with African populations. This finding confirms the retention of ancestral Y-STR haplotypes among males of African descent and reflects the transatlantic slave trade's impact on global Y-chromosomal variation. More than 9 million Africans were forcibly brought to America from the 16th century to the mid-19th century (Gouveia *et al*., 2020). The introduction of enslaved Africans transformed the genetic landscape of America, resulting in the deep genetic roots shared between African populations and those in North America and South America. This also highlights the genetic differentiation shaped by historical migrations, intercontinental gene flow, and recent patterns of admixture (Tishkoff *et al*., 2009; Bryc *et al*., 2010; Campbell *et al*., 2014).

**Middle Eastern and North African Paternal Continuity**

Populations that speak Afroasiatic languages in North Africa, particularly in Libya and Algeria, form a distinct group alongside Middle Eastern populations from Iraq, Lebanon, Israel, Kuwait, and the United Arab Emirates. This close connection suggests historical gene flow mediated by males across the Mediterranean and the Arabian Peninsula, which aligns with the dispersal of Afroasiatic language family (Cruciani *et al*., 2007; Fadhlaoui-Zid *et al*., 2013). The populations of Libya and Algeria have been found to exhibit a mixed Y-STR signature, suggesting a combination of North African and sub-Saharan male lineages. According to Bekada *et al*. (2013), people in North Africa carry a mixture of sub-Saharan African and Middle Eastern genetic influences, which include individuals from Libya and Algeria.

**Asian Populations and Lineage Clustering**

Populations from China, India, and the Philippines form a separate branch, consistent with region-specific Asian lineages (Underhill *et al*., 2001; Kivisild *et al*., 2003; Karafet *et al*., 2008). India (Dravidian) is slightly apart from East and Southeast Asian populations, reflecting its complex demographic history and ancient separation of South Asian paternal lineages (Sahoo *et al*., 2006).

**European Ancestry and Male Lineages in Latin America**

European populations, including those from Germany, Italy, Spain, and the United Kingdom, tend to cluster closely. This suggests they share relatively homogeneous Y-STR profiles consistent with Indo-European ancestry, as Underhill *et al*. (2001) reported. Interestingly, Brazil and Peru also cluster within this group, despite being located in Latin America, due to their significant European paternal ancestry. This ancestry stems from male-biased migration during the colonial period, as noted by Rodrigues *et al*. (2022). Rescue *et al*. (2016) found that over 90% of Y-chromosome lineages in Brazilian males are of European origin, while Native American and African ancestries account for the remainder.

The AMOVA results obtained here revealed significant genetic variation among these major continental regions. This finding is typical for Y-chromosome markers, which are characterised by paternal inheritance and a lack of recombination (Jobling and Tyler-Smith, 2003; Kayser *et al*., 2003). Moderate to low genetic differentiation was observed among populations within each continent and among individuals within specific populations. All Φ-statistics are significant ($P < 0.001$), confirming that the observed genetic structure is statistically meaningful and not random. Y-STRs are located in the non-recombining segment of the Y chromosome and are passed down exclusively via the paternal lineage. This makes them particularly valuable for tracing male-specific population structures, migration patterns, and deep ancestry (Ballantyne *et al*., 2014).

## 6.6 CONCLUSION

Africa is widely regarded as the origin of the human species and contains the most remarkable diversity of human DNA lineages worldwide. Because of its wide range of genetic diversity, Africa has a key role in human history. This diversity supports the idea that early humans moved from Africa to various regions across the globe. It is essential to clarify that "African American" can be somewhat imprecise. The term mainly refers to Sub-Saharan Africans who were taken from Africa during the Transatlantic Slave Trade. Most of these individuals were black. However, it does not include North Africans, such as Berbers and Sahrawis. Some Indigenous African groups have lighter skin and curly hair, rather than the dark skin and tightly curled hair often expected. Africa has many different ethnic groups, each with skin colours, facial features, body, and hair.

Analysing autosomal STR datasets with Nei and $F_{ST}$ genetic distances and methods like Multi-Dimensional Scaling (MDS), Neighbour-Joining (NJ), and *STRUCTURE* helps us understand large regional patterns and more minor details of mixing within populations. The study revealed strong

regional continuity across Africa, Asia, and Europe while highlighting the genetic effects of colonial and transatlantic migrations from Africa to America. It emphasized the link between geographical locations and bi-parental genetic profiles and showcased the uniqueness of North African populations globally. This analysis shows how useful STRs are for understanding complex population histories influenced by ancient splits and recent human movements. The research also places Nigeria's genetic distance alongside other global populations on a continental scale. The Analysis of Molecular Variance (AMOVA) confirmed an established pattern in human population genetics: most autosomal STR variation occurs within populations, with moderate genetic structuring observable across regions and between populations. These findings indicate that STRs are valuable tools in forensics and anthropology, allowing us to differentiate between population groups at both regional and continental levels.

Genetic distances from $R_{ST}$ and methods like Multi-Dimensional Scaling (MDS) and Neighbour-Joining (NJ) using Y-STRs show how effective paternal markers are in uncovering complex and deep-rooted relationships among populations. This research clarified the genetic distance of the Igbo, Yoruba, and Hausa-Fulani ethnic groups within West Africa, alongside other African populations and worldwide. This research examined deep paternal divergences in Africa, focusing on ethnolinguistic groups and identifying ancient lineages among the speakers of Afroasiatic, Khoisan, Niger-Congo (Atlantic, Bantu, Mende, and Gur), and Nilo-Saharan (Nilotic) language families. It also uncovered shared Y-chromosomal signatures among West African diaspora populations. Additionally, the research established a continuity of Middle Eastern and North African ancestry, reflecting shared Afroasiatic paternal lineages. The study showed a distinct clustering of European and Asian Y-haplogroups, with Latin American populations exhibiting European male-line admixture. The NJ tree results mirror ancient human population structures and recent historical migrations, providing a global high-resolution perspective on paternal ancestry. The outcomes of the AMOVA analysis support the notion that while human populations share a substantial portion of their genetic diversity, there are apparent regional and intra-regional differences, particularly influenced by continental separations, population bottlenecks, and genetic drift. Y-STR markers, which are paternally inherited and non-recombining, are particularly effective in revealing the structure of male lineage populations.

# CHAPTER SEVEN

# GENERAL DISCUSSION

## 7.1 THE NIGERIAN POPULATION GENETICS AND THE FORENSIC IMPLICATIONS

The Nigerian population, represented by three major ethnolinguistic groups — Hausa-Fulanis, Yorubas, and Igbos — shows genetic proximity due to shared Niger-Congo ancestry. Despite all, it is also distinguishable based on paternal Y-STR lineages. Understanding this internal substructure is essential for forensic and anthropological applications. This project evaluated forensic genetic markers in Nigeria, analysing two autosomal STR kits—the Qiagen Investigator ESSplex SE QS Kit and the GlobalFiler Express PCR Amplification Kit—along with a Y-STR dataset from the Promega PowerPlex Y23 System Kit. The study focused on unrelated individuals from the three ethnic groups. The findings recommended the most preferred STR kit for establishing an offender DNA database and forensic DNA analysis in Nigeria. The study also provided allele frequency data and explored genetic relationships between the Nigerian and global populations through various methods. Africa has more than 2,000 ethnolinguistic groups exhibiting significant genetic diversity. Studies have examined the genetic connections within and between Nigerian populations, considering geographic factors.

In Nigeria, autosomal STRs present a high potential for individualisation, as most variation is found within populations. The autosomal STR profiles generated from Nigeria (Chapters 3 and 4) are valuable in global databases and can be effectively utilised in forensic comparisons involving African or diaspora samples. This highlights the forensic significance of the Nigerian population, demonstrating a strong discriminatory power for identification purposes using either the 24-locus GlobalFiler Express Kit or the 17-locus QIAGEN Investigator ESSplex SE QS Kit. However, the 21 autosomal loci provided by the GlobalFiler™ Express Kit offer more markers for analysis, resulting in greater discriminatory power, improved likelihood ratios, and lower random match probabilities than the 16 loci available in the QIAGEN Investigator ESSplex SE QS Kit. Additionally, the information generated from the Promega PowerPlex Y23 System Kit (Chapter 5) gave valuable insights into the ancestry patterns of male individuals analysed in the Nigerian population. This study emphasises the importance of incorporating Y-STR data specific to the Nigerian population into global forensic databases. Creating ethnic-specific allele frequency tables for the Igbo, Yoruba, and Hausa-Fulani groups can improve the accuracy of forensic comparisons in Nigeria and beyond. This data is helpful for paternity testing, criminal investigations, and identifying victims of disasters. Understanding both deep ancestral markers and recent population changes makes Nigerian groups important references for studying human genetic diversity and

forensic genetics. STR markers are essential tools for understanding human history and helping achieve justice in forensic cases (Tishkoff *et al*., 2009; Kayser, 2017).

Regarding Chapter 5 of this project, Bryc *et al*. (2010) analysed the genetic structure of the West African population. They found minimal genetic differentiation between the Igbos and Yorubas, as indicated by their low genetic distance. Unlike the earlier study, which grouped the Hausas and Fulanis as a single sub-population, Bryc *et al.* (2010) collected separate samples for each ethnic group. Their findings revealed that the Hausas are genetically closer to the Igbos and Yorubas, while the Fulanis are genetically distinct from all three groups. Although Bryc *et al*. (2010) primarily focused on evaluating autosomal short tandem repeats (STRs), a similar outcome was observed in the Y-STR analysis conducted in this study. The study of genetic relationships among Nigerian populations revealed that a segment of the Hausa-Fulani subpopulation is genetically distant from the Igbo and Yoruba subpopulations, as indicated by the Y-STR Principal Component Analysis (PCA) performed in Chapter 5 of this research. Additionally, a global population Multi-Dimensional Scaling (MDS) plot, which assessed genetic relationships between Nigeria and other populations worldwide, positioned the Hausa-Fulani as genetically distinct from the Igbos and Yorubas. However, they remain relatively close to the West African genetic cluster (Chapter 6).
Given that the Hausa-Fulani ethnic group has mixed ancestry, the distinct genetic differentiation observed in the PCA (Figure 5.9) may suggest that those identified as genetically distinct could be the Fulanis. Historical records support this, as the Fulanis exhibit significant differences in male ancestry compared to the Igbos, Yorubas, and Hausas. The Fulanis historically migrated eastward into northern Nigeria from Futa Toro in Lower Senegal, pursuing a pastoral and nomadic lifestyle (Lovejoy, 2016). The Neighbour-Joining output of the Y-STR analysis (Figure 6.8) confirmed this claim by demonstrating a shared ancestral relationship between the Niger-Congo Mende-speaking population in Senegal and the Hausa-Fulanis, similar to the genetic connection that the Hausa-Fulani share with the Igbos and Yorubas. This suggests that within the mixed ancestry of the Hausa-Fulani ethnic group, the Hausas are more genetically related to the Igbos and Yorubas. At the same time, the Fulanis are more genetically connected to the Niger-Congo population in Senegal. The jihad movement initiated in the 19th century by Usman Dan Fodio—a Fulani spiritual leader, thinker, and transformative reformer—targeted the Hausa Kingdoms in what became known as the Fulani War (Hiskett, 1994). This conflict led to the establishment of the Sokoto Caliphate. The Hausas and the Fulanis interacted and intermarried over time, resulting in a new ethnic group called the Hausa-Fulani, who mainly speak different dialects of the Hausa language. This interaction also brought together their cultures and languages (Lovejoy, 2016; Fortes-Lima *et al*., 2025). The Fulani people speak Fulfulde (also known as Fula), a language that belongs to the Niger-Congo language

family, specifically within the Atlantic branch (Vicente *et al*., 2019; Fortes-Lima *et al*., 2025). This linguistic classification helps explain why the Hausa-Fulanis remain within the Niger-Congo-Mende cluster despite the Hausa language being classified under the Afroasiatic language family. The assertion that the Igbo people are a "lost tribe of Israel" has been thoroughly examined and clarified in recent research. Some members of the Igbo community in southeastern Nigeria, particularly those advocating for the sovereignty of Biafra, have historically proposed a potential genetic or ancestral connection to ancient Israel (Olasupo, 2014; Shragg, 2015; Agbo, 2024). However, this claim mainly comes from cultural, religious, and political beliefs, and no scientific evidence supports it (Ejiofor, 2022). Historical and genetic studies indicate that the connections are weak and unclear, with insufficient evidence to support the idea that the Igbo people are descendants of the Israelites. Additionally, it is essential to note that Israel has not officially recognised the Igbo community as Jewish by descent (Chiluwa and Chiluwa, 2020). This study reveals that the Igbo people are more closely related, both genetically and ancestrally, to the Yoruba and other groups in West Africa and sub-Saharan Africa than to the Israeli groups in the Middle East (Figures 6.7 and 6.8). Moreover, populations in North Africa, such as those from Algeria and Libya, exhibit greater genetic affinities to Israeli populations compared to any groups in sub-Saharan Africa.

The genetic research evaluating 7,664 unrelated individuals using autosomal STR markers and 7,348 unrelated individuals using Y-STR markers provided valuable insights into the Nigerian population structure at the global level (Chapter 6), which is essential for forensic studies. The study employed various genetic distance measures, including Nei, $F_{ST}$, and $R_{ST}$, along with methods such as Multi-Dimensional Scaling (MDS), Neighbour Joining (NJ), and Analysis of Molecular Variance (AMOVA). These methods helped to examine patterns of genetic diversity, population grouping, and ancestry. Based on the Y-STR results, Nigerian populations align with other West African groups (Niger-Congo) at the continental level. They are genetically distinguishable from North African (Afroasiatic), East African (Nilotic), and Indigenous Central African (Pygmy, Khoisan) populations. Autosomal STR results place Nigeria firmly within the African continental genetic clade but with closer affinities to African diaspora populations in North and South America, reflecting evidence of the Transatlantic Slave Trade, during which millions of West Africans—including many from present-day Nigeria—were forcibly relocated to the Americas. Nigerian populations, particularly those from Yoruba and Igbo ethnic groups, have significantly influenced the genetic background of African Americans and Afro-Caribbean populations, as demonstrated by comparative autosomal STR data (Tishkoff *et al*., 2009; Bryc *et al*., 2010).

This research underlines some forensic implications and highlights the need to consider population differences. The moderate differentiation among Nigerian ethnic groups suggests that population substructure should be regarded in forensic calculations, such as random match probabilities and likelihood ratios. Additionally, the Hardy-Weinberg and linkage equilibrium assumptions may not always be valid when combining these diverse Nigerian groups. While Hardy-Weinberg equilibrium may hold for individual subpopulations, combining them could still introduce bias related to population structure. The findings of the AMOVA analysis indicate that the majority of autosomal STR differentiation occurs within populations, with moderate differentiation observed among various populations and geographic regions. Y-STR markers help trace paternal ancestry because they are passed down from father to son and do not undergo recombination along most of the Y chromosome, which helps keep specific genetic traits associated with each lineage. They provide a more straightforward way to identify a male lineage and its evolution. The high resolution of Y-STRs enables the effective tracing of male lineages in various contexts, including criminal investigations, paternity testing, and ancestry research. Nigerian populations can help create reference Y-STR databases, which aid in distinguishing between suspects from different ethnic or regional backgrounds. Research findings support the inclusion of allele frequencies specific to the Igbo, Yoruba, and Hausa-Fulani populations in forensic STR databases, such as the Y-STR Haplotypic Reference Database (YHRD). This will improve quality assurance, forensic identification, population databasing, and standardisation as it enhances confidence during forensic investigation.

## 7.2 FUTURE WORK

This thesis, along with other published research, has contributed to the development of a national STR allele frequency database for the Igbo, Yoruba, and Hausa-Fulani ethnolinguistic groups. However, establishing population-specific autosomal and Y-STR allele frequency databases for other Nigerian ethnolinguistic groups, such as Ijaw, Benin, Tiv, Ibibio, and Nupe, would enhance the calculation of forensic random match probabilities and improve international casework collaborations. While Nigeria is home to over 300 ethnic groups, only a few—namely the Igbo, Yoruba, Hausa, and Fulani—have been genetically characterised. Future research should include minority and indigenous populations to capture the full spectrum of Nigeria's genetic diversity. This approach will curtail the likelihood of bias due to population differences and promote more inclusivity in forensic science. The application of other genetic markers, such as Rapidly Mutating Y-STRs (RM-YSTRs), can be helpful, especially in cases involving closely related male individuals. Incorporating such makers could be essential. RM-YSTRs mutate more often than standard Y-STRs, which can help to effectively discriminate against closely associated males, like

fathers, sons, brothers, and cousins. Using Y-STRs can help solve cases involving close male relatives in criminal investigations and paternity tests. It is also a good idea to look into microhaplotypes and SNP panels for tracing ancestry and identifying individuals in complex cases, especially those with degraded samples.

Future work should prioritise the study of female lineage and biogeographic inference. Mitochondrial DNA (mtDNA) sequencing can effectively trace maternal ancestry and improve our understanding of family history when used with Y-STR data. Future studies should utilise mtDNA sequencing in conjunction with Y-STR data to aid in tracing maternal family lines and historical migration paths. This combined approach can significantly support efforts to determine the origin of individuals, particularly in cases involving unidentified remains or missing persons. Using Next-Generation Sequencing (NGS) techniques can significantly enhance these efforts in forensic science. NGS allows for various analyses beyond short tandem repeats (STRs). It can simultaneously analyse autosomal STRs, X-STRs, Y-STRs, mitochondrial DNA (mtDNA) and single nucleotide polymorphisms (SNPs). This method is helpful for future investigations because it can identify unknown individuals, predict DNA traits, and untangle complex mixtures of DNA from several sources.

Additionally, it is advisable to focus on forensic validation and standardisation. Validating STR kits in local settings and aligning with international guidelines, such as ISO 17025 and ISFG, will promote effective forensic practices in Nigeria. To ensure their reliability, validation studies on STR kits, such as the GlobalFiler™ Express Kit and the Promega PowerPlex® Y23 System Kit, should be conducted under local laboratory conditions. Policy and capacity development must be taken seriously. The Nigerian legislature should prioritise advocacy for a national forensic infrastructure, training programs, and ethical guidelines to advance forensic genetic practice and research in Nigeria. The importance of creating a national forensic genetics institute or unit cannot be overemphasised. This should involve partnerships between the government and academic institutions. The government must also improve forensic DNA analysis, bioinformatics, and population genetics training programs. Ethical frameworks for sample collection, informed consent, and data protection in forensic and research contexts should also be promoted.

## REFERENCES


AbdEl-Hafez, A. F., El-Alfy, S. H., Swelim, H. H., & Hassan, N. H. (2019). Genetic variation at 15 autosomal STR loci among seven Egyptian populations. *Biochemical Genetics*, 57, 170-191. https://doi.org/10.1007/s10528-018-9879-0

Abdurrahman, U. (2012). Religion and language in the transformation of education in northern Nigeria during British colonial rule, 1900-1960. *Intellectual Discourse*, 20(2): 165-188. https://doi.org/10.31436/id.v20i2.299

Abimbola, K. (2006). *Yoruba Culture: A Philosophical Account*. Iroko Academic Publishers.

Aborisade, R. A., Adegoke, N., Adeleke, O. A., Ebobo, U. C., Ogunmefun, F. M., Chineyemba, L. I., & Adedayo, S. S. (2024). Policing rape and serious sexual offences in Nigeria: Officers' experiences and appraisal of police investigative approaches. *Police Practice and Research*, 25(3), 251-268. https://doi.org/10.1080/15614263.2023.2222870

Adebanwi, W., & Obadare, E. (2013). Introducing Nigeria at fifty: The nation in narration. In *Nigeria at Fifty* (pp. 1-28). Routledge.

Adebisi, T. A., & Oni, C. S. (2012). Assessment of relevance of the national directorate of employment (NDE) training programmes to the needs of the trainees in Southwestern Nigeria. *International Journal of Vocational and Technical Education*, 4(3), 29-37. https://academicjournals.org/journal/IJVTE/article-abstract/48904EC1250

Adeyemo, A. A., Chen, G., Chen, Y., & Rotimi, C. (2005). Genetic structure in four West African population groups. *BMC Genetics*, 6, 1-9. https://doi.org/10.1186/1471-2156-6-38

Agbaire, J. J., & Dunne, M. (2024). Nigerian higher education catchment policy: exclusions and the absent presence of ethnicity. *Higher Education Policy*, 37(4), 692-709. https://doi.org/10.1057/s41307-023-00324-1

Agbo, B. U., Ebuehi, O. A. T., & Osuntoki, A. A. (2017). Ethnic Origin of Crime Scene Evidential Materials Determination in Three Main Ethno-Linguistic Population Groups in Nigeria. *Annual Research & Review in Biology*, 12(4), 1-8. https://doi.org/10.9734/ARRB/2017/32783

Agbo, P. O. (2024). 'We shall triumph like the Jews': unveiling the implicit side of IPOB's armed separatism in Southeast Nigeria. *Security Journal*, 37(3), 1144-1163. https://doi.org/10.1057/s41284-023-00412-2

Ajibade, T. B., Ayinde, O. E., & Abdoulaye, T. (2019). Discovery of Maize Price and Food Crop Market Dynamics in Nigeria. *Review of Agricultural and Applied Economics (RAAE)*, 22(1340-2019-781), 51-64. https://doi.org/10.22004/ag.econ.285930

Akhiwu, O. W., & Obaseki, D. (2014). Knowledge of Nigerian policemen about modern forensics. *Annals of Biomedical Sciences,* 13(2), 79–83. https://www.ajol.info/index.php/abs/article/view/105613.

Akinyemi, A., Adedini, S., Hounton, S., Akinlo, A., Adedeji, O., Adonri, O., Friedman, H., Shiferaw, S., Maïga, A., Amouzou, A. & Barros, A.J., (2015). Contraceptive use and distribution of high-risk births in Nigeria: a sub-national analysis. *Global Health Action*, 8(1), 29745. https://doi.org/10.3402/gha.v8.29745

Akpan, U.O.U., Amusa, O.D., Adebesin, O.A., Adebajo, O.A., Akande, T.E., Onaja, P., Ayuba, P., Udoh, K., Akpan, H.B., Garba, N. & Abafras, F.P. (2024). Genetic Relationships and Population Structure among Nigerian Ethnic Groups (Ibibio, Igbo, Hausa, Tiv and Yoruba) Based on Nine DNA Loci. *Nigerian Journal of Basic and Clinical Sciences*, 1, 10-4103. https://doi.org/10.4103/njbcs.njbcs_68_24

Akpan-Idiok, A. U., & Ofem, K. I. (2014). Physicochemical Characteristics, Degradation Rate and Vulnerability Potential of Obudu Cattle Ranch Soils in Southeast Nigeria. *Open Journal of Soil Science*, 4(02), 57. https://doi.org/10.4103/njbcs.njbcs_68_24

Alberts, B., Bray, D., Hopkin, K., Johnson, A.D., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2015). *Essential Cell Biology*. Garland Science.

Alghafri, R. (2020). *Y chromosome short tandem repeats typing*. In P. Shrivastava, H. R. Dash, J. A. Lorente, & J. Imam (Eds.), *Forensic DNA typing: Principles, Applications and Advancements* (pp. 277–300). Springer.

Al-Snan, N. R., Messaoudi, S., R. Babu, S., & Bakhiet, M. (2019). Population genetic data of the 21 autosomal STRs included in GlobalFiler kit of a population sample from the Kingdom of Bahrain. *PloS One*, 14(8), e0220620. https://doi.org/10.1371/journal.pone.0220620

Alves, S. I. A., Dantas, C. W. D., Macedo, D. B., & Ramos, R. T. J. (2024). What are microsatellites and how to choose the best tool: a user-friendly review of SSR and 74 SSR mining tools. *Frontiers in Genetics*, 15, 1474611. https://doi.org/10.3389/fgene.2024.1474611

Amankwaa, A. O. (2020). Trends in forensic DNA database: transnational exchange of DNA data. *Forensic Sciences Research*, 5(1), 8-14. https://doi.org/10.1080/20961790.2019.1565651

Amankwaa, A. O., & McCartney, C. (2018). The UK national DNA database: implementation of the protection of freedoms act 2012. *Forensic Science International*, 284, 117-128. https://doi.org/10.1016/j.forsciint.2017.12.041

Amankwaa, A. O., & McCartney, C. (2019). The effectiveness of the UK national DNA database. *Forensic Science International: Synergy*, 1, 45-55. https://doi.org/10.1016/j.fsisyn.2019.03.004

Amelung, N., & Machado, H. (2019). Affected for good or for evil: The formation of issue-publics that relate to the UK National DNA Database. *Public Understanding of Science*, 28(5), 590-605. https://doi.org/10.1177/0963662519836346

Arauna, L.R., Mendoza-Revilla, J., Mas-Sandoval, A., Izaabel, H., Bekada, A., Benhamamouch, S., Fadhlaoui-Zid, K., Zalloua, P., Hellenthal, G. & Comas, D. (2016). Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. *Molecular Biology and Evolution*, 34(2), 318-329. https://doi.org/10.1093/molbev/msw218

Ardlie, K. G., Kruglyak, L., & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*, 3(4), 299-309. https://doi.org/10.1038/nrg777

Arowolo, D. (2010). The effects of western civilisation and culture on Africa. *Afro Asian Journal of Social Sciences*, 1(1), 1-13. http://onlineresearchjournals.com/aajoss/art/53.pdf

Aslam, M., Naeem, F., Seher, R., Shabbir, M. Z., Shehzad, W., & Imran, M. (2023). Effect of storage temperature and duration on direct PCR amplification of various feather types and DBS matrices. *Gene*, 854, 147116. https://doi.org/10.1016/j.gene.2022.147116

Babatunde, S. A. (2019*). Livestock Production at the Nexus of Resources Competition and Ethno-Religious Cynicism in Nigeria – Implicative Analysis on Food Security*. OSF Preprints. https://doi.org/10.31219/osf.io/rj2ct

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., & Wishart, D. S. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Research*, 44(W1), W147-W153. https://doi.org/10.1093/nar/gkw419

Bakare, M. O. (2015). Demography and Medical Education among Nigerian Final Year Medical Students-Implication for Regional and Human Resource Development. *Journal of Health Education Research & Development*, 3(3), 1-5. https://www.hilarispublisher.com/open-access/demography-and-medical-education-among-nigerian-final-year-medicalstudentsimplication-for-regional-and-human-resource-development-2380-5439-1000150.pdf

Bakhtiar, H. S. (2024). The Evolution of Scientific Evidence Theory in Criminal Law: A Transformative Insight. *Media Iuris*, 7(2), 221-240. https://doi.org/10.20473/mi.v7i2.51095

Balding, D. J. (1999). When can a DNA profile be regarded as unique? *Science & Justice: Journal of the Forensic Science Society*, 39(4), 257-260. https://doi.org/10.1016/S1355-0306(99)72057-5

Balding, D. J., & Donnelly, P. (1995a). Inference in forensic identification. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 158(1), 21-40. https://doi.org/10.2307/2983402

Balding, D. J., & Donnelly, P. (1995b). Inferring identify from DNA profile evidence. *Proceedings of the National Academy of Sciences*, 92(25), 11741-11745. https://doi.org/10.1073/pnas.92.25.11741

Balding, D. J., & Nichols, R. A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International*, 64(2-3), 125-140. https://doi.org/10.1016/0379-0738(94)90397-2

Balding, D. J., Greenhalgh, M., & Nichols, R. A. (1996). Population genetics of STR loci in Caucasians. *International Journal of Legal Medicine*, 108(6), 300-305. https://doi.org/10.1007/BF02432124

Ballantyne, K.N., Ralf, A., Aboukhalid, R., Achakzai, N.M., Anjos, M.J., Ayub, Q., Balažic, J., Ballantyne, J., Ballard, D.J., Berger, B. & Bobillo, C. (2014). Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. *Human Mutation*, 35(8), 1021-1032. https://doi.org/10.1002/humu.22564

Balloux, F., & Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, 11(2), 155-165. https://doi.org/10.1046/j.1365-294X.2002.01411.x

Bär, W., Brinkmann, B., Lincoln, P., Mayr, W. & Rossi, U. (1993). Editorial: Statement by DNA Commission of the International Society for Forensic Haemogenetics concerning the National Academy of Sciences report on DNA Technology in Forensic Science in the USA. *Forensic Science International*, 59(1), 1-2. https://doi.org/10.1016/0379-0738(93)90201-Y

Barbaro, A., Falcone, G., La Marca, A., & Barbaro, A. (2024). Use of the Investigator ESSplex SE QS Kit (QIAGEN) at half PCR Reaction volumes for the analysis of forensic samples. *Forensic Sciences*, 4(1), 152-163. https://doi.org/10.3390/forensics4010011

Barrio, P.A., Martín, P., Alonso, A., Müller, P., Bodner, M., Berger, B., Parson, W., Budowle, B. & Dnaseqex Consortium (2019). Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power. *Forensic Science International: Genetics*, 42, 49-55. https://doi.org/10.1016/j.fsigen.2019.05.006

Batibo, H. M. (2005). *Language Decline and Death in Africa: Causes, Consequences, and Challenges* (Vol. 132). Multilingual Matters.

Batini, C., Ferri, G., Destro-Bisol, G., Brisighelli, F., Luiselli, D., Sánchez-Diz, P., Rocha, J., Simonson, T., Brehm, A., Montano, V. & Elwali, N.E. (2011). Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Molecular Biology and Evolution*, 28(9), 2603-2613. https://doi.org/10.1093/molbev/msr033

Batini, C., Hallast, P., Zadik, D., Delser, P.M., Benazzo, A., Ghirotto, S., Arroyo-Pardo, E., Cavalleri, G.L., De Knijff, P., Dupuy, B.M. & Eriksen, H.A. (2015). Large-scale recent expansion of European patrilineages shown by population resequencing. *Nature Communications*, 6(1), 7152. https://doi.org/10.1038/ncomms8152

Bauer, M. (2007). RNA in forensic science. *Forensic Science International: Genetics*, 1(1), 69-74. https://doi.org/10.1016/j.fsigen.2006.12.006

Becker, N. S., Verdu, P., Froment, A., Le Bomin, S., Pagezy, H., Bahuchet, S., & Heyer, E. (2011). Indirect evidence for the genetic determination of short stature in African Pygmies. *American Journal of Physical Anthropology*, 145(3), 390-401. https://doi.org/10.1002/ajpa.21421

Bekada, A., Fregel, R., Cabrera, V. M., Larruga, J. M., Pestano, J., Benhamamouch, S., & González, A. M. (2013). Introducing the Algerian mitochondrial DNA and Y-chromosome profiles into the North African landscape. *PLoS One*, 8(2), e56775. https://doi.org/10.1371/journal.pone.0056775

Bellanova, R. (2008). *The "Prüm Process": The Way Forward for Police Cooperation and Data Exchange?* In E. Guild & F. Geyer (Eds.), *Security Versus Justice? Police and Judicial Cooperation in the European Union* (pp. 203–221). Ashgate

Bellanova, R. (2009). *Prüm: A Model "Prêt-à-Exporter"? The 2008 German–US Agreement on Data Exchange* (CEPS Challenge Paper No. 13). Centre for European Policy Studies. https://researchportal.vub.be/en/publications/a-model-pr%C3%AAt-%C3%A0-exporter-the-2008-germanus-agreement-on-data-excha

Bello, O.B., Ganiyu, O.T., Wahab, M.K.A., Afolabi, M.S., Oluleye, F., Mahmud, J., Azeez, M.A. & Abdulmaliq, S.Y. (2012). Evidence of climate change impacts on agriculture and food security in Nigeria. *International Journal of Agriculture and Forestry*, 2(2), 49-55. https://doi.org/10.5923/j.ijaf.20120202.08

Bender, M. L. (2000). "Nilo-Saharan." In B. Heine & D. Nurse (Eds.), *African Languages: An Introduction* (pp. 43–73). Cambridge University Press.

Bhardwaj, J., Goyal, K., Malsawmzuali, C., & Narula, A. (2025). Revolutionizing Forensic Science: The Role of Artificial Intelligence in Evidence Analysis. *International Journal of Interdisciplinary Approaches in Psychology*, *3*(1), 101-116. https://doi.org/10.61113/ijiap.v3i1.643

Bienen, H. (2013). *Political Conflict and Economic Change in Nigeria*. Routledge.

Binuomoyo, O. K. (2016). Socio-Political Reform and Entity Identity: Nigeria's Struggles towards Stability. *International Journal of Law and Peace Works,*1 (1), 1-10. https://www.kwpublisher.com/paper/socio-political-reform-and-entity-identity-nigeria-struggles-towards-stability

Blakemore, B., & Blake, C. (2012). Can the national DNA database be effective and comply with human rights legislation? *The Police Journal*, 85(3), 191-202. https://doi.org/10.1350/pojo.2012.85.3.573

Bodner, M., Bastisch, I., Butler, J.M., Fimmers, R., Gill, P., Gusmão, L., Morling, N., Phillips, C., Prinz, M., Schneider, P.M. & Parson, W. (2016). Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). *Forensic Science International: Genetics*, 24, 97-102. https://doi.org/10.1016/j.fsigen.2016.06.008

Bondarenko, D. M., & Roese, P. M. (1999). Benin prehistory: The origin and settling down of the Edo. *Anthropos*, (H. 4. /6), 542-552. https://www.jstor.org/stable/40465021

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics*, 18, 78-89. https://doi.org/10.1016/j.fsigen.2015.02.002

Bossart, J. L., & Prowell, D. P. (1998). Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution*, 13(5), 202-206. https://doi.org/10.1016/S0169-5347(98)01372-9

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), 314. https://doi.org/10.1016/0002-9297(80)90018-1

Bozzo, W. R., Canteros, M., Rabitti, L., Furman, N., Gagliardi, F., & Piñero, M. H. (2019). Y-STR chromosome structure variations: Incidence and implications. *Forensic Science International: Genetics Supplement Series*, 7(1), 591-593. https://doi.org/10.1016/j.fsigss.2019.10.102

Bradbury, R. E. (2017). *The Benin Kingdom and the Edo-Speaking Peoples of South-Western Nigeria: Western Africa Part XIII*. Routledge. (Original work published 1957) https://doi.org/10.4324/9781315293851

Brisighelli, F., Capelli, C., Boschi, I., Garagnani, P., Lareu, M. V., Pascali, V. L., & Carracedo, A. (2009). Allele frequencies of fifteen STRs in a representative sample of the Italian population. *Forensic Science International: Genetics*, 3(2), e29-e30. https://doi.org/10.1016/j.fsigen.2008.05.002

Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe, C., Tishkoff, S.A. & Bustamante, C.D. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences*, 107(2), 786-791. https://doi.org/10.1073/pnas.0909559107

Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe, C., Tishkoff, S.A. & Bustamante, C.D. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences*, 107(2), 786-791. https://doi.org/10.1073/pnas.0909559107

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *The American Journal of Human Genetics*, 96(1), 37-53. https://doi.org/10.1016/j.ajhg.2014.11.010

Buckleton, J., Curran, J., Goudet, J., Taylor, D., Thiery, A., & Weir, B. S. (2016). Population-specific FST values for forensic STR markers: A worldwide survey. *Forensic Science International: Genetics*, *23*, 91-100. https://doi.org/10.1016/j.fsigen.2016.03.004

Budowle, B. (1995). Technical working group on DNA analysis methods. *Guidelines for a Quality Assurance Program for DNA Analysis. Crime Lab Digest*, 22, 21-43. https://www.ojp.gov/pdffiles1/Digitization/153914NCJRS.pdf

Budowle, B., Moretti, T. R., Baumstark, A. L., Defenbaugh, D. A., & Keys, K. M. (1999). Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *Journal of Forensic Science*, 44(6), 1277-1286. https://doi.org/10.1520/JFS14601J

Budowle, B., Onorato, A.J., Callaghan, T.F., Manna, A.D., Gross, A.M., Guerrieri, R.A., Luttman, J.C. & McClure, D.L. (2009). Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences*, 54(4), 810-821. https://doi.org/10.1111/j.1556-4029.2009.01046.x

Bullock, K., Hadfield, A., Keningale, P., Mallett, E., Tong, S., & Wellings, F. (2025). United Kingdom–European Union policing and law enforcement cooperation in the post-Brexit era. *Policing and Society*, 35, 1–14. https://doi.org/10.1080/10439463.2024.2442714

Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences*, 51(2), 253-265. https://doi.org/10.1111/j.1556-4029.2006.00046.x

Butler, J. M. (2009). *Fundamentals of Forensic DNA Typing*. Academic Press.

Butler, J. M. (2011). *Advanced Topics in Forensic DNA Typing: Methodology*. Academic Press.

Butler, J. M. (2014). *Advanced Topics in Forensic DNA Typing: Interpretation*. Academic Press.

Butler, J. M. (2015a). *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*. (2nd ed.). Elsevier Academy Press.

Butler, J. M. (2015b). The future of forensic DNA analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1674), 20140252. https://doi.org/10.1098/rstb.2014.0252

Butler, J. M. (2023). Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019–2022. *Forensic Science International: Synergy*, 6, 100311. https://doi.org/10.1016/j.fsisyn.2023.100311

Byard, R. W., James, H., Berkata, J. & Health, K. (2016). Locard's principle of exchange, dental examination, and fragments of skin. *Journal of Forensic Sciences*, 61(2), 545-547. https://doi.org/10.1111/1556-4029.12970

Calafell, F., & Comas, D. (2021). The Y Chromosome. In *Evolution of the Human Genome II: Human Evolution Viewed from Genomes* (pp. 121-136). Tokyo: Springer Japan.

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics.*, 9(1), 403-433. https://doi.org/10.1146/annurev.genom.9.081307.164258

Campbell, M. C., Hirbo, J. B., Townsend, J. P., & Tishkoff, S. A. (2014). The peopling of the African continent and the diaspora into the new world. *Current Opinion in Genetics & Development*, 29, 120-132. https://doi.org/10.1016/j.gde.2014.09.001

Carey, L., & Mitnik, L. (2002). Trends in DNA forensic analysis. *Electrophoresis*, 23(10), 1386-1397. https://doi.org/10.1002/1522-2683(200205)23:10<1386::AID-ELPS1386>3.0.CO;2-M

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

Chakraborty, R. (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Human Biology*, 64(2),141-159. https://www.jstor.org/stable/41464266

Chibuike, U. C., & Eme, O. I. (2019). Terrorism & its Socio-Economic Effects in Nigeria. *Journal of Contemporary Research in Social Sciences*, 1(1), 97-113. https://learning-gate.com/index.php/2641-0249/article/view/29/26

Chidozie, A. K., & Isaac, A. (2015). Genetic diversity between two Igbo men from Owerri senatorial province as determined by autosomal short tandem repeats, Y-chromosomal short tandem repeats and mitochondrial DNA typing methods. *Nigerian Journal of Experimental and Clinical Biosciences*, 3(1), 29-35. https://doi.org/10.4103/2348-0149.158164

Chien, A., Edgar, D. B., & Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *Journal of Bacteriology*, 127(3), 1550-1557. https://doi.org/10.1128/jb.127.3.1550-1557.1976

Chiluwa, I., & Chiluwa, I. M. (2020). Separatists or terrorists? Jews or Nigerians? Media and cyber discourses on the complex identity of the "Biafrans". *Journal of Language and Politics*, 19(4), 583-603. https://doi.org/10.1075/jlp.19041.chi

Colaco, S., & Modi, D. (2018). Genetics of the human Y chromosome and its association with male infertility. *Reproductive Biology and Endocrinology*, 16, 1-24. https://doi.org/10.1186/s12958-018-0330-5

Cole, S. A. (2013). *De-neutralizing Identification: S. & Marper v. United Kingdom, Biometric Databases, Uniqueness, Privacy and Human Rights.* In J. Bennett Moses, B. McSherry, A. Norrie, & S. Bronitt (Eds.), *Identification and Registration Practices in Transnational Perspective* (pp. 77–97). Palgrave Macmillan.

Cole-Showers, C. L. (2014). Population structure and demographics in Nigerian populations utilizing Y-chromosome markers. Ph.D. Thesis, Department of Biotechnology, University of the Western Cape, South Africa. http://hdl.handle.net/11394/5326

Collins, F. S., & Mansoura, M. K. (2001). The human genome project: revealing the shared inheritance of all humankind. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 91(S1), 221-225. https://doi.org/10.1002/1097-0142(20010101)91:1+<221::AID-CNCR8>3.0.CO;2-9

Cooper, G. M., & Adams, K. (2022). *The Cell: A Molecular Approach*. Oxford University Press.

Cotton, E.A., Allsop, R.F., Guest, J.L., Frazier, R.R.E., Koumi, P., Callow, I.P., Seager, A. & Sparkes, R.L. (2000). Validation of the AMPFlSTR® SGM Plus™ system for use in forensic casework. *Forensic Science International*, 112(2-3), 151-161. https://doi.org/10.1016/S0379-0738(00)00182-1

Council of the European Union. (2016). *Implementation of the Provisions on Information Exchange of the Prüm Decisions: Overview of Documents and Procedures, Overview of Declarations, and State of Play of Implementation of Automated Data Exchange*. https://www.eumonitor.eu/9353000/1/j9vvik7m1c3gyxp/vk3t7m2qejz7

Council of the European Union (2019). *Outcome of the Council Meeting*. Justice and Home Affairs 3697th Council Meeting. http://www.statewatch.org/news/2019/jun/eu-jha-council-6-jun.prel.pdf

Crow, J. F. (2017). *An Introduction to Population Genetics Theory*. Scientific Publishers.

Cruciani, F., La Fratta, R., Trombetta, B., Santolamazza, P., Sellitto, D., Colomb, E.B., Dugoujon, J.M., Crivellaro, F., Benincasa, T., Pascone, R. & Moral, P. (2007). Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Molecular Biology and Evolution*, 24(6), 1300-1311. https://doi.org/10.1093/molbev/msm049

Curtin, D. E. (2017). *Brexit and the EU Area of Freedom, Security, and Justice: Bespoke Bits and Pieces.* In F. Fabbrini (Ed.), *The Law & Politics of Brexit* (pp. 183–201). Oxford University Press. https://academic.oup.com/book/26188/chapter/194316971

D'Amato, M. E., Ehrenreich, L., Cloete, K., Benjeddou, M., & Davison, S. (2010). Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. *Forensic Science International: Genetics*, 4(2), 104-110. https://doi.org/10.1016/j.fsigen.2009.06.011

Dagba, B. I., Sambe, L. N., & Shomkegh, S. A. (2013). Totemic beliefs and biodiversity conservation among the Tiv People of Benue State, Nigeria. *Journal of Natural Sciences Research*, 3(3), 145-149. https://www.iiste.org/Journals/index.php/JNSR/article/view/6919/7016

Dash, H. R., Elkins, K. M., & Al-Snan, N. R. (2023). *Role of Forensic DNA Databases in Criminal Identification.* In P. Shrivastava & J. A. Lorente (Eds.), *Advancements in Forensic DNA Analysis* (pp. 119–127). Singapore: Springer Nature Singapore.

Dash, H. R., Shrivastava, P., & Das, S. (2020). *Principles and Practices of DNA Analysis: A Laboratory Manual for Forensic DNA Typing*. Humana Press.

De Filippo, C., Barbieri, C., Whitten, M., Mpoloka, S.W., Gunnarsdóttir, E.D., Bostoen, K., Nyambe, T., Beyer, K., Schreiber, H., de Knijff, P. and Luiselli, D. (2011). Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Molecular Biology and Evolution*, 28(3),1255-1269. https://doi.org/10.1093/molbev/msq312

Dedrickson, K. (2017). Universal DNA databases: a way to improve privacy? *Journal of Law and the Biosciences*, 4(3), 637. https://doi.org/10.1093/jlb/lsx041

Deinert, L., Hossen, S., Ikoyi, I., Kwapinksi, W., Noll, M., & Schmalenberger, A. (2024). Poultry litter biochar soil amendment affects microbial community structures, promotes phosphorus cycling and growth of barley (Hordeum vulgare). *European Journal of Soil Biology*, 120, 103591. https://doi.org/10.1016/j.ejsobi.2023.103591

Delémont, O., Bitzer, S., Jendly, M., & Ribaux, O. (2017). *The Practice of Crime Scene Examination in An Intelligence-Based Perspective*. In Q. Rossy, D. Décary-Hétu, O. Delémont & M. Mulone (Eds.), *The Routledge International Handbook of Forensic Intelligence and Criminology* (pp. 86–101). Routledge.

Delicado, P., & Pachón-García, C. (2024). Multidimensional scaling for big data. *Advances in Data Analysis and Classification*, 1-22. Available from: https://doi.org/10.1007/s11634-024-00591-9

Destro-Bisol, G., Donati, F., Coia, V., Boschi, I., Verginelli, F., Caglia, A., Tofanelli, S., Spedini, G. & Capelli, C. (2004). Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Molecular Biology and Evolution*, 21(9), 1673-1682. https://doi.org/10.1093/molbev/msh177

Deubel, T. F. (2020). Ethnography of the Western Sahara: An Overview of Anthropological Research in Sahrawi Communities. *L'Ouest Saharien*, 12(2), 33-53. https://doi.org/10.3917/ousa.202.0033

Dimmendaal, G. J. (2008). "Language Ecology and Linguistic Diversity on the African Continent." *Language and Linguistics Compass*, 2(5), 840–858. https://doi.org/10.1111/j.1749-818X.2008.00085.x

Direct Amplification of DNA using the Investigator® ESSPlex SE QS Kit (February 2021) Available from: https://www.qiagen.com/us/resources/download.aspx?id=00c548f6-e1e8-40dc-8ecd-d8f506fea449&lang=en

Dirks, E., & Leibold, J. (2020). *Genomic Surveillance: Inside China's DNA Dragnet*. Australian Strategic Policy Institute. https://xjdp.aspi.org.au/explainers/genomic-surveillance/

Doleac, J. L. (2017). The effects of DNA databases on crime. *American Economic Journal: Applied Economics*, 9(1), 165-201. https://doi.org/10.1257/app.20150043

Doveri, S., Lee, D., Maheswaran, M., & Powell, W. (2024). *Molecular Markers — History, Features and Applications*. In *Principles and Practices of Plant Genomics,* Vol. 1 (pp. 23–67). CRC Press. https://doi.org/10.1201/9781003579298-2

Du, W., Peng, Z., Feng, C., Zhu, B., Wang, B., Wang, Y., Liu, C. and Chen, L. (2017). Forensic efficiency and genetic variation of 30 InDels in Vietnamese and Nigerian populations. *Oncotarget*, 8(51), p.88934. https://doi.org/10.18632/oncotarget.21494

Durnal, E. W. (2010). Crime scene investigation (as seen on TV). *Forensic Science International*, 199(1-3), 1-5. https://doi.org/10.1016/j.forsciint.2010.02.015

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359-361. https://doi.org/10.1007/s12686-011-9548-7

Edwards, A. L., Hammond, H. A., Jin, L., Caskey, C. T., & Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, 12(2), 241-253. https://doi.org/10.1016/0888-7543(92)90371-X

Ehret, C. (1995). *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary* (Vol. 126). Univ of California Press.

Ejiofor, P. F. (2022). Jewishness without Jews? Ontological security, ethnonationalism, and the social power of analogical reasoning in postcolonial Nigeria. *Nationalities Papers*, 1, 1-29. https://doi.org/10.1017/nps.2022.70

Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, 12(1), 14683. https://doi.org/10.1038/s41598-022-14395-4

Elkins, K. M. (2012). *Forensic DNA Biology: A Laboratory Manual*. Academic Press.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5(6), 435. https://doi.org/10.1038/nrg1348

Elmrghni, S., Dixon, R. A., Coulson-Thomas, Y. M., & Williams, D. R. (2012). Genetic data provided by 15 autosomal STR loci in the Libyan population living in Benghazi. *Forensic Science International: Genetics*, 6(3), e93-e94. https://doi.org/10.1016/j.fsigen.2011.07.006

Erbaş, R. (2017). The tension between genome privacy and criminal justice in the wake of DNA databases. *Ceza Hukuku ve Kriminoloji Dergisi/Journal of Penal Law & Criminology*, 5(2), 163-178. https://doi.org/10.26650/JPLC360271

Esoimeme, E. E. (2019). Using the lie detector test to curb corruption in the Nigerian Police Force. *Journal of Financial Crime*, 26(3), 874-880. https://doi.org/10.1108/JFC-06-2018-0058

Etebong, P. C. (2018). Demography in Nigeria: Problems and Prospects. *Biostatistics and Biometrics Open Access Journal*, 5(1): 1-5. https://doi.org/10.19080/BBOAJ.2018.05.555654

Etin-Osa, D., & Etin-Osa, C. E. (2019). Forensic Science and The Nigerian Society. *Journal of Nuclear Sciences*, 6(1), 17-21. https://doi.org/10.1501/nuclear.2023.49

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Evett, I. W., & Weir, B. S. (1998). *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists.* Sunderland, MA: Sinauer Associates.

Evett, I. W., Gill, P. D., Scrange, J. K., & Weir, B. S. (1996). Establishing the robustness of short-tandem-repeat statistics for forensic applications. *American Journal of Human Genetics*, 58(2), 398. PMID: 8571967; PMCID: PMC1914534

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1, 47–50. https://doi.org/10.1177/117693430500100003

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479-491. https://doi.org/10.1093/genetics/131.2.479

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A. & Tress, M.L. (2014). Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Human Molecular Genetics*, 23(22), 5866-5878. https://doi.org/10.1093/hmg/ddu309

Fadare, D. A. (2009). Modelling of solar energy potential in Nigeria using an artificial neural network model. *Applied Energy*, 86(9), 1410-1422. https://doi.org/10.1016/j.apenergy.2008.12.005

Fadhlaoui-Zid, K., Haber, M., Martínez-Cruz, B., Zalloua, P., Benammar Elgaaied, A., & Comas, D. (2013). Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS One*, 8(11), e80293. https://doi.org/10.1371/journal.pone.0080293

Fafunwa, A. B. (2018). *History of Education in Nigeria*. Routledge.

Fakorede, S. T., Adekoya, K. O., Akpan, U. U., & Ogunlusi, O. V. (2019). Allele frequencies and haplotype diversities of five Y-chromosome short tandem repeat loci in a random sample of Yoruba population in Lagos, Nigeria. *FUW Trends in Science and Technology*, 4(2): 577-581. https://ir.unilag.edu.ng/handle/123456789/4337

Fakorede, S. T., Adekoya, K. O., Iroanya, O. O., Kasu, M., & D'Amato, M. E. (2024). Forensic parameters of 10 Y-STR markers in the Nigerian Hausa, Igbo, and Yoruba populations. *UMYU Scientifica*, 3(1), 103-112. https://doi.org/10.56919/usci.2431.012

Falola, T., & Aderinto, S. (2010). *Nigeria, Nationalism, and Writing History* (Vol. 46). University Rochester Press.

Falola, T., & Ezekwem, O. (Eds.). (2016). *Writing the Nigeria-Biafra War*. Boydell & Brewer.

Falola, T., & Heaton, M. M. (2008). *A History of Nigeria*. Cambridge University Press.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567-1587. https://doi.org/10.1093/genetics/164.4.1567

Fandade, V., Singh, P., Singh, D., Sharma, H., Thakur, G., Saini, S., Kumar, P., Mantri, S., Bishnoi, O.P. & Roy, J. (2024). Genome-wide identification of microsatellites for mapping, genetic diversity and cross-transferability in wheat (*Triticum* spp). *Gene*, 896, 148039. https://doi.org/10.1016/j.gene.2023.148039

Federal Bureau of Investigation. (2024). *CODIS—NDIS statistics*. FBI Laboratory Services. https://le.fbi.gov/science-and-lab/biometrics-and-fingerprints/codis/codis-ndis-statistics

Fernandes, A. T., Gonçalves, R., Gomes, S., Filon, D., Nebel, A., Faerman, M., & Brehm, A. (2011). Y-chromosomal STRs in two populations from Israel and the Palestinian Authority Area: Christian and Muslim Arabs. *Forensic Science International: Genetics*, 5(5), 561-562. https://doi.org/10.1016/j.fsigen.2010.08.005

Ferrara, L. (2024). *Examining DNA databases.* In *Ethical Reasoning in Forensic Science* (pp. 45–60). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-58392-6_5

Fiodorova, A. (2014). DNA for crime investigation: European co-operation model. *Recent Advances in DNA & Gene Sequences (Formerly Recent Patents on DNA & Gene Sequences)*, 8(2), 126-133. https://doi.org/10.2174/2352092209666150218003440

Fisher, R. A. (1949). *Theory of Inbreeding*: Oliver and Boyd.

Fisher, R. A. (1951). Standard calculations for evaluating a blood-group system. *Heredity*, 5(1), 95-102. https://doi.org/10.1038/hdy.1951.5

Flores, S., Sun, J., King, J., & Budowle, B. (2014). Internal validation of the GlobalFiler™ Express PCR Amplification Kit for the direct amplification of reference DNA samples on a high-throughput automated workflow. *Forensic Science International: Genetics*, 10, 33-39. https://doi.org/10.1016/j.fsigen.2014.01.005

Foreman, L. A., & Evett, I. W. (2001). Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system. *International Journal of Legal Medicine*, 114(3), 147-155. https://doi.org/10.1007/s004140000138

Foreman, L. A., & Lambert, J. A. (2000). Genetic differentiation within and between four UK ethnic groups. *Forensic Science International*, 114(1), 7-20. https://doi.org/10.1016/S0379-0738(00)00271-1

Foreman, L. A., Champod, C., Evett, I. W., Lambert, J. A., & Pope, S. (2003). Interpreting DNA evidence: A review. *International Statistical Review*, 71(3), 473-495. https://doi.org/10.1111/j.1751-5823.2003.tb00207.x

Foreman, L. A., Lambert, J. A., & Evett, I. W. (1998). Regional genetic variation in Caucasians. *Forensic Science International*, 95(1), 27-37. https://doi.org/10.1016/S0379-0738(98)00079-6

Forouzesh, M., Irani, S., Soleimani, A., & Monabati, S. J. (2022). Application of Y-STR, DIP-STR and SNP-STR markers in interpretation of forensic genetic profiling: A narrative review. *Iranian Journal of Public Health,* 51(7), 1538–1545. https://doi.org/10.18502/ijph.v51i7.10087

Fortes-Lima, C. A., Diallo, M. Y., Janoušek, V., Černý, V., & Schlebusch, C. M. (2025). Population history and admixture of the Fulani people from the Sahel. *The American Journal of Human Genetics*, 112(2), 261-275. https://doi.org/10.1016/j.ajhg.2024.12.015

Fortes-Lima, C., Brucato, N., Croze, M., Bellis, G., Schiavinato, S., Massougbodji, A., Migot-Nabias, F. and Dugoujon, J.M. (2015). Genetic population study of Y-chromosome markers in Benin and Ivory Coast ethnic groups. *Forensic Science International: Genetics*, 19, 232-237. https://doi.org/10.1016/j.fsigen.2015.07.021

Frajzyngier, Z., & Shay, E. (2019). *The Afroasiatic Languages*. Cambridge University Press.

Frederiks, M. T. (2010). Let Us Understand Our Differences: Current Trends in Christian-Muslim Relations in Sub-Saharan Africa. *Transformation*, 27(4), 261-274. https://doi.org/10.1177/0265378810378562

Fregel, R., Méndez, F.L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M.D., Santana, J., Morales, J., Ávila-Arcos, M.C., Underhill, P.A., Shapiro, B. & Wojcik, G. (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proceedings of the National Academy of Sciences*, 115(26), 6774-6779. https://doi.org/10.1073/pnas.1800851115

Fujii, K., Watahiki, H., Mita, Y., Iwashima, Y., Miyaguchi, H., Kitayama, T., Nakahara, H., Mizuno, N. & Sekiguchi, K. (2016). Next-generation sequencing analysis of off-ladder alleles due to migration shift caused by sequence variation at D12S391 locus. *Legal Medicine*, 22, 62-67. https://doi.org/10.1016/j.legalmed.2016.08.003

Geppert, M., Edelmann, J., & Lessig, R. (2009). The Y-chromosomal STRs DYS481, DYS570, DYS576 and DYS643. *Legal Medicine*, 11, S109-S110. https://doi.org/10.1016/j.legalmed.2009.01.063

Gettings, K. B., Aponte, R. A., Vallone, P. M., & Butler, J. M. (2015). STR allele sequence variation: current knowledge and future issues. *Forensic Science International: Genetics*, 18, 118-130. https://doi.org/10.1016/j.fsigen.2015.06.005

Gill, P. (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *International Journal of Legal Medicine*, 114(4), 204-210. https://doi.org/10.1007/s004149900117

Gill, P. (2002). Role of short tandem repeat DNA in forensic casework in the UK—past, present, and future perspectives. *Biotechniques*, 32(2), 366-385. https://doi.org/10.2144/02322rv01

Gill, P., & Evett, I. (1995). *Population Genetics of Short Tandem Repeat (STR) Loci.* In B. S. Weir (Ed.), *Human Identification: The Use of DNA Markers* (pp. 69–87). Springer, Dordrecht. https://doi.org/10.1007/978-0-306-46851-3_9

Gill, P., Bleka, Ø., Hansson, O., Benschop, C., & Haned, H. (2020). *Forensic Practitioner's Guide to the Interpretation of Complex DNA Profiles*. Academic Press.

Gill, P., Fereday, L., Morling, N., & Schneider, P. M. (2006). The evolution of DNA databases—recommendations for new European STR loci. *Forensic Science International*, 156(2-3), 242-244. https://doi.org/10.1016/j.forsciint.2005.05.036

Gill, P., Werrett, D. J., Budowle, B., & Guerrieri, R. (2004). An assessment of whether SNPs will replace STRs in national DNA databases-joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). *Science & Justice: Journal of the Forensic Science Society*, 44(1), 51-53. https://doi.org/10.1016/S1355-0306(04)71685-8

Gillespie, J. H. (2004). *Population Genetics: A Concise Guide*. JHU Press.

Gjertson, D.W., Brenner, C.H., Baur, M.P., Carracedo, A., Guidet, F., Luque, J.A., Lessig, R., Mayr, W.R., Pascali, V.L., Prinz, M. & Schneider, P.M. (2007). ISFG: recommendations on biostatistics in paternity testing. *Forensic Science International: Genetics*, 1(3-4), 223-231. https://doi.org/10.1016/j.fsigen.2007.06.006

GlobalFiler™ Express PCR Amplification Kit User Guide (2020). Available at: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0030237-GlobalFilerExpressPCRAmpKit-UG.pdf

Gomes, C., Amorim, A., Okolie, V.O., Keshinro, S.O., Starke, A., Vullo, C., Gusmão, L. & Gomes, I. (2019). Genetic insight into Nigerian population groups using an X-chromosome decaplex system. *Forensic Science International: Genetics Supplement Series*, *7*(1), 501-503. https://doi.org/10.1016/j.fsigss.2019.10.067

Good, B. H. (2022). Linkage disequilibrium between rare mutations. *Genetics*, 220 (4), iyac004. https://doi.org/10.1093/genetics/iyac004

Good, J. (2017). *Niger-Congo Languages.* In R. Hickey (Ed.), *The Cambridge Handbook of Areal Linguistics* (pp. 471–499). Cambridge University Press. https://doi.org/10.1017/9781107279872.018

Gouveia, M.H., Borda, V., Leal, T.P., Moreira, R.G., Bergen, A.W., Kehdy, F.S., Alvim, I., Aquino, M.M., Araujo, G.S., Araujo, N.M. and Furlan, V. (2020). Origins, admixture dynamics, and homogenization of the African gene pool in the Americas. *Molecular Biology and Evolution*, 37(6), 1647-1656. https://doi.org/10.1093/molbev/msaa033

Gouy, A., & Zieger, M. (2017). STRAF—a convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics*, 30, 148-151. https://doi.org/10.1016/j.fsigen.2017.07.007

Graham, E. A. (2007). DNA reviews: the national DNA database of the United Kingdom. *Forensic Science, Medicine, and Pathology*, 3, 285-288. https://doi.org/10.1007/s12024-007-9014-8

Greely, H. T., Riordan, D. P., Garrison, N. A., & Mountain, J. L. (2006). Family ties: the use of DNA offender databases to catch offenders' kin. *The Journal of Law, Medicine & Ethics*, 34(2), 248-262. https://doi.org/10.1111/j.1748-720X.2006.00031.x

Guanglin, H., Lan-Hai, W., & Mengge, W. (2023). Forensic investigative genetic genealogy and fine-scale structure of human populations. *Frontiers in Genetics*, 13, 1067865. https://doi.org/10.3389/fgene.2022.1067865

Güldemann, T. (2014). '*Khoisan' Linguistic Classification Today.* In T. Güldemann & A.-M. Fehn (Eds.), *Beyond 'Khoisan': Historical relations in the Kalahari Basin* (pp. 1–40). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.330.01gul

Guo, F., Shen, H., Tian, H., Jin, P., & Jiang, X. (2014). Development of a 24-locus multiplex system to incorporate the core loci in the Combined DNA Index System (CODIS) and the European Standard Set (ESS). *Forensic Science International: Genetics*, 8(1), 44-54. https://doi.org/10.1016/j.fsigen.2013.07.007

Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48, 361-372. https://doi.org/10.2307/2532296

Guyon, L., Guez, J., Toupance, B., Heyer, E., & Chaix, R. (2024). Patrilineal segmentary systems provide a peaceful explanation for the post-Neolithic Y-chromosome bottleneck. *Nature Communications*, 15(1), 3243. https://doi.org/10.1038/s41467-024-47618-5

Haddish, K., Chierto, E., Di Vella, G., Lacerenza, D., Raddi, S., Aneli, S., Bogale, A.L., Kidane, E., Yizengaw, A., Getaneh, Y. & Tewelemedhin, G. (2022). A reference database of forensic autosomal and gonosomal STR markers in the Tigray population of Ethiopia. *Forensic Science International: Genetics*, 56, 102618. https://doi.org/10.1016/j.fsigen.2021.102618

Hallast, P., Ebert, P., Loftus, M., Yilmaz, F., Audano, P.A., Logsdon, G.A., Bonder, M.J., Zhou, W., Höps, W., Kim, K. & Li, C. (2023). Assembly of 43 human Y chromosomes reveals extensive complexity and variation. *Nature*, 621(7978), .355-364. https://doi.org/10.1038/s41586-023-06425-6

Hameed, I. H., Ommer, A. J., Murad, A. F., & Mohammed, G. J. (2015). Allele frequency data of 21 autosomal short tandem repeat loci in Mesan and Basra provinces in South Iraq. *Egyptian Journal of Forensic Sciences*, *5*(4), 150-156. https://doi.org/10.1016/j.ejfs.2014.10.003

Hammer, M. F. (1995). A recent common ancestry for human Y chromosomes. *Nature*, 378(6555), 376-378. https://doi.org/10.1038/378376a0

Harrel, M., Mayes, C., Houston, R., Holmes, A. S., Gutierrez, R., & Hughes, S. (2021). The performance of quality controls in the Investigator® Quantiplex® Pro RGQ and Investigator® 24plex STR kits with a variety of forensic samples. *Forensic Science International: Genetics*, *55*, 102586. https://doi.org/10.1016/j.fsigen.2021.102586

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J. & Kern, R. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

Hartl, D. L. & Clark, A. G. (1997). *Principles of Population Genetics* (Vol. 116). Sunderland, MA: Sinauer Associates.

Hartshorne, D., Roeder, A., Elsmore, P., McDonald, A., & Greenham, J. (2024). *The Challenges of Introducing Massively Parallel Sequencing into the UK Forensic Market.* In S. Francese & R. S. P. King (Eds.), *Driving Forensic Innovation in the 21st Century: Crossing the Valley of Death* (pp. 219–237). Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-56556-4_10

Hayward, R. J. (2000). "*Afroasiatic*." In B. Heine & D. Nurse (Eds.), *African Languages: An Introduction* (pp. 74–98). Cambridge University Press.

Hedrich, H. (Ed.). (2004). *The Laboratory Mouse*. Academic Press.

Hedrick, P. W. (1987). Gametic disequilibrium measures: proceed with caution. *Genetics*, 117(2), 331-341. https://doi.org/10.1093/genetics/117.2.331

Heine, B., & Nurse, D. (Eds.). (2000). *African Languages: An Introduction*. Cambridge University Press.

Henke, J., Henke, L., Chatthopadhyay, P., Kayser, M., Dulmer, M., Cleef, S., Poche, H. & Felske-Zech, H. (2001). Application of Y-chromosomal STR haplotypes to forensic genetics. *Croatian Medical Journal*, 42(3), 292-297.

Henn, B.M., Botigué, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J. & Bustamante, C.D. (2012). Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genetics*, 8(1), .e1002397. https://doi.org/10.1371/journal.pgen.1002397

Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A. & Lin, A.A. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proceedings of the National Academy of Sciences*, 108(13), 154-5162. https://doi.org/10.1073/pnas.1017511108

Herrera-Paz, E. F., García, L. F., Aragon-Nieto, I., & Paredes, M. (2008). Allele frequencies distributions for 13 autosomal STR loci in 3 Black Carib (Garifuna) populations of the Honduran Caribbean coasts. *Forensic Science International: Genetics*, 3(1), e5-e10. https://doi.org/10.1016/j.fsigen.2008.02.004

Hill, C. R., Duewer, D. L., Kline, M. C., Coble, M. D., & Butler, J. M. (2013). US population data for 29 autosomal STR loci. *Forensic Science International: Genetics*, *7*(3), e82-e83. https://doi.org/10.1016/j.fsigen.2012.12.004

Hill, C. R., Duewer, D. L., Kline, M. C., Sprecher, C. J., McLaren, R. S., Rabbach, D. R., Krenke, B. E., Ensenberger, M. G., Fulmer, P. M., Storts, D. R. & Butler, J. M. (2011). Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. *Forensic Science International: Genetics*, 5(4), 269-275. https://doi.org/10.1016/j.fsigen.2010.03.014

Hill, C. R., Kline, M. C., Mulero, J. J., Lagace, R. E., Chang, C. W., Hennessy, L. K., & Butler, J. M. (2007). Concordance study between the AmpFℓSTR® MiniFilerTM PCR amplification kit and conventional STR typing kits. *Journal of Forensic Sciences*, 52(4), 870-873. https://doi.org/10.1111/j.1556-4029.2007.00491.x

Hiskett, M. (1994). *The Sword of Truth: The Life and Times of the Shehu Usuman dan Fodio*. Northwestern University Press.

Hodgson, J. A., Mulligan, C. J., Al-Meeri, A., & Raaum, R. L. (2014). Early back-to-Africa migration into the Horn of Africa. *PLoS Genetics*, 10(6), e1004393. https://doi.org/10.1371/journal.pgen.1004393

Hoelzel, A. R., Goldsworthy, S. D., & Fleischer, R. C. (2002). *Population Genetic Structure*. In: Hoelzel, A. R. (Ed.), *Marine Mammal Biology: An Evolutionary Approach*, (pp. 325-352) John Wiley & Sons.

Hohoff, C., Schürenkamp, M., & Brinkmann, B. (2009). Meiosis study in a population sample from Nigeria: allele frequencies and mutation rates of 16 STR loci. *International Journal of Legal Medicine*, 123(3), 259-261. https://doi.org/10.1007/s00414-008-0307-6

Holland, B. S. (2000). Genetics of marine bioinvasions. *Hydrobiologia*, 420(1), 63-71.

Hollfelder, N., Schlebusch, C. M., Günther, T., Babiker, H., Hassan, H. Y., & Jakobsson, M. (2017). Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations. *PLoS Genetics*, 13(8), e1006976. https://doi.org/10.1371/journal.pgen.1006976

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating, and interpreting $F_{ST}$. *Nature Reviews Genetics*, 10(9), 639-650. https://doi.org/10.1038/nrg2611

Home Office (2015) *Prüm Business and Implementation Case*. UK Government https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/

Home Office. (2024a). *Forensic Information Databases Annual Report 2023–2024*. UK Government. https://www.gov.uk/government/publications/forensic-information-databases-annual-report-2023-to-2024/forensic-information-databases-annual-report-2023-to-2024-accessible

Home Office. (2024b). *National DNA Database Statistics*. UK Government. https://www.gov.uk/government/statistics/national-dna-database-statistics

Hopen, C. E. (2018). *The Pastoral Fulbe Family in Gwandu*. Routledge.

Hopwood, A. J., Hurth, C., Yang, J., Cai, Z., Moran, N., Lee-Edghill, J. G., Nordquist, A., Lenigk, R., Estes, M. D., Haley, J. P. & McAlister, C. R. (2010). Integrated microfluidic system for rapid forensic DNA analysis: sample collection to DNA profile. *Analytical Chemistry*, 82(16), 6991-6999. https://doi.org/10.1021/ac101355r

Hopwood, A. J., Puch-Solis, R., Tucker, V. C., Curran, J. M., Skerrett, J., Pope, S. & Tully, G. (2012). Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts. *Science & Justice*, 52(3), 185-190. https://doi.org/10.1016/j.scijus.2012.05.005

Huang, S., Jin, X., Zhang, H., Jin, H., Ren, Z., Wang, Q., Liu, Y., Ji, J., Yang, M., Zhang, H. & Zheng, X. (2022). Developmental validation of the novel five-dye-labeled multiplex autosomal STR panel and its forensic efficiency evaluation. *Frontiers in Genetics*, 13, .897650. https://doi.org/10.3389/fgene.2022.897650

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(03), 90-95. https://doi.org/10.1109/MCSE.2007.55

Hyman, L. M., Rolle, N., Sande, H., Clem, E., Jenks, P. S. E., Lionnet, F., Merrill, J., & Baier, N. (2019). *Niger-Congo Linguistic Features and Typology.* In H. E. Wolff (Ed.), *The Cambridge Handbook of African Linguistics* (pp. 191–245). Cambridge University Press. https://doi.org/10.1017/9781108283991.009

Ibelema, M. (2000). *Nigeria: The Politics of Marginalization. Current History,* 99(637), 211–214. https://doi.org/10.1525/curh.2000.99.637.211

Igiede, A. I. (2013). *Cultural Division among Diasporic Nigerians: A Systemic Dilemma. Race, Gender & Class, 20*(1/2), 344–352. https://doi.org/10.2979/reseafrilite.50.1.07

Institute for Economics & Peace. (2019). *Global Peace Index 2019: Measuring Peace in a Complex World*. http://visionofhumanity.org/reports

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature,* 409, 860–921. https://doi.org/10.1038/35057062

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945. https://doi.org/10.1038/nature03001

International Monetary Fund. (2024). *World Economic Outlook: Global Economy and Regional Trends*. https://www.imf.org/en/Publications/WEO

Investigator® ESSplex SE QS Handbook (February 2021). For multiplex amplification of the European standard set of loci, plus SE33 and Amelogenin. Available from: https://www.qiagen.com/us/resources/download.aspx?id=1fce64fd-9cf1-4750-a4d0-e7aa64113d7b&lang=en

Investigator® Quantiplex Handbook. For quantification of human and male DNA in forensic samples. (June 2018). Available from: https://www.qiagen.com/us/resources/download.aspx?id=57497d59-7a43-4eaf-8c94-086e88742e86&lang=en

Inyang, A. A., & Bassey, M. E. (2014). Imperial treaties and the origins of British colonial rule in Southern Nigeria, 1860-1890. *Mediterranean Journal of Social Sciences*, 5(20), 1946-1953. https://doi.org/10.5901/mjss.2014.v5n20p1946

Iorliam, A. (2018). Impact of Forensic Science and Bodies that Need Forensic Science in Nigeria. In *Fundamental Computing Forensics for Africa* (pp. 85-91). Springer, Cham.

Jacewicz, R. (2018). Standardization of research in the field of forensic genetics–the current state in the world and introduction to the guidelines in Poland with reference to the work of the Expert Team for Standards and Assessment in Forensic Genetics of the Polish Speaking Working Group of the International Society for Forensic Genetics (ISFG-PL). *Archiwum Medycyny Sądowej i Kryminologii/Archives of Forensic Medicine and Criminology*, 68(4), 215-231. https://doi.org/10.5114/amsik.2018.84531

Jain, M., Kalsi, A.K., Kumar, P., Halder, A. (2017). *The Human Y Chromosome*. In: Kumar, A., Sharma, M. (eds) *Basics of Human Andrology*. (pp. 77-98) Springer, Singapore.

James, S. H., & Nordby, J. J. (2002). *Forensic Science: An Introduction to Scientific and Investigative Techniques*. CRC Press.

Jeffreys, A. J., Brookfield, J. F., & Semeonoff, R. (1985a). Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, 317(6040), 818-819. https://doi.org/10.1038/317818a0

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985b). Individual-specific 'fingerprints' of human DNA. *Nature*, 316(6023), 76-79. https://doi.org/10.1038/316076a0

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985c). Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314(6006), 67-73. https://doi.org/10.1038/314067a0

Jiang, B., Qu, W., Wang, F., Zhang, L., Rong, H., Li, J., Wen, D., Zeye, M. M. J., He, W., Wang, C., Xing, H., Zhang, T., Jin, C., Chen, L., Liu, Y., Cai, J. & Zha, L. (2021). Development and validation of novel 8-dye short tandem repeat multiplex system for forensic applications. *International Journal of Legal Medicine*, 135, 2263-2274. https://doi.org/10.1007/s00414-021-02695-9

Jin, Y., Schäffer, A. A., Sherry, S. T., & Feolo, M. (2017). Quickly identifying identical and closely related subjects in large databases using genotype data. *PloS one*, 12(6), e0179106. https://doi.org/10.1371/journal.pone.0179106

Joannin, P. (2017). *Elements of Differentiation Within the Schengen Acquis and the Prüm Convention.* In: *Flexibility in the EU and Beyond* (pp. 123-138). Nomos Verlagsgesellschaft mbH & Co. KG.

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, 5(10), 739. https://doi.org/10.1038/nrg1455

Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics*, 4(8), 598-612. https://doi.org/10.1038/nrg1124

Jobling, M. A., & Tyler-Smith, C. (2017). Human Y-chromosome variation in the genome-sequencing era. *Nature Reviews Genetics*, 18(8), 485-497. https://doi.org/10.1038/nrg.2017.36

Johnson, B. T. (2024). Advances in postmortem fingerprinting: Applications in disaster victim identification. *Journal of Forensic Sciences*, 69(5), 1681-1689. https://doi.org/10.1111/1556-4029.15513

Johnson, P. & William, R. (2004b). DNA and Crime Investigation: Scotland and the UK National Database. *Scottish Journal of Criminal Justice Studies: The Journal of the Scottish Association for the Study of Delinquency*, 10, 71-84. https://pubmed.ncbi.nlm.nih.gov/16557290/

Johnson, P., & Williams, R. (2004a). Post-conviction DNA testing: the UK's first 'exoneration'case? *Science & Justice: Journal of the Forensic Science Society*, 44(2), 77. https://doi.org/10.1016/S1355-0306(04)71692-5

Kanaka, K.K., Sukhija, N., Goli, R.C., Singh, S., Ganguly, I., Dixit, S.P., Dash, A. & Malik, A.A. (2023). On the concepts and measures of diversity in the genomics era. *Current Plant Biology*, 33, 100278. https://doi.org/10.1016/j.cpb.2023.100278

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L. & Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research*, 18(5), 830-838. https://doi.org/10.1101/gr.7172008

Kaye, J. (2006). Police collection and access to DNA samples. *Genomics, Society and Policy*, 2(1), 16-27. https://www.lancaster.ac.uk/fss/journals/gsp/vol2no1/kayeabstract.htm

Kayser, M. (2017). Forensic use of Y-chromosome DNA: a general overview. *Human Genetics*, 136(5), 621-635. https://doi.org/10.1007/s00439-017-1776-9

Kayser, M., Brauer, S., Schädlich, H., Prinz, M., Batzer, M.A., Zimmerman, P.A., Boatin, B.A. & Stoneking, M. (2003). Y chromosome STR haplotypes and the genetic structure of US populations of African, European, and Hispanic ancestry. *Genome Research*, 13(4), 624-634. https://doi.org/10.1101/gr.463003

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., & Honda, K. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *International Journal of Legal Medicine*, 110, 125-133. https://doi.org/10.1007/s004140050051

Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A.C., Mohyuddin, A., Mehdi, S.Q., Rosser, Z., Stoneking, M., Jobling, M.A. & Sajantila, A. (2004). A comprehensive survey of human Y-chromosomal microsatellites. *The American Journal of Human Genetics*, 74(6), 1183-1197. https://doi.org/10.1086/421531

Keefe, C. (2024). *Molecular Methods in Hominin Sex Estimation* (Master's Thesis). The George Washington University. Retrieved from: https://www.proquest.com/openview/66d7c2db5ad1dab8274dcffd70548f18/1?pq-origsite=gscholar&cbl=18750&diss=y

Kelty, S. F., Joshua, P. R., & Robertson, J. (2024). Building on the critical skills framework of top crime scene examiners to recruit high-caliber crime scene investigation candidates. *Wiley Interdisciplinary Reviews: Forensic Science*, e1519. https://doi.org/10.1002/wfs2.1519

Kierkegaard, S. (2008). The Prüm decision–An uncontrolled fishing expedition in 'Big Brother'Europe. *Computer Law & Security Review*, 24(3), 243-252. https://doi.org/10.1016/j.clsr.2008.03.002

Kimpton, C. P., Gill, P., Walton, A., Urquhart, A., Millican, E. S., & Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *Genome Research*, 3(1), 13-22. https://doi.org/10.1101/gr.3.1.13

Kitause, R. H., & Achunike, H. C. (2013). Religion in Nigeria from 1900-2013. *Religion*, 3(18), 45-57. https://www.iiste.org/Journals/index.php/RHSS/article/view/8950

Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V. and Gölge, M. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *The American Journal of Human Genetics*, 72(2), 313-332. https://doi.org/10.1086/346068

Kobilinsky, L. F., Liotti, T. F., & Oeser-Sweat, J. (2005). *DNA: Forensic and Legal Applications* (p. 155). Wiley-Interscience.

Kockum, I., Huang, J., & Stridh, P. (2023). Overview of genotyping technologies and methods. *Current Protocols*, 3(4), e727. https://doi.org/10.1002/cpz1.727

Kofi, A.E., Hakim, H.M., Khan, H.O., Ismail, S.A., Ghansah, A., Haslindawaty, A.R.N., Shamsuddin, S., Aziz, M.Y., Chambers, G.K. & Edinur, H.A. (2020). Population dataset for 21 simple tandem repeat loci in the Akan population of Ghana. *Data in Brief*, 31, 105746. https://doi.org/10.1016/j.dib.2020.105746

Kohler, N. S. (2023). What are Jews: interrogating genetic studies and the reification of race. *Journal of Anthropological Sciences= Rivista di Antropologia: JASS,* 101, 185-199. https://doi.org/10.4436/JASS.10001

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179-1191. https://doi.org/10.1111/1755-0998.12387

Krambrich, J., Bringeland, E., Hesson, J. C., Hoffman, T., Lundkvist, Å., Lindahl, J. F., & Ling, J. (2022). Usage of FTA® Classic Cards for Safe Storage, Shipment, and Detection of Arboviruses. *Microorganisms*, 10(7), 1445. https://doi.org/10.3390/microorganisms10071445

Kumar, L. S. (1999). DNA markers in plant improvement: an overview. *Biotechnology Advances*, 17(2-3), 143-182. https://doi.org/10.1016/S0734-9750(98)00018-4

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Kumar, S., Tamura, K., & Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics*, 10(2), 189-191. https://doi.org/10.1093/bioinformatics/10.2.189

Kutranov, S. (2011). Comparison study of four different 16-locus "expanded ESS" STR kits. *Forensic Science International: Genetics Supplement Series*, 3(1), e196-e197. https://doi.org/10.1016/j.fsigss.2011.08.098

Larnane, A., Pierlé, S.A., Letexier, M., Gibert, J., Soucies, C., Santucci, J., Ghosh, D., Hubac, S., Hermitte, F. & Deleuze, J.F. (2024). An innovative approach for low input forensic DNA sample analysis using the GlobalFiler™ IQC PCR amplification Kit on the Magelia® platform. *Forensic Science International: Genetics*, 72,103093. https://doi.org/10.1016/j.fsigen.2024.103093

Leat, N., Ehrenreich, L., Benjeddou, M., Cloete, K., & Davison, S. (2007). Properties of novel and widely studied Y-STR loci in three South African populations. *Forensic Science International*, 168(2-3), 154-161. https://doi.org/10.1016/j.forsciint.2006.07.009

Lee, D. (2004). Report on the current activities of the Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee. *Forensic Science Communications,* 6(3). https://www.ojp.gov/ncjrs/virtual-library/abstracts/report-current-activities-scientific-working-group-dna-analysis

Lewontin, R. C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics,* 49(1), 49–67. https://doi.org/10.1093/genetics/49.1.49

Li, C. (2018). Forensic Genetics. *Forensic Science Research*, 3(2), 103-104. https://doi.org/10.1080/20961790.2018.1489445

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. & Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319(5866), 1100-1104. https://doi.org/10.1126/science.1153717

Li, N., Yang, Y., Xu, F., Chen, X., Wei, R., Li, Z., Pan, W. & Zhang, W. (2022). Genetic diversity and population structure analysis of Castanopsis hystrix and construction of a core collection using phenotypic traits and molecular markers. *Genes*, 13(12), 2383. https://doi.org/10.3390/genes13122383

Linacre, A. M. T. (2003). The UK national DNA database. *Lancet*, *361*(9372), 1841-1842. https://doi.org/10.1016/S0140-6736(03)13539-8

Lipson, M., Sawchuk, E.A., Thompson, J.C., Oppenheimer, J., Tryon, C.A., Ranhorn, K.L., De Luna, K.M., Sirak, K.A., Olalde, I., Ambrose, S.H. and Arthur, J.W. (2022). Ancient DNA and deep population structure in sub-Saharan African foragers. *Nature*, 603(7900), 290-296. https://doi.org/10.1038/s41586-022-04430-9

Lokken, P. (2013). From the "Kingdoms of Angola" to Santiago de Guatemala: The Portuguese Asientos and Spanish Central America, 1595–1640. *Hispanic American Historical Review*, 93(2), 171-203. https://doi.org/10.1215/00182168-2077126

Lovejoy, P. E. (2016). *Jihād in West Africa During the Age of Revolutions*. Ohio University Press.

Lygo, J.E., Johnson, P.E., Holdaway, D.J., Woodroffe, S., Kimpton, C.P., Gill, P., Whitaker, J.P. & Clayton, T.M. (1994). The validation of short tandem repeat (STR) loci for use in forensic casework. *International Journal of Legal Medicine*, 107(2), 77-89. https://doi.org/10.1007/BF01225493

Machado, H., & Granja, R. (2020). Forensic Genetics and Governance of Transnational Criminality. In *Forensic Genetics in the Governance of Crime* (pp. 71-84). Palgrave Pivot, Singapore.

Maguire, C. N., McCallum, L. A., Storey, C., & Whitaker, J. P. (2014). Familial searching: A specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives—The UK experience. *Forensic Science International: Genetics*, 8(1), 1-9. https://doi.org/10.1016/j.fsigen.2013.07.00

Malhotra, S., & Jamir, L. (2024). Application of Forensic Biology in Criminal Justice System. In *Forensic Justice* (pp. 367-386). Routledge.

Margoliash, J. B. (2024). *Ubiquitous Genome-Wide Variation at Short Tandem Repeats is Causally Linked to Changes in Gene Expression, Blood Cell Counts and Serum Biomarkers in Human Populations* (Doctoral Dissertation). University of California, San Diego. https://escholarship.org/uc/item/4tf2589s

Martín, P., de Simón, L. F., Luque, G., Farfán, M. J., & Alonso, A. (2014). Improving DNA data exchange: validation studies on a single 6 dye STR kit with 24 loci. *Forensic Science International: Genetics*, 13, 68-78. https://doi.org/10.1016/j.fsigen.2014.07.002

Martinez, B., Catelli, L., Romero, M., Okolie, V. O., Keshinro, S. O., Carvalho, E. F., Vullo, C. & Gusmão, L. (2017). Forensic evaluation of 27 y-str haplotypes in a population sample from nigeria. *Forensic Science International: Genetics Supplement Series*, 6, e289-e291. https://doi.org/10.1016/j.fsigss.2017.09.138

Martire, K. A., Kemp, R. I., Watkins, I., Sayle, M. A., & Newell, B. R. (2013). The expression and interpretation of uncertain forensic science evidence: Verbal equivalence, evidence strength, and the weak evidence effect. *Law and Human Behavior*, 37(3), 197. https://doi.org/10.1037/lhb0000027

McCartney, C. (2004). Forensic DNA sampling and the England and Wales National DNA Database: a sceptical approach. *Critical Criminology*, 12(2), 157-178. https://doi.org/10.1023/B:CRIT.0000040255.29101.7a

McCartney, C. (2005). The DNA expansion programme and criminal investigation. *British Journal of Criminology*, 46(2), 175-192. https://doi.org/10.1093/bjc/azi094

McCartney, C. (2013). *Forensic Identification and Criminal Justice*. Willan Publishing.

McDonald, J., & Lehman, D. C. (2012). Forensic DNA analysis. *Clinical Laboratory Science*, 25(2), 109. https://doi.org/10.29074/ascls.25.2.109

McKiernan, H. E., & Danielson, P. B. (2017). Molecular diagnostic applications in forensic science. In *Molecular Diagnostics* (pp. 371-394). Academic Press.

McKinney, W. (2010). Data structures for statistical computing in Python. *SciPy*, 445(1), 51-56. https://doi.org/10.25080/Majora-92bf1922-00a

Mehar, P. A., Bhoyar, L. Z., & Mahakalkar, A. L. (2024). Integration of Bite Mark Microbiome Analysis with Forensic DNA Profiling: Advancements, Challenges, and Synergistic Approaches. *Rambam Maimonides Medical Journal*, 15(3), e0014. https://doi.org/10.5041/RMMJ.10528

Mertens, G., Rand, S., Moordtgat, K., Cardoen, E., De Bruyn, I., Mommers, N., Leijnen, G., Jehaes, E. & Jacobs, W. (2011). Population genetic analysis of Moroccans residing in Belgium using 16 autosomal STRs of the PowerPlex ESI 17 multiplex. *Forensic Science International: Genetics*, 5(4), 352-353. https://doi.org/10.1016/j.fsigen.2010.01.008

Millo, T., Jaiswa, A. K., & Behera, C. (2008). Collection, preservation and forwarding of biolgical samples for toxicological analysis in medicolegal autopsy cases: A review. *Journal of Indian Academic Forensic Medicine*, 30(2), 96-100. https://doi.org/10.1177/0971097320080211

Mnookin, J. L. (2001). Fingerprint evidence in an age of DNA profiling. *Brooklyn Law Review,* 67(1), 13–70. https://brooklynworks.brooklaw.edu/blr/vol67/iss1/3

Muñoz, H. S., & Fiodorova, A. (2014). DNA and law enforcement in the European Union: Tools and human rights protection. *Utrecht Law Review***,** 10(1), 149–162. https://doi.org/10.18352/ulr.26222

Myers, B. A., King, J. L., & Budowle, B. (2012). Evaluation and comparative analysis of direct amplification of STRs using PowerPlex® 18D and Identifiler® Direct systems. *Forensic Science International: Genetics*, 6(5), 640-645. https://doi.org/10.1016/j.fsigen.2012.02.005

National Population Commission (NPC), Federal Ministry of Health and Social Welfare of Nigeria, & ICF. (2024). *Nigeria Demographic and Health Survey 2023–24: Key Indicators Report* (PR157). Abuja, Nigeria, and Rockville, MD, USA: NPC, FMoHSW, & ICF. https://dhsprogram.com/pubs/pdf/PR157/PR157.pdf

National Research Council, (1997). *Evaluating Human Genetic Diversity*. National Academy Press, Washington DC.

National Research Council. (1992). DNA typing: statistical basis for interpretation. In *DNA Technology in Forensic Science*. National Academies Press (US).

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, 106(949), 283-292. https://doi.org/10.1086/282771

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.

Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, 98(3), 625-640. https://doi.org/10.1093/genetics/98.3.625

Neuvonen, A. (2017). *Finnish Population Genetics in a Forensic Context*. An Academic Dissertation Presented to the Faculty of Medicine, University of Helsinki, Finland.

Newman, P. (2009). Hausa and the Chadic languages. In *The Major Languages of South Asia, the Middle East and Africa*. 2nd ed. (pp. 162-177). Routledge.

NicDaeid, N., & White, P. C. (Eds.). (2024). *Crime Scene to Court: The Essentials of Forensic Science*. Royal Society of Chemistry.

Norrgard, K. (2008). Forensics, DNA fingerprinting, and CODIS. *Nature Education*, 1(1), 35. https://www.nature.com/scitable/topicpage/forensics-dna-fingerprinting-and-codis-736

Nte, N. D. (2012). An evaluation of the challenges of forensic investigation and unsolved murders in Nigeria. *African Journal of Criminology and Justice Studies: AJCJS*, 6(1/2), 143–162. https://digitalscholarship.tsu.edu/ajcjs/vol6/iss1/9/

Nydick, C. (2009). The British invasion (of privacy): DNA databases in the United Kingdom and United States in the wake of the Marper case. *Emory International Law Review,* 23, 609–650.

Obafunwa, J. O., Faduyile, F. A., Soyemi, S. S., Eze, U. O., Nwana, E. J., & Odesanmi, W. O. (2015). Forensic investigation of mass disasters in Nigeria: A review. *Nigerian Medical Journal: Journal of the Nigeria Medical Association*, 56(1), 1–5. https://doi.org/10.4103/0300-1652.149162

Obleščuk, I., Makar, A., & Ledić, A. (2024). Forensic DNA Database Management. In Primorac, D. (Ed.). *Forensic Science and Molecular Anthropology - Topics Selected from 12th ISABS Conference on Forensic and Anthropological Genetics.* (pp. 53-64) IntechOpen.

Ogbaa, K. (2003). *The Nigerian Americans*. Greenwood Publishing Group.

Ogundiran, A. (2009). Material life and domestic economy in a frontier of the Oyo Empire during the mid-Atlantic age. *The International Journal of African Historical Studies*, 42(3), 351–385. http://www.jstor.org/stable/40646774

Okolie, V. O., Cisana, S., Schanfield, M. S., Adekoya, Oyedeji, O. A. & Podini, D. (2018). Population data of 21 autosomal STR loci in Hausa, Igbo and Yoruba people of Nigeria. *International Journal of Legal Medicine*, 132, 735-737. https://doi.org/10.1007/s00414-017-1722-3

Okonkwo, C. O., & Smith, A. A. (2018). Personality and Ethnicity as Psychosocial Factors in Understanding Corruption in Nigeria. *Practicum Psychologia*, 8(1), 76-98. https://journals.aphriapub.com/index.php/PP/article/view/583

Olasupo, F. A. (2014). Black African Jews, the Nigerian Question and the Lost Ten Tribes of Israel: A Comparison of Igbo and Yoruba Claims to Jewish and Judaic Traditions. *OIDA International Journal of Sustainable Development*, 7(04), 49-62. https://ssrn.com/abstract=2490087

Omaka, A. O. (2014). The forgotten victims: Ethnic minorities in the Nigeria-Biafra war, 1967-1970. *Journal of Retracing Africa*, 1(1), 25-40. https://encompass.eku.edu/jora/vol1/iss1/2

Onimisi, T. (2014). The Politics of State Creation in Nigeria and the Economic Viability of the Existing 36 States. *International Journal of Social Sciences and Management*, 1(2), 64-68. https://doi.org/10.3126/ijssm.v1i2.10009

Onyeiwu, S. (2024). Population, Youth Bulge and Economic Development. In *Emerging Issues in Contemporary African Economies* (pp. 123-141). Palgrave Macmillan, Cham.

Osaghae, E. E., & Suberu, R. T. (2005). *A History of Identities, Violence and Stability in Nigeria* (Vol. 6). Oxford: Centre for Research on Inequality, Human Security and Ethnicity, University of Oxford.

Osinuga, O. O. O. (2015). *Nigeria's Sexual Offences Bill 2013, Matters Arising*. (July 21, 2015). Available from SSRN: https://ssrn.com/abstract=2634134

Otu, N., & Elechi, O. O. (2018). The Nigeria police forensic investigation failure. *Journal of Forensic Science and Criminal Investigation*, 9(1), 1-8. https://doi.org/10.19080/JFSCI.2018.09.555752

Oweibia, M., Elemuwa, U.G., Akpan, E., Daniel, E.T., Oruikor, G.J., Tarimobowei, E., Okoho, E.E., Elemuwa, C.O., Raimi, M.O. & Babatunde, A.A. (2024). Analyzing Nigeria's Journey Towards Sustainable Development Goals: A Comprehensive Review from Inception to Present. *F1000Research*, 13, 984. https://doi.org/10.12688/f1000research.148020.1

Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Romero, I.G., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D. & Bekele, E. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *The American journal of Human Genetics*, 91(1), 83-96. https://doi.org/10.1016/j.ajhg.2012.05.015

Park, H. C., Ahn, E. R., & Shin, S. C. (2021). Comparative analysis of allele variation using allele frequencies according to sample size in Korean population. *Genes & Genomics*, 43(11), 1301-1305. https://doi.org/10.1007/s13258-021-01159-z

Park, S. J., Kim, J. Y., Yang, Y. G., & Lee, S. H. (2008). Direct STR amplification from whole blood and blood-or saliva-spotted FTA® without DNA purification. *Journal of Forensic Sciences*, 53(2), 335-341. https://doi.org/10.1111/j.1556-4029.2008.00666.x

Parks, S. A. (2000). Compelled DNA Testing in Rape Caes: Illustrating the Necessity of an Exception to the Self-Incrimination Clause. *The William & Mary Journal of Women and the Law*, 7, 499. https://scholarship.law.wm.edu/wmjowl/vol7/iss2/7/

Parson, W., Ballard, D., Budowle, B., Butler, J.M., Gettings, K.B., Gill, P., Gusmão, L., Hares, D.R., Irwin, J.A., King, J.L. and de Knijff, P. (2016). Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics*, 22,.54-63. https://doi.org/10.1016/j.fsigen.2016.01.009

Pascali, V. L., Dobosz, M., & Brinkmann, B. (1998). Coordinating Y-chromosomal STR research for the Courts. *International Journal of Legal Medicine*, 112, 1-1. https://doi.org/10.1007/s004140050188

Paschou, P., Drineas, P., Lewis, J., Nievergelt, C.M., Nickerson, D.A., Smith, J.D., Ridker, P.M., Chasman, D.I., Krauss, R.M. & Ziv, E. (2008). Tracing sub-structure in the European American population with PCA-informative markers. *PLoS Genetics*, 4(7), e1000114. https://doi.org/10.1371/journal.pgen.1000114

Paschou, P., Ziv, E., Burchard, E. G., Choudhry, S., Rodriguez-Cintron, W., Mahoney, M. W., & Drineas, P. (2007). PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genetics*, 3(9), e160. https://doi.org/10.1371/journal.pgen.0030160

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Pazoki, N., Salehi, M., Angaji, S. A., & Abdollahpour-Alitappeh, M. (2024). Elucidating the impact of Y chromosome microdeletions and altered gene expression on male fertility in assisted reproduction. *Human Molecular Genetics*, 33(17), 1540-1553. https://doi.org/10.1093/hmg/ddae086

Peakall, R. O. D., & Smouse, P. E. (2006). GenAIEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6(1), 288-295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall, R..O. D., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pedro, J. C. and Campos, A. D. (2103). The Treaty of Prüm and Unauthorised Migration. In. *Social Control and Justice* (p.317). Eleven International Publishing, The Netherlands.

Perlin, M. W. (2010, December). Explaining the likelihood ratio in DNA mixture interpretation. In *Promega's Twenty First International Symposium on Human Identification*. Retrieved from: https://www.promega.ee/~/media/files/resources/conference%20proceedings/ishi%2021/oral%20presentations/perlin.pdf

Perry, G.H., Foll, M., Grenier, J.C., Patin, E., Nédélec, Y., Pacis, A., Barakatt, M., Gravel, S., Zhou, X., Nsobya, S.L. & Excoffier, L. (2014). Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. Proceedings of the National Academy of Sciences, 111(35),.E3596-E3603. https://doi.org/10.1073/pnas.1402875111

Petr, M., Hajdinjak, M., Fu, Q., Essel, E., Rougier, H., Crevecoeur, I., Semal, P., Golovanova, L.V., Doronichev, V.B., Lalueza-Fox, C., & De la Rasilla, M. (2020). The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science*, 369(6511), 1653-1656. https://doi.org/10.1126/science.abb6460

Pierce, B. A. (2012). *Genetics: A Conceptual Approach*. Macmillan.

Plourd, C. J. (2010). *Science, the Law, and Forensic Identification*. In: *Forensic Dentistry*, Second Edition (eds. Senn, D. R. and Stimson, P. G.), CRC Press, Boca Raton.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of *STRUCTURE*: applications, parameter settings, and supporting software. *Frontiers in genetics*, 4, 98. https://doi.org/10.3389/fgene.2013.00098

Pray, L. A. (2008). Discovery of DNA structure and function: Watson and Crick. *Nature Education*, 1(1), 100. https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397

Primorac, D., & Schanfield, M. (Eds.). (2023). *Forensic DNA Applications: An Interdisciplinary Perspective*. CRC Press.

Pritchard, J. K., & Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics*, 69(1), 1-14. https://doi.org/10.1086/321275

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959. https://doi.org/10.1093/genetics/155.2.945

Prohaska, A., Racimo, F., Schork, A.J., Sikora, M., Stern, A.J., Ilardo, M., Allentoft, M.E., Folkersen, L., Buil, A., Moreno-Mayar, J.V. & Korneliussen, T. (2019). Human disease variation in the light of population genomics. *Cell*, 177(1), 115-131. https://doi.org/10.1016/j.cell.2019.01.052

Prokop, J. W., & Deschepper, C. F. (2015). Chromosome Y genetic variants: impact in animal models and on human disease. *Physiological Genomics*, 47(11), 525-537. https://doi.org/10.1152/physiolgenomics.00074.2015

Promega Technical Manual (2023). *PowerPlex® Y23 System for Use on the Applied Biosystems® Genetic Analyzers*: Instructions for Use of Products DC2305 and DC2320. Available from: https://www.promega.co.uk/-/media/files/resources/protocols/technical-manuals/tmd/powerplex-y23-system-protocol.pdf?rev=61cdeda3ef334beea013630d90f13cf4&sc_lang=en

Puch-Solis, R., & Pope, S. (2021). Interpretation of DNA data within the context of UK forensic science—evaluation. *Emerging Topics in Life Sciences*, 5(3), 405-413. https://doi.org/10.1042/ETLS20200340

Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., Angustia, S.M., Santos, L.H., Anslinger, K., Bayer, B. and Ayub, Q. (2014). A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International: Genetics*, 12, 12-23. https://doi.org/10.1016/j.fsigen.2014.04.008

Python Software Foundation. (2024). *Python Language Reference (version 3.12)*. https://docs.python.org/3/reference/

QIAamp® DNA Investigator Handbook: For purification of total (genomic and mitochondrial) DNA from surface and buccal swabs, FTA® and Guthrie cards, body fluid stains, chewing gum, cigarette butts, nail clippings, hair, paper and similar materials, small volumes of blood or saliva, tissues, laser-microdissected specimens, bones, teeth, and sexual assault specimens (January 2020). Available from: https://www.qiagen.com/us/resources/download.aspx?id=26ef8f2c-7c2a-49e6-b2d2-39d4e130b3cc&lang=en

Ralf, A., Lubach, D., Kousouri, N., Winkler, C., Schulz, I., Roewer, L., Purps, J., Lessig, R., Krajewski, P., Ploski, R. & Dobosz, T. (2020). Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers. *Human Mutation*, 41(9), 1680-1696. https://doi.org/10.1002/humu.24068

Rands, C. M., Meader, S., Ponting, C. P., & Lunter, G. (2014). 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genetics*, 10(7), e1004525. https://doi.org/10.1371/journal.pgen.1004525

Rashid, M.N.A., Mahat, N.A., Khan, H.O., Wahab, R.A., Maarof, H., Ismail, D., Alwi, A.R. & SyedHassan, S.R. (2020). Population data of 21 autosomal STR loci in Malaysian populations for human identification. *International Journal of Legal Medicine*, 134, 1675-1678. https://doi.org/10.1007/s00414-020-02279-z

Rath, K., Rothe, J., Saulich, F., Nagy, M., & Melisch, C. (2024). Genetic analysis of multiple burials from the medieval churchyard from the St. Peters church of old Cölln/Berlin. *Journal of Archaeological Science: Reports*, 53, 104337. https://doi.org/10.1016/j.jasrep.2023.104337

Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S. & Altshuler, D. (2002). Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genetics*, 32(1), 135-142. https://doi.org/10.1038/ng947

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489-494. https://doi.org/10.1038/nature08365

Resque, R., Gusmao, L., Geppert, M., Roewer, L., Palha, T., Alvarez, L., Ribeiro-dos-Santos, A. & Santos, S. (2016). Male lineages in Brazil: Intercontinental admixture and stratification of the European background. *PLoS One*, 11(4), .e0152573. https://doi.org/10.1371/journal.pone.0152573

Rhie, A., Nurk, S., Cechova, M., Hoyt, S.J., Taylor, D.J., Altemose, N., Hook, P.W., Koren, S., Rautiainen, M., Alexandrov, I.A. & Allen, J. (2023). The complete sequence of a human Y chromosome. *Nature*, 621(7978), 344-354. https://doi.org/10.1038/s41586-023-06457-y

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution,* 43(1), 223–225. https://doi.org/10.1111/j.1558-5646.1989.tb04220.x

Rodrigues, E. L., Machado, F. B., Arruda, M. M., de Moura-Neto, R. S., & Medina-Acosta, E. (2009). Genetic data on 15 STR autosomal loci for a sample population of the Northern Region of the State of Rio de Janeiro, Brazil. *Forensic Science International: Genetics*, 4(1), e25-e26. https://doi.org/10.1016/j.fsigen.2009.01.008

Rodrigues, P., Velázquez, I.F., Ribeiro, J., Simão, F., Amorim, A., Carvalho, E.F., Bravi, C.M., Basso, N.G., Real, L.E., Galli, C. & González, A.D.C. (2022). Tierra del Fuego: What is left from the precolonial male lineages? *Genes*, 13(10), 1712. https://doi.org/10.3390/genes13101712

Roewer, L. (2013). DNA fingerprinting in forensics: past, present, future. *Investigative Genetics*, 4, 1-10. https://doi.org/10.1186/2041-2223-4-22

Roewer, L. (2019). Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. *Wiley Interdisciplinary Reviews: Forensic Science*, 1(4), e1336. https://doi.org/10.1002/wfs2.1336

Roewer, L., & Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Science International*, 53(2), 163-171. https://doi.org/10.1016/0379-0738(92)90193-Z

Roewer, L., Amemann, J., Spurr, N. K., Grzeschik, K. H., & Epplen, J. T. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Human Genetics*, 89, 389-394. https://doi.org/10.1007/BF00194309

Roman-Santos, C. (2010). Concerns associated with expanding DNA databases. *Hastings Science & Technology Law Journal*, 2, 267. https://repository.uclawsf.edu/hastings_science_technology_law_journal/vol2/iss2/5/

Romeika, J. M., & Yan, F. (2013). Recent advances in forensic DNA analysis. *Journal of Forensic Research*, 12(001). https://doi.org/10.4172/2157-7145.S12-001

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298(5602), 2381-2385. https://doi.org/10.1126/science.1078311

Rubi-Castellanos, R., Anaya-Palafox, M., Mena-Rojas, E., Bautista-España, D., Muñoz-Valle, J. F., & Rangel-Villalobos, H. (2009). Genetic data of 15 autosomal STRs (Identifiler kit) of three Mexican Mestizo population samples from the States of Jalisco (West), Puebla (Center), and Yucatan (Southeast). *Forensic Science International: Genetics*, 3(3), e71-e76. https://doi.org/10.1016/j.fsigen.2008.07.006

Rupp, N., Ameje, J., & Breunig, P. (2005). New studies on the Nok culture of central Nigeria. *Journal of African Archaeology*, 3(2), 283-290. https://doi.org/10.3213/1612-1651-10056

Saeed, N., Nam, H., Haq, M. I. U., & Muhammad Saqib, D. B. (2018). A survey on multidimensional scaling. *ACM Computing Surveys (CSUR)*, 51(3), 1-25. https://doi.org/10.1145/3178155

Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S., Trivedi, R., Endicott, P., Kivisild, T., Metspalu, M. & Villems, R. (2006). A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proceedings of the National Academy of Sciences*, 103(4), 843-848. https://doi.org/10.1073/pnas.0507714103

Salas, A., Carracedo, Á., Richards, M., & Macaulay, V. (2005). Charting the ancestry of African Americans. *The American Journal of Human Genetics*, 77(4), 676-680. https://doi.org/10.1086/491675

Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V. & Carracedo, Á. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *The American Journal of Human Genetics*, 74(3), 454-465. https://doi.org/10.1086/382194

Salem, N., van de Loosdrecht, M.S., Sümer, A.P., Vai, S., Hübner, A., Peter, B., Bianco, R.A., Lari, M., Modi, A., Al-Faloos, M.F.M. & Turjman, M. (2025). Ancient DNA from the Green Sahara reveals ancestral North African lineage. *Nature* 641, 144–150. https://doi.org/10.1038/s41586-025-08793-7

Sánchez-Diz, P., Acosta, M.A., Fonseca, D., Fernández, M., Gomez, Y., Jay, M., Alape, J., Lareu, M.V., Carracedo, A. & Restrepo, C.M. (2009). Population data on 15 autosomal STRs in a sample from Colombia. *Forensic Science International: Genetics*, 3(3), e81-e82. https://doi.org/10.1016/j.fsigen.2008.08.002

Santos, F. (2016). *Overview of the Implementation of the Prüm Decisions* (Technical Report, November 2016). Available from: https://estudogeral.uc.pt/bitstream/10316/41091/1/Overview%20of%20the%20implementation%20of%20the%20Pr%C3%BCm%20Decisions.pdf

Santos, F., Machado, H., & Silva, S. (2013). Forensic DNA databases in European countries: is size linked to performance? *Life Sciences, Society and Policy*, 9(1), 12. https://doi.org/10.1186/2195-7819-9-12

Sarno, S., Boattini, A., Carta, M., Ferri, G., Alù, M., Yao, D.Y., Ciani, G., Pettener, D. & Luiselli, D. (2014). An ancient Mediterranean melting pot: investigating the uniparental genetic structure and population history of sicily and southern Italy. *PLoS One*, 9(4), e96074. https://doi.org/10.1371/journal.pone.0096074

Schlebusch, C. M., & Jakobsson, M. (2018). Tales of human migration, admixture, and selection in Africa. *Annual Review of Genomics and Human Genetics*, 19(1), 405-428. https://doi.org/10.1146/annurev-genom-083117-021759

Schneider, P. M. (2007). Scientific standards for studies in forensic genetics. *Forensic Science International*, 165(2-3), 238-243. https://doi.org/10.1016/j.forsciint.2006.06.067

Scottish Legal News. (2015, February 3). SPA scientists now able to identify DNA in weaker samples. Scottish Legal News. Retrieved from https://www.scottishlegal.com/articles/spa-scientists-now-able-identify-dna-weaker-samples

Seldin, M.F., Shigeta, R., Villoslada, P., Selmi, C., Tuomilehto, J., Silva, G., Belmont, J.W., Klareskog, L. & Gregersen, P.K. (2006). European population substructure: clustering of northern and southern populations. *PLoS Genetics*, 2(9), e143. https://doi.org/10.1371/journal.pgen.0020143

Semo, A.C., Carvalho, M.R., Bogas, V., Serra, A., Lopes, V., Brito, P., Sá, F.B., Porto, M.J., Gonçalves, I.M.T. & Corte-Real, F., (2017). Allelic frequencies of 15 autosomal STRs from two main population groups (Makua and Changana) in Mozambique. *Forensic Science International: Genetics Supplement Series*, 6, e286-e288. https://doi.org/10.1016/j.fsigss.2017.09.139

Servent, A. R., & MacKenzie, A. (2017). Eroding Germany's commitment to data protection: policy entrepreneurs and coalition politics in EU passenger name records. *German Politics*, 26(3), 398-413. https://doi.org/10.1080/09644008.2016.1250889

Shaw, I., & Sandiford, A. (2024). *Forensic Science: The Science Behind the Truth*. Royal Society of Chemistry.

Shete, S., Tiwari, H., & Elston, R. C. (2000). On estimating the heterozygosity and polymorphism information content value. *Theoretical Population Biology*, 57(3), 265-271. https://doi.org/10.1006/tpbi.2000.1452

Shinagawa, J., Moteki, H., Nishio, S. Y., Noguchi, Y., & Usami, S. I. (2020). Haplotype analysis of GJB2 mutations: founder effect or mutational hot spot?. *Genes*, 11(3), 250. https://doi.org/10.3390/genes11030250

Shragg, L. D. (2015). *Songs of a lost tribe: An investigation and analysis of the musical properties of the Igbo Jews of Nigeria*. The University of Arizona.

Shrivastava, P., Jain, T., & Kumawat, R. K. (2021). Direct PCR amplification from saliva sample using non-direct multiplex STR kits for forensic DNA typing. *Scientific Reports*, 11(1), 7112. https://doi.org/10.1038/s41598-021-86633-0

Shuman, D. W., & Greenberg, S. A. (2003). The expert witness, the adversary system, and the voice of reason: Reconciling impartiality and advocacy. *Professional Psychology: Research and Practice,* 34(3), 219–224. https://doi.org/10.1037/0735-7028.34.3.219

Siegel, J. A., & Saukko, P. J. (2012). *Encyclopaedia of Forensic Sciences*. Academic Press.

Silva, D. S. B. S. (2024). Validation of NGS for casework at forensic DNA laboratories. In *Next Generation Sequencing (NGS) Technology in DNA Analysis* (pp. 485-509). Academic Press.

Sim, J. E., Park, S. J., Lee, H. C., Kim, S. Y., Kim, J. Y., & Lee, S. H. (2013). High-throughput STR analysis for DNA database using direct PCR. *Journal of Forensic Sciences*, 58(4), 989-992. https://doi.org/10.1111/1556-4029.12166

Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell*, 177(1), 26-31. https://doi.org/10.1016/j.cell.2019.02.048

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139(1), 457-462. https://doi.org/10.1093/genetics/139.1.457

Slatkin, M. (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6), 477-485. https://doi.org/10.1038/nrg2361

Smith, M. G. (2018). The beginnings of Hausa society, AD 1000–1500. In: *The Historian in Tropical Africa* (pp. 339-357). Routledge.

Sobiah, R., Syeda, R.H., Zunaira, E., Nageen, Z., Maria, K., Syeda, A.Z., Shahana, S.M., Akifa, M., Abdul, J. & Muhammad, R.K. (2018). Implications of targeted next generation sequencing in forensic science. *Journal of Forensic Research*, 9(2), 1-8. https://doi.org/10.4172/2157-7145.1000416

Sobrino, B., Bríon, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, 154(2-3), 181-194. https://doi.org/10.1016/j.forsciint.2004.10.020

Sood, R., & Singh, V. (2024). DNA sequencing technologies in accelerating molecular breeding. *IntechOpen.* https://doi.org/10.5772/intechopen.1006294

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology,* 98(3), 503-517. https://doi.org/10.1016/S0022-2836(75)80083-0

Sparkes, R., Kimpton, C., Gilbard, S., Carne, P., Andersen, J., Oldroyd, N., Thomas, D., Urquhart, A. & Gill, P. (1996a). The validation of a 7-locus multiplex STR test for use in forensic casework. (II), Artefacts, casework studies and success rates. *International Journal of Legal Medicine*, 109(4), 195-204. https://doi.org/10.1007/BF01225518

Sparkes, R., Kimpton, C., Watson, S., Oldroyd, N., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A. & Gill, P. (1996b). The validation of a 7-locus multiplex STIR test for use in forensic casework. *International Journal of Legal Medicine*, 109(4), 186-194. https://doi.org/10.1007/BF01225517

Stanciu, F., Cuțăr, V., Vladu, S., Stoian, I.M., Rădulescu, A., Cotolea, A., Păunache, M.A. and Cucoș, A. (2024). Unlocking Forensic Potential: CODIS Implementation and DNA Data Exchange in Romania. *GENOMICA*, 1(1), 3-9. https://doi.org/10.5281/zenodo.11069588

Steele, C. D. (2016). *Statistical issues surrounding the analysis of forensic low-template DNA samples* (Doctoral dissertation, UCL (University College London)).

Steele, C. D., Court, D. S., & Balding, D. J. (2014). Worldwide Estimates Relative to Five Continental-Scale Populations. *Annals of Human Genetics*, 78(6), 468-477. https://doi.org/10.1111/ahg.12081

Stringer, C. (2016). The origin and evolution of Homo sapiens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1698), 20150237. https://doi.org/10.1098/rstb.2015.0237

Syndercombe Court, D. (2021). The Y chromosome and its use in forensic DNA analysis. *Emerging Topics in Life Sciences*, 5(3), 427-441. https://doi.org/10.1042/ETLS20200339

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 38(7), 3022-3027. https://doi.org/10.1093/molbev/msab120

Tan, J.Y.Y., Tan, Y.P., Ng, S., Tay, A.S., Phua, Y.H., Tan, W.J., Ong, T.Y.R., Chua, L.M. & Syn, C.K.C. (2017). A preliminary evaluation study of new generation multiplex STR kits comprising of the CODIS core loci and the European Standard Set loci. *Journal of Forensic and Legal Medicine*, 52, 16-23. https://doi.org/10.1016/j.jflm.2017.07.017

Taqi, Z., Alenizi, M., Alenizi, H., Ismael, S., Dukhyil, A.A.B., Nazir, M., Sanqoor, S., Al Harbi, E., Al-Jaber, J., Theyab, J. and Moura-Neto, R.S. (2015). Population genetics of 23 Y-STR markers in Kuwaiti population. *Forensic Science International: Genetics*, 16, 203-204. https://doi.org/10.1016/j.fsigen.2015.01.007

Than, K.Z., Muisuk, K., Woravatin, W., Suwannapoom, C., Srikummool, M., Srithawong, S., Lorphengsy, S. & Kutanan, W. (2022). Genetic structure and forensic utility of 23 autosomal STRs of the ethnic Lao groups from Laos and Thailand. *Frontiers in Genetics*, 13, 954586. https://doi.org/10.3389/fgene.2022.954586

Thompson, J.M., Ewing, M.M., Frank, W.E., Pogemiller, J.J., Nolde, C.A., Koehler, D.J., Shaffer, A.M., Rabbach, D.R., Fulmer, P.M., Sprecher, C.J. & Storts, D.R. (2013). Developmental validation of the PowerPlex® Y23 System: a single multiplex Y-STR analysis system for casework and database samples. *Forensic Science International: Genetics*, *7*(2), 240-250. https://doi.org/10.1016/j.fsigen.2012.10.013

Thompson, W. C., Taroni, F., & Aitken, C. G. (2003). How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Science*, 48(1), 1-8. https://doi.org/10.1520/JFS2001171

Tian, C., Kosoy, R., Lee, A., Ransom, M., Belmont, J. W., Gregersen, P. K., & Seldin, M. F. (2008). Analysis of East Asia genetic substructure using genome-wide SNP arrays. *PLoS One*, 3(12), e3862. https://doi.org/10.1371/journal.pone.0003862

Tilanus, M. G. J. (2006). Short tandem repeat markers in diagnostics: what's in a repeat? *Leukemia,* 20, 1353-1355. https://doi.org/10.1038/sj.leu.2404273

Tillmar, A. (2010). *Populations and Statistics in Forensic Genetics* (Doctoral Dissertation). Linköping University Electronic Press. https://www.diva-portal.org/smash/get/diva2:309703/fulltext01.pdf

Tillmar, A. O., Bäckström, G., & Montelius, K. (2009). Genetic variation of 15 autosomal STR loci in a Somali population. *Forensic Science International: Genetics*, 4(1), e19-e20. https://doi.org/10.1016/j.fsigen.2009.01.004

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. (2009). The genetic structure and history of Africans and African Americans. *Science*, 324(5930), 1035-1044. https://doi.org/10.1126/science.1172257

Tiwari, A., & Kusum, S. (2024). Role of Forensic Criminology in Access to Justice—A Critical Analysis. In *Forensic Justice* (pp. 42-67). Routledge.

Toom, V., Granja, R., & Ludwig, A. (2019). The Prüm Decisions as an aspirational regime: reviewing a decade of cross-border exchange and comparison of forensic DNA data. *Forensic Science International: Genetics*, 41, 50-57. https://doi.org/10.1016/j.fsigen.2019.03.023

Triki-Fendri, S., Sanchez-Diz, P., Rey-González, D., Ayadi, I., Alfadhli, S., Rebai, A., & Carracedo, Á. (2013). Population genetics of 17 Y-STR markers in West Libya (Tripoli region). *Forensic Science International: Genetics*, *7*(3), e59-e61. https://doi.org/10.1016/j.fsigen.2013.02.002

Uberoi, D., Palmour, N., & Joly, Y. (2024). The advent of forensic DNA databases: it's time to agree on some international governance principles. *Forensic Science International: Genetics*, 103095. https://doi.org/10.1016/j.fsigen.2024.103095

UK Parliament, House of Commons Library. (2025, March 5). *Resetting the UK's Relationship with the European Union* (Research Briefing, CBP-10207). https://commonslibrary.parliament.uk/research-briefings/cbp-10207/

Ukeje, C., & Adebanwi, W. (2008). Ethno-nationalist claims in southern Nigeria: Insights from Yoruba and Ijaw nationalisms since the 1990s. *Ethnic and Racial Studies*, 31(3), 563-591. https://doi.org/10.1080/01419870701491978

Umar, H., Abolagba, E., Giroh, D., & Lalabe, B. (2011). Costs and returns analysis of Gum Arabic and some selected tree crops production in Adamawa and Yobe states, Nigeria: an implication for poverty alleviation. *World Rural Observations*, 3(2), 49-54. https://www.sciencepub.net/rural/rural0302/07_5004rural0302_49_54.pdf

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Lahr, M.M., Foley, R.A., Oefner, P.J. & Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics*, 65(1), 43-62. https://doi.org/10.1046/j.1469-1809.2001.6510043.x

Uthman, O. A., Aiyedun, V., & Yahaya, I. (2011). Exploring variations in under-5 mortality in Nigeria using league table, control chart and spatial analysis. *Journal of Public Health*, 34(1), 125-130. https://doi.org/10.1093/pubmed/fdr050

Varsha, V. (2006). DNA fingerprinting in the criminal justice system: An overview. *DNA and Cell Biology*, 25(3), 181-188. https://doi.org/10.1089/dna.2006.25.181

Veerappa, A. M., Padakannaya, P., & Ramachandra, N. B. (2013). Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. *Functional & Integrative Genomics*, 13, 285-293. https://doi.org/10.1007/s10142-013-0323-6

Vicente, M., Priehodová, E., Diallo, I., Podgorná, E., Poloni, E. S., Černý, V., & Schlebusch, C. M. (2019). Population history and genetic adaptation of the Fulani nomads: inferences from genome-wide data and the lactase persistence trait. *BMC Genomics*, 20, 1-12. https://doi.org/10.1186/s12864-019-6296-7

Vizmanos, J. L., Cross, N. C., & Novo, F. J. (2020). The nature of traits, genes and variation. In *Principles of Nutrigenetics and Nutrigenomics* (pp. 3-9). Academic Press.

Vossen, R. (Ed.). (2013). *The Khoesan Languages.* London & New York: Routledge.

Wahlund, S. (1928). Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, 11(1), 65-106. https://doi.org/10.1111/j.1601-5223.1928.tb02483.x

Wallace, H. (2006). The UK national DNA database. *EMBO Reports*, 7(S1), S26-S30. https://doi.org/10.1038/sj.embor.7400727

Wallace, H. M., Jackson, A. R., Gruber, J., & Thibedeau, A. D. (2014). Forensic DNA databases–Ethical and legal standards: A global review. *Egyptian Journal of Forensic Sciences*, 4(3), 57-63. https://doi.org/10.1016/j.ejfs.2014.04.002

Wang, C.C., Yeh, H.Y., Popov, A.N., Zhang, H.Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., Kim, A.M. & Mallick, S. (2021). Genomic insights into the formation of human populations in East Asia. *Nature*, 591(7850), 413-419. https://doi.org/10.1038/s41586-021-03336-2

Wang, D. Y., Chang, C. W., Lagacé, R. E., Oldroyd, N. J., & Hennessy, L. K. (2011). Development and Validation of the AmpFℓSTR® Identifiler® Direct PCR Amplification Kit: A Multiplex Assay for the Direct Amplification of Single-Source Samples. *Journal of Forensic Sciences*, 56(4), 835-845. https://doi.org/10.1111/j.1556-4029.2011.01757.x

Wang, D. Y., Gopinath, S., Lagacé, R. E., Norona, W., Hennessy, L. K., Short, M. L., & Mulero, J. J. (2015). Developmental validation of the GlobalFiler® Express PCR Amplification Kit: a 6-dye multiplex assay for the direct amplification of reference samples. *Forensic Science International: Genetics*, 19, 148-155. https://doi.org/10.1016/j.fsigen.2015.07.013

Wang, M., Cai, J., Chen, J., Liu, J., Geng, X., Yu, X., & Yang, J. (2024). *PCR Techniques and their Clinical Applications*. In M. Aycan & M. Yildiz (Eds.), *Polymerase Chain Reaction Research* (pp. 75–78). IntechOpen. https://doi.org/10.5772/intechopen.110220

Ward, T., Edmond, G., Martire, K., & Wortley, N. (2017). Forensic science, reliability and scientific validity: Advice from America. *Criminal Law Review*, 5, 357-378. Retrieved from: https://researchportal.northumbria.ac.uk/en/publications/forensic-science-reliability-and-scientific-validity-advice-from

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open-Source Software*, 6(60), 3021. https://doi.org/10.21105/joss.03021

Watson, J. D. (2012). *The Polymerase Chain Reaction*. Springer Science & Business Media.

Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738. https://doi.org/10.1038/171737a0

Weir, B. S. (1992). Population genetics in the forensic DNA debate. *Proceedings of the National Academy of Sciences*, 89(24), 11654-11659. https://doi.org/10.1073/pnas.89.24.11654

Weir, B. S. (1994). The effects of inbreeding on forensic calculations. *Annual Review of Genetics*, 28(1), 597-621. https://doi.org/10.1146/annurev.ge.28.120194.003121

Weir, B. S. (1996). *Genetic Data Analysis II. Methods for Discrete Population Genetic Data*. Sinauer Associates, Inc. Publishers.

Weir, B. S. (2012). Estimating F-statistics: A historical view. *Philosophy of Science*, 79(5), 637-643. https://doi.org/10.1086/667904

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution,* 38(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Westemeier-Rice, E. S., Winters, M. T., Rawson, T. W., & Martinez, I. (2024). More than the SRY: the non-coding landscape of the y chromosome and its importance in human disease. *Non-Coding RNA*, 10(2), 21. https://doi.org/10.3390/ncrna10020021

Williams, R., & Johnson, P. (2005). Inclusiveness, effectiveness and intrusiveness: issues in the developing uses of DNA profiling in support of criminal investigations. *The Journal of Law, Medicine & Ethics*, 33(3), 545-558. https://doi.org/10.1111/j.1748-720X.2005.tb00517.x

Williams, R., & Johnson, P. (2013). *Genetic Policing: The Uses of DNA in Police Investigations*. Willan.

Williamson, K., & Blench, R. (2000). *Niger-Congo.* In: B. Heine & D. Nurse (Eds.), *African Languages: An Introduction* (pp. 11–42). Cambridge University Press

Willuweit, S., Roewer, L., & International Forensic Y Chromosome User Group. (2007). Y chromosome haplotype reference database (YHRD): update. *Forensic Science International: Genetics*, 1(2), 83-87. https://doi.org/10.1016/j.fsigen.2007.01.017

Wood, E.T., Stover, D.A., Ehret, C., Destro-Bisol, G., Spedini, G., McLeod, H., Louie, L., Bamshad, M., Strassmann, B.I., Soodyall, H. & 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(7), 867-876. https://doi.org/10.1038/sj.ejhg.5201408

World Population Review. (2025). *Crime Rate by Country 2025*. Retrieved May 23, 2025, from https://worldpopulationreview.com/country-rankings/crime-rate-by-country/

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, 56(645), 330-338. https://doi.org/10.1086/279872

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97–159. https://doi.org/10.1093/genetics/16.2.97

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, 15(1), 323-354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Yamamoto, F. I., & Hakomori, S. I. (1990). Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *Journal of Biological Chemistry*, 265(31), 19257-19262. https://doi.org/10.1016/S0021-9258(17)30652-X

You, K., Dal Bianco, S., Lin, Z., & Amankwah-Amoah, J. (2019). Bridging technology divide to improve business environment: Insights from African nations. *Journal of Business Research*, 97, 268-280. https://doi.org/10.1016/j.jbusres.2018.01.015

Zarza, Á. G. (2015). EU Information Systems and Databases. In *Exchange of Information and Data Protection in Cross-border Criminal Proceedings in Europe* (pp. 117-139). Springer, Berlin, Heidelberg.

Zeller, C., & Elkins, K. M. (2020). Simulation of population sampling and allele frequency, linkage equilibrium, and random match probability calculations. *The Journal of Forensic Science Education*, 2(1)1-5. https://jfse-ojs-tamu.tdl.org/jfse/index.php/jfse/article/view/18

Zhu, B., & Shen, C. (2024). *Tools and techniques of using NGS platforms in forensic population genetic studies.* In: A. Editor & B. Editor (Eds.), *Next Generation Sequencing (NGS) Technology in DNA Analysis* (pp. 121–171). Academic Press.

**APPENDIX 1**
**PARTICIPANT INFORMATION SHEET**

# Participant Information Sheet

**Name of department:** Department of Pure and Applied Chemistry

**Title of the study:** Comparative Population Genetic Study of some STR Kits in the Nigerian Population

**Introduction**


**Chief Investigator:**
Name: Dr Penny Haddrill
Status: PhD supervisor
Tel.: 0141 548 4337
E-mail: penny.haddrill@strath.ac.uk

**Co-Investigator:**
Name: Daniel Awomukwu
Status: PhD Student
Tel.: UK: +44 (0) 751 689 3262
Nigeria: +234 806 365 7688
E-mail: daniel.awomukwu@strath.ac.uk


**What is the purpose of this investigation?**
The aim of this research is to analyse the population genetics of unrelated Nigerian individuals from three different ethnic groups using a 17 locus STR kit, 24 locus STR kit, and Y-STR kit in order to determine which would be the preferred STR kit for forensic analysis and the development of a national DNA database in Nigeria.

**Why have you been invited to take part?**
The aim of the project is to generate allele and genotype frequencies for the different STR markers in the different population sub-groups of Nigeria. It is essential that these samples are collected from volunteers who originate from Nigerian populations; these samples cannot be simulated. All Nigerian participants over 18 in age are welcomed to provide blood samples.

**Do you have to take part?**
This study involves participants donating blood samples. Participation in this research is entirely voluntary and you have the right to refuse to participate without giving a reason; refusing to participate will not negatively affect you in any way. You also have the right to withdraw from this research at any time up to the completion of the project (June 2020) without detriment and without giving a reason and ask for your samples and data to be destroyed.

**What will you do in the project?**
You will be required to provide small blood sample, collected by a trained investigator using a sterile lancet to make a 'finger prick', and blood deposited onto a collection card. DNA will be extracted from the samples and analysed at the University of Strathclyde in the UK. No payments will be provided for taking part in this research. When you sign the consent form, you will also be asked to provide information about your ethnicity, which will allow the researchers to look for genetic associations between different population groupings.

**What are the potential risks to you in taking part?**
The collection of blood using a lancet carries a small risk of infection to the donor, which will be minimised with the use of alcohol wipes and participants will be given a plaster following donation. In addition, participants will only be asked to handle their own samples. In this way, the chances of infection will be substantially reduced and will be negligible.

**What happens to the information in the project?**
If you volunteer to take part in the project, we will collect a blood sample from you, and the DNA will be extracted from this and analysed to produce a DNA profile. We will also take your name, which will be associated with your sample using a code, until the end of the project when we will delete your name so that we no longer have your identity (see below). The only other information we will collect is your ethnicity, which we will associate with your sample so that we can compare samples from different population groups within Nigeria.

**Who will have access to the information?**
All participants' information will be kept confidential. In order to preserve anonymity, samples will be labelled with a code that does not contain any information allowing the participant to be identified, except for by the investigators. This will allow a participant's data to be identified, removed and destroyed should they subsequently wish to withdraw from the study. Once the project is completed, the codes linking identity with the samples will be deleted and after this point participants will no longer be able to withdraw from the study, as the destruction of the codes will mean that their data will be fully anonymised.

**Where will the information be stored and how long will it be kept for?**
Collected blood samples on FTA cards will be kept in a laboratory with restricted access in the Regional Centre for Bioresources and carefully returned from Nigeria to the Centre for Forensic Science laboratory, University of Strathclyde, Glasgow, where it will be stored in secure restricted access laboratory in the forensic DNA suite (R6.24). They will be kept in a Multi-Barrier Pouch with a desiccant packet. Access to the FTA cards in the laboratory will be restricted to only the research investigators, Dr. Penny Haddrill and Daniel Awomukwu.

All data outputs will be stored electronically on password-protected computers only accessible by the investigators. Electronic data may be retained indefinitely in this form, and no information will be put onto any databases. All blood samples will be securely disposed within one year of the end date of the research, which is estimated to be in June 2020

If you agree to participate and sign a consent form to indicate this, your signed consent form will be retained at the University of Strathclyde for a period of five years.

The outcomes of this study will be written into a PhD thesis by co-investigator Daniel Awomukwu. In addition, it is envisaged that the outcomes of this study will be written into journal/conference publication(s). In neither form of publication will any information be included that could allow the participants to be identified.

Thank you for reading this information – please ask any questions if you are unsure about what is written here.

Please also read our Privacy Notice for Research Participants, a copy of which will be provided to you by the researcher with this information sheet.

**What happens next?**
If you are happy to be involved in the project, please sign the consent form provided to confirm this. Please note that participants will not be informed of the specific results of the tests but will be supplied with electronic copies of any resulting publications upon request.

If you do not want to be involved in the project, we would like to thank you for your attention.



Appendix 1 (continued)

**Researcher contact details:**

Daniel Awomukwu
Centre for Forensic Science
Department of Pure and Applied Chemistry
University of Strathclyde
Royal College, 204 George Street, Glasgow, G1 1XW
Telephone: +44 (0)751 689 3262

**Chief Investigator details:**

Dr Penny Haddrill,
Centre for Forensic Science
Department of Pure and Applied Chemistry
University of Strathclyde
Royal College, 204 George Street, Glasgow, G1 1XW
Telephone: 0141 548 4377
E-mail: penny.haddrill@strath.ac.uk

This investigation was granted ethical approval by the Department of Pure and Applied Chemistry Ethics Committee.

If you have any questions/concerns, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or further information may be sought from, please contact:

Secretary to the University Ethics Committee
Research & Knowledge Exchange Services
University of Strathclyde
Graham Hills Building
50 George Street
Glasgow
G1 1QE

Telephone: 0141 548 3707
Email: ethics@strath.ac.uk



Appendix 1 (continued)

**APPENDIX 2**

**CONSENT FORM**

# Consent Form

**Name of department: Pure and Applied Chemistry**

**Title of the study: Comparative Population Genetic Study of some STR kits in the Nigerian Population**

- I confirm that I have read and understood the information sheet for the above project and the researcher has answered any queries to my satisfaction.
- I confirm that I have read and understood the Privacy Notice for Participants in Research Projects and understand how my personal information will be used and what will happen to it (i.e. how it will be stored and for how long).
- I understand that my participation is voluntary and that I am free to withdraw from the project at any time, up to the point of completion, without having to give a reason and without any consequences. If I exercise my right to withdraw and I don't want my data to be used, any data which have been collected from me will be destroyed. However, I understand that upon the completion of the study I will no longer be able to withdraw.
- I understand that any information recorded in the investigation will remain confidential and no information that identifies me will be made publicly available.
- I consent to being a participant in the project.
- I understand that I will be asked to donate a blood sample, collected by the investigators using a sterile lancet.
- I consent to the DNA in my sample being analysed.
- I understand that all of my biological samples will be securely destroyed within one year of the end of the project, which is estimated to be on the 1st June 2020
- I consent to the taking of biological samples from me and understand that they will be the property of the University of Strathclyde.

| | |
|---|---|
| (PRINT NAME) | |
| Signature of Participant: | Date: |

# APPENDIX 3

# SAMPLE COLLECTION REQUEST LETTER

PROFESSIONAL SERVICES | STUDENT EXPERIENCE & ENHANCEMENT SERVICES

20th November 2020

To Whom It May Concern,

Dear Sir/Ma,

**Request for Collection of Blood Samples from Voluntary Donors for the Development of a Non-Offenders DNA Database for Forensic Investigation in Nigeria**

I humbly wish to request permission to collect some blood samples on an FTA card from voluntary donors in your reputable organisation for the completion of the research project involving the population genetics analysis of 500 unrelated Nigerian individuals from three different ethnic groups (Igbo, Yoruba, Hausa-Fulani). The project is aimed to develop an allele frequency database for forensic investigation in Nigeria. The data will be returned to the United Kingdom for DNA profiling and data processing.

The data collection was initially ongoing across some states last month but was halted by the EndSARS protest, which has posed a minimal time for completion. Your establishment will provide a fertile ground for the completion in a record time. The ethical clearance for the research has already been carried out at the Centre for Forensic Science, Glasgow, United Kingdom.

I shall be most grateful if my request is granted as soon as possible to enable a quick sample collection from the various ethnic groups within this short period.

Should you require any additional information, don't hesitate to contact me at +234806365788 or e-mail daniel.awomukwu@strath.ac.uk.

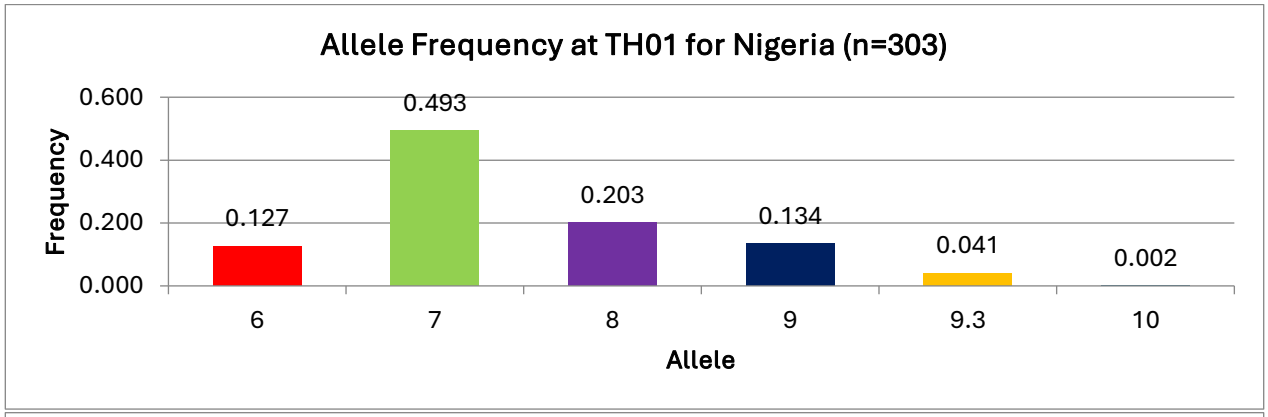
I look forward to hearing from you.

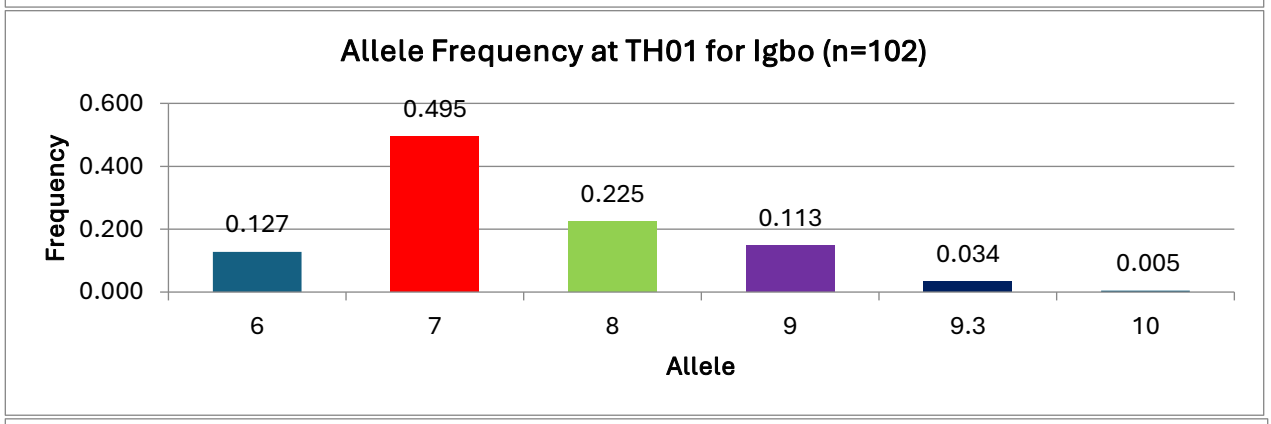
Kindly accept my assurance of goodwill.

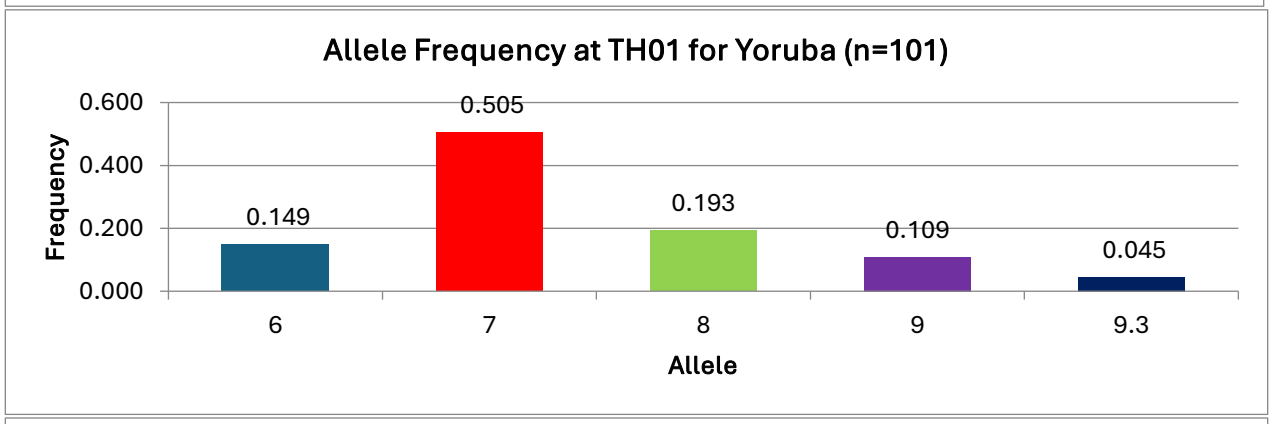
Thank you, and kind regards,

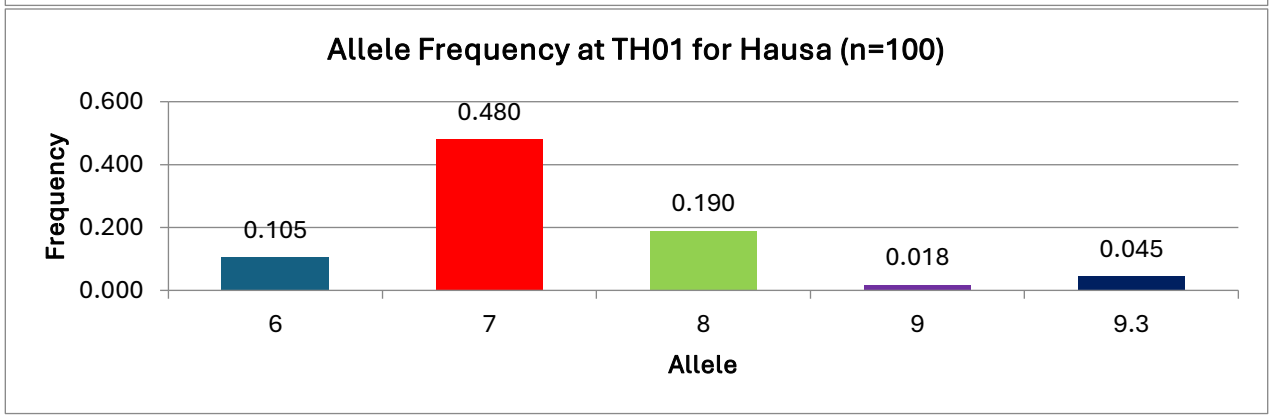
**Daniel Awomukwu**
***PhD Researcher,***
***Centre for Forensic Science,***
***University of Strathclyde,***
***Glasgow, United Kingdom.***

## APPENDIX 4
## ALLELE FREQUENCIES AT TH01 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



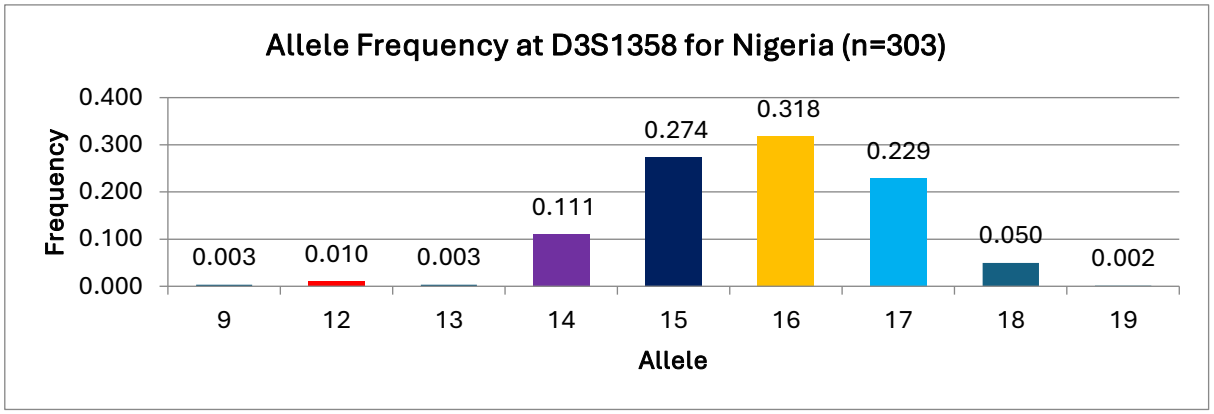



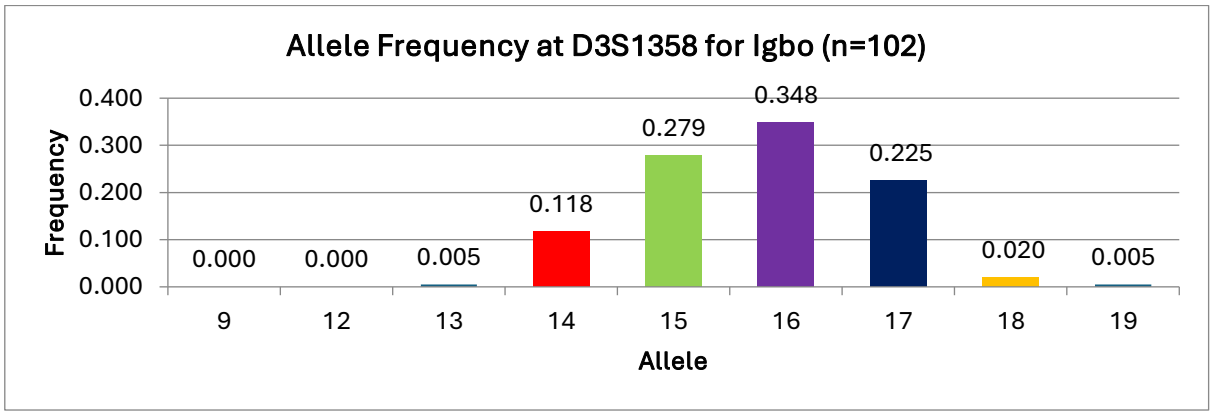



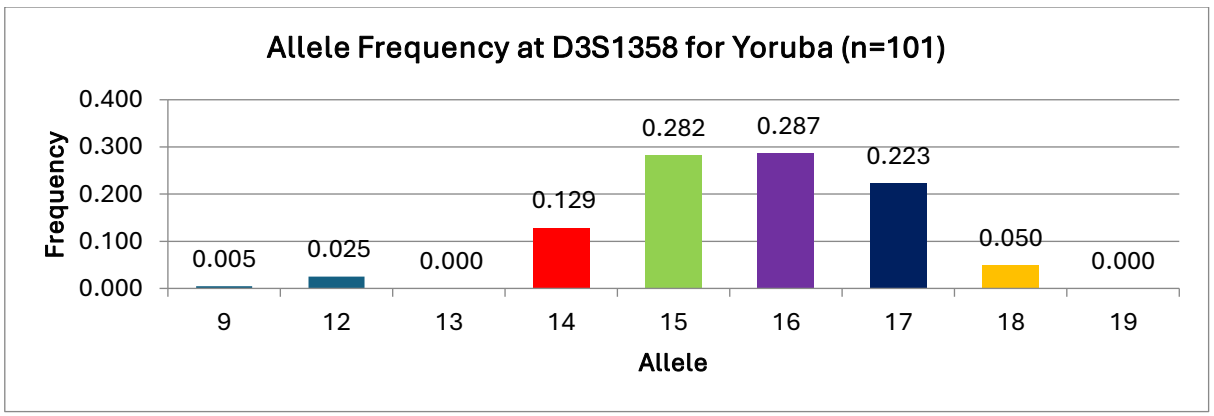



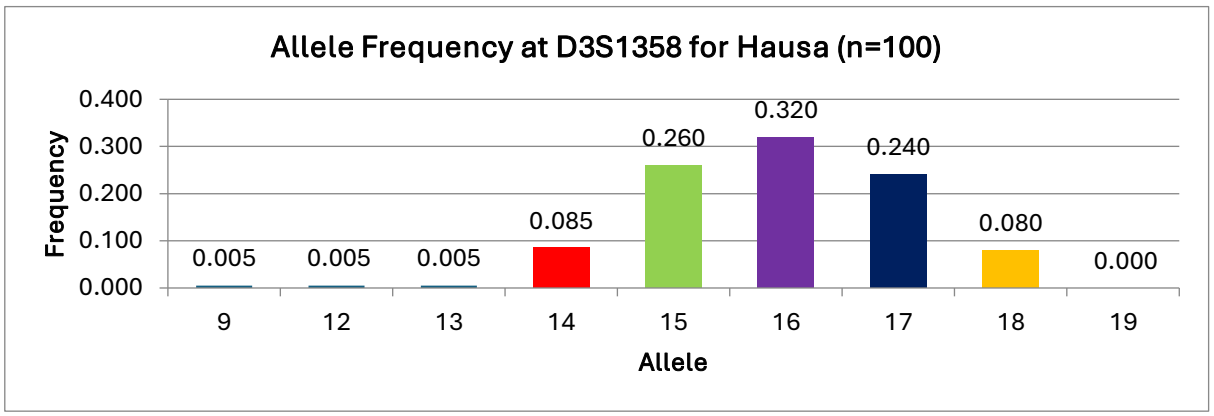
## APPENDIX 5

## ALLELE FREQUENCIES AT D3S1358 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



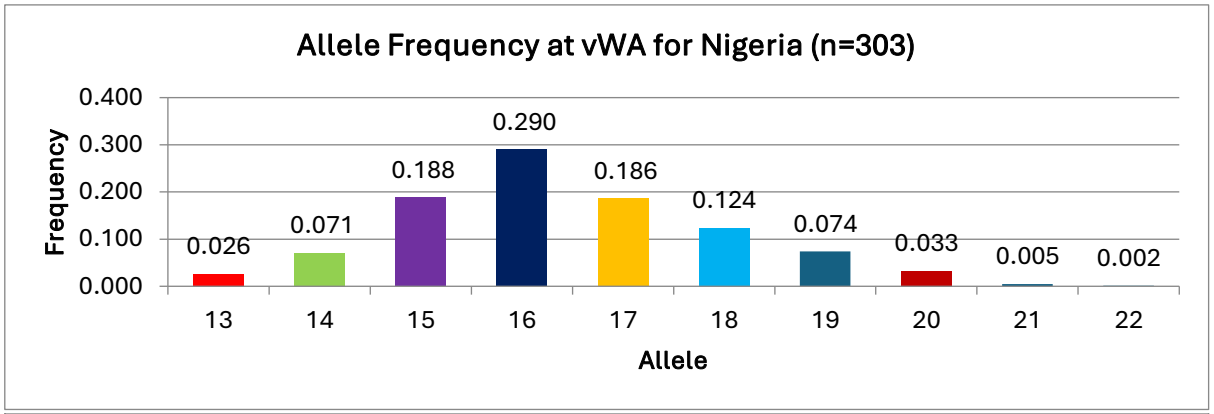



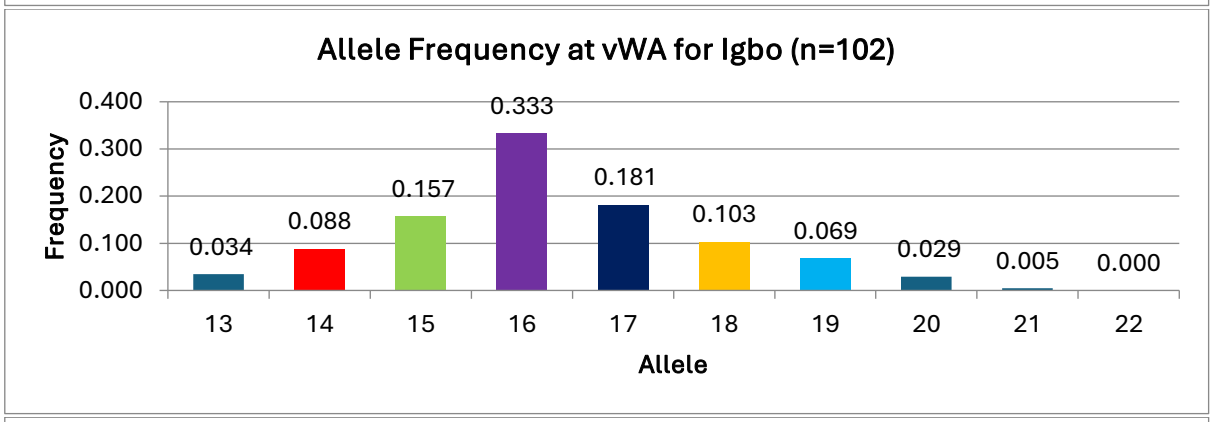



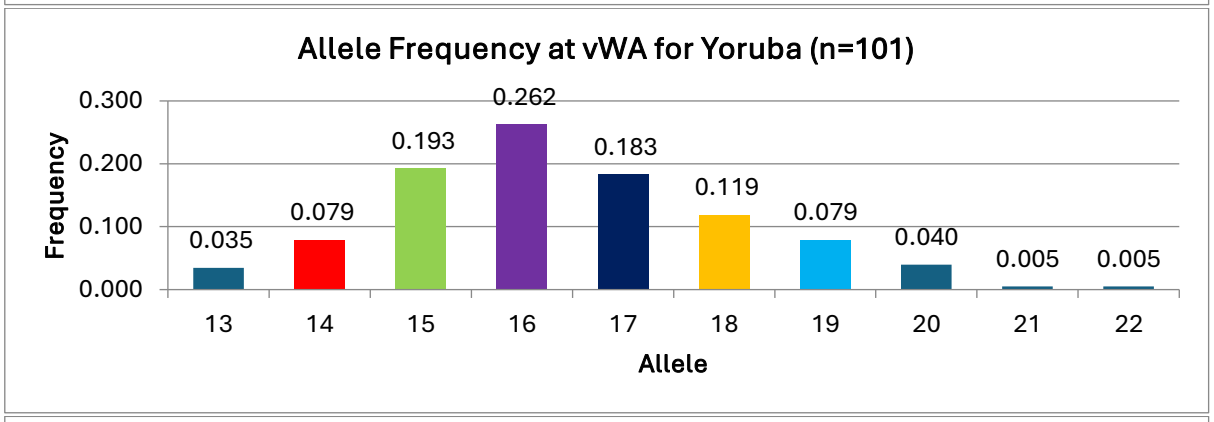



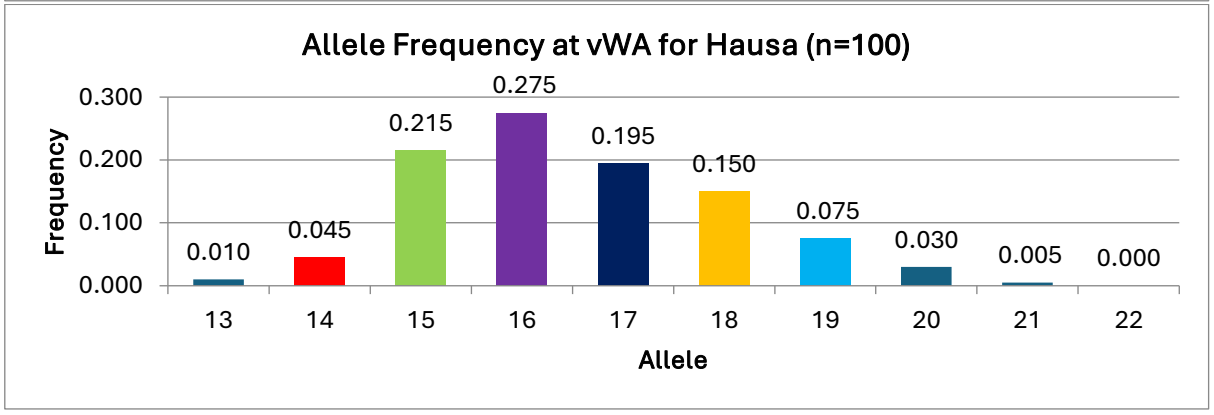
## APPENDIX 6

## ALLELE FREQUENCIES AT vWA WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



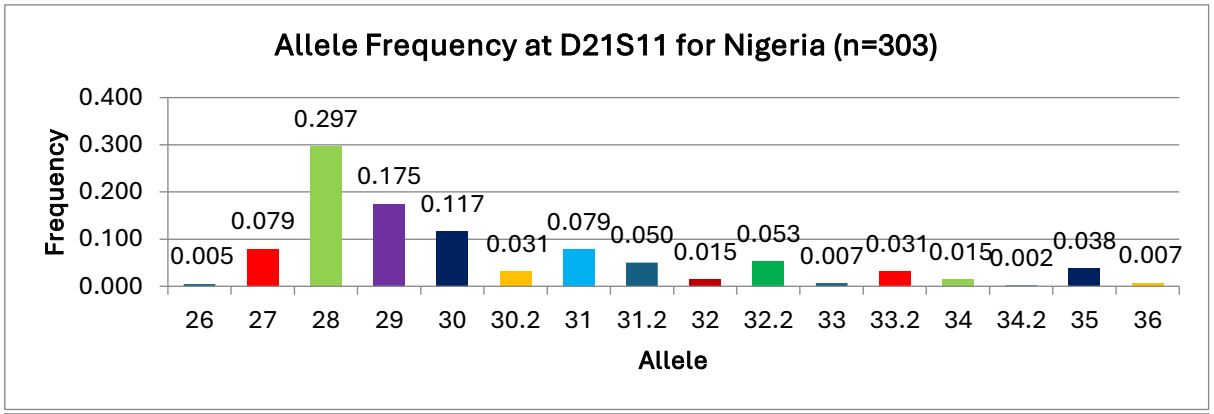



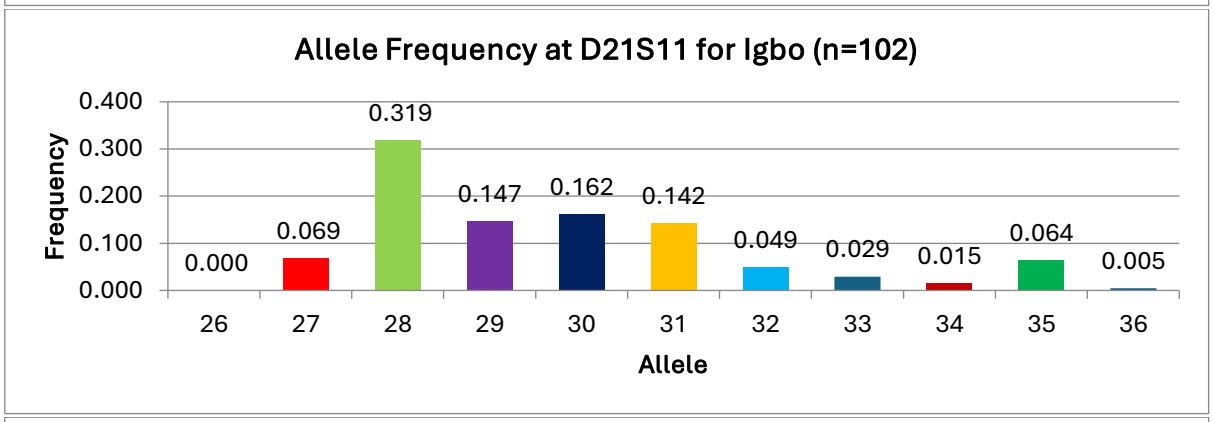



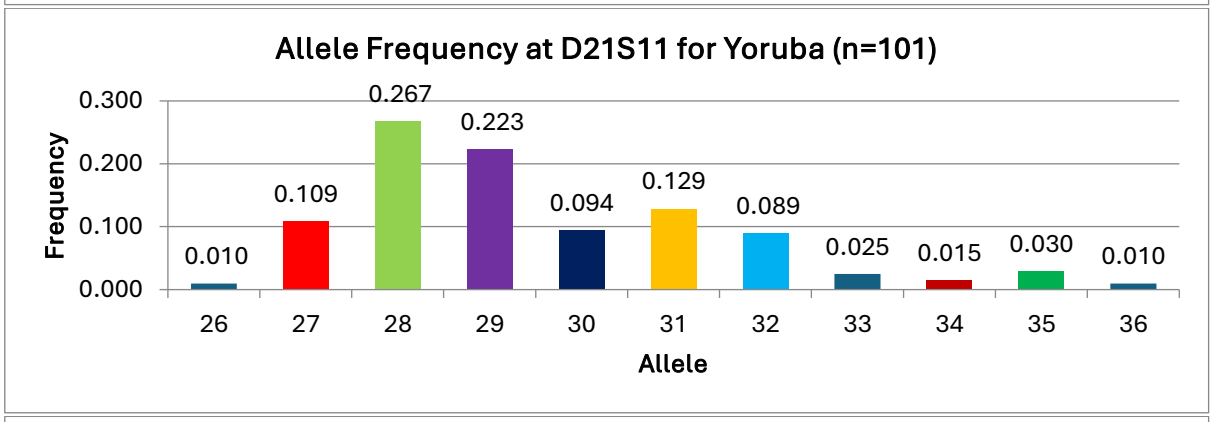



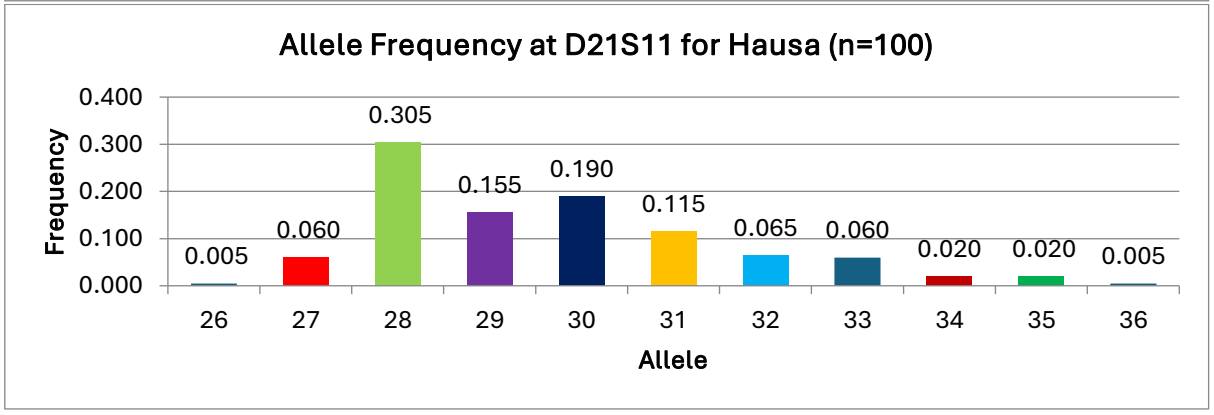
## APPENDIX 7

## ALLELE FREQUENCIES AT D21S11 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



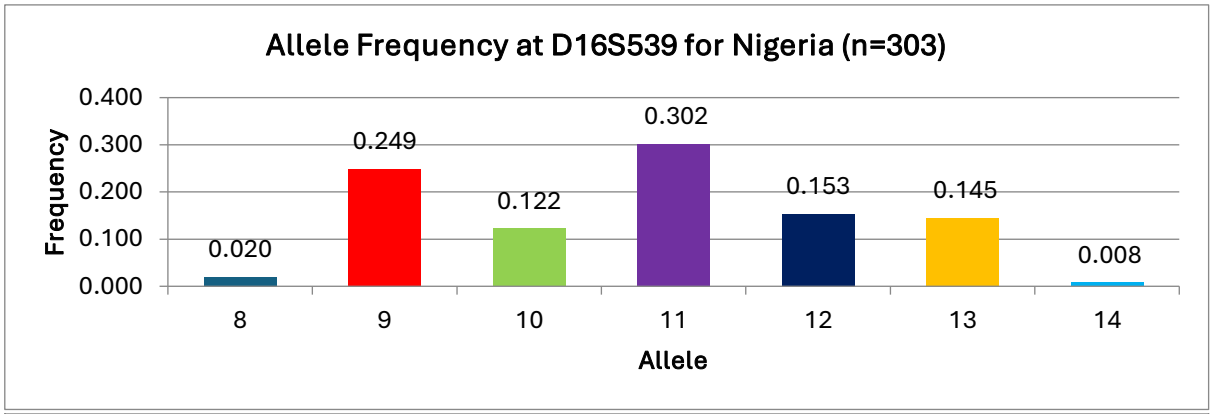



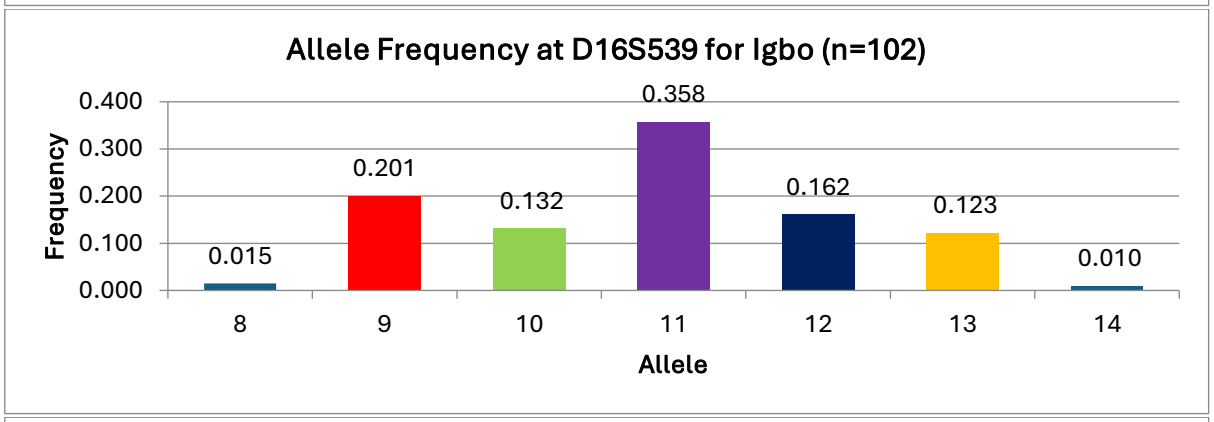



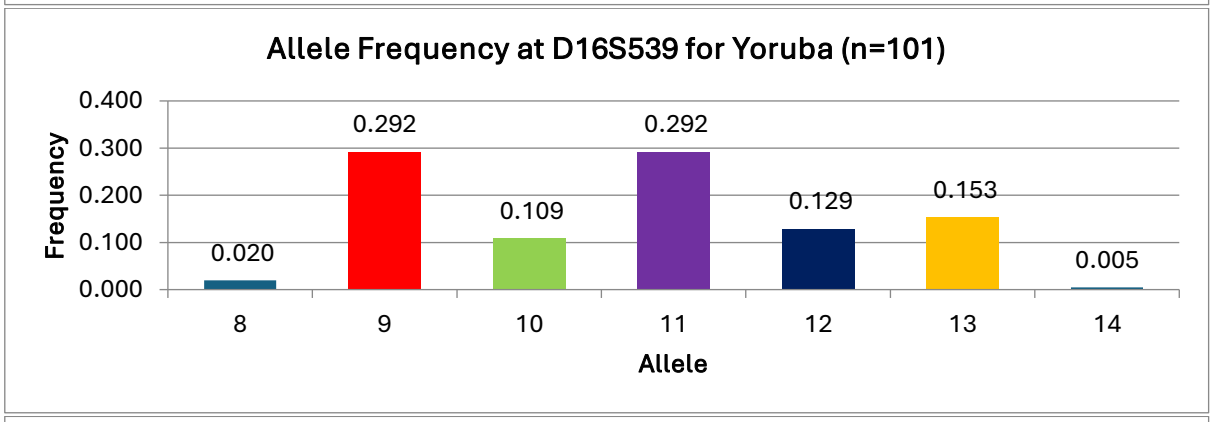



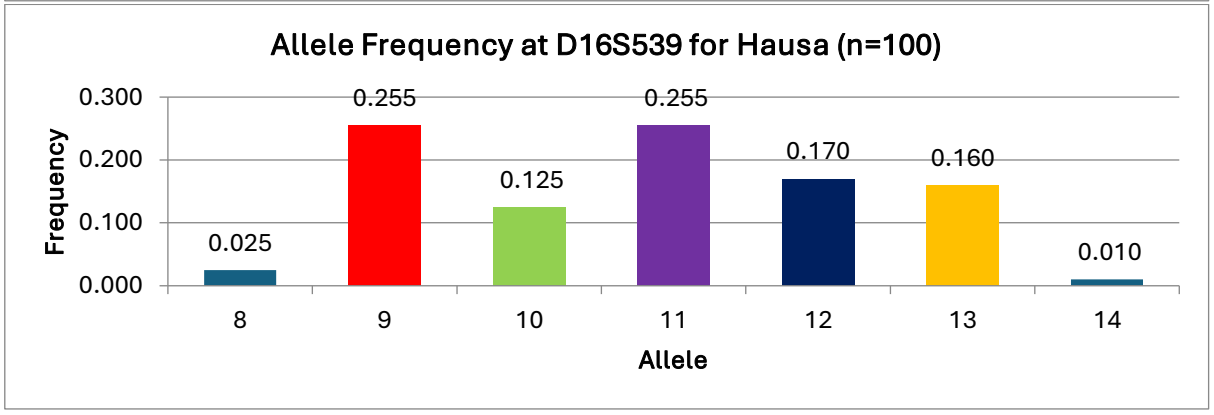
## APPENDIX 8

## ALLELE FREQUENCIES AT D16S539 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



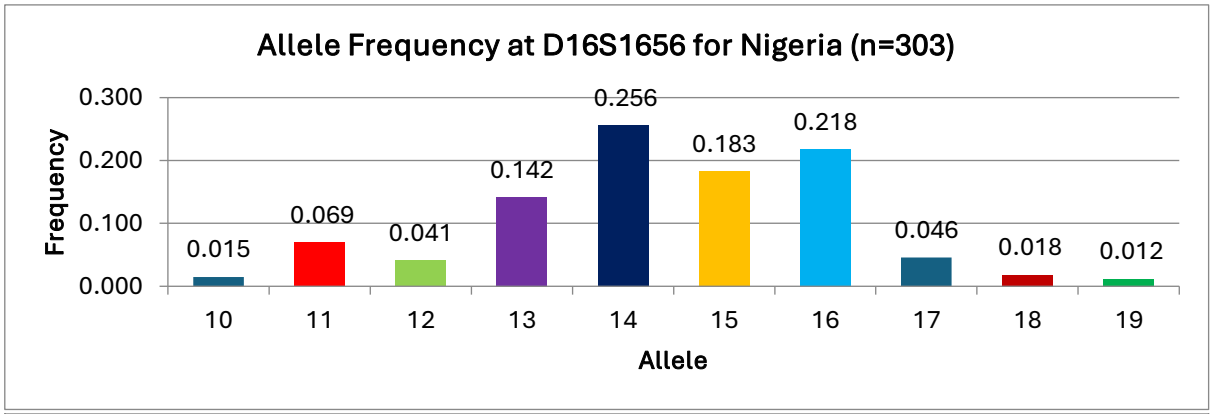



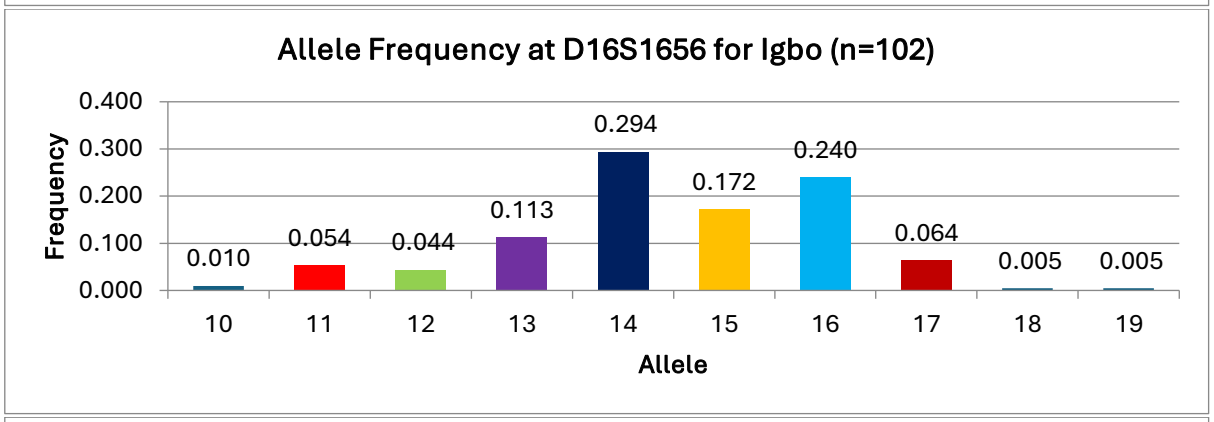



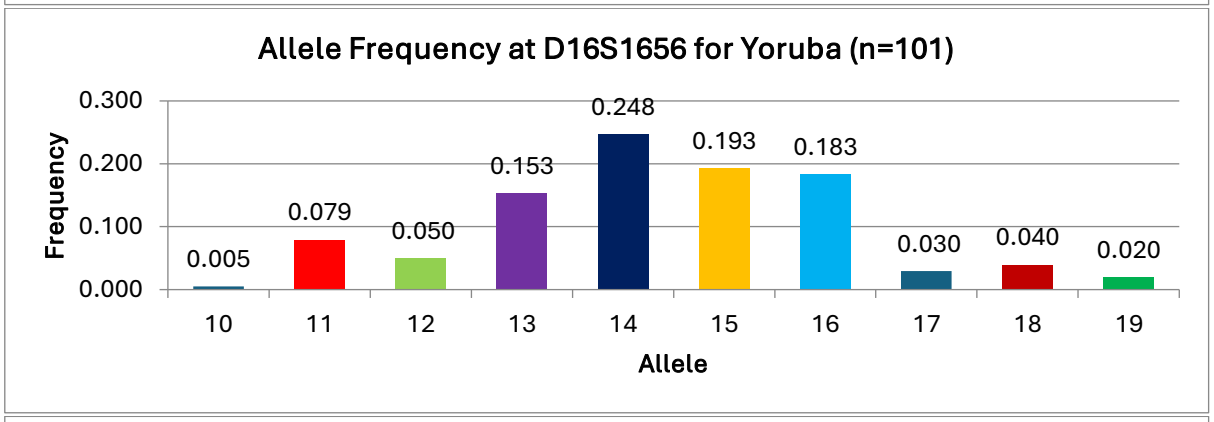



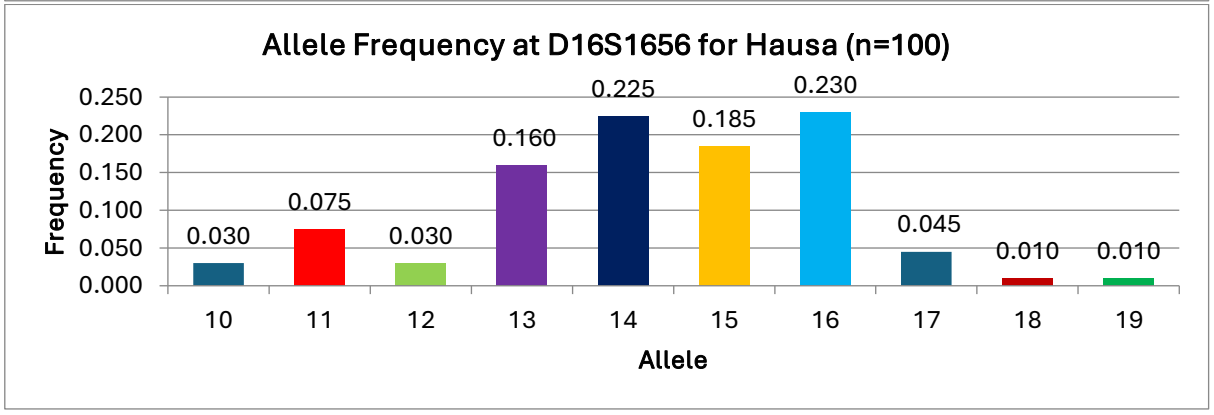
## APPENDIX 9

## ALLELE FREQUENCIES AT D16S1656 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



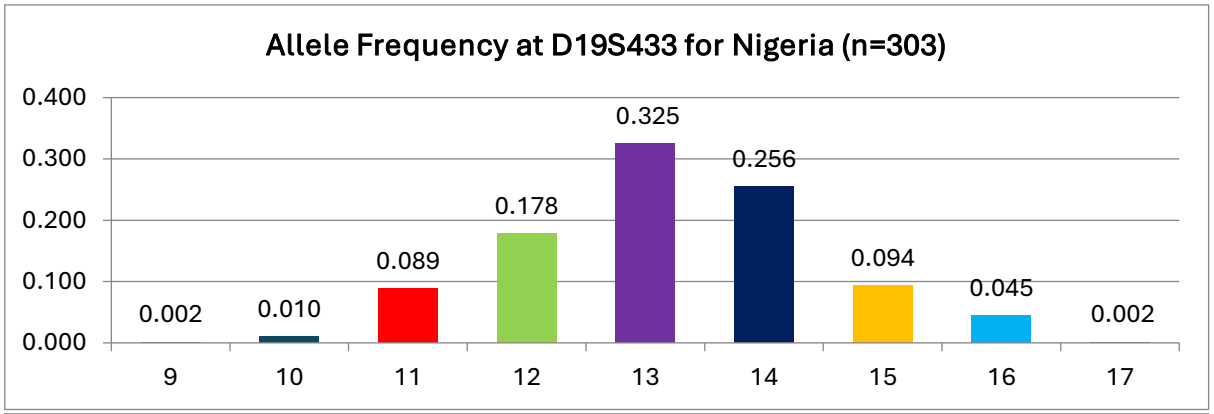



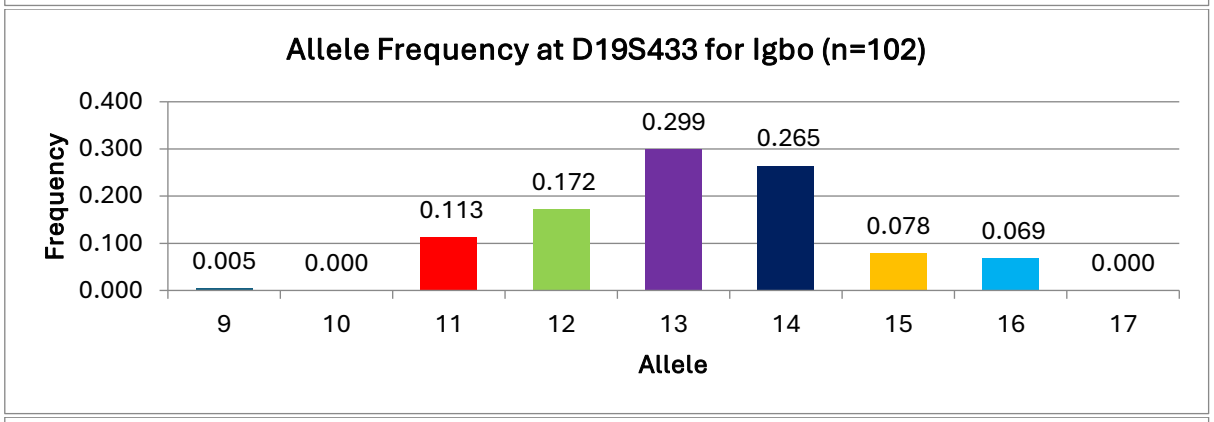



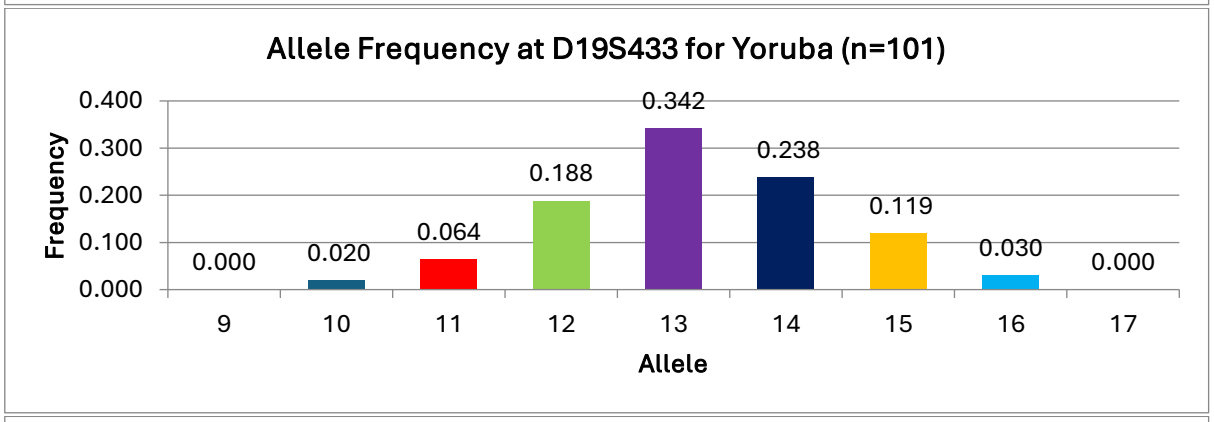



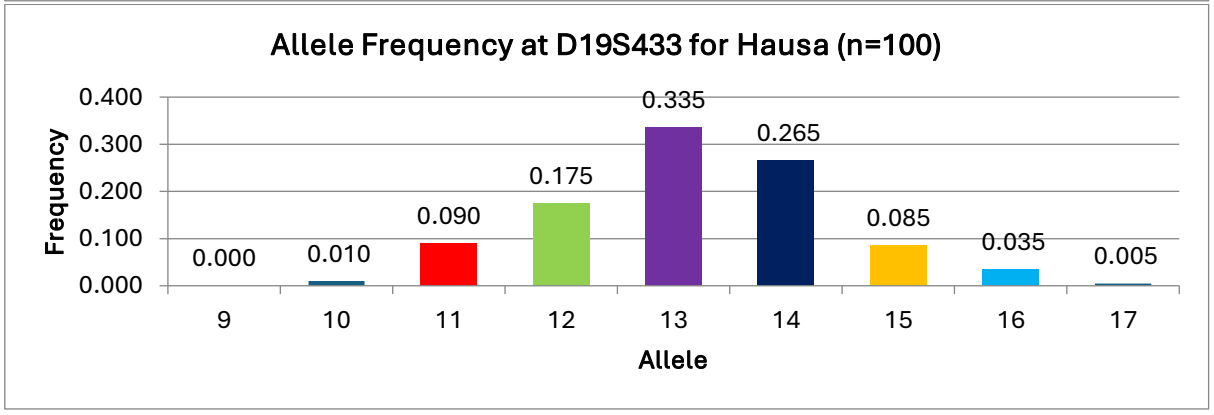
## APPENDIX 10

## ALLELE FREQUENCIES AT D19433 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



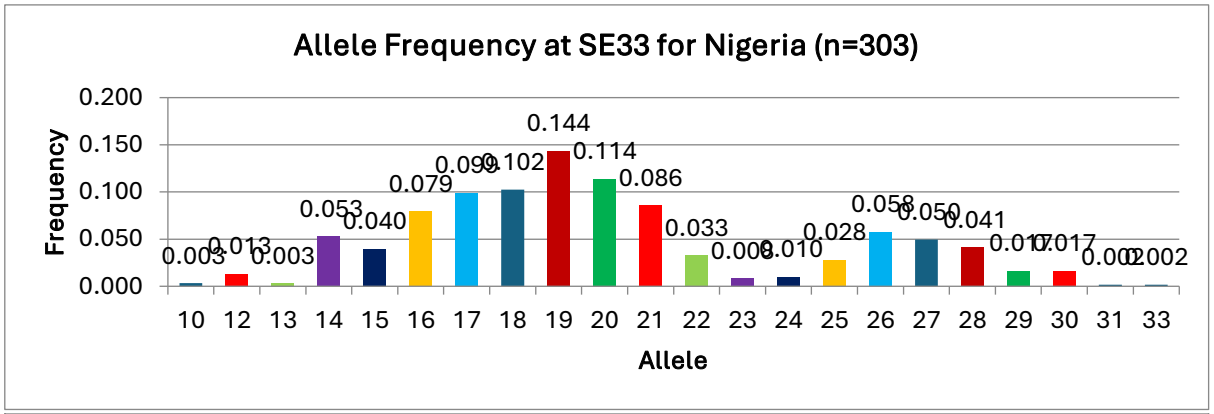



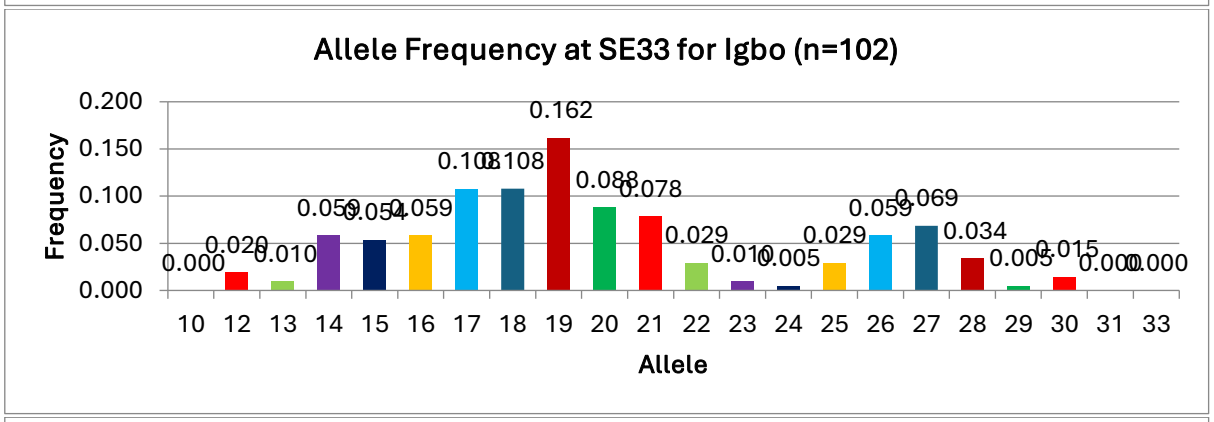



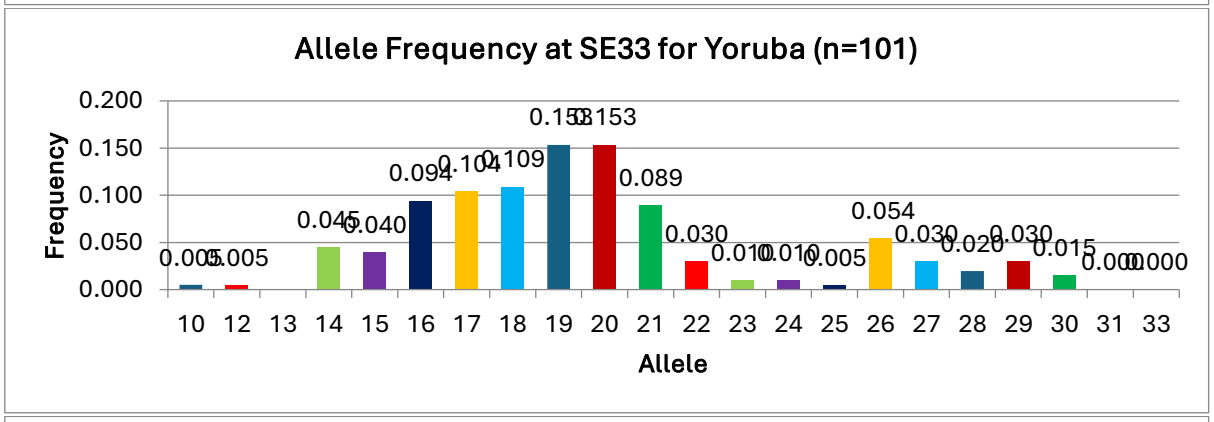



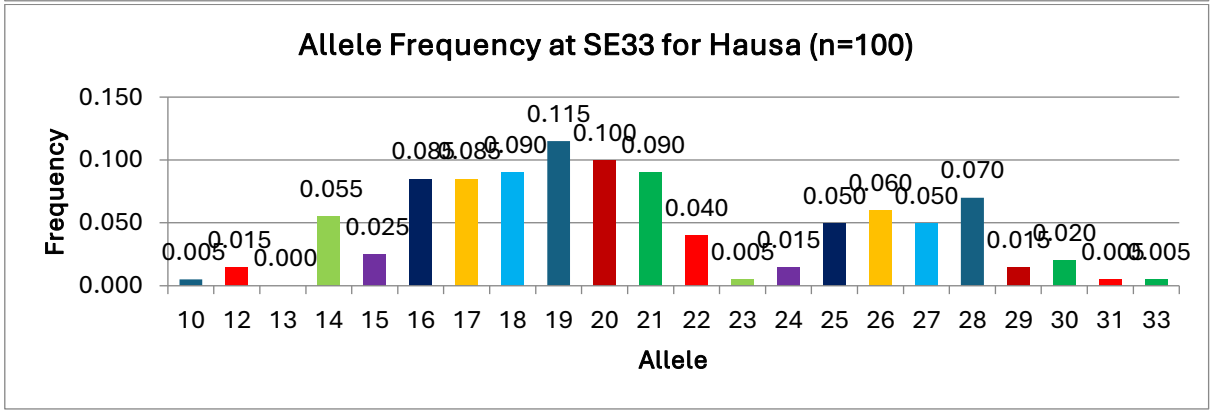
## APPENDIX 11

## ALLELE FREQUENCIES AT SE33 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



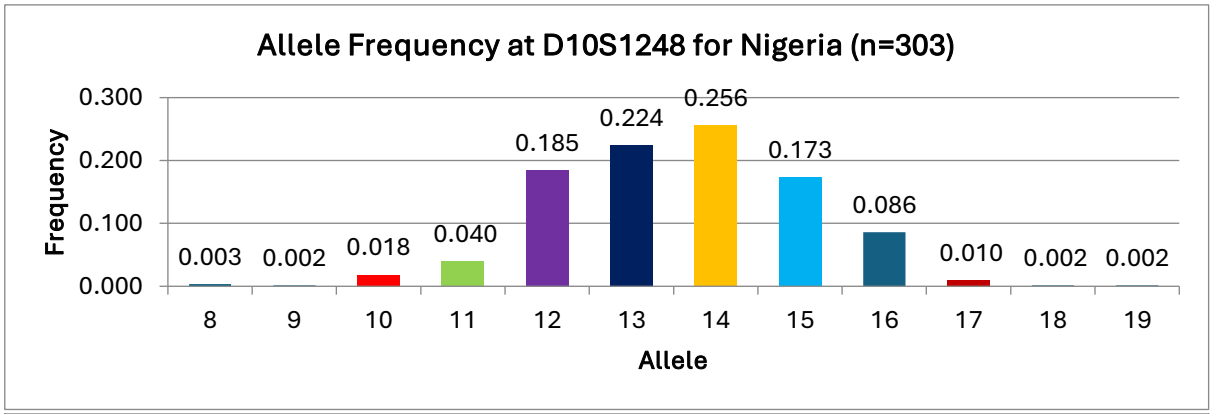



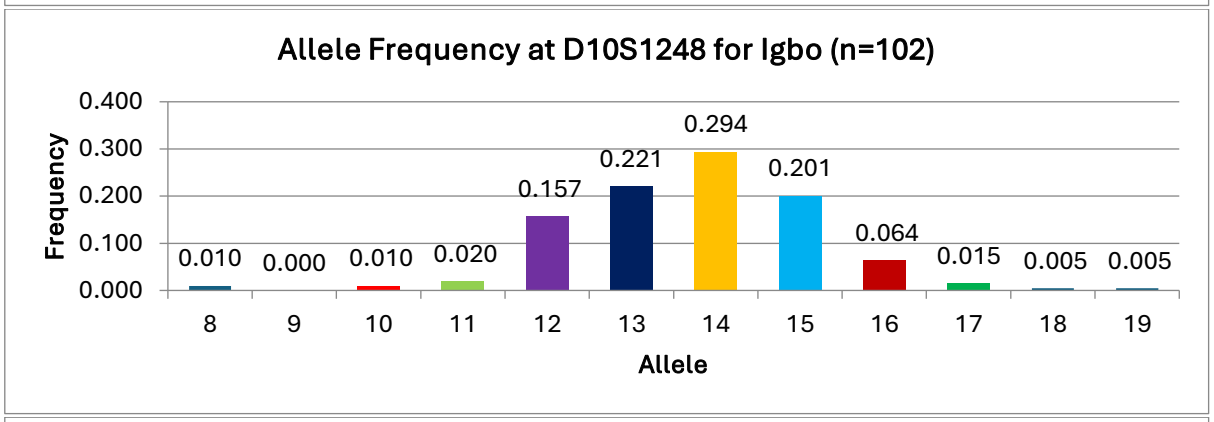



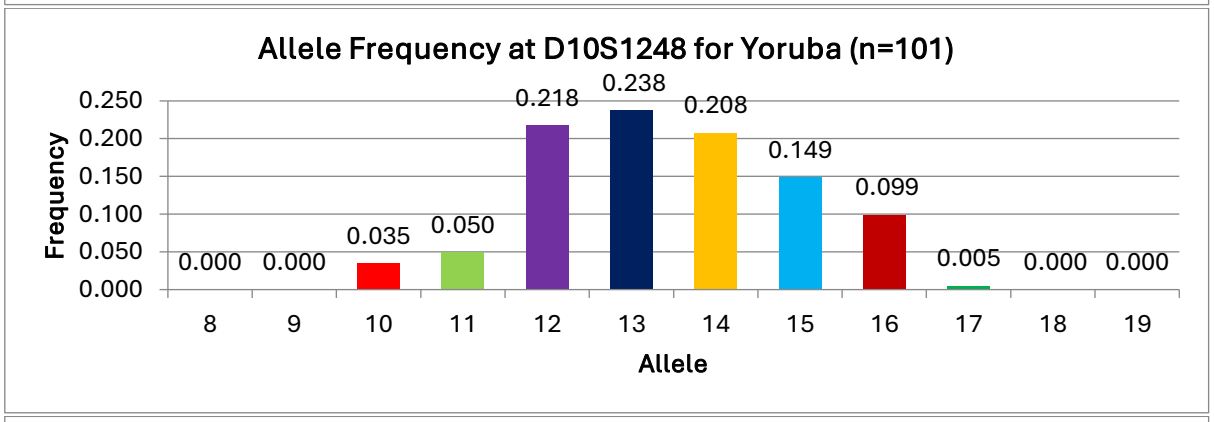



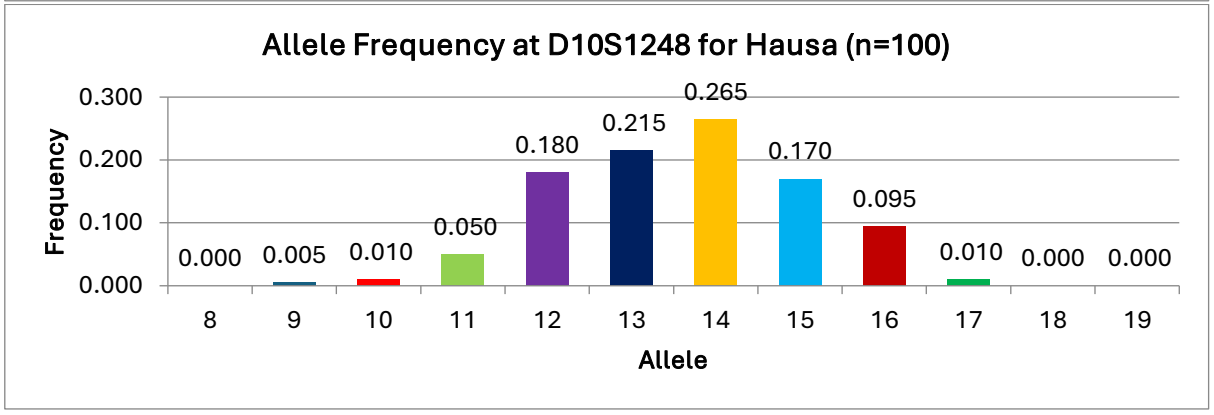
## APPENDIX 12

## ALLELE FREQUENCIES AT D10S1248 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



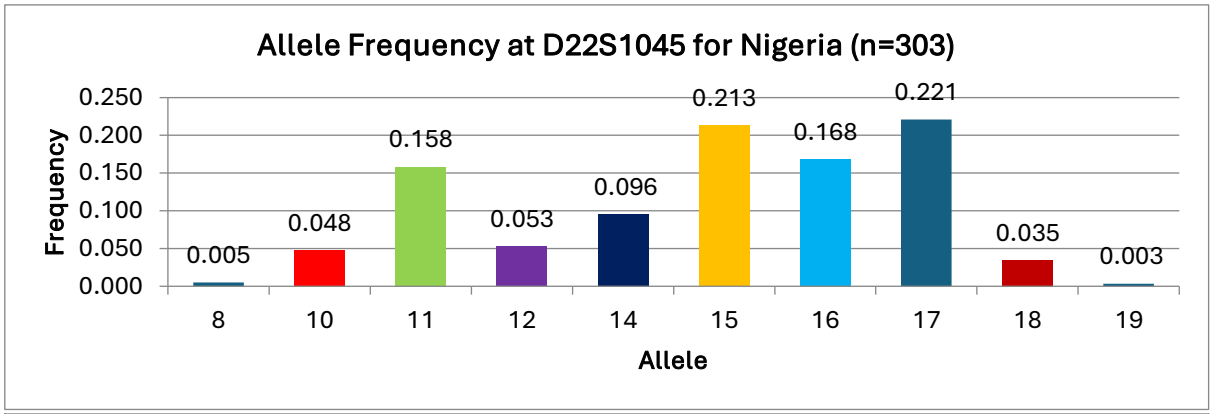



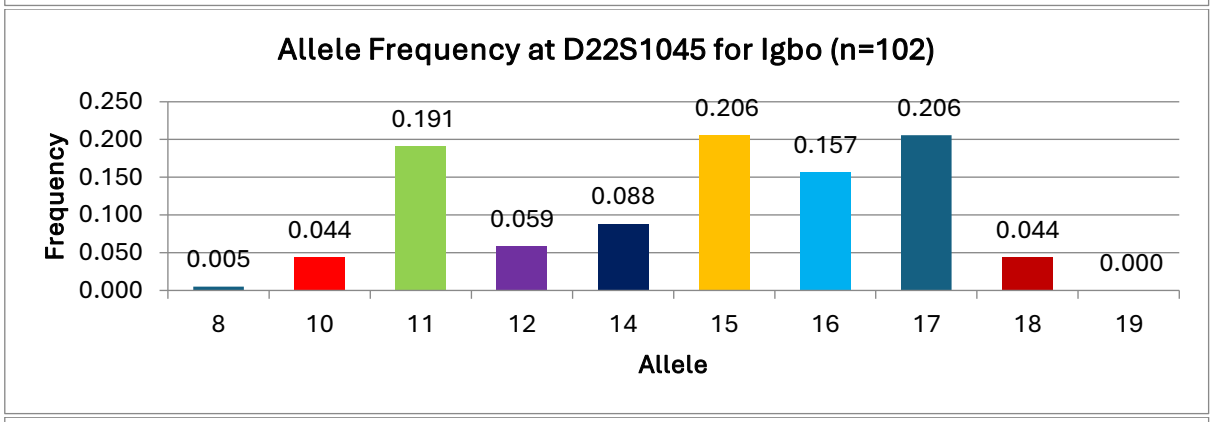



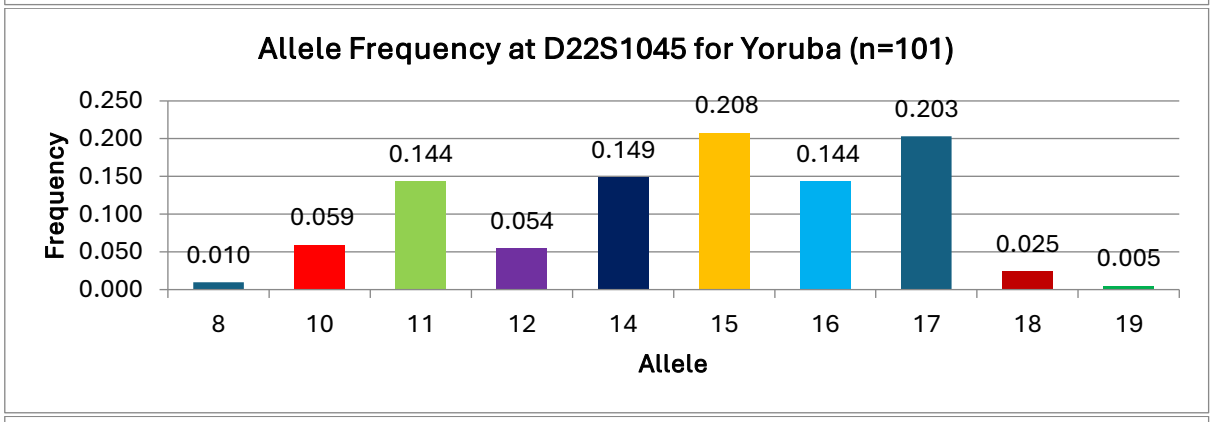



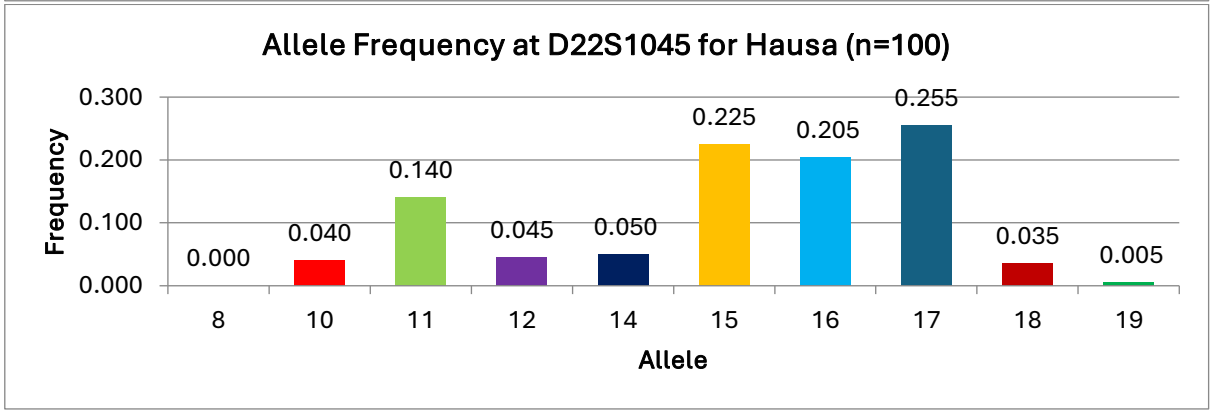
## APPENDIX 13

## ALLELE FREQUENCIES AT D22S1045 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



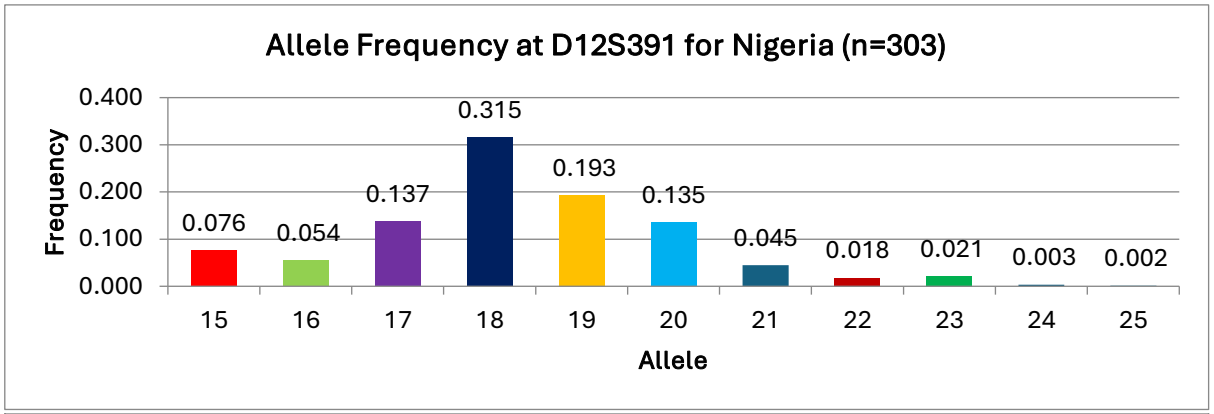



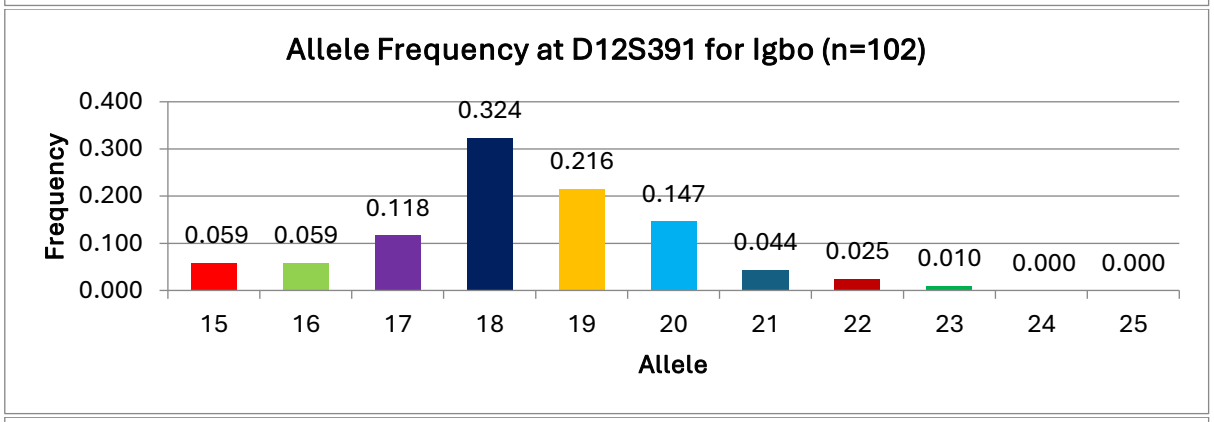



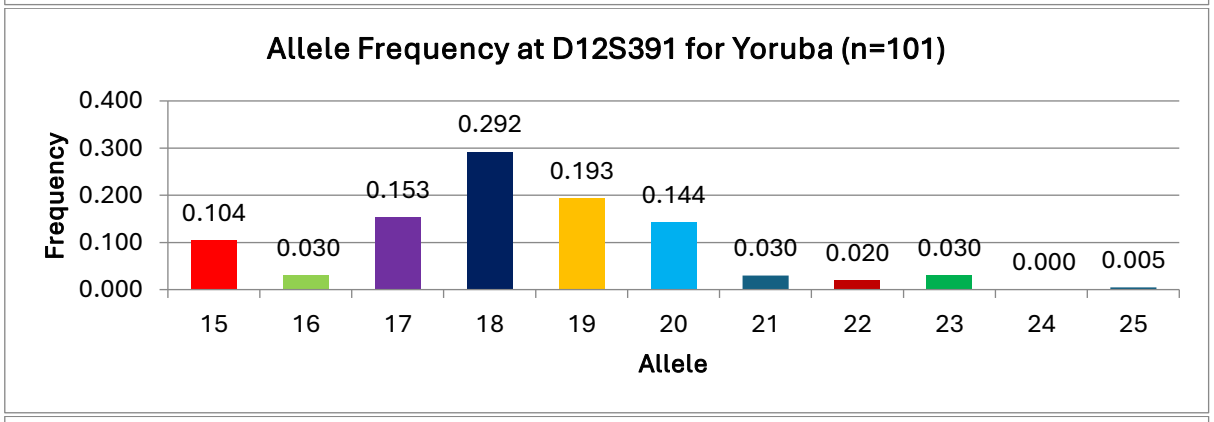



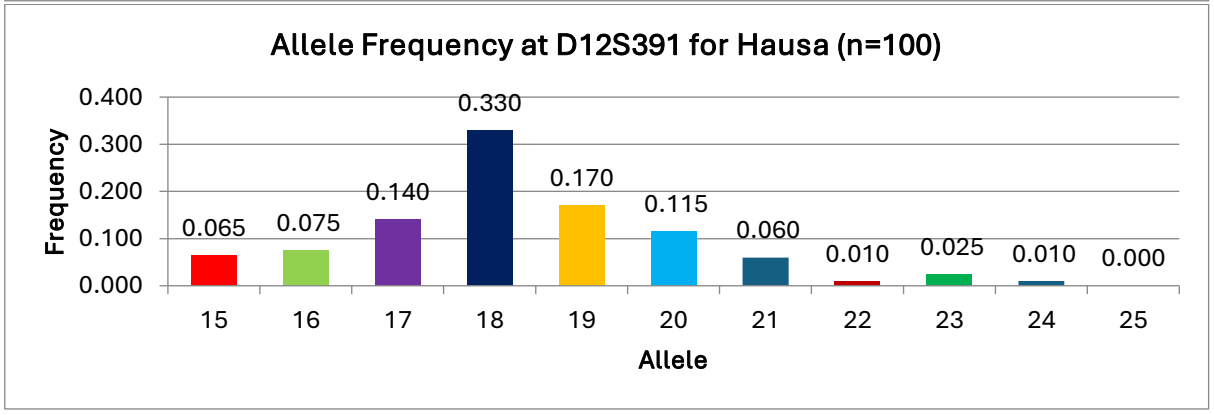
## APPENDIX 14

## ALLELE FREQUENCIES AT D12S391 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



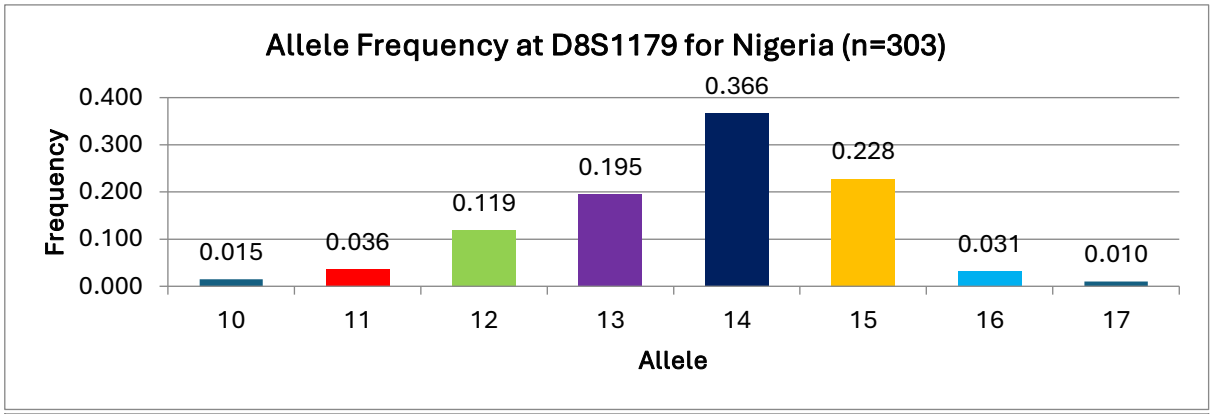



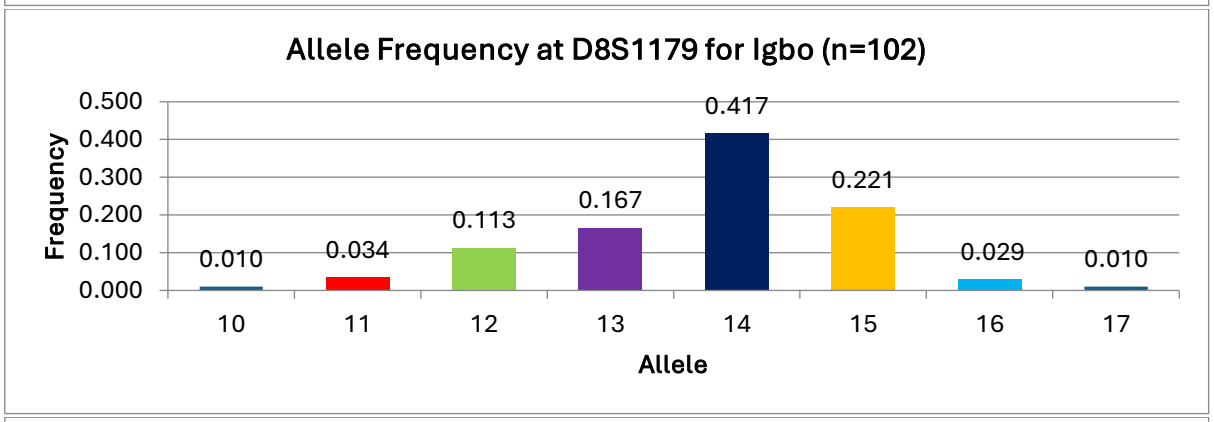



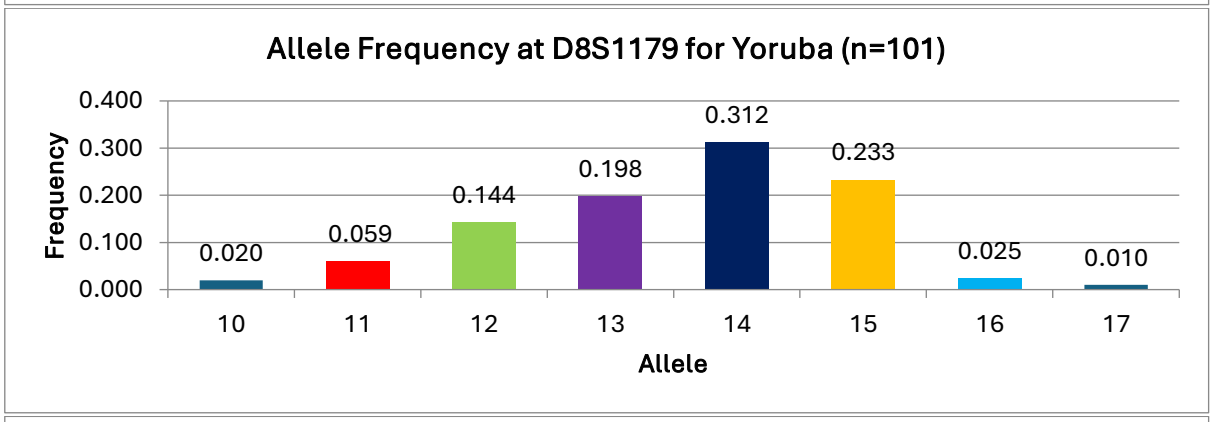



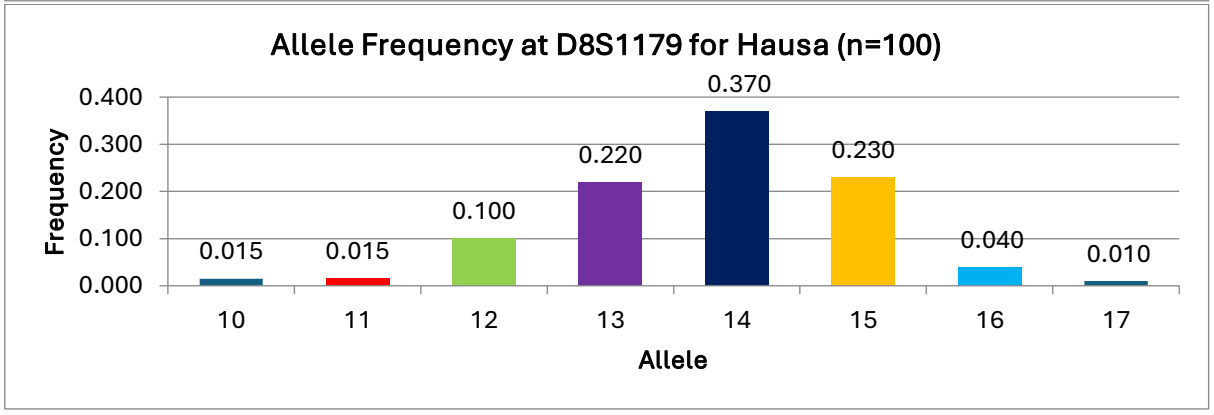
## APPENDIX 15

## ALLELE FREQUENCIES AT D8S1179 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



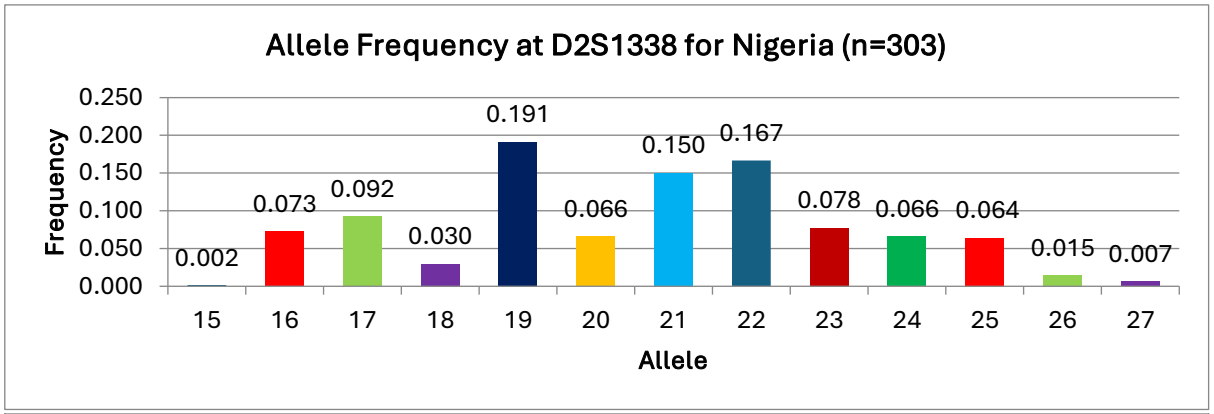



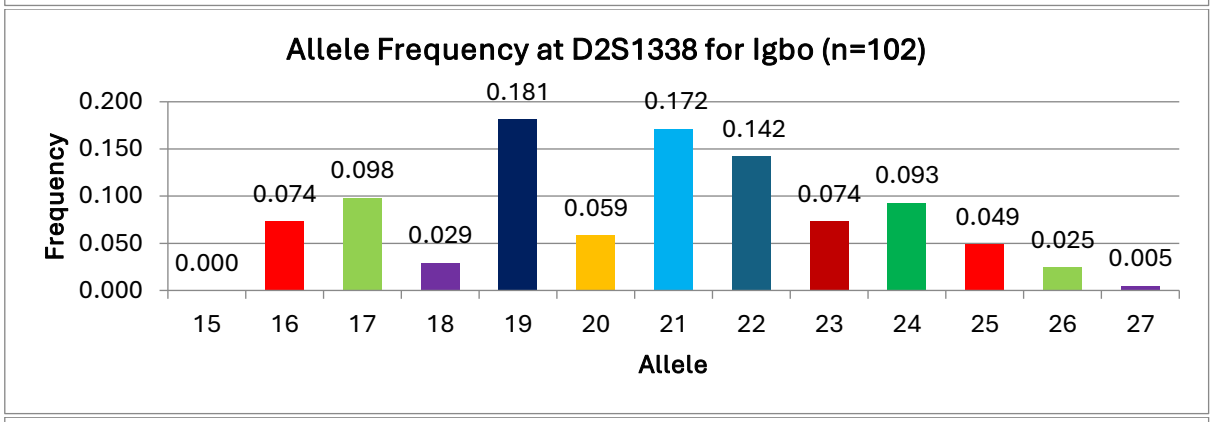



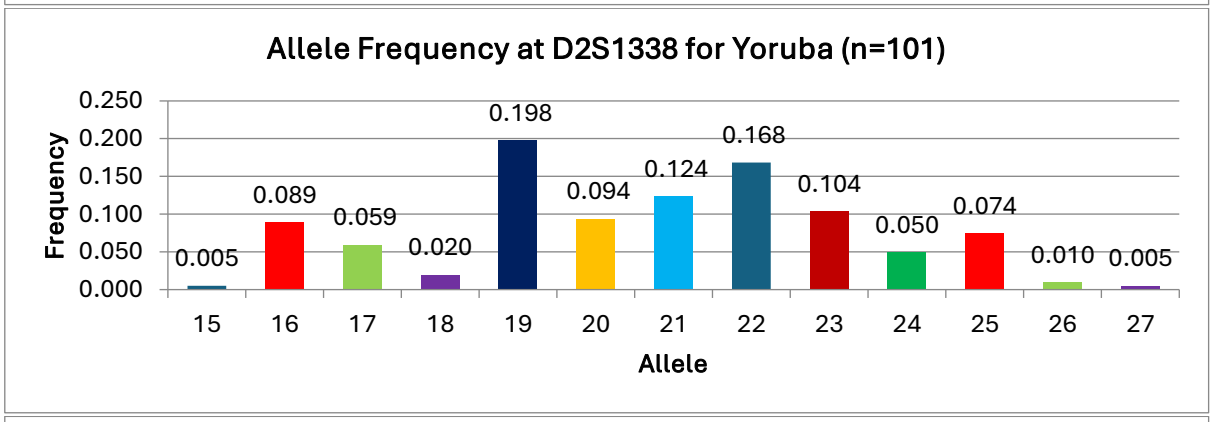



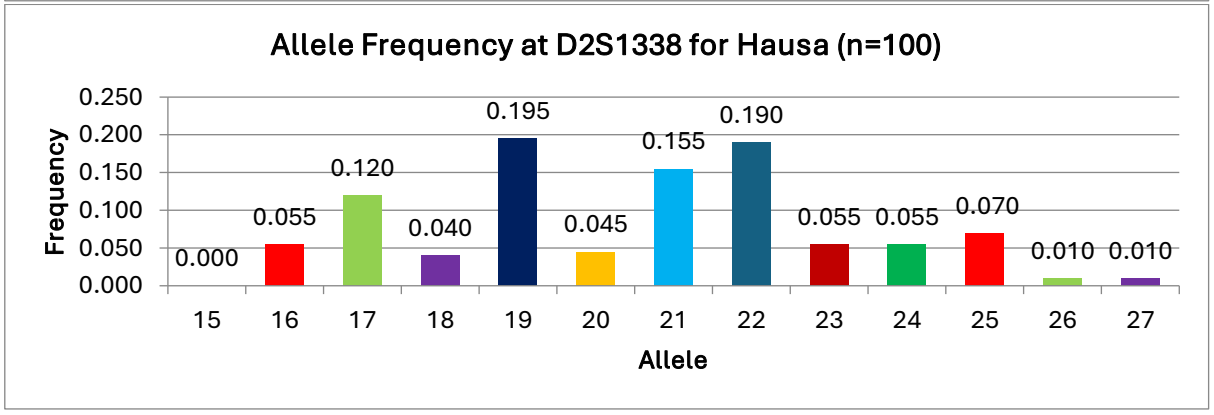
## APPENDIX 16

## ALLELE FREQUENCIES AT D2S1338 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



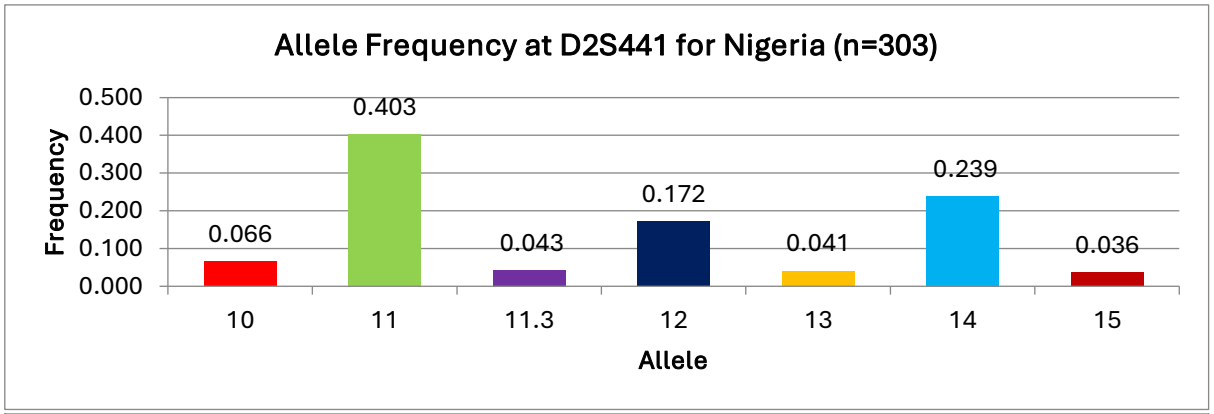



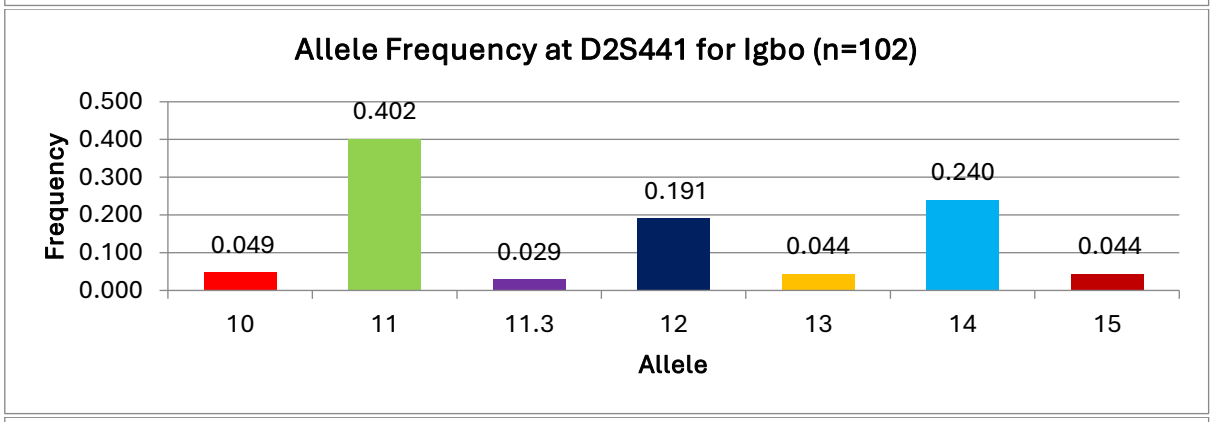



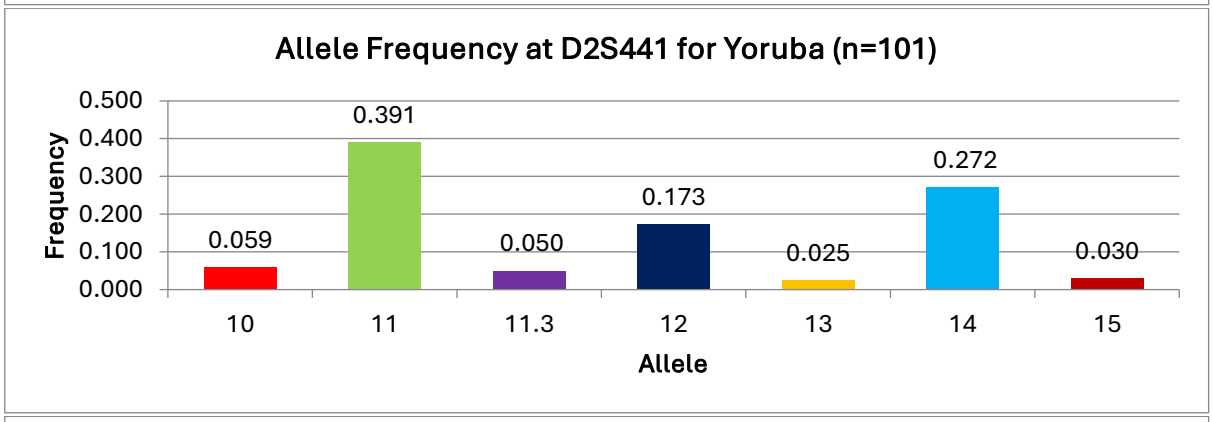



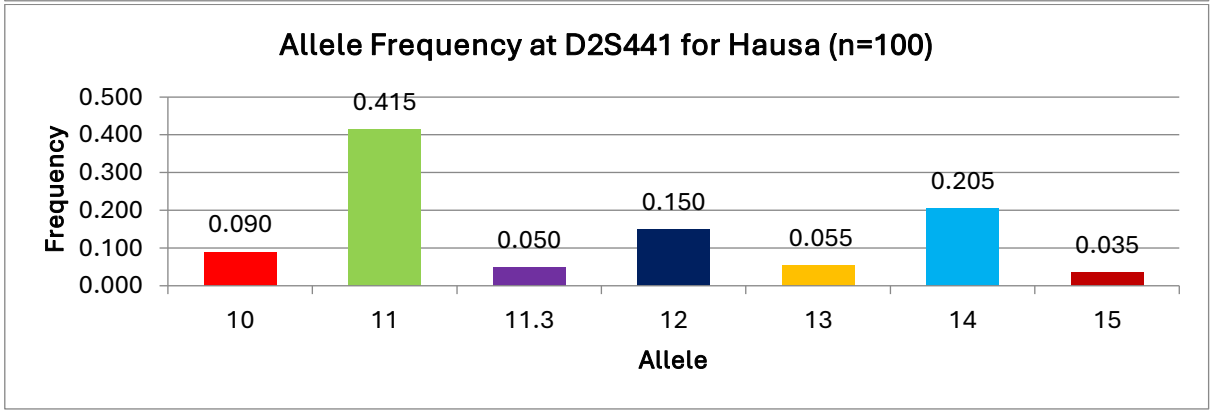
## APPENDIX 17

## ALLELE FREQUENCIES AT D2S441 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



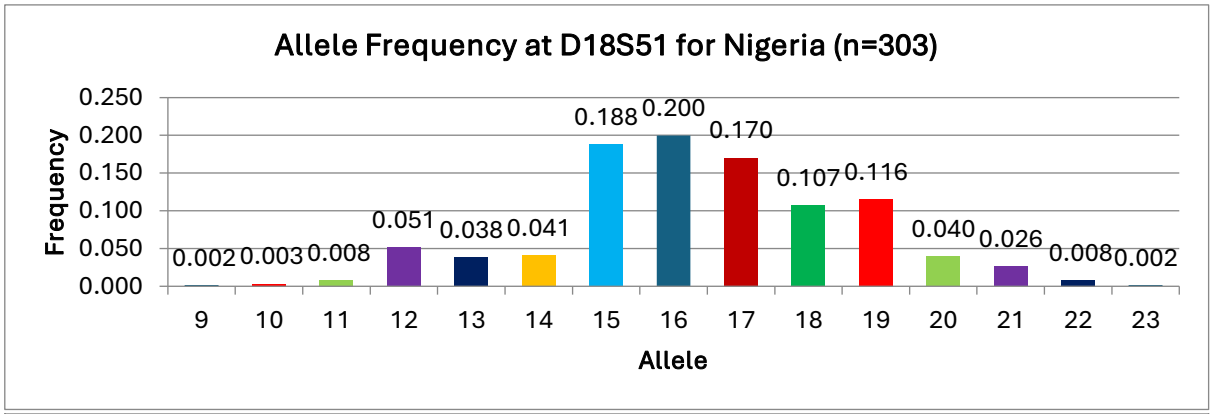



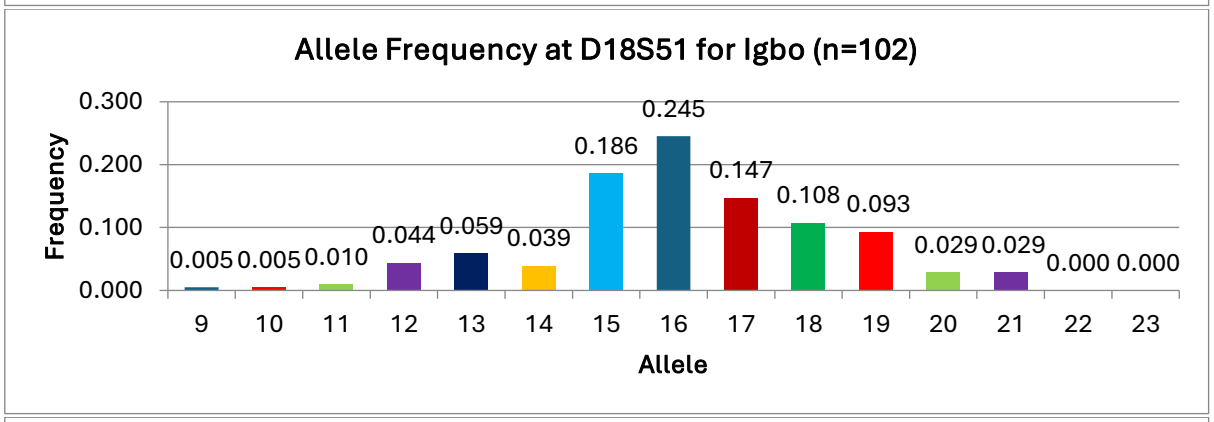



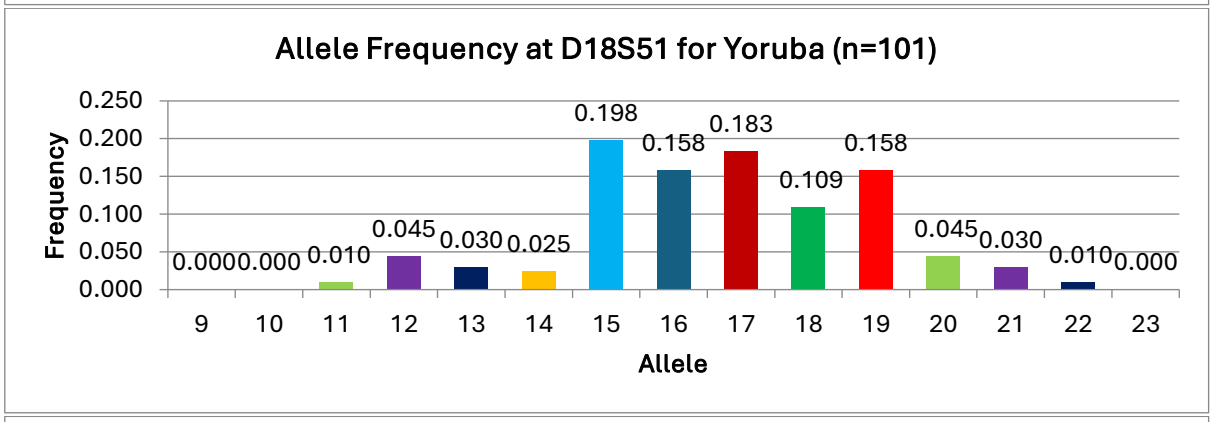



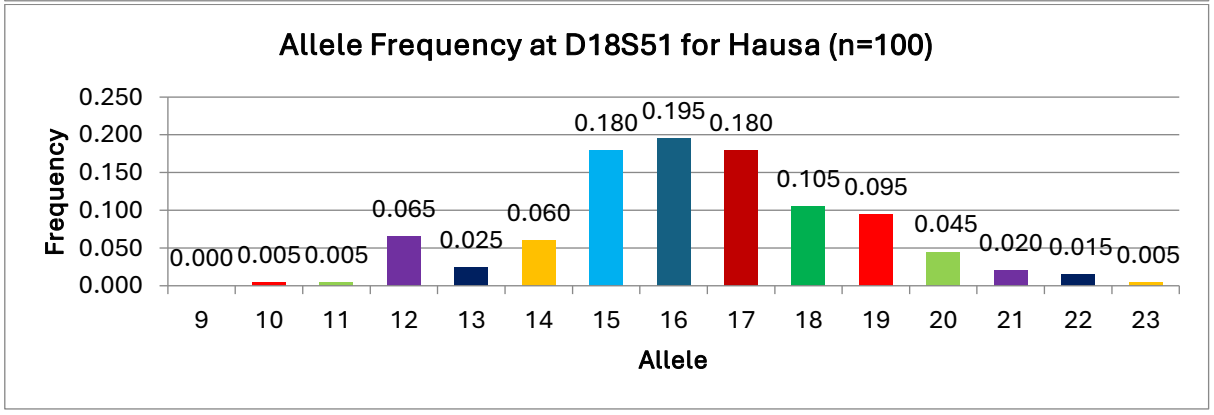
## APPENDIX 18

## ALLELE FREQUENCIES AT D18S51 WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



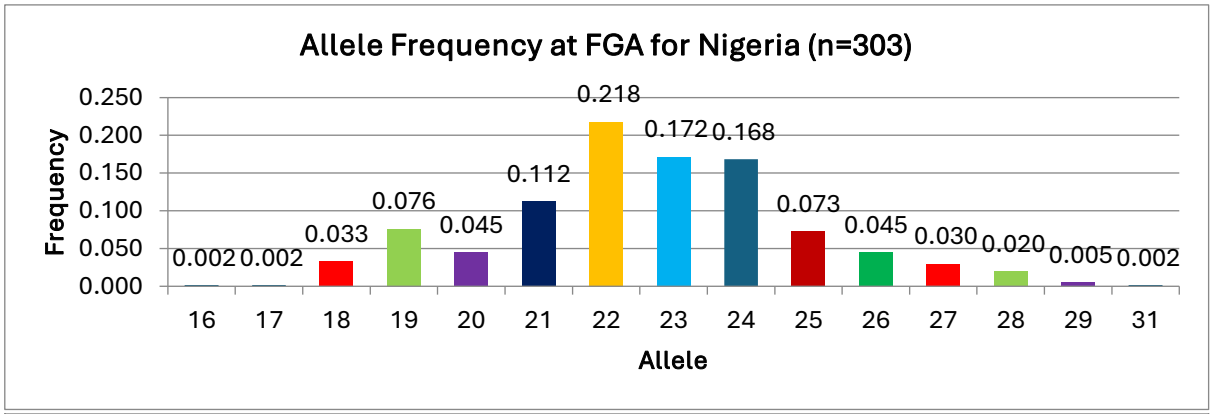



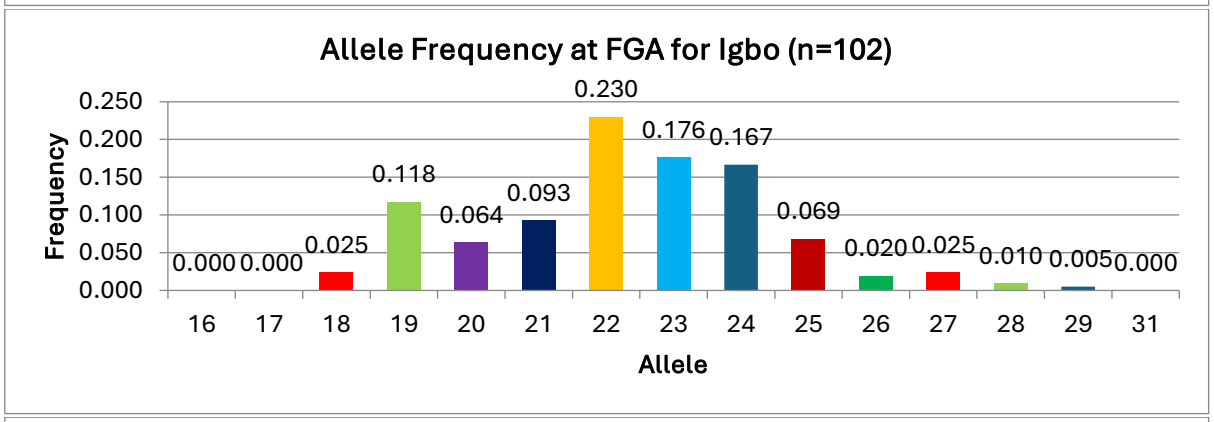



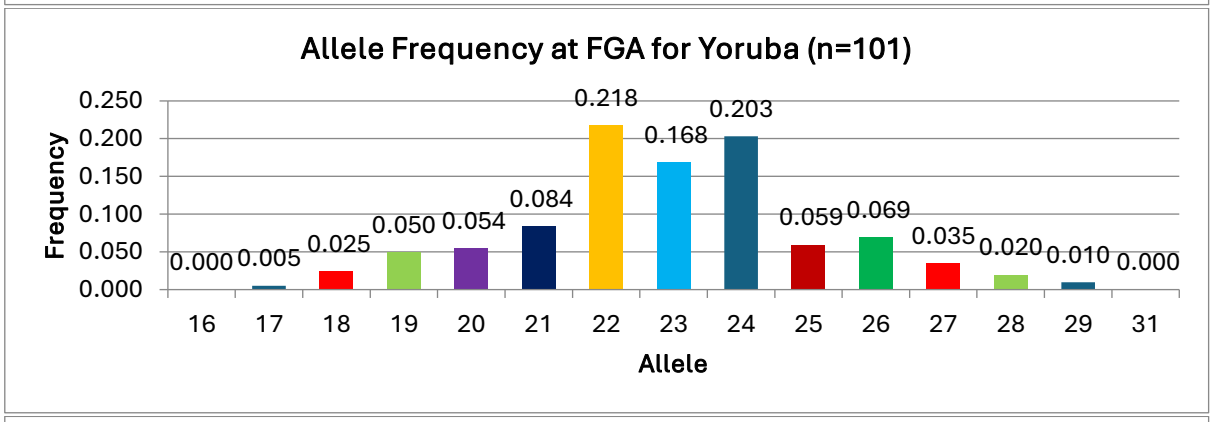



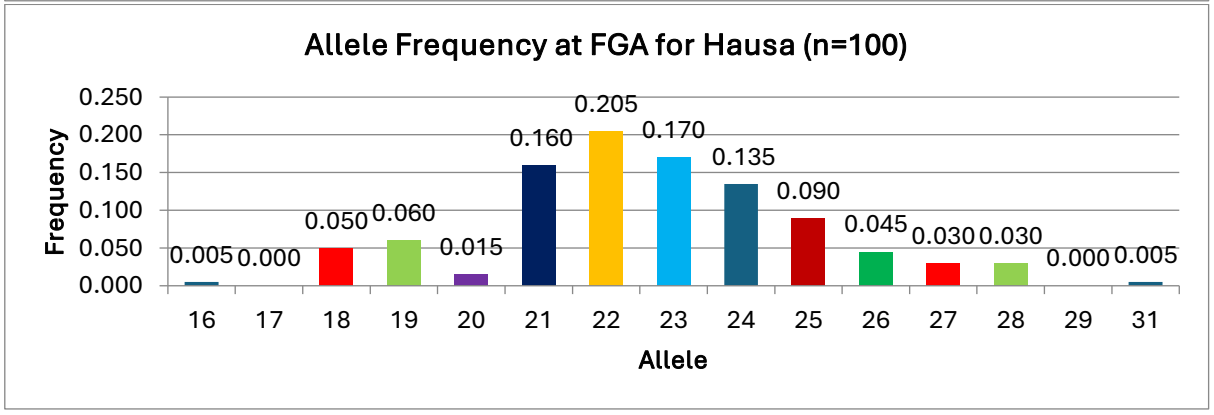
## APPENDIX 19

## ALLELE FREQUENCIES AT FGA WITH GRAPHS BY POPULATION FOR QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT



## APPENDIX 20

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM EXACT TEST ANALYSES FOR THE NIGERIAN POPULATION USING THE QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Nigerians | TH01 | 303 | 6 | 12.82405 | 0.976605 | 1 | False | False |
| Nigerians | D3S1358 | 303 | 9 | 7.28673 | 0.995001 | 1 | False | False |
| Nigerians | vWA | 303 | 10 | 30.54965 | 0.784243 | 1 | False | False |
| Nigerians | D21S11 | 303 | 16 | 80.368 | 0.063187 | 1 | False | False |
| Nigerians | D16S539 | 303 | 7 | 9.203718 | 0.9976 | 1 | False | False |
| Nigerians | D1S1656 | 303 | 16 | 40.32146 | 0.443511 | 1 | False | False |
| Nigerians | D19S433 | 303 | 14 | 48.23306 | 0.442511 | 1 | False | False |
| Nigerians | SE33 | 303 | 34 | 270.9083 | 0.15237 | 1 | False | False |
| Nigerians | D10S1248 | 303 | 12 | 115.0732 | 0.763647 | 1 | False | False |
| Nigerians | D22S1045 | 303 | 10 | 9.33361 | 0.786043 | 1 | False | False |
| Nigerians | D12S391 | 303 | 14 | 34.72459 | 0.838632 | 1 | False | False |
| Nigerians | D8S1179 | 303 | 8 | 7.397778 | 0.988802 | 1 | False | False |
| Nigerians | D2S1338 | 303 | 13 | 57.97766 | 0.09818 | 1 | False | False |
| Nigerians | D2S441 | 303 | 9 | 13.40862 | 0.886623 | 1 | False | False |
| Nigerians | D18S51 | 303 | 18 | 74.38051 | 0.829234 | 1 | False | False |
| Nigerians | FGA | 303 | 19 | 75.43546 | 0.458908 | 1 | False | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 21

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM EXACT TEST ANALYSES FOR THE IGBO SUBPOPULATION USING THE QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Igbo | TH01 | 102 | 6 | 12.56558 | 0.673665 | 1 | False | False |
| Igbo | D3S1358 | 102 | 7 | 9.178982 | 0.869626 | 1 | False | False |
| Igbo | vWA | 102 | 9 | 15.0321 | 0.329934 | 1 | False | False |
| Igbo | D21S11 | 102 | 14 | 84.04559 | 0.054189 | 0.867027 | False | False |
| Igbo | D16S539 | 102 | 7 | 12.61546 | 0.604479 | 1 | False | False |
| Igbo | D1S1656 | 102 | 13 | 42.11756 | 0.241152 | 1 | False | False |
| Igbo | D19S433 | 102 | 11 | 8.336365 | 0.995401 | 1 | False | False |
| Igbo | SE33 | 102 | 24 | 125.243 | 0.212158 | 1 | False | False |
| Igbo | D10S1248 | 102 | 11 | 81.37624 | 0.095781 | 1 | False | False |
| Igbo | D22S1045 | 102 | 9 | 9.864757 | 0.59888 | 1 | False | False |
| Igbo | D12S391 | 102 | 9 | 22.48101 | 0.690462 | 1 | False | False |
| Igbo | D8S1179 | 102 | 8 | 12.29534 | 0.680864 | 1 | False | False |
| Igbo | D2S1338 | 102 | 12 | 45.57405 | 0.175765 | 1 | False | False |
| Igbo | D2S441 | 102 | 7 | 17.77555 | 0.422316 | 1 | False | False |
| Igbo | D18S51 | 102 | 15 | 23.75232 | 0.242951 | 1 | False | False |
| Igbo | FGA | 102 | 14 | 38.54851 | 0.227355 | 1 | False | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 22

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM EXACT TEST ANALYSES FOR THE YORUBA SUBPOPULATION USING THE QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Yoruba | TH01 | 101 | 5 | 5.382998 | 0.897021 | 1 | False | False |
| Yoruba | D3S1358 | 101 | 7 | 11.72803 | 0.586083 | 1 | False | False |
| Yoruba | vWA | 101 | 10 | 37.38235 | 0.238352 | 1 | False | False |
| Yoruba | D21S11 | 101 | 15 | 117.0436 | 0.010398 | 0.166367 | True | False |
| Yoruba | D16S539 | 101 | 7 | 12.3066 | 0.408118 | 1 | False | False |
| Yoruba | D1S1656 | 101 | 15 | 31.7913 | 0.154369 | 1 | False | False |
| Yoruba | D19S433 | 101 | 12 | 31.98178 | 0.40012 | 1 | False | False |
| Yoruba | SE33 | 101 | 27 | 101.3085 | 0.40052 | 1 | False | False |
| Yoruba | D10S1248 | 101 | 8 | 8.798508 | 0.881624 | 1 | False | False |
| Yoruba | D22S1045 | 101 | 10 | 11.7162 | 0.091182 | 1 | False | False |
| Yoruba | D12S391 | 101 | 11 | 20.01557 | 0.517696 | 1 | False | False |
| Yoruba | D8S1179 | 101 | 8 | 8.853617 | 0.955409 | 1 | False | False |
| Yoruba | D2S1338 | 101 | 13 | 33.26308 | 0.389722 | 1 | False | False |
| Yoruba | D2S441 | 101 | 8 | 5.690595 | 0.904419 | 1 | False | False |
| Yoruba | D18S51 | 101 | 13 | 30.90531 | 0.925015 | 1 | False | False |
| Yoruba | FGA | 101 | 17 | 32.15183 | 0.579284 | 1 | False | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 23

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM EXACT TEST ANALYSES FOR THE HAUSA-FULANI SUBPOPULATION USING THE QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Hausa-Fulani | TH01 | 100 | 5 | 2.156385 | 0.991602 | 1 | False | False |
| Hausa-Fulani | D3S1358 | 100 | 8 | 6.010559 | 0.544491 | 1 | False | False |
| Hausa-Fulani | vWA | 100 | 9 | 18.66617 | 0.337333 | 1 | False | False |
| Hausa-Fulani | D21S11 | 100 | 16 | 85.01735 | 0.024195 | 0.387123 | True | False |
| Hausa-Fulani | D16S539 | 100 | 7 | 7.997039 | 0.910818 | 1 | False | False |
| Hausa-Fulani | D1S1656 | 100 | 17 | 52.47816 | 0.268946 | 1 | False | False |
| Hausa-Fulani | D19S433 | 100 | 8 | 32.47814 | 0.176941 | 1 | False | False |
| Hausa-Fulani | SE33 | 100 | 28 | 180.5768 | 0.943011 | 1 | False | False |
| Hausa-Fulani | D10S1248 | 100 | 9 | 7.630565 | 0.829634 | 1 | False | False |
| Hausa-Fulani | D22S1045 | 100 | 9 | 13.57248 | 0.494701 | 1 | False | False |
| Hausa-Fulani | D12S391 | 100 | 12 | 15.42037 | 0.882424 | 1 | False | False |
| Hausa-Fulani | D8S1179 | 100 | 8 | 5.151131 | 0.990802 | 1 | False | False |
| Hausa-Fulani | D2S1338 | 100 | 12 | 94.52539 | 0.033989 | 0.543819 | True | False |
| Hausa-Fulani | D2S441 | 100 | 8 | 6.439795 | 0.876825 | 1 | False | False |
| Hausa-Fulani | D18S51 | 100 | 15 | 60.97538 | 0.733453 | 1 | False | False |
| Hausa-Fulani | FGA | 100 | 15 | 58.96937 | 0.614277 | 1 | False | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 24

## NEI'S GENETIC DISTANCE PAIRWISE POPULATION MATRIX FOR THE QIAGEN™ INVESTIGATOR® ESSPLEX SE QS KIT

| Sub-Population | Igbo | Yoruba | Hausa-Fulani | Egypt |
|---|---|---|---|---|
| **Igbo** | 0.000 | | | |
| **Yoruba** | 0.004 | 0.000 | | |
| **Hausa-Fulani** | 0.006 | 0.005 | 0.000 | |
| **Egypt** | 0.088 | 0.087 | 0.086 | 0.000 |

***Egypt is included as an outgroup.**

## APPENDIX 25

## NEI'S GENETIC IDENTITY PAIRWISE POPULATION MATRIX FOR THE QIAGEN™ INVESTIGATOR® ESSPLEX SE QS KIT

| Sub-Population | Igbo | Yoruba | Hausa-Fulani | Egypt |
|---|---|---|---|---|
| **Igbo** | 1.000 | | | |
| **Yoruba** | 0.996 | 1.000 | | |
| **Hausa-Fulani** | 0.994 | 0.995 | 1.000 | |
| **Egypt** | 0.912 | 0.913 | 0.914 | 1.000 |

***Egypt is included as an outgroup.**

## APPENDIX 26

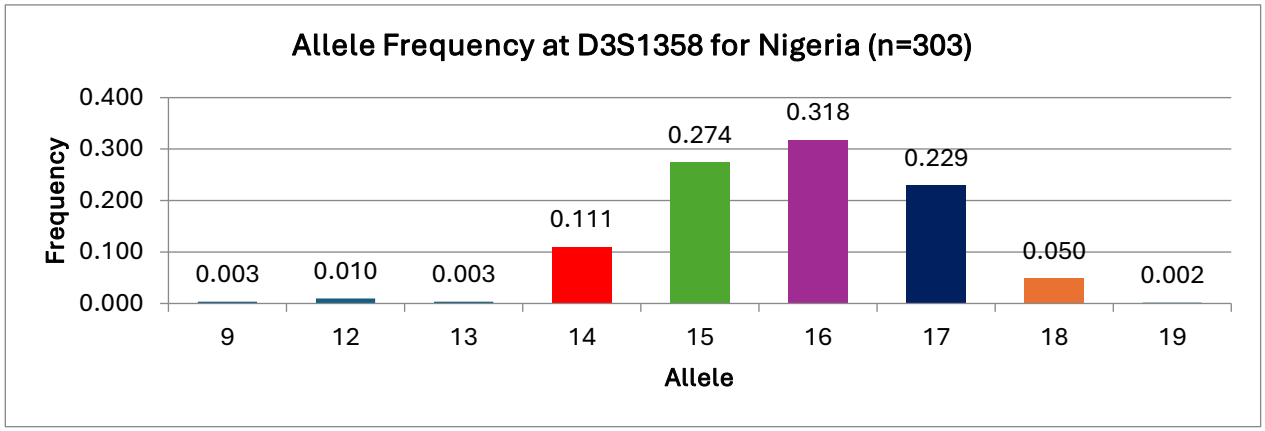
## MEAN LnP(K)±SD AND DELTA *K* VALUES FOR THE QIAGEN™ INVESTIGATOR ESSPLEX SE QS KIT

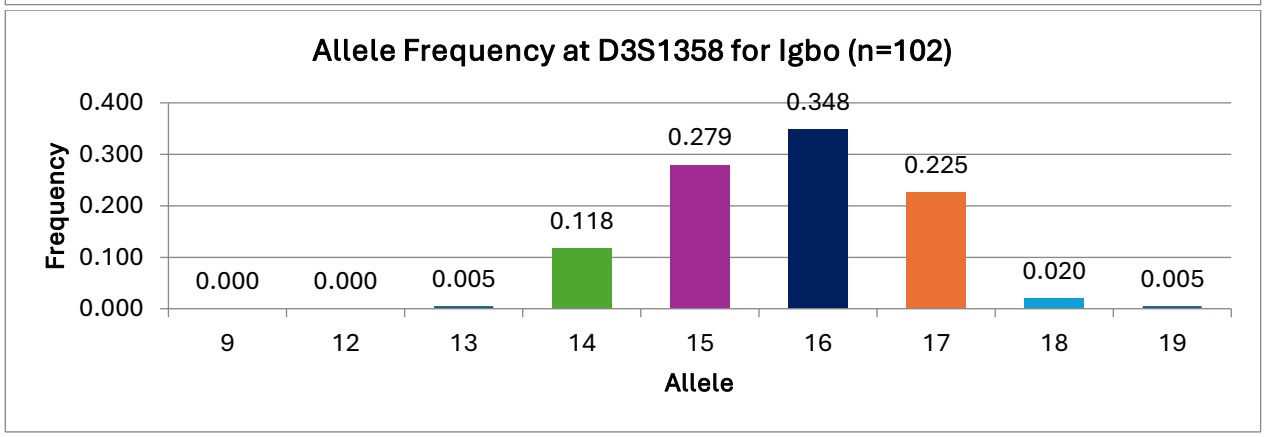
| *K* | Reps | Mean LnP(*K*) | SD LnP(*K*) | Ln'(*K*) | \|Ln''(K)\| | Delta *K* |
|---|---|---|---|---|---|---|
| 1 | 5 | -19120.9 | 0.75166 | NA | NA | NA |
| 2 | 5 | -19191.1 | 18.32389 | -70.2 | 10.92 | 0.59594 |
| 3 | 5 | -19272.2 | 58.03548 | -81.12 | 129.14 | 2.22519 |
| 4 | 5 | -19482.5 | 141.7526 | -210.26 | 236.12 | 1.66572 |
| 5 | 5 | -19456.6 | 212.6896 | 25.86 | 610.8 | 2.87179 |
| 6 | 5 | -20041.6 | 429.2856 | -584.94 | 830.56 | 1.93475 |
| 7 | 5 | -19795.9 | 643.0521 | 245.62 | 234 | 0.36389 |
| 8 | 5 | -19784.3 | 478.9833 | 11.62 | 378.8 | 0.79084 |
| 9 | 5 | -20151.5 | 1068.673 | -367.18 | 631.7 | 0.59111 |
| 10 | 5 | -19887 | 552.3188 | 264.52 | NA | NA |

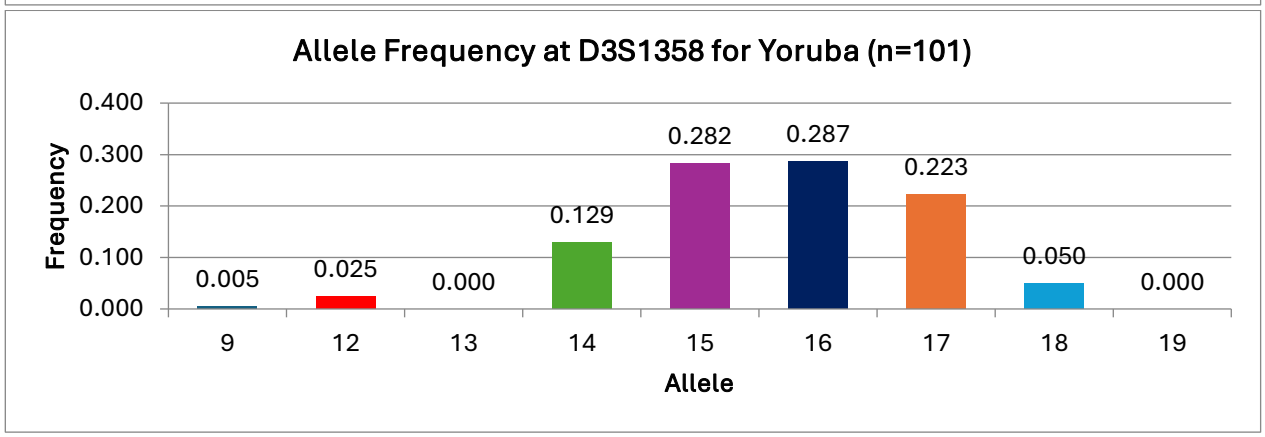
The peak of Delta *K* (*K* = 5), indicating the optimal number of clusters, is highlighted in red.

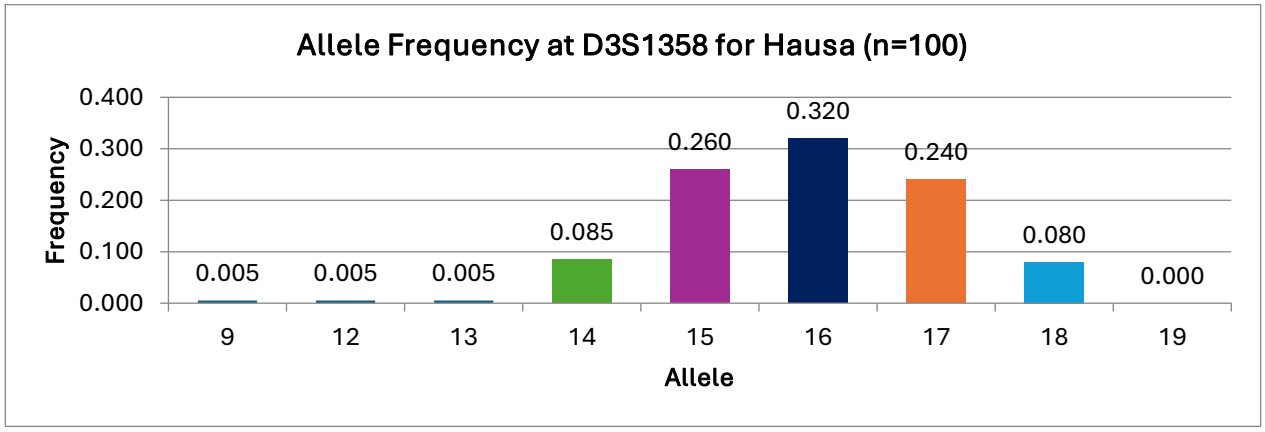
## APPENDIX 27

## ALLELE FREQUENCY DISTRIBUTION AT D3S1358 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



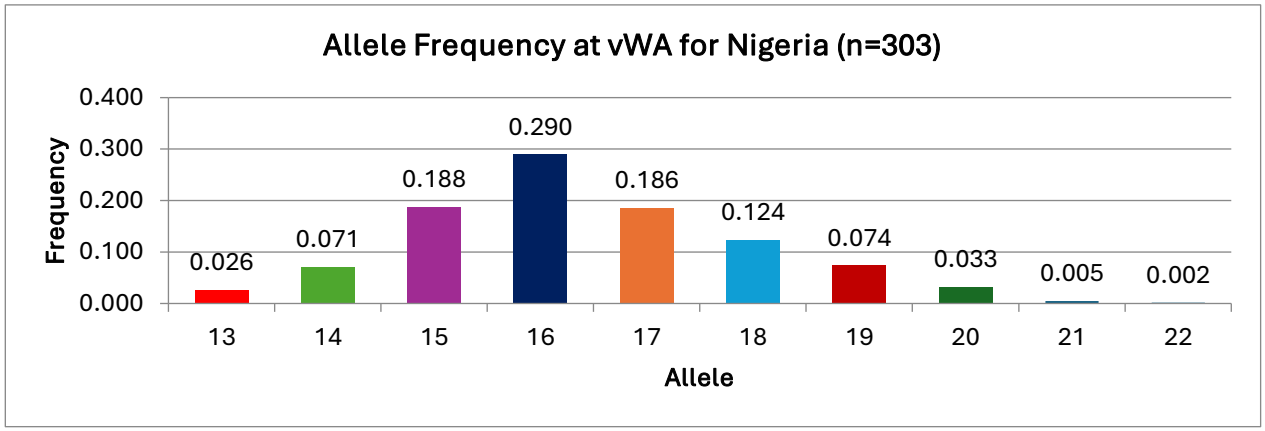



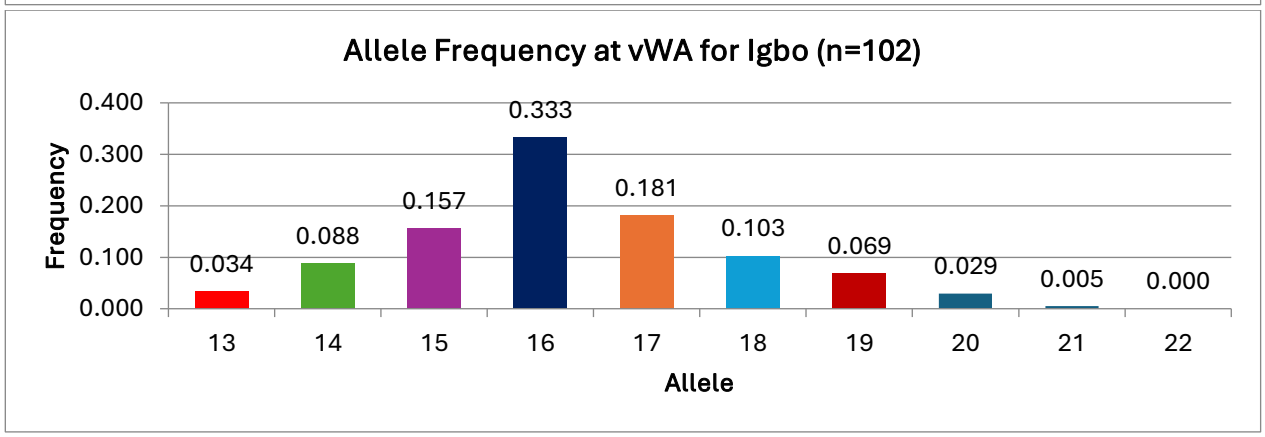



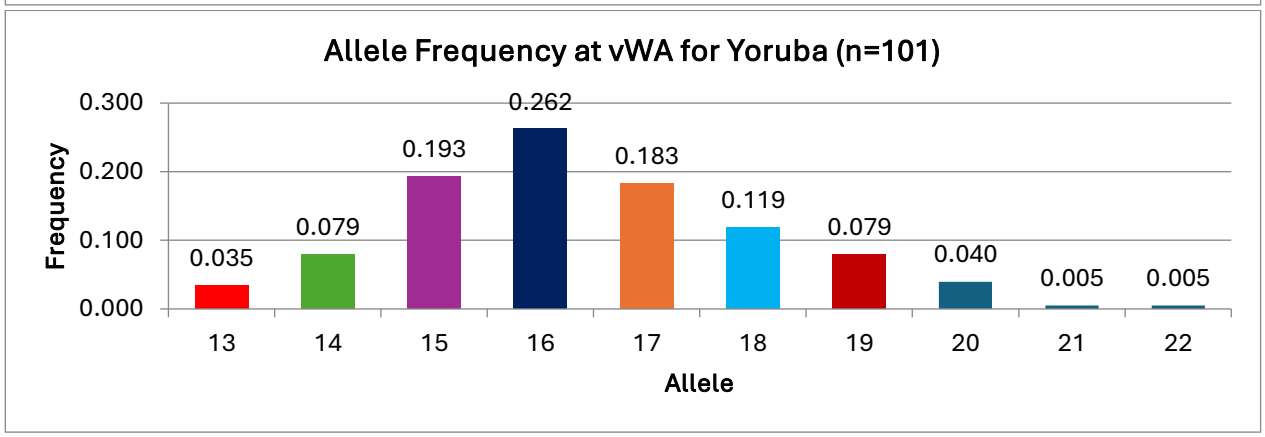



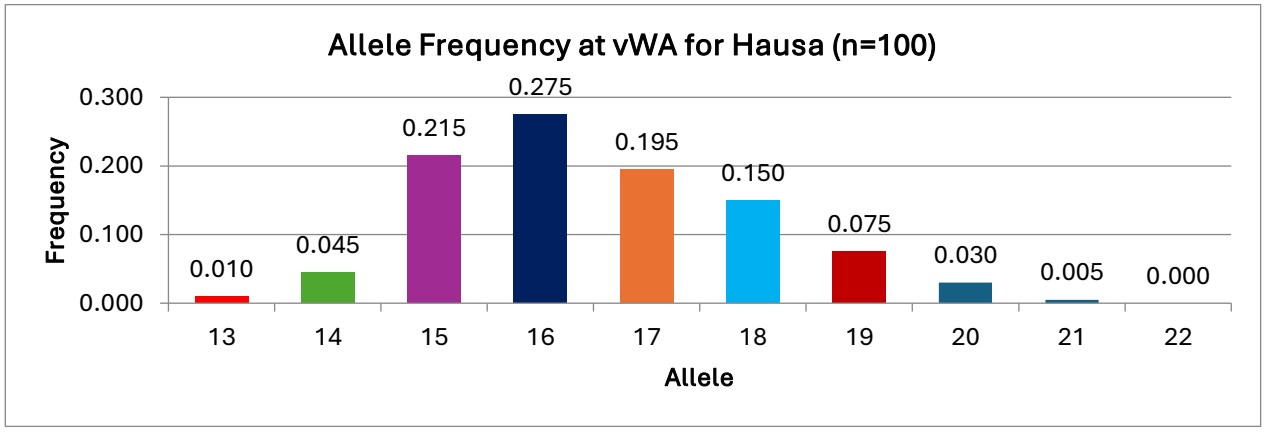
## APPENDIX 28

## ALLELE FREQUENCY DISTRIBUTION AT vWA BY POPULATION USING GLOBALFILER™ EXPRESS KIT



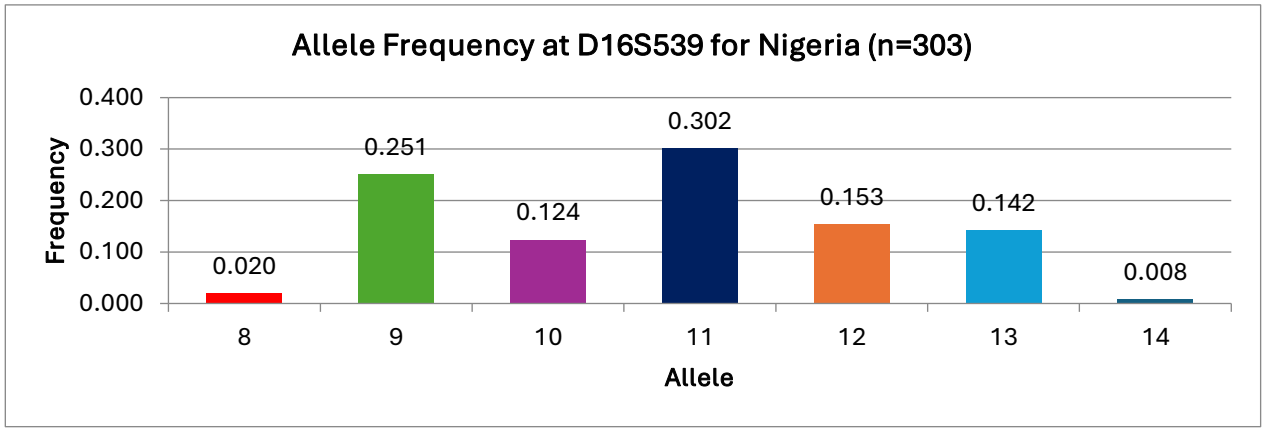



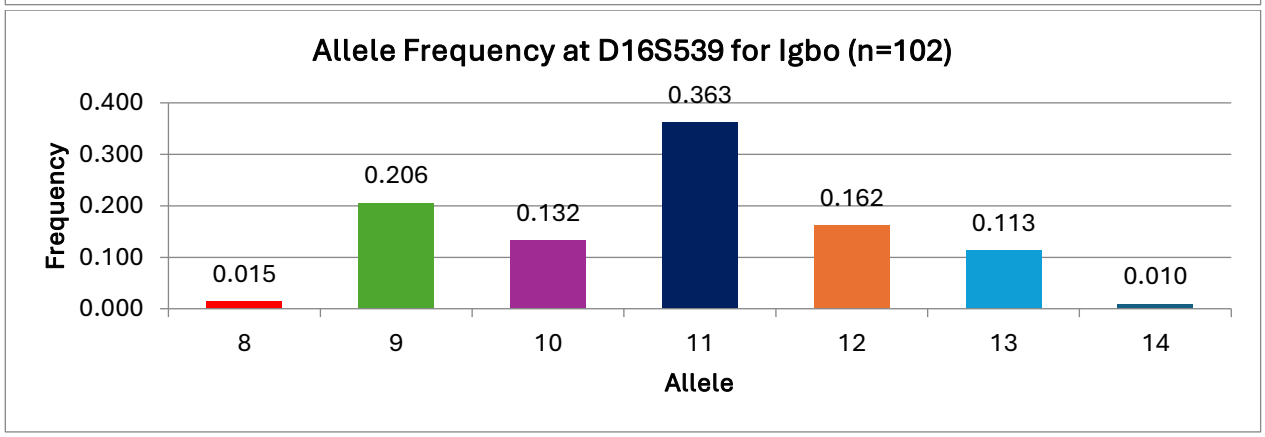



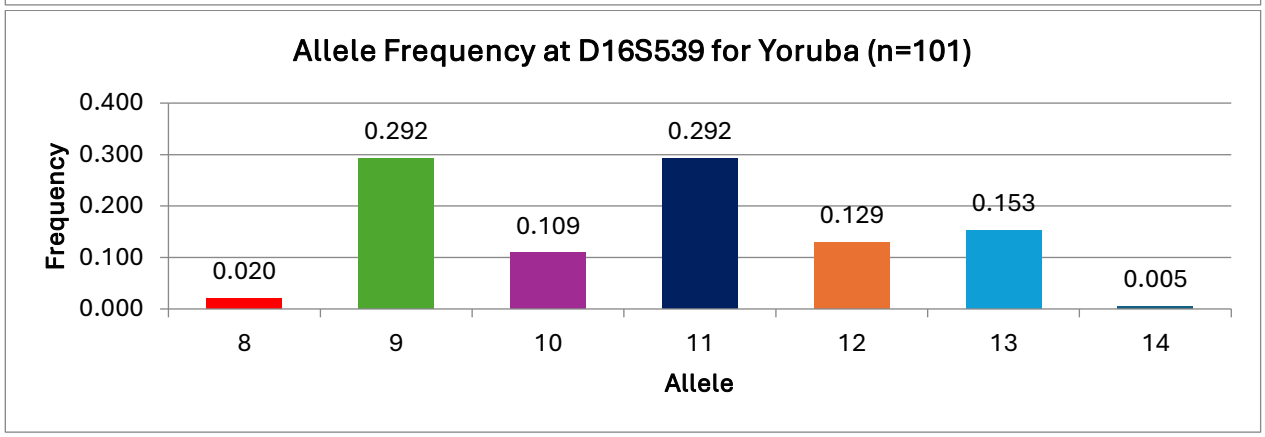



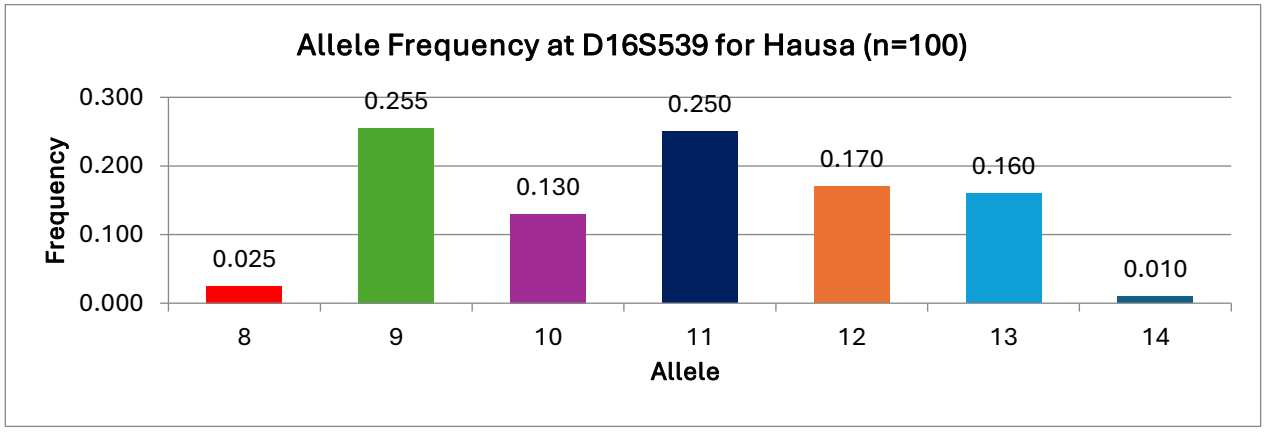
## APPENDIX 29

## ALLELE FREQUENCY DISTRIBUTION AT D16S539 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



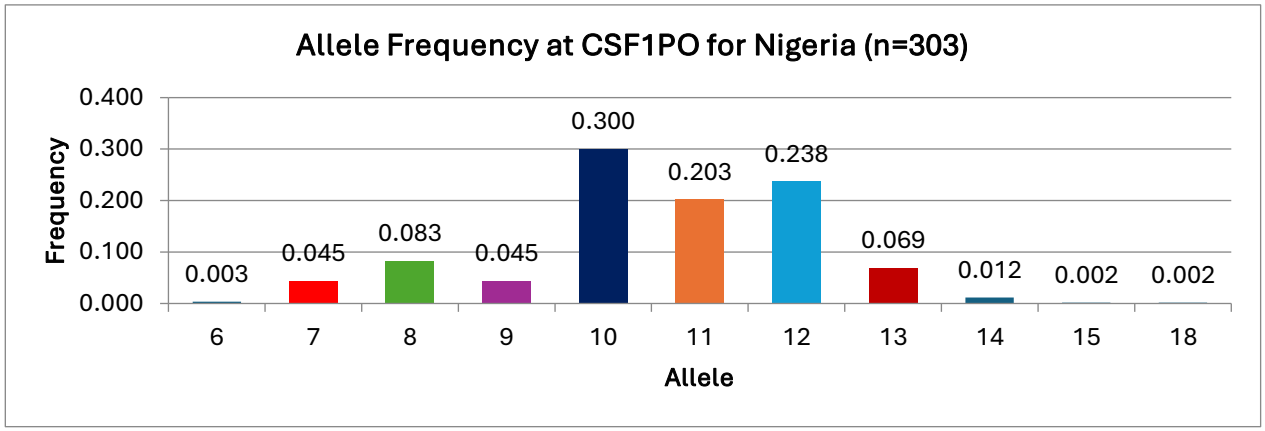



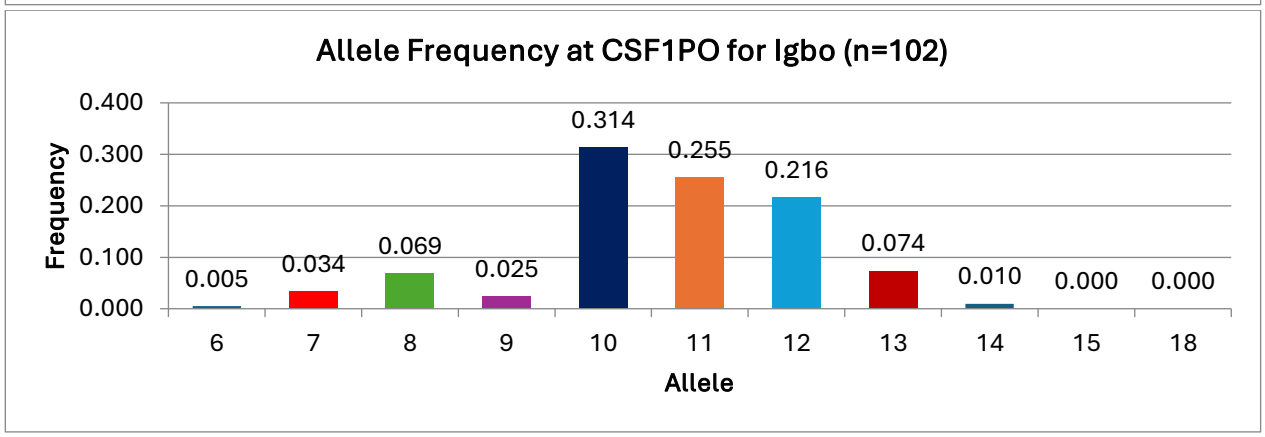



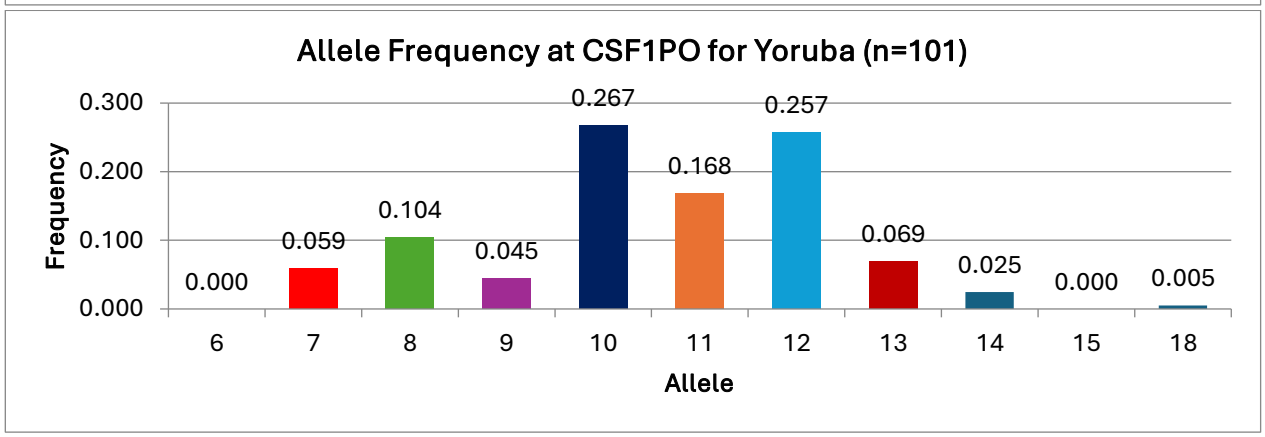



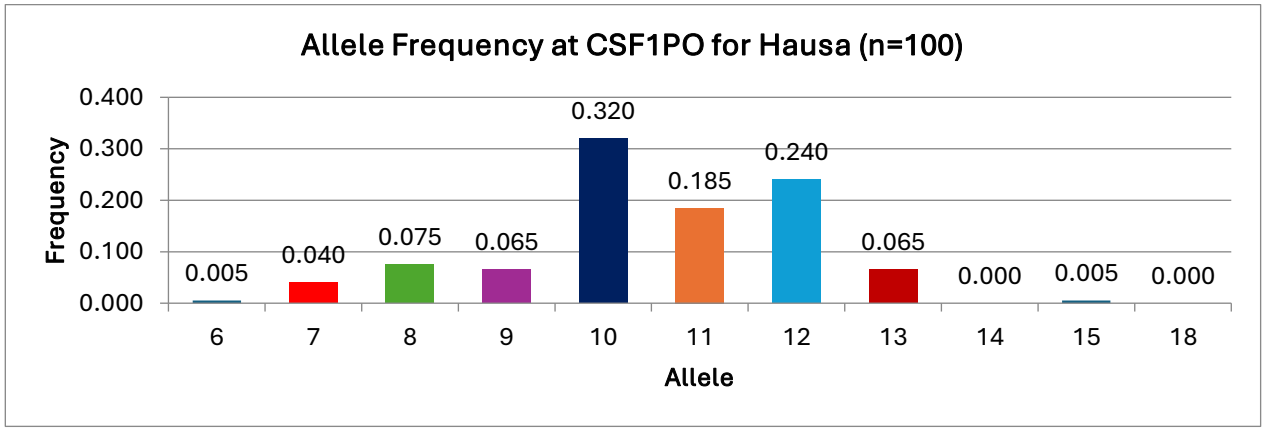
## APPENDIX 30

## ALLELE FREQUENCY DISTRIBUTION AT CSF1PO BY POPULATION USING GLOBALFILER™ EXPRESS KIT



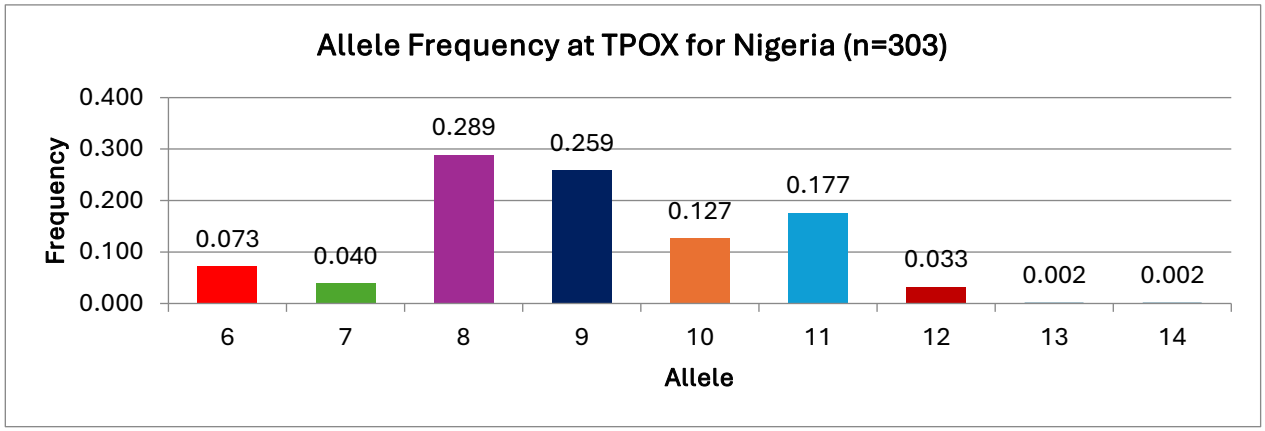



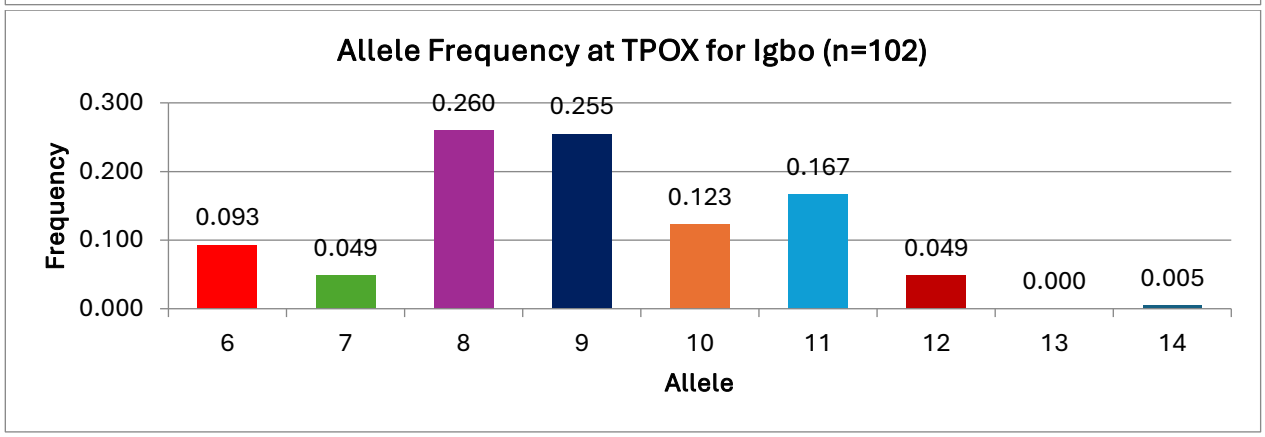



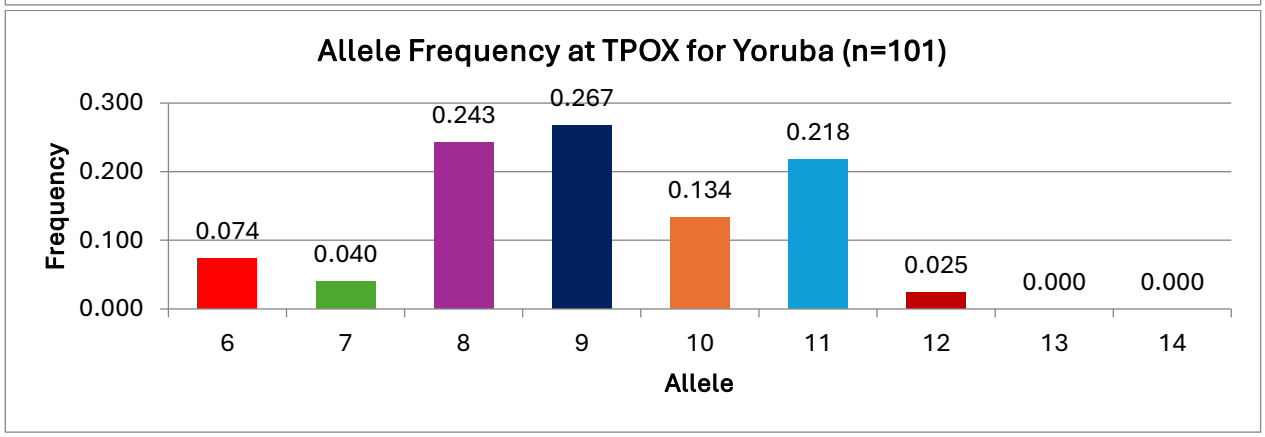



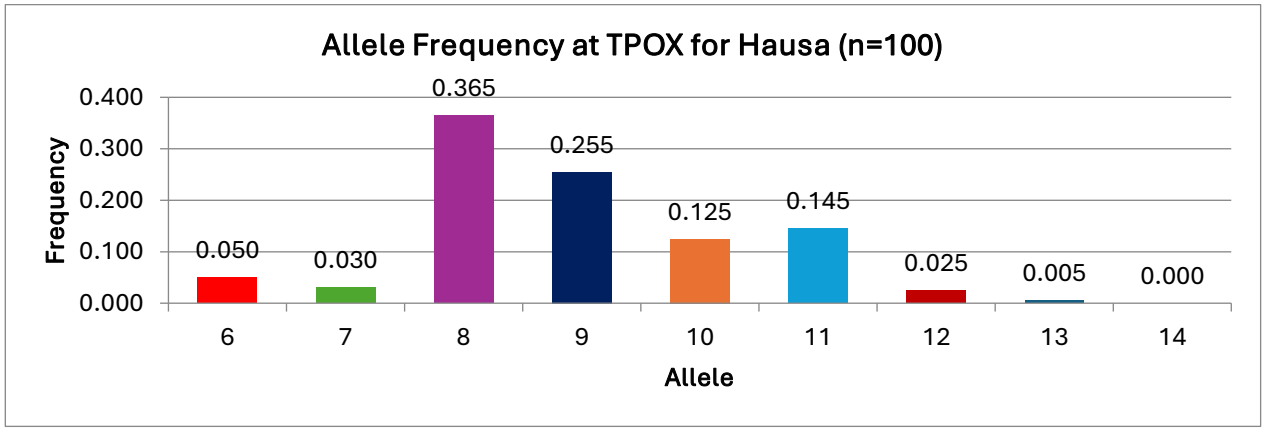
## APPENDIX 31

## ALLELE FREQUENCY DISTRIBUTION AT TPOX BY POPULATION USING GLOBALFILER™ EXPRESS KIT



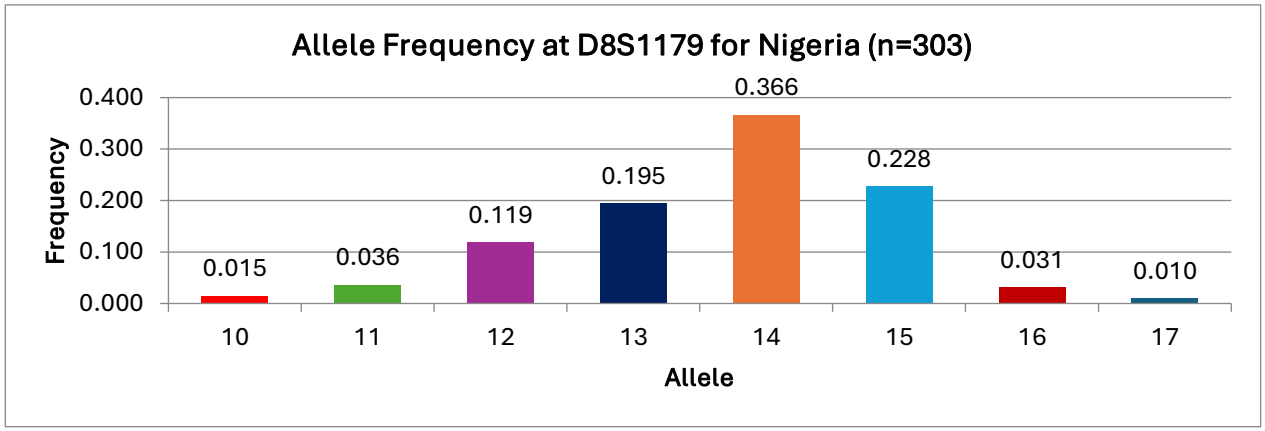



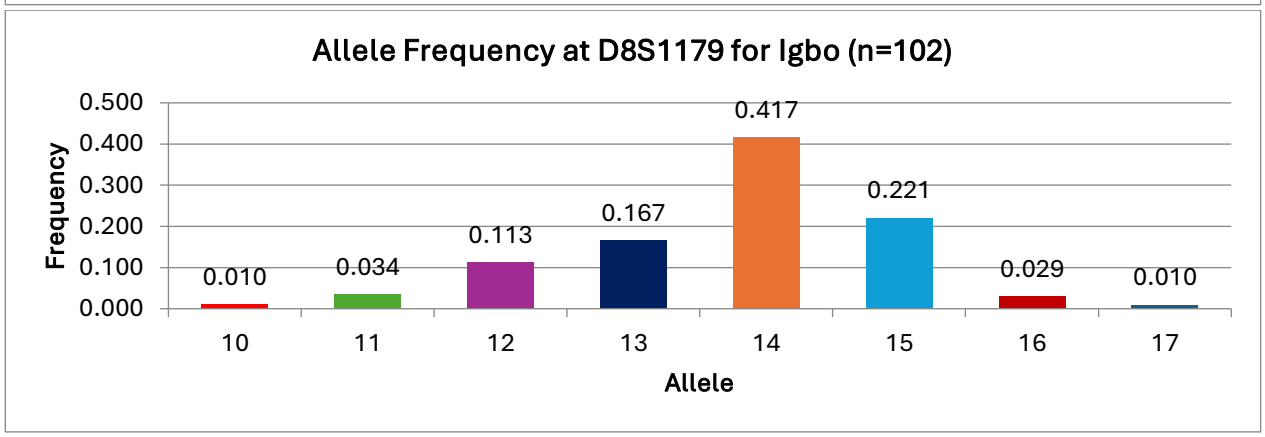



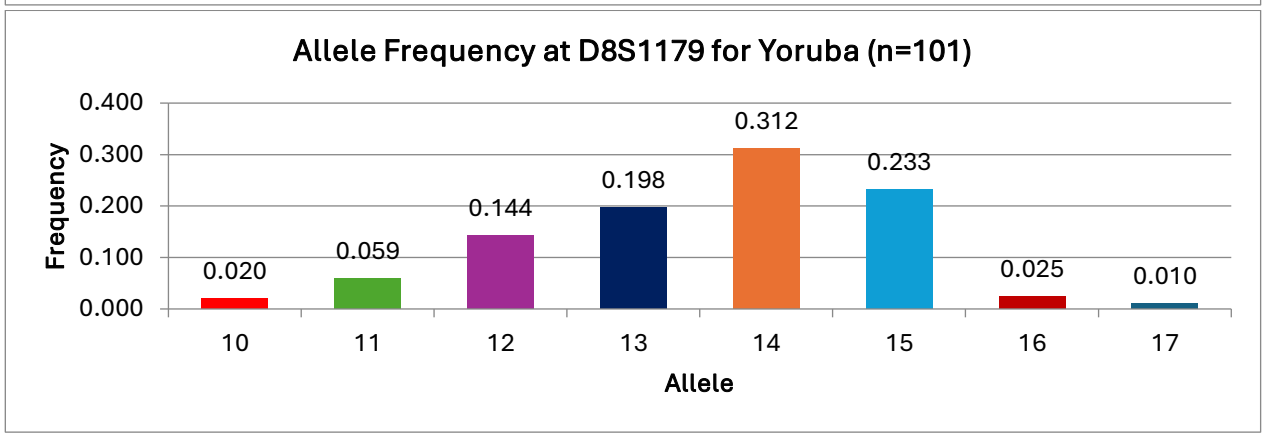



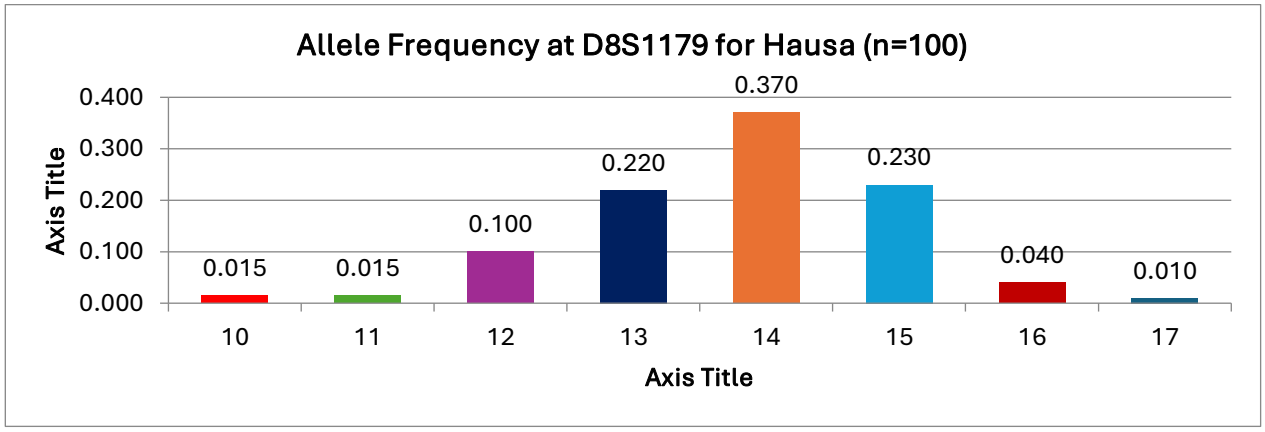
## APPENDIX 32

## ALLELE FREQUENCY DISTRIBUTION AT D8S1179 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



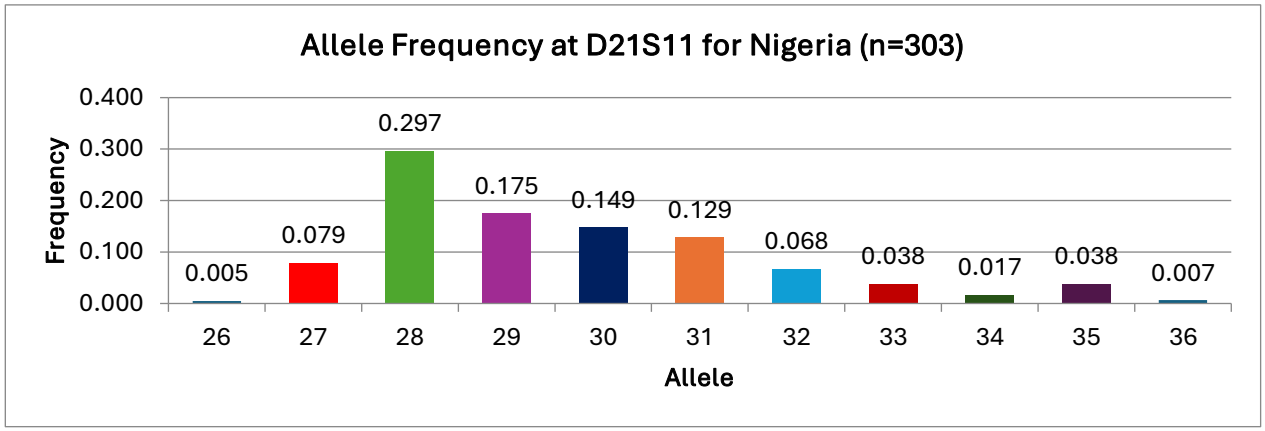



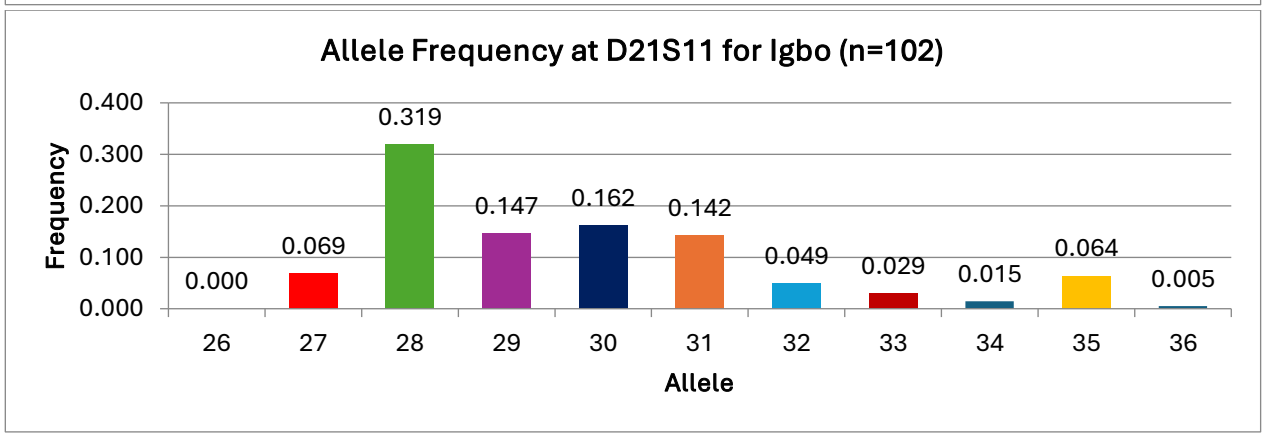



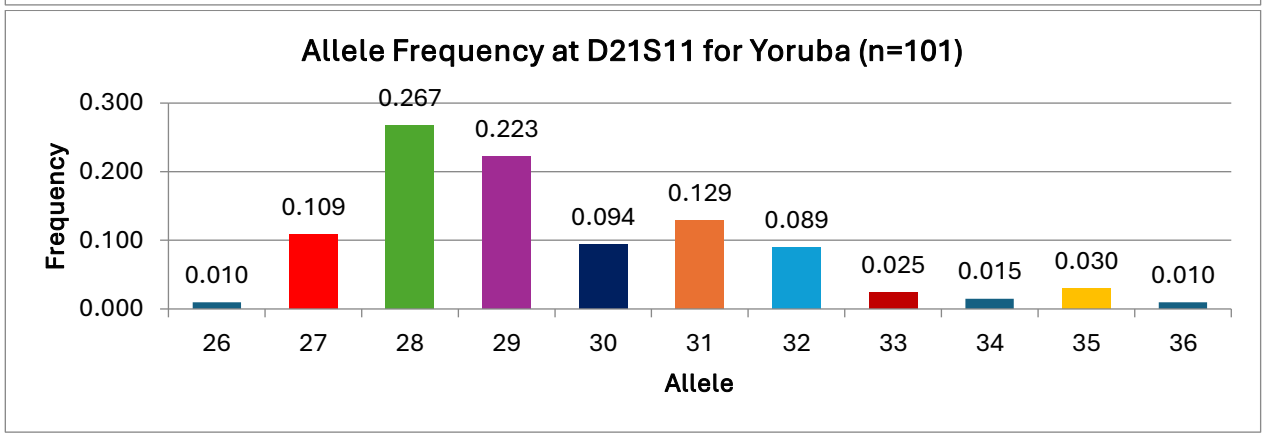



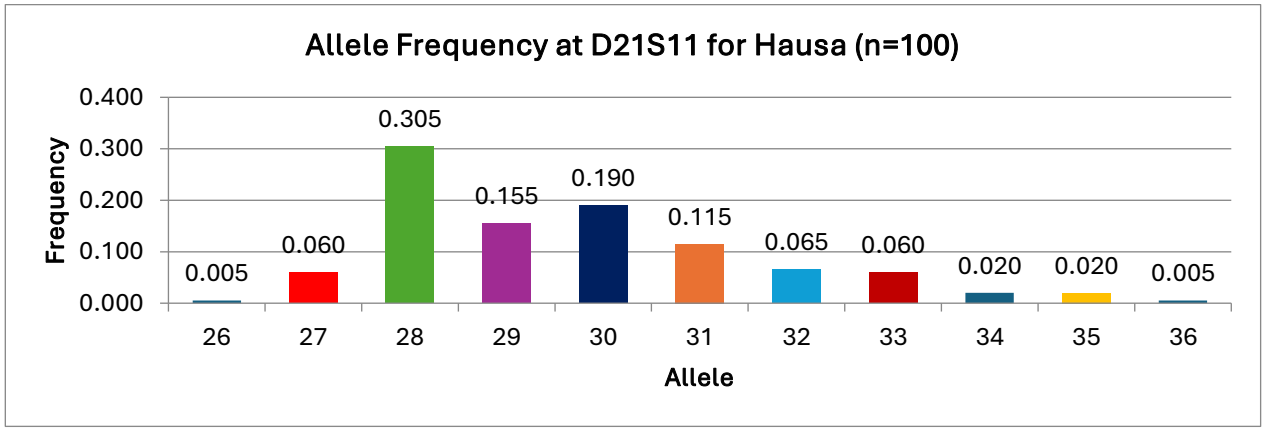
## APPENDIX 33

## ALLELE FREQUENCY DISTRIBUTION AT D21S11 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



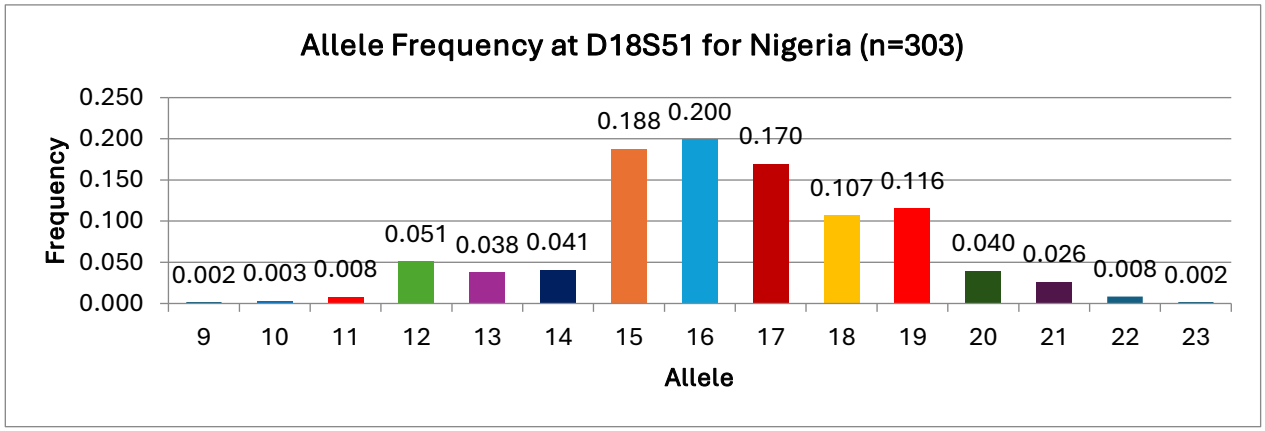



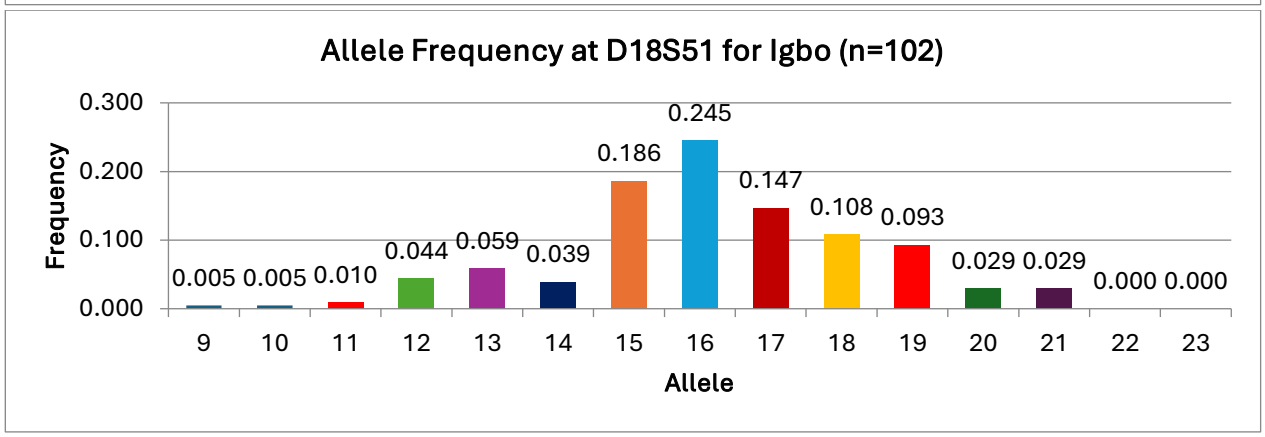



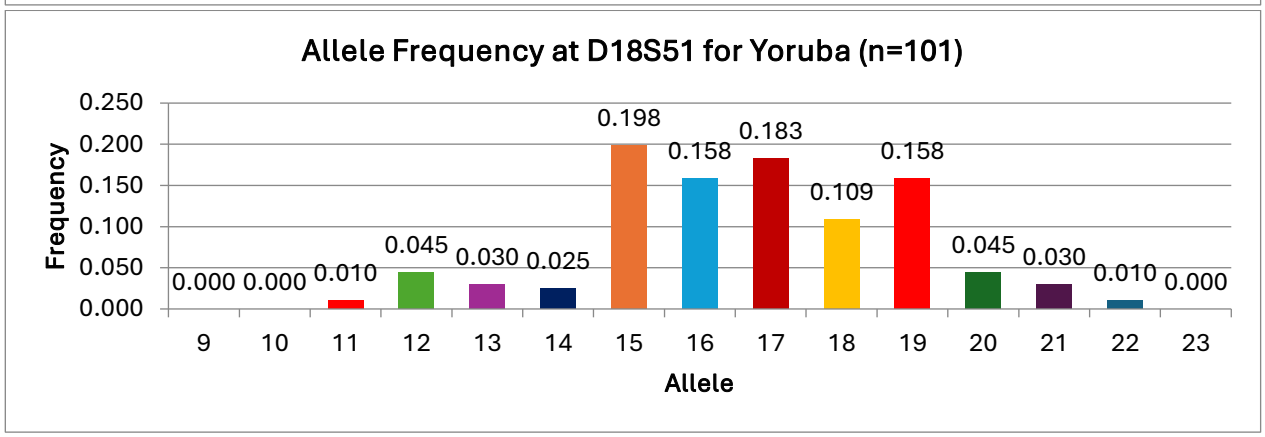



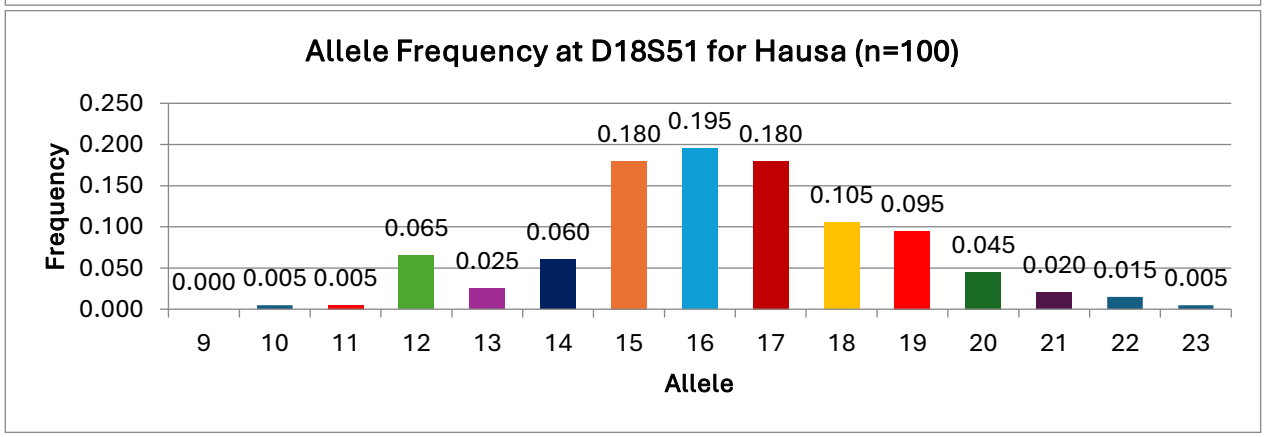
## APPENDIX 34

## ALLELE FREQUENCY DISTRIBUTION AT D18S51 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



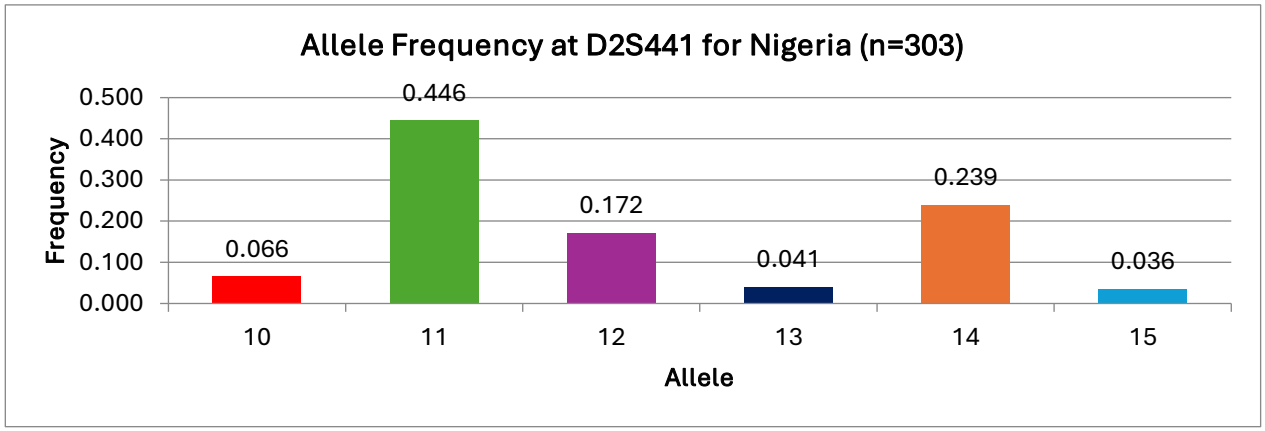



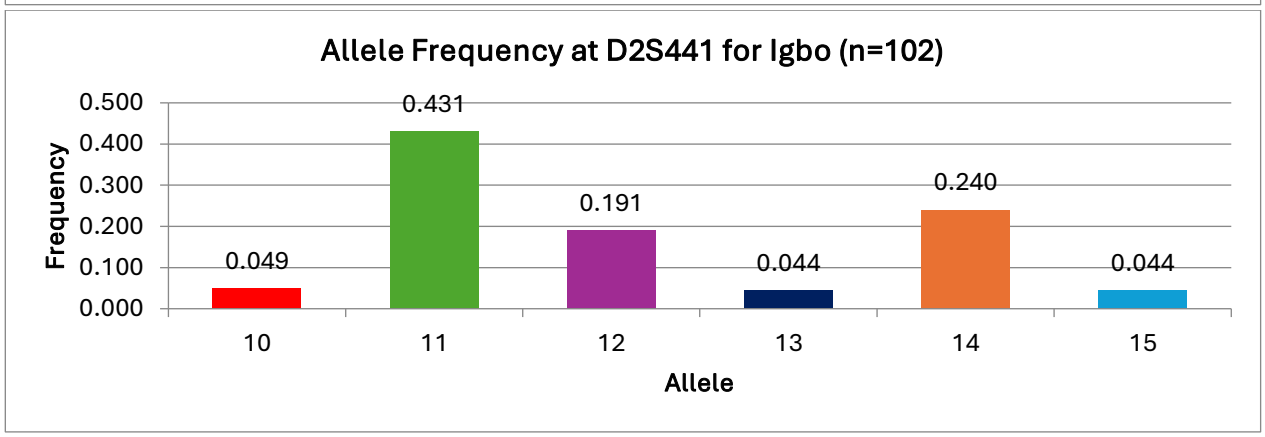



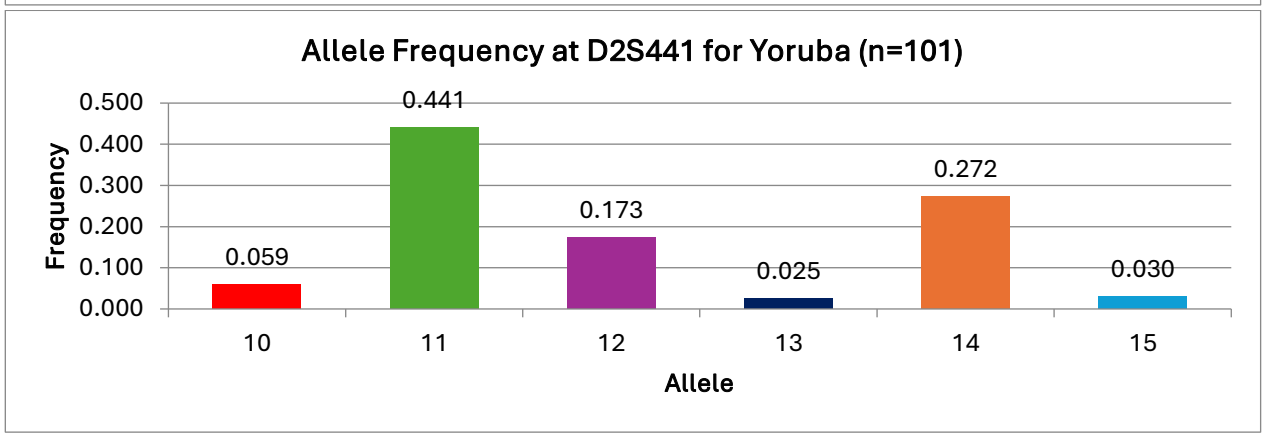



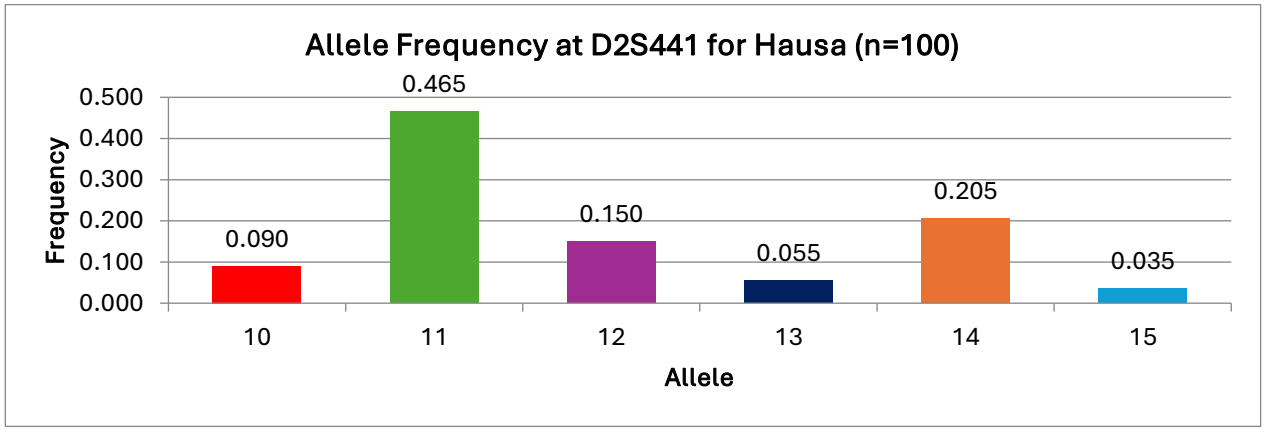
## APPENDIX 35

## ALLELE FREQUENCY DISTRIBUTION AT D2S441 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



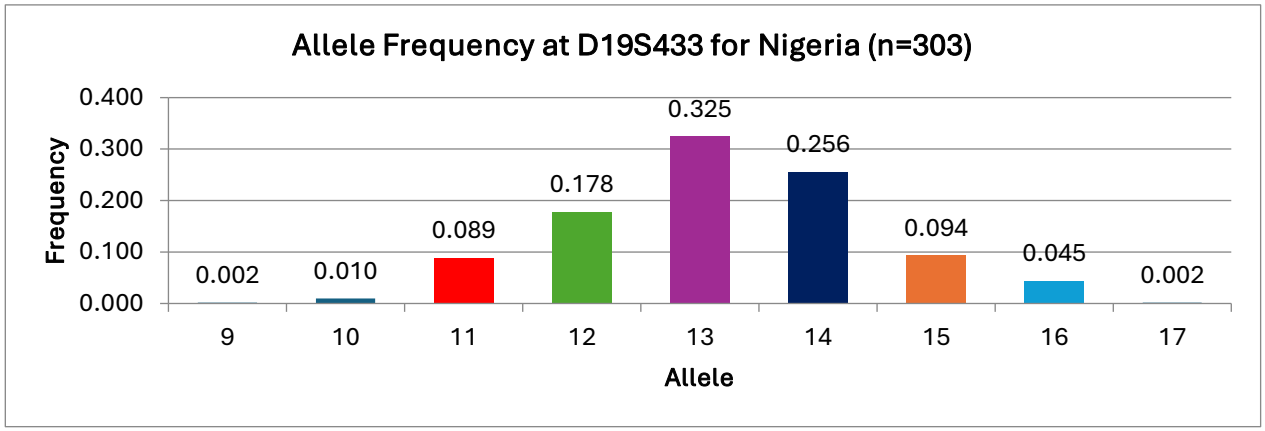



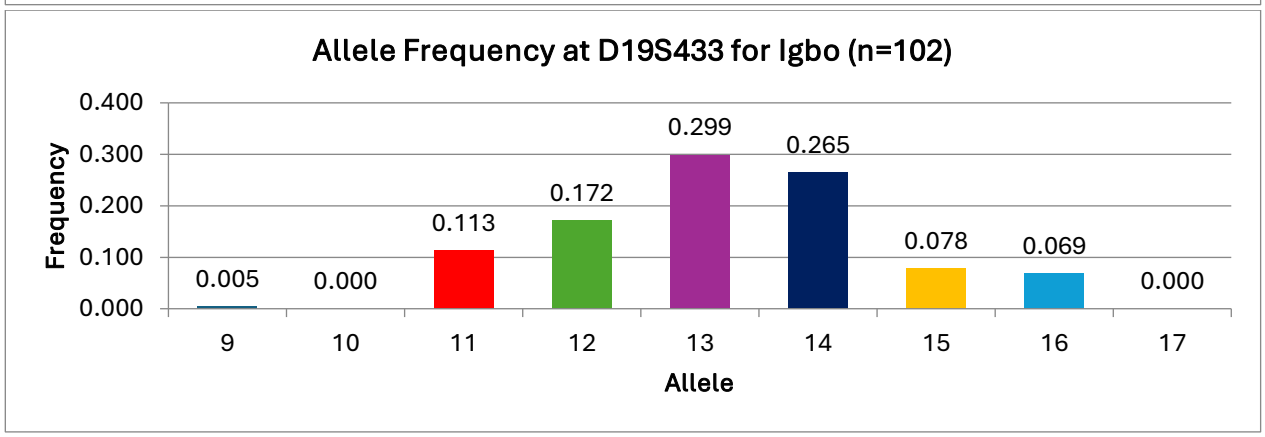



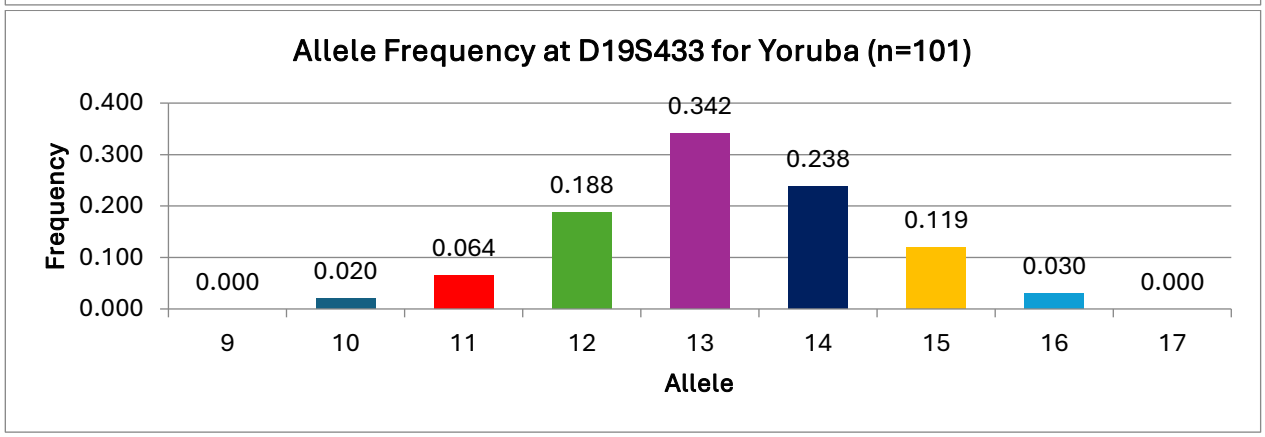



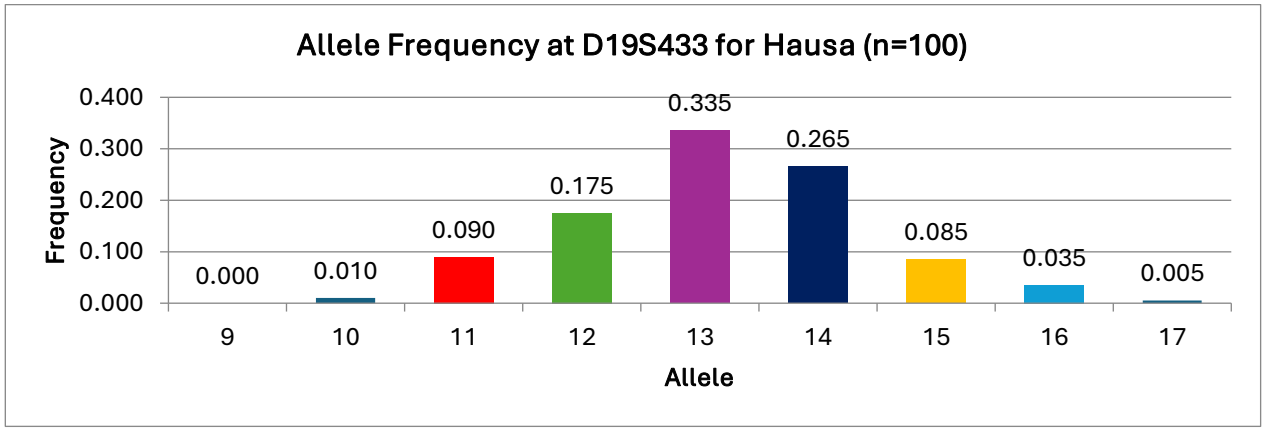
## APPENDIX 36

## ALLELE FREQUENCY DISTRIBUTION AT D19S433 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



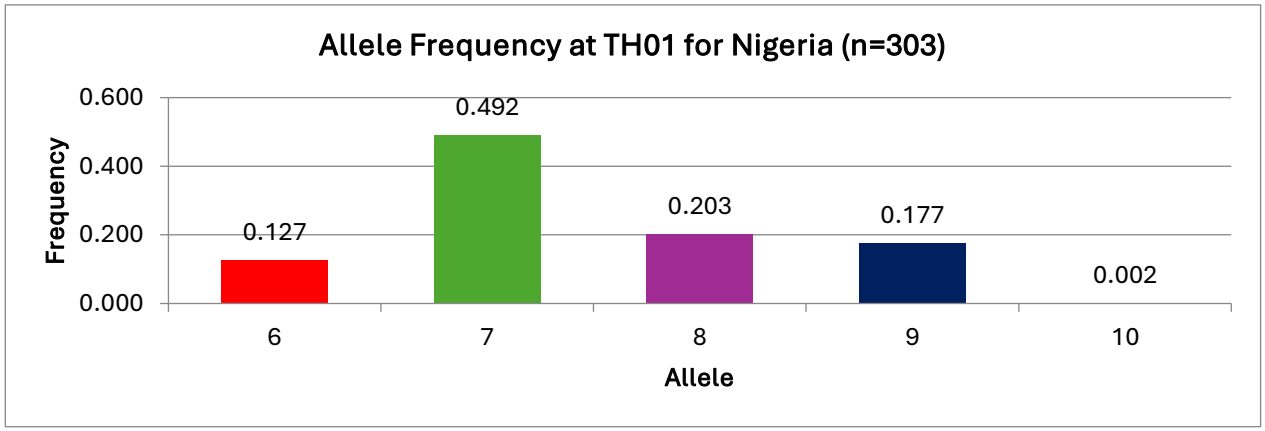



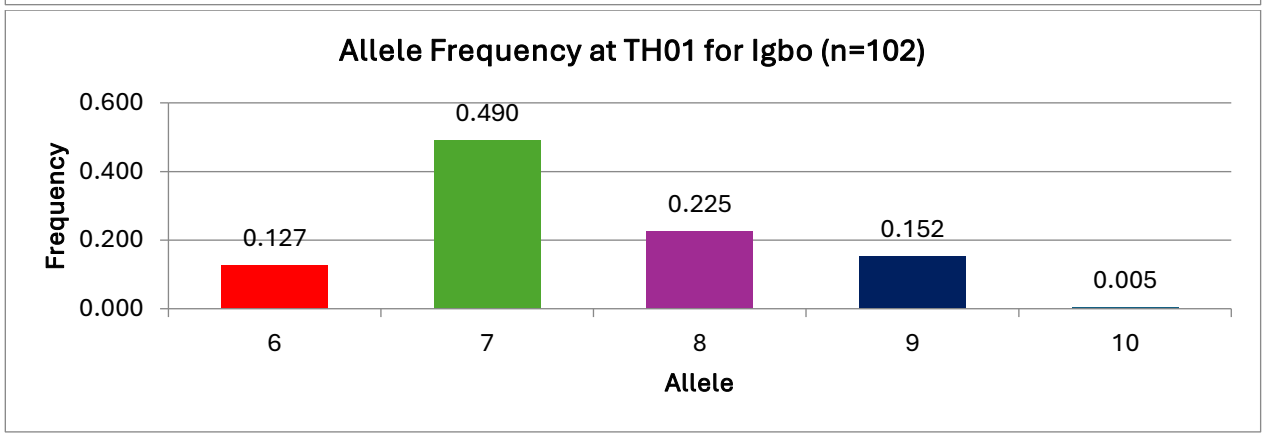



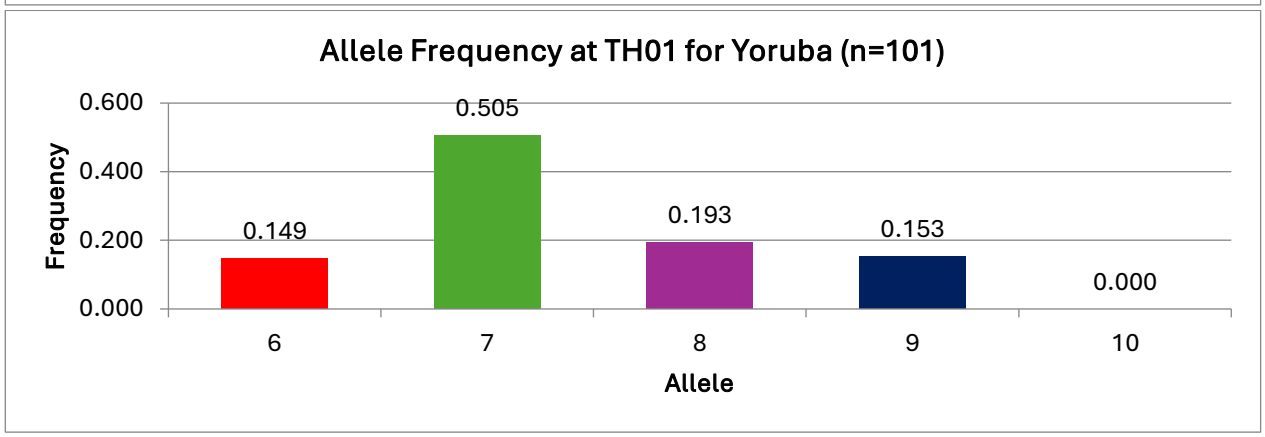



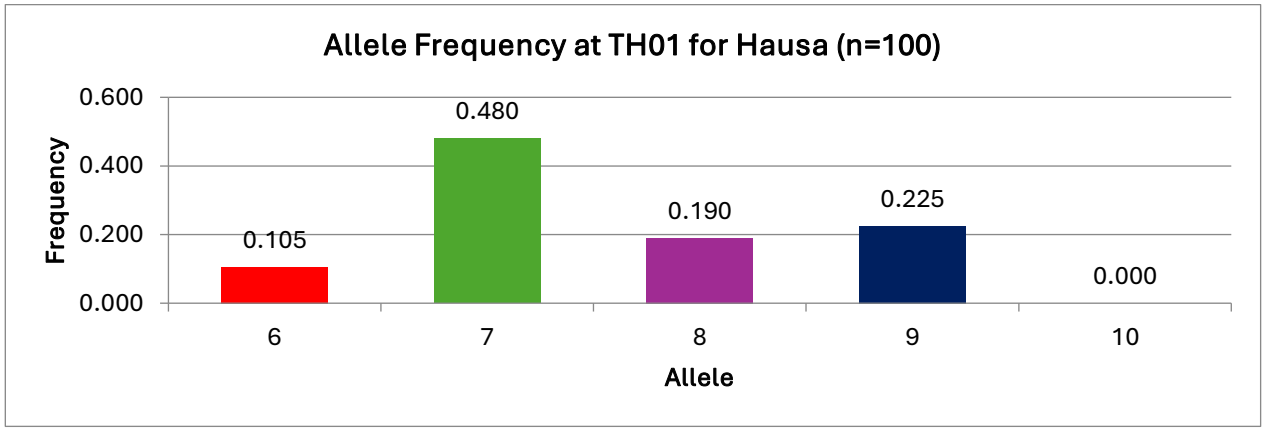
## APPENDIX 37

## ALLELE FREQUENCY DISTRIBUTION AT TH01 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



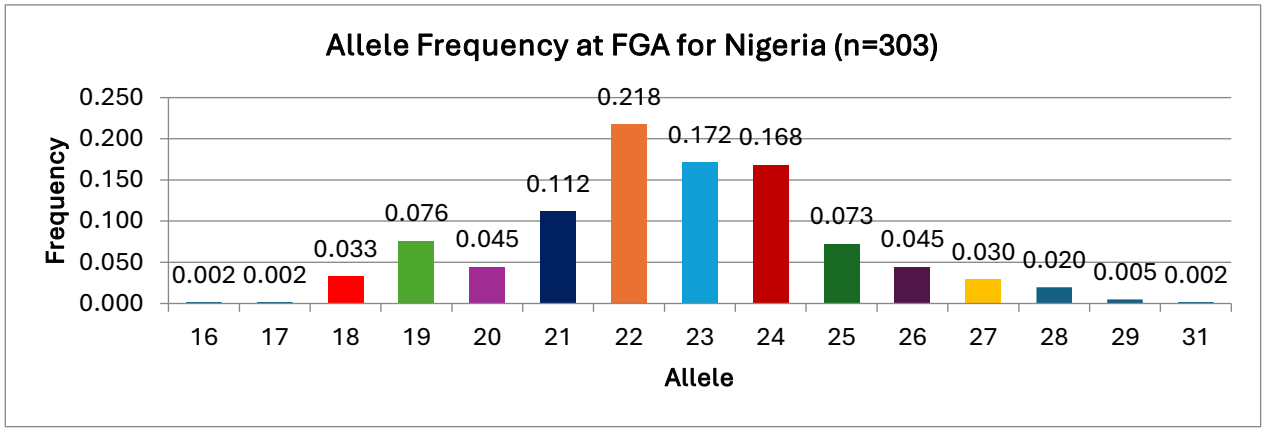



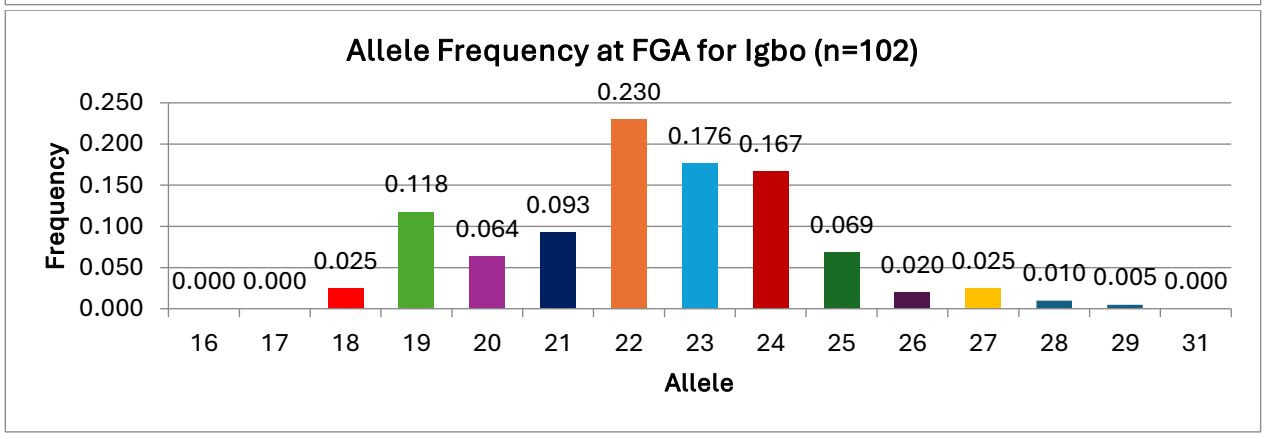



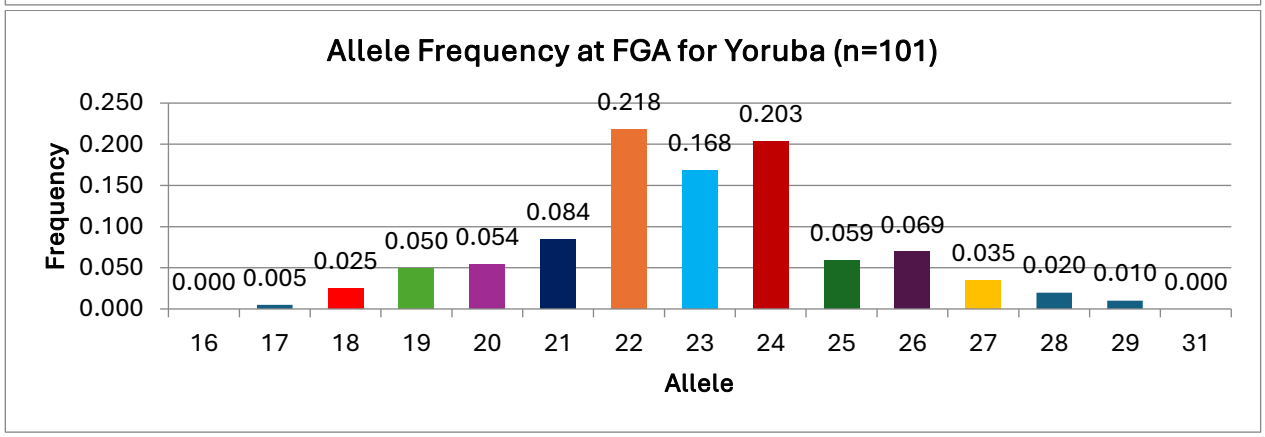



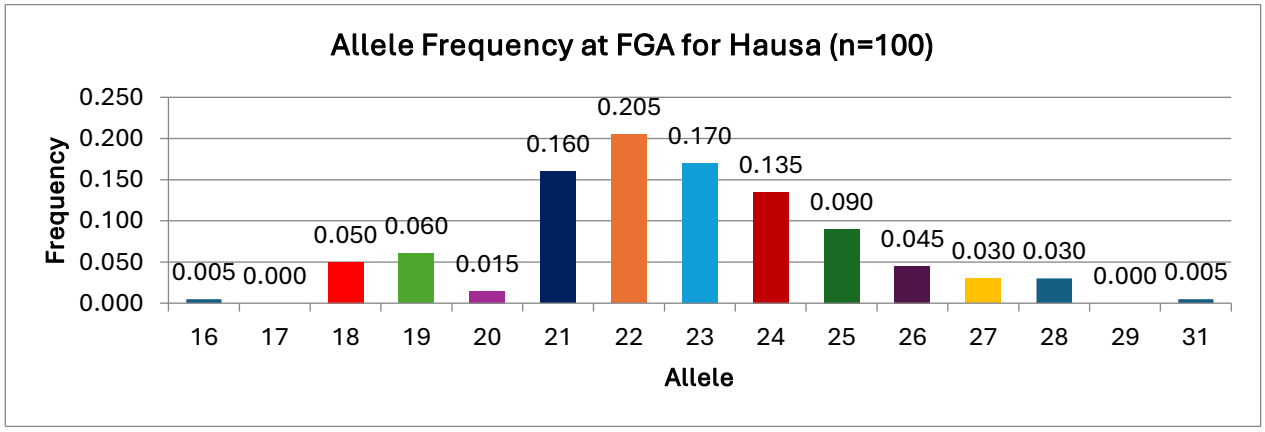
## APPENDIX 38

## ALLELE FREQUENCY DISTRIBUTION AT FGA BY POPULATION USING GLOBALFILER™ EXPRESS KIT



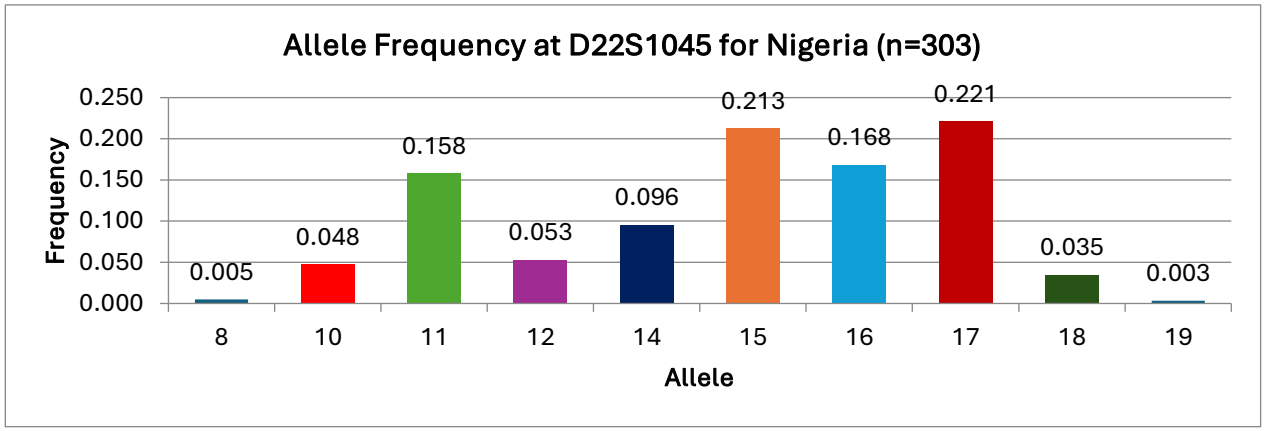



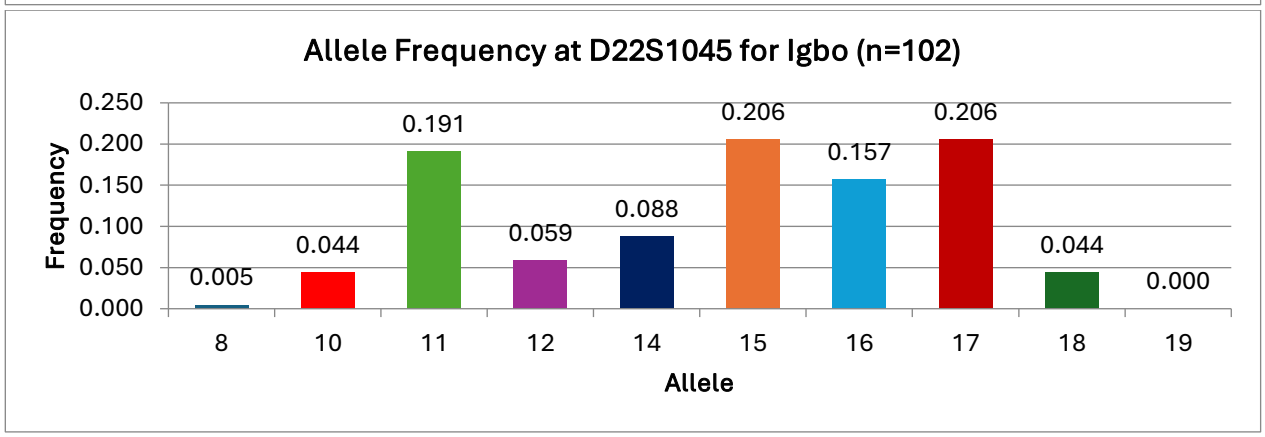



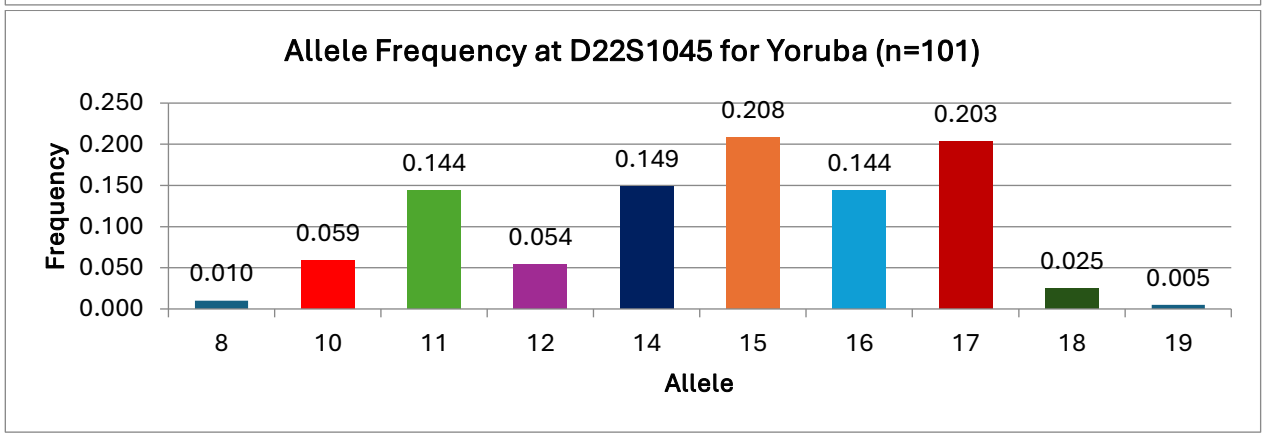



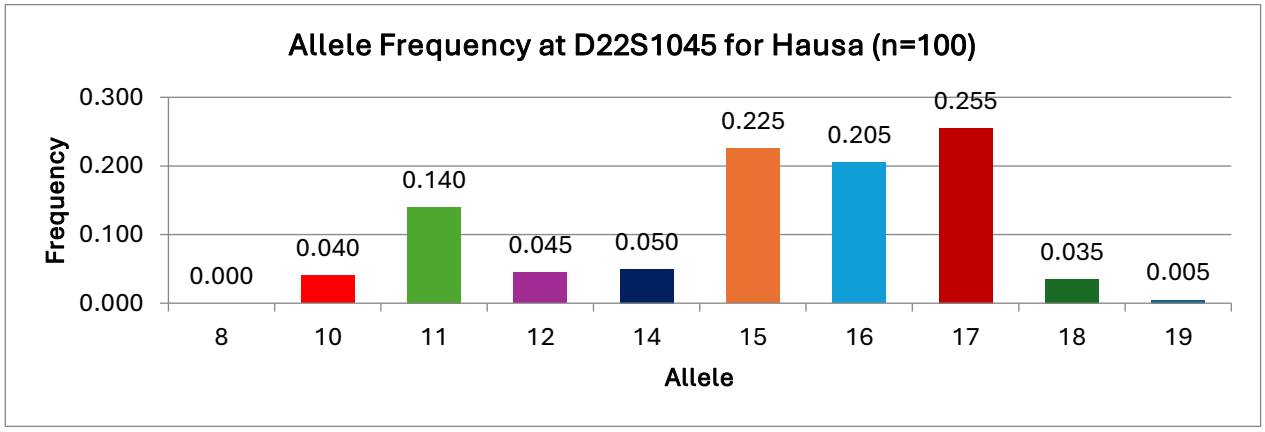
## APPENDIX 39

## ALLELE FREQUENCY DISTRIBUTION AT D22S1045 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



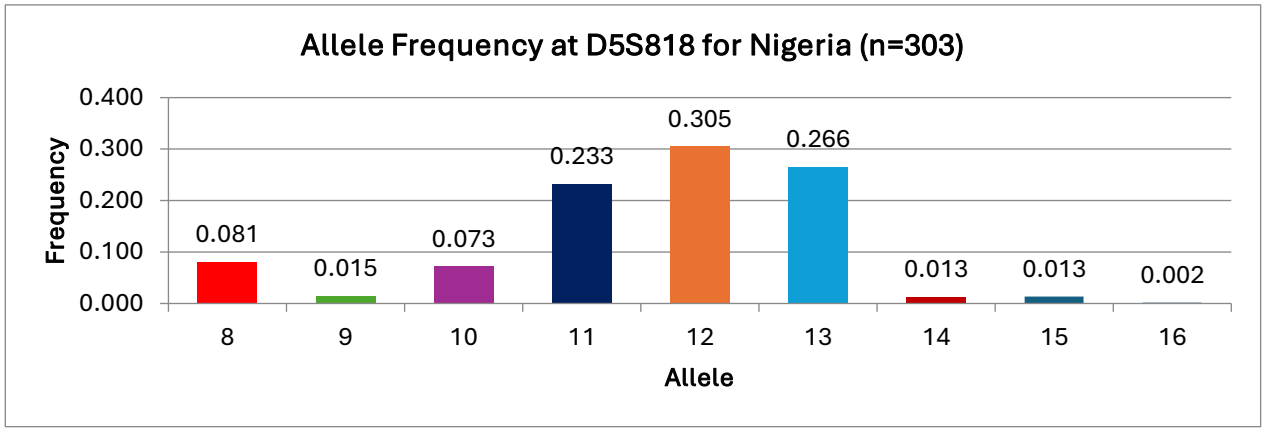



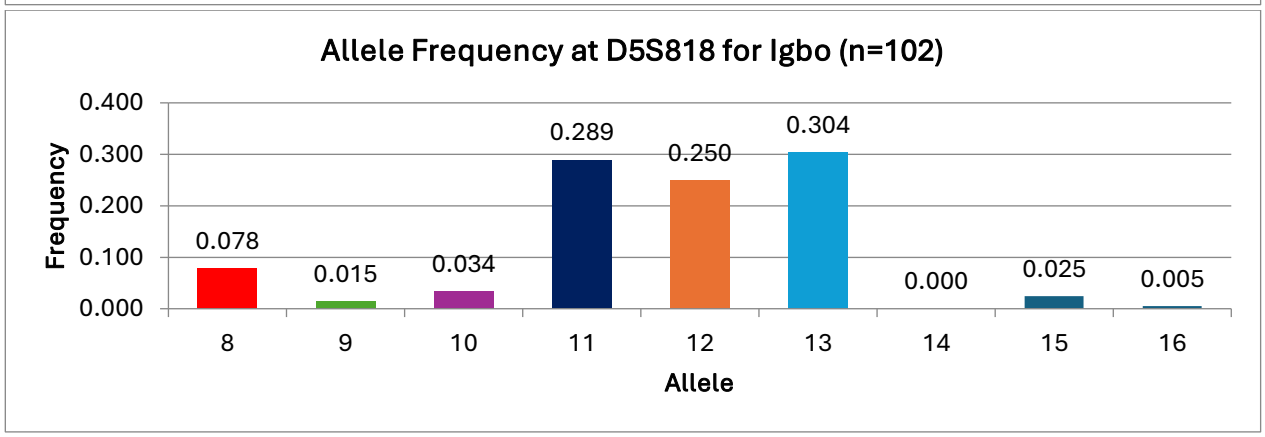



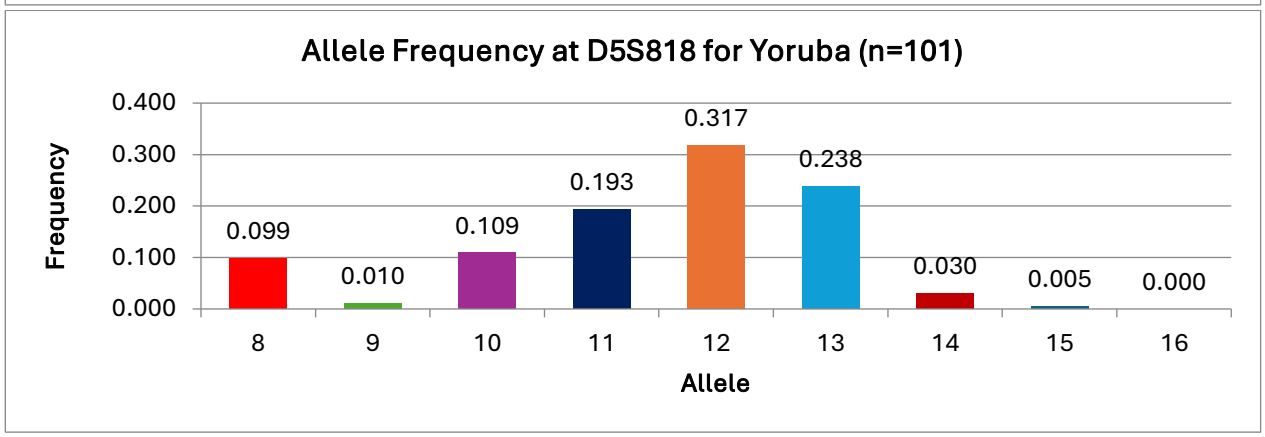



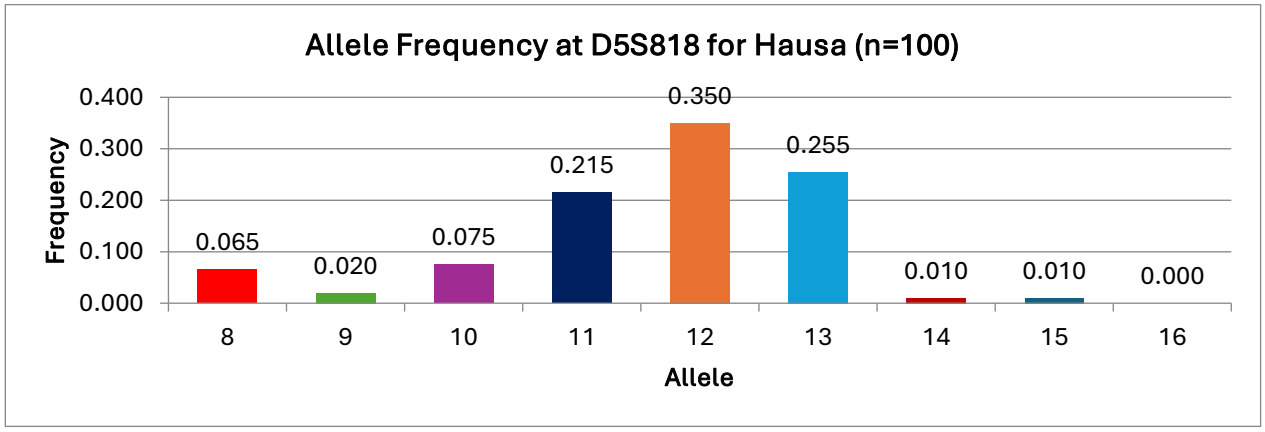
## APPENDIX 40

## ALLELE FREQUENCY DISTRIBUTION AT D5S818 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



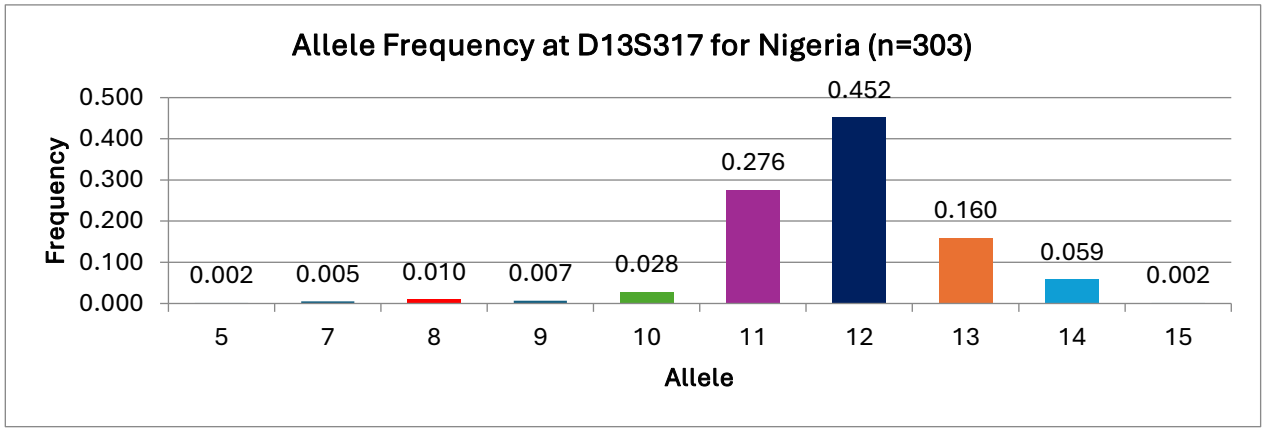



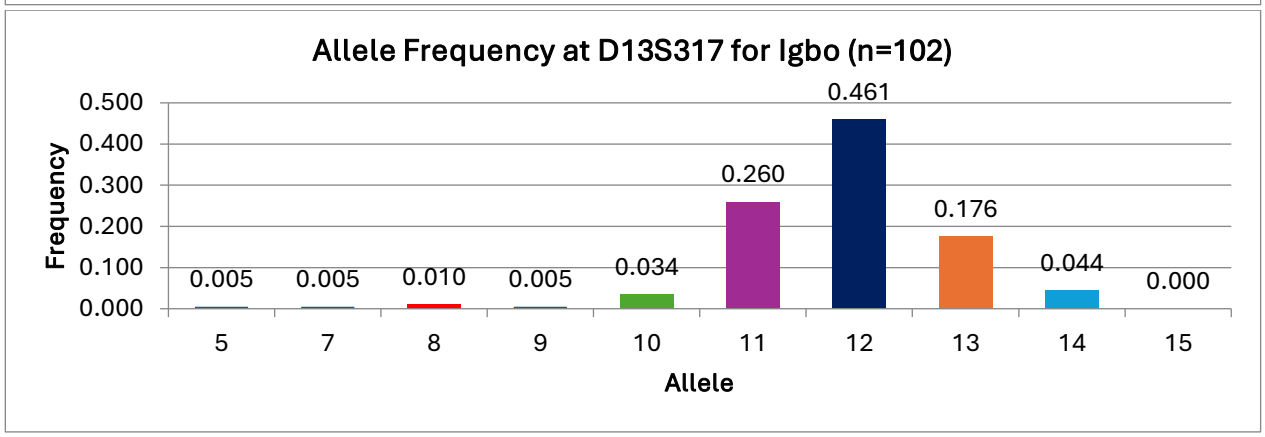



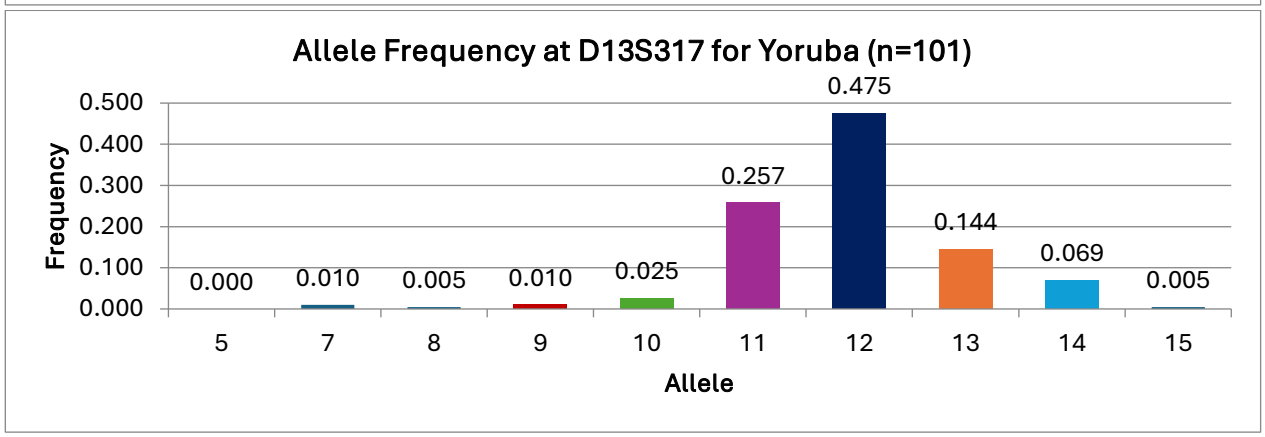



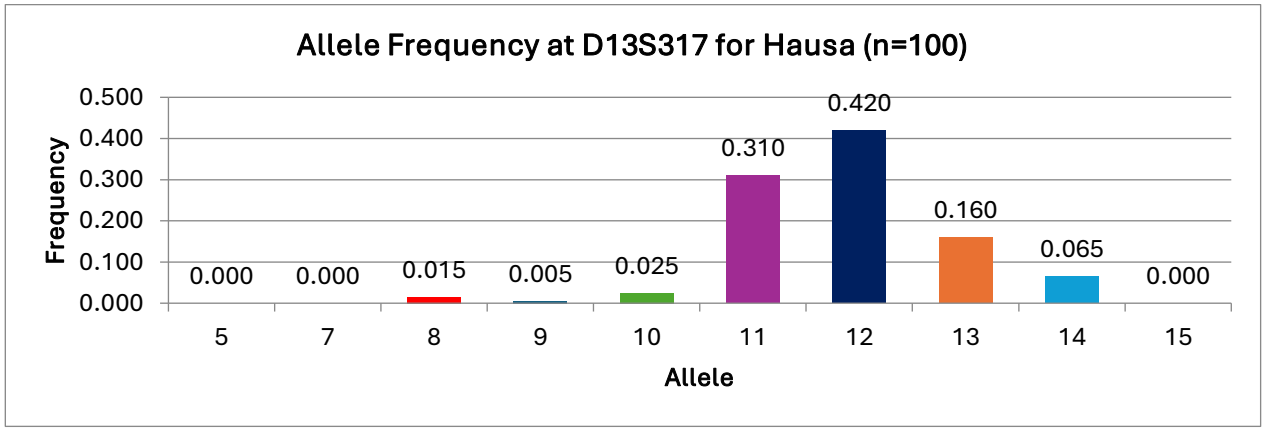
## APPENDIX 41

## ALLELE FREQUENCY DISTRIBUTION AT D13S317 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



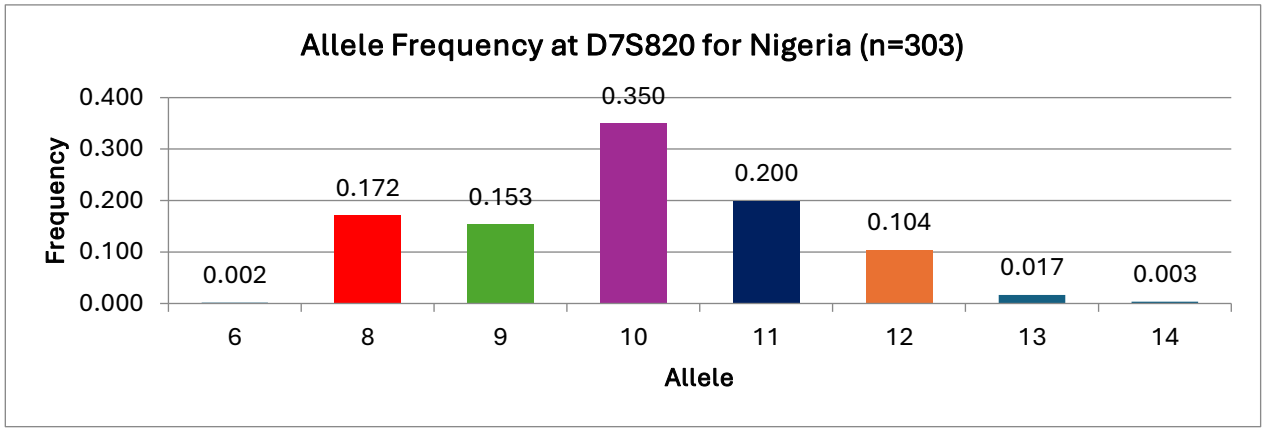



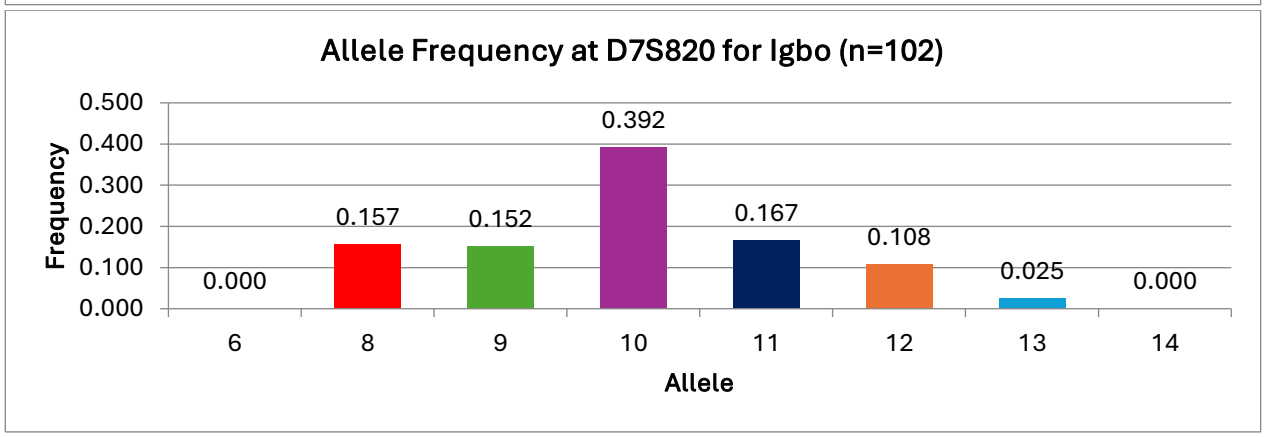



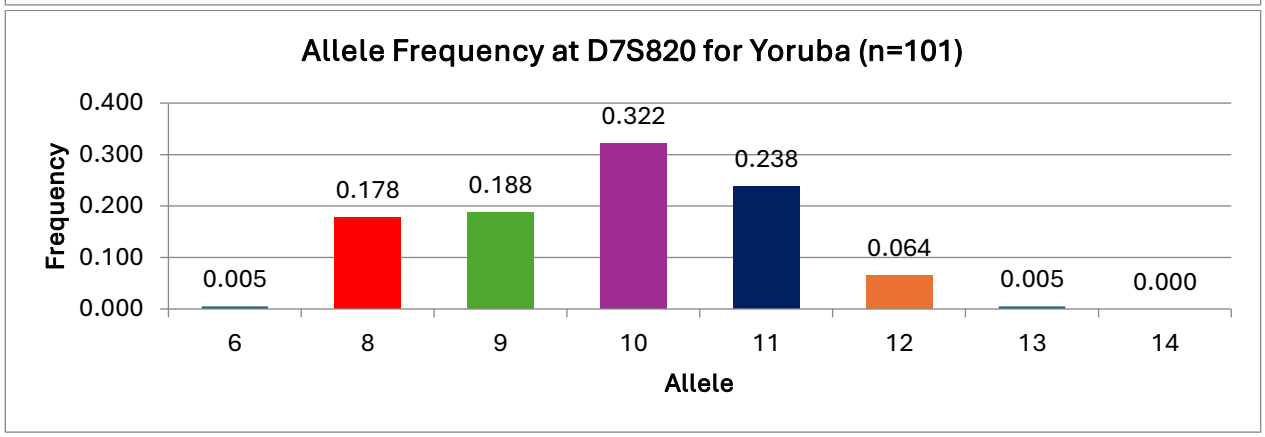



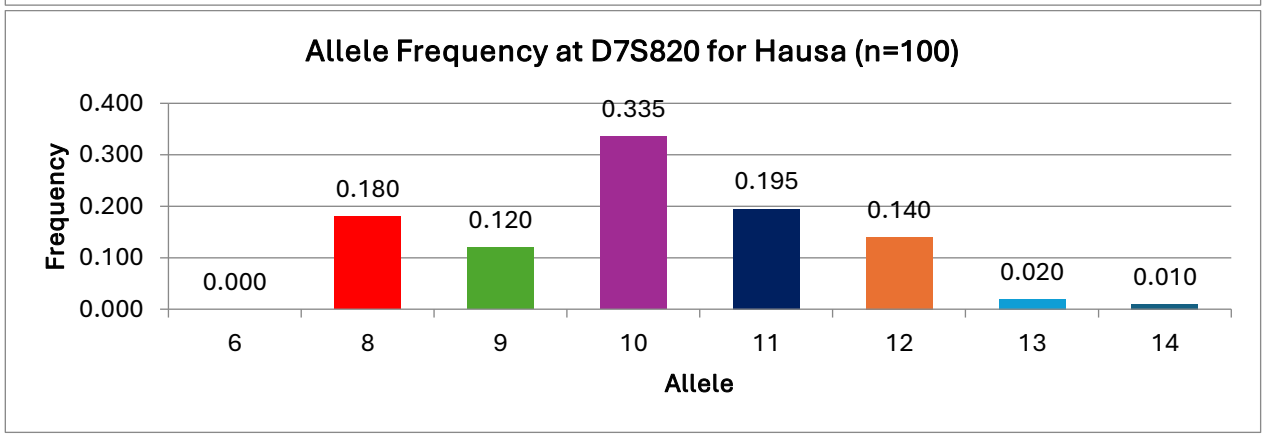
## APPENDIX 42

## ALLELE FREQUENCY DISTRIBUTION AT D7S820 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



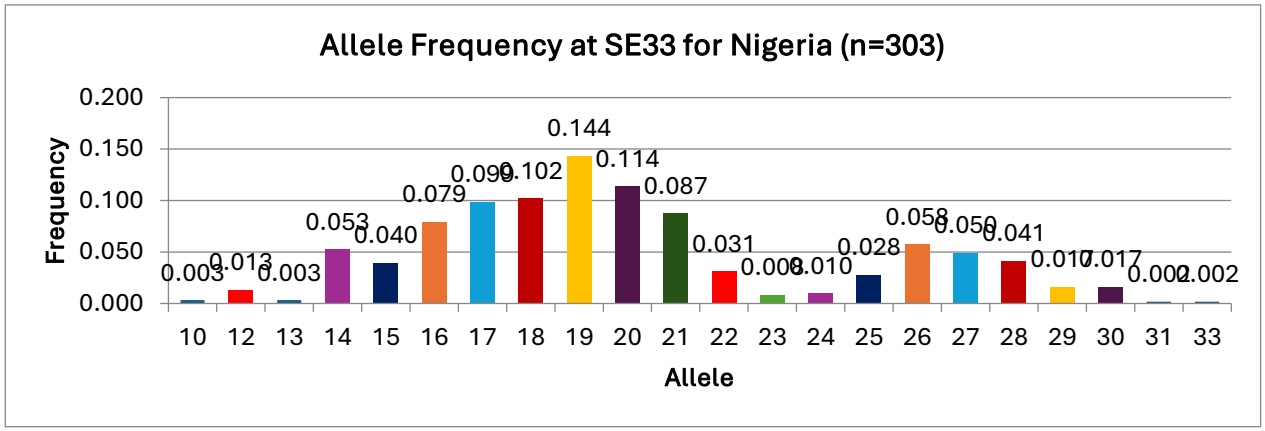



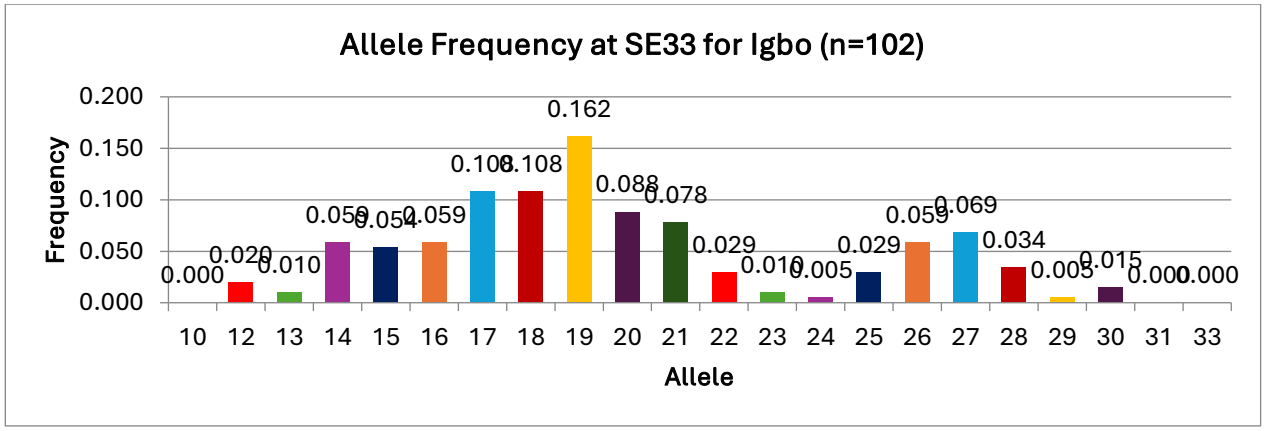



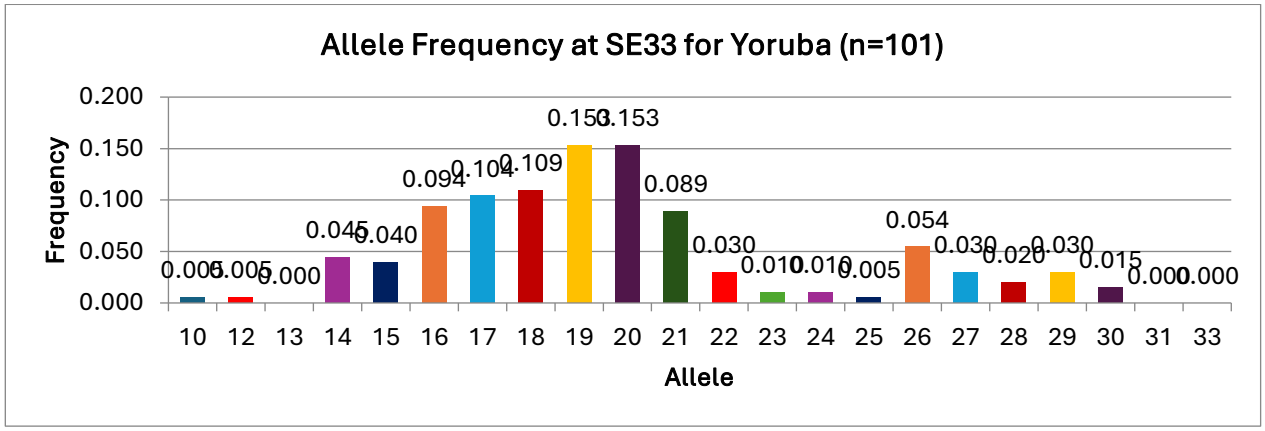



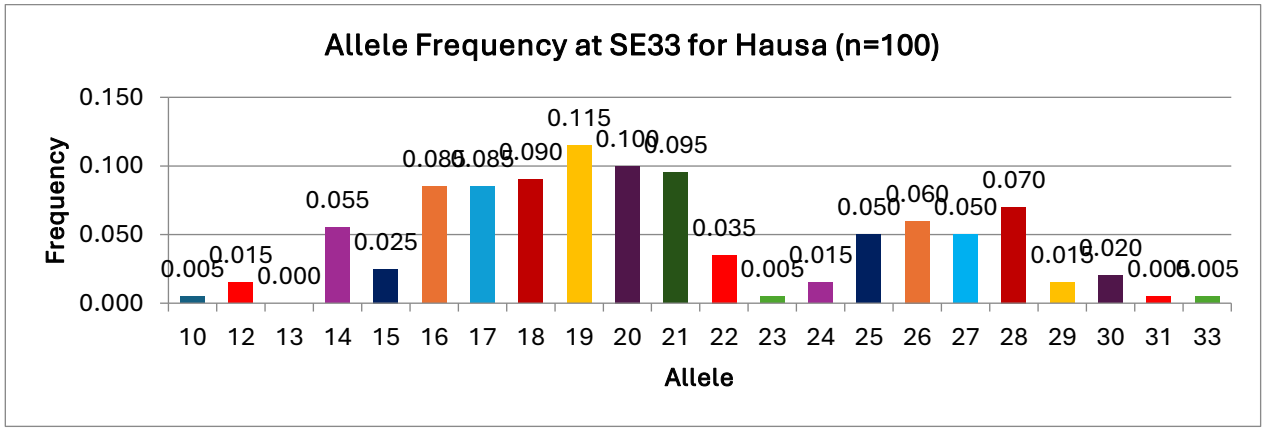
## APPENDIX 43

## ALLELE FREQUENCY DISTRIBUTION AT SE33 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



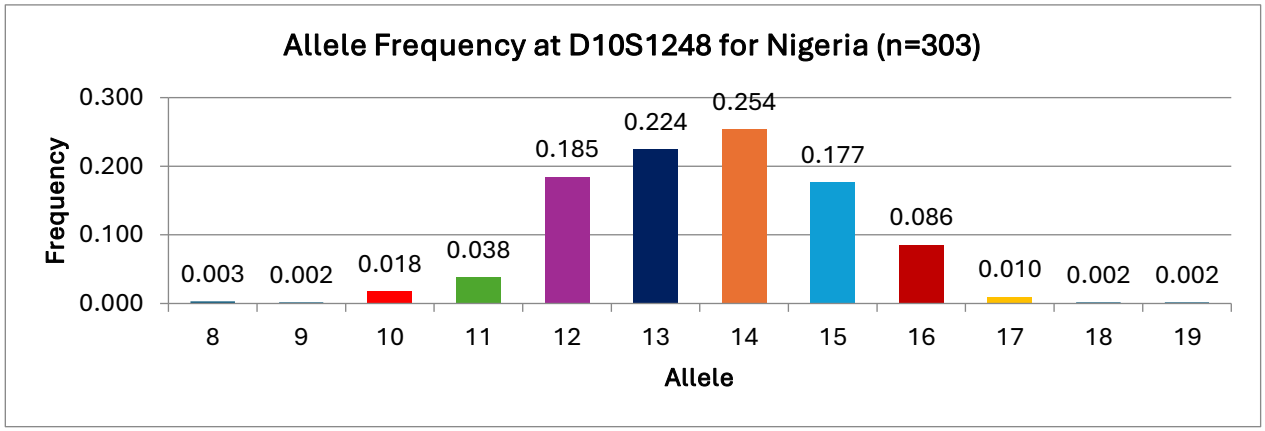



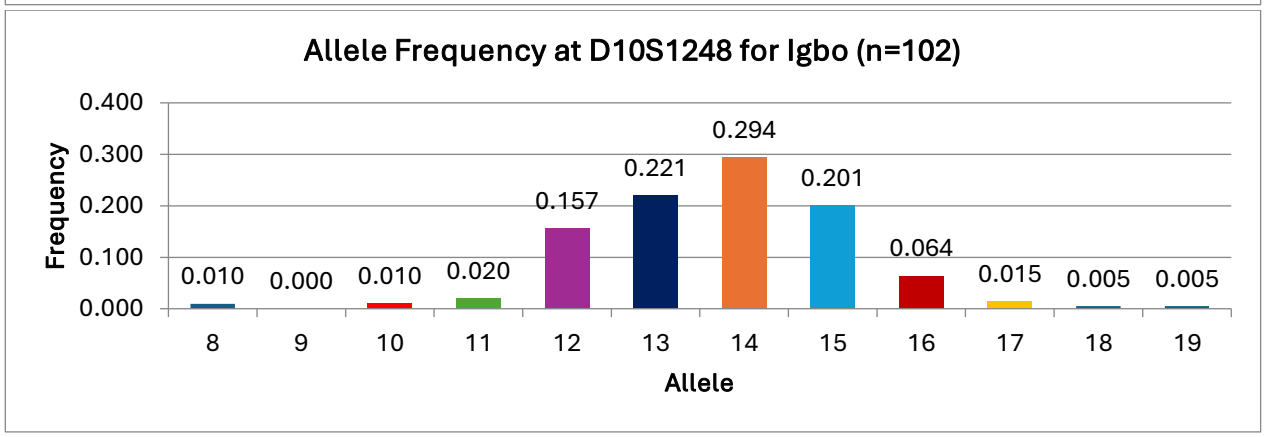



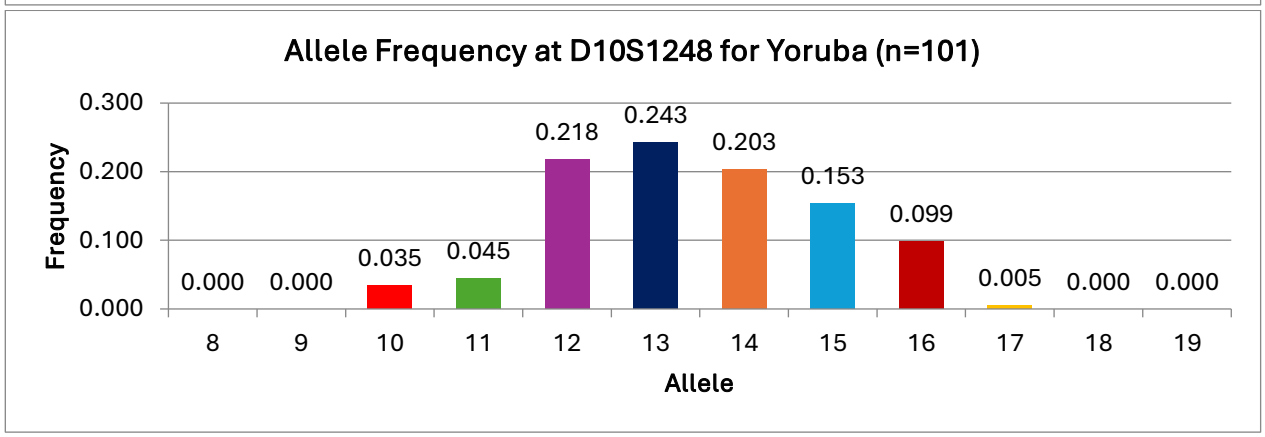



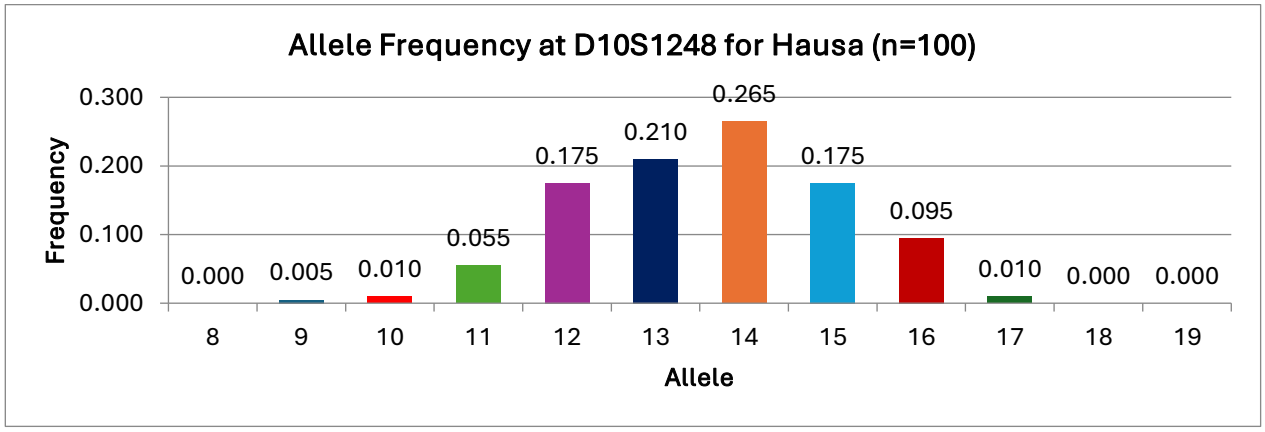
## APPENDIX 44

## ALLELE FREQUENCY DISTRIBUTION AT D10S1248 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



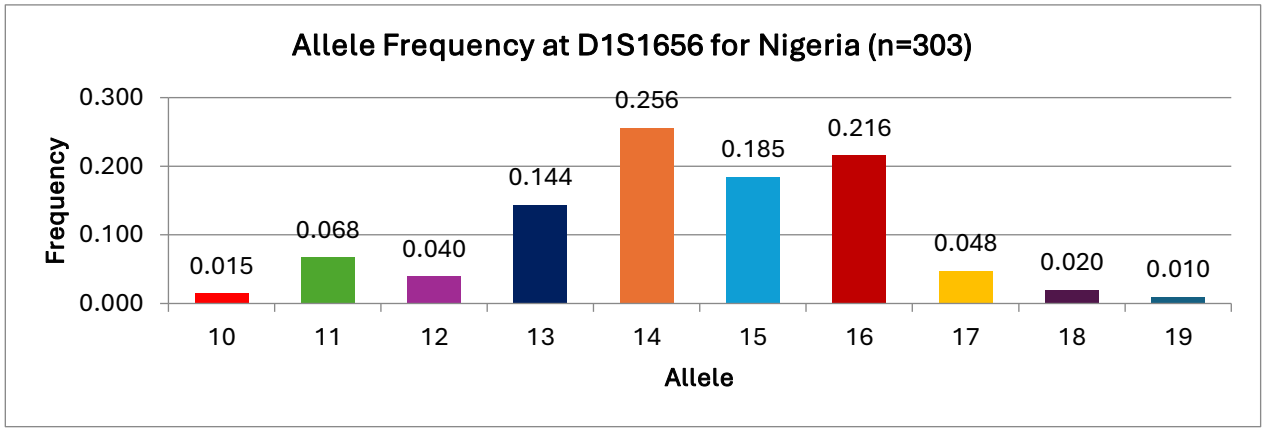



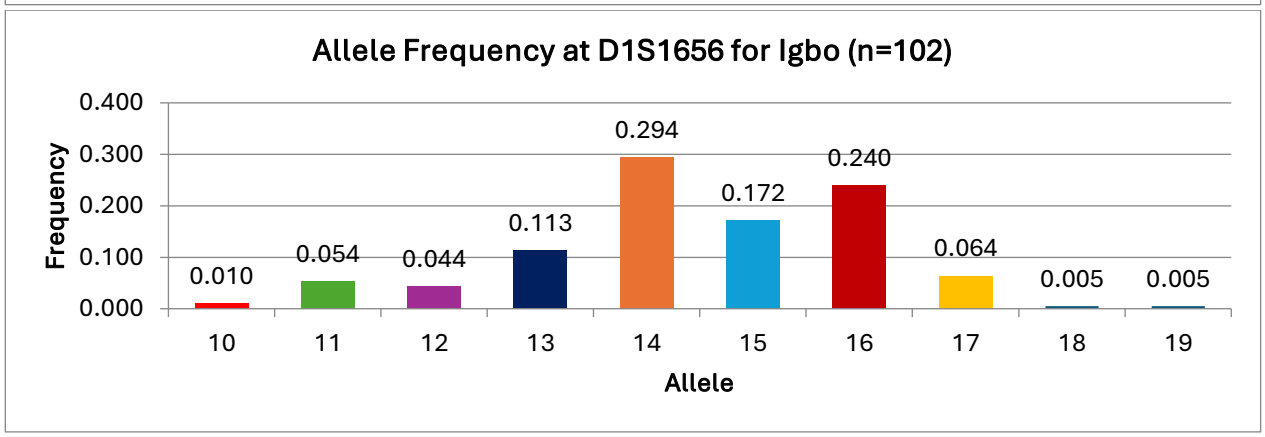



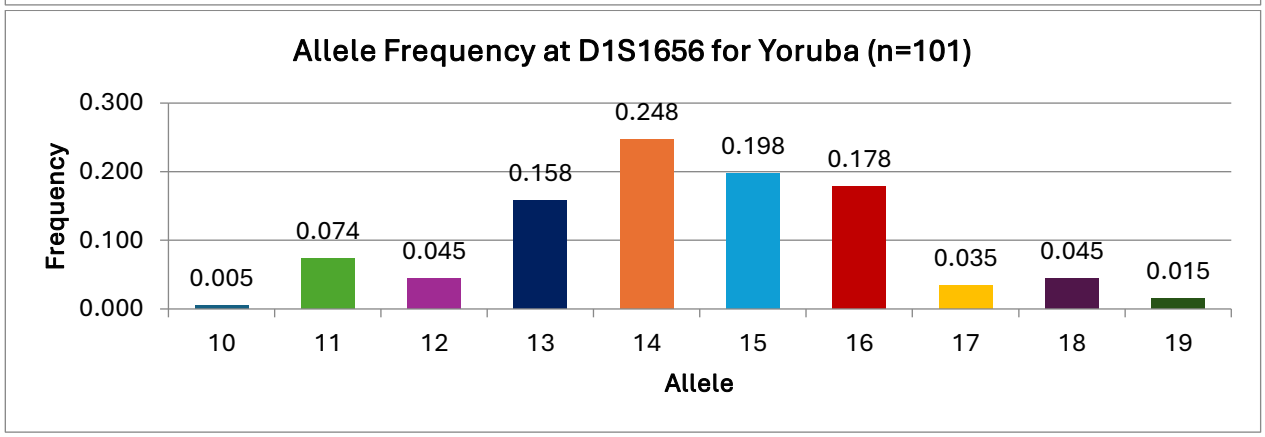



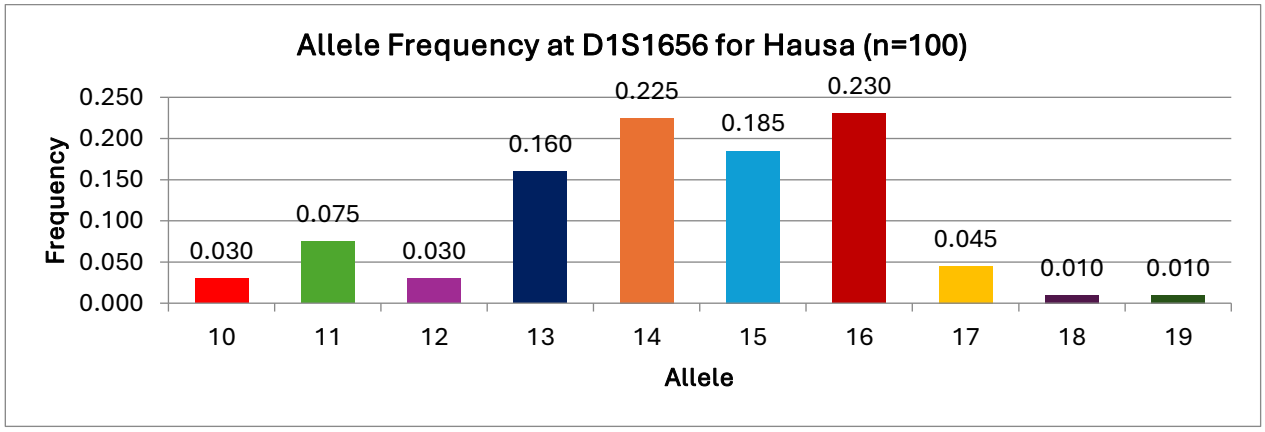
## APPENDIX 45

## ALLELE FREQUENCY DISTRIBUTION AT D1S1656 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



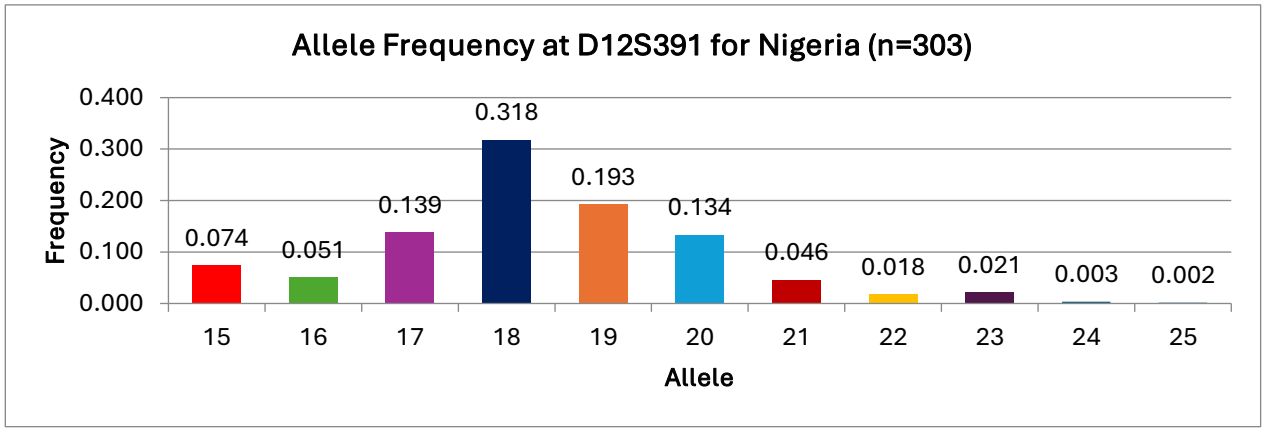



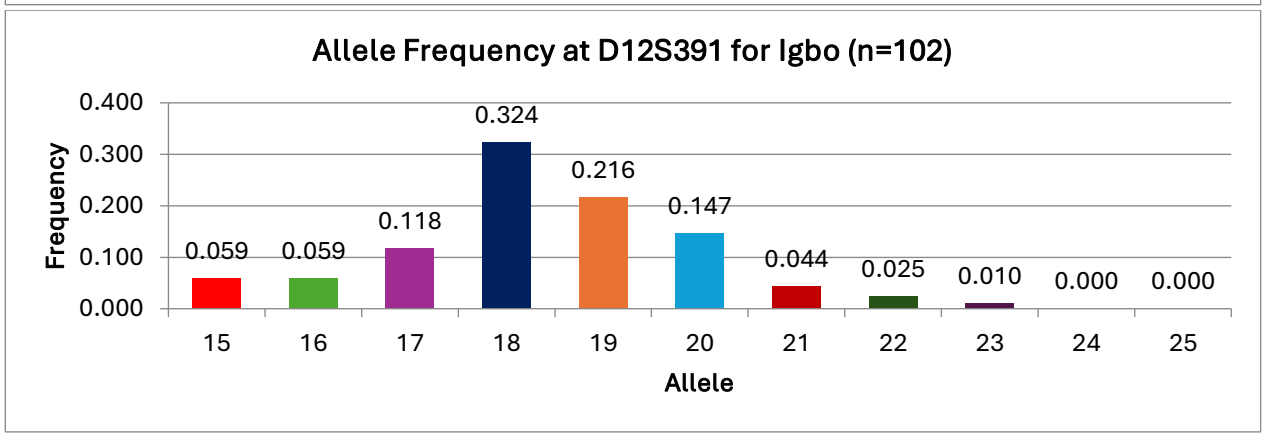



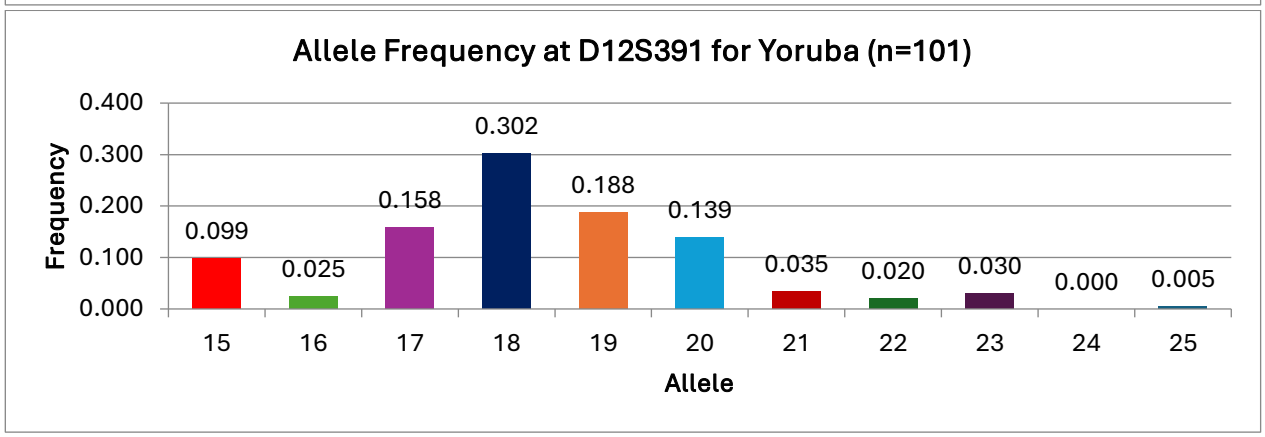



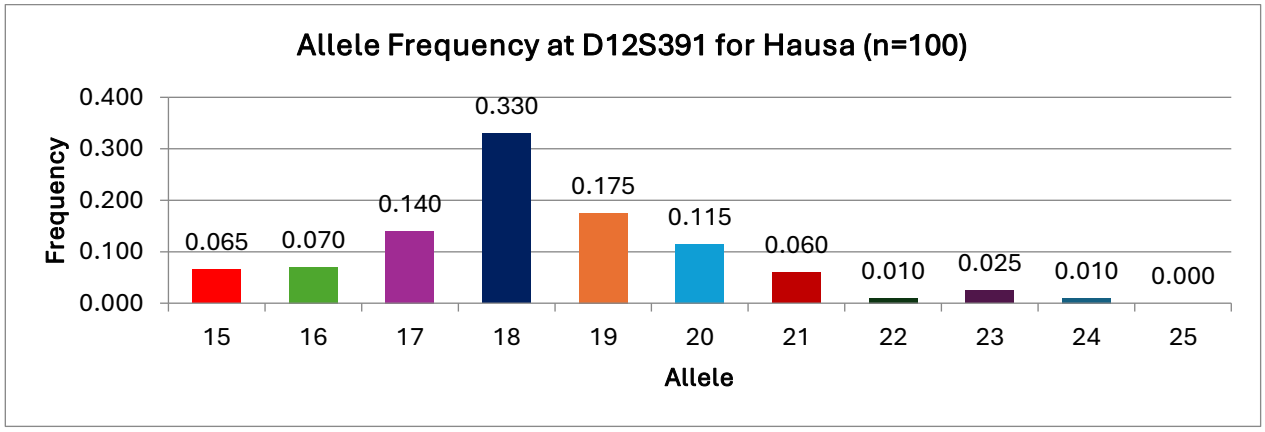
## APPENDIX 46

## ALLELE FREQUENCY DISTRIBUTION AT D12S391 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



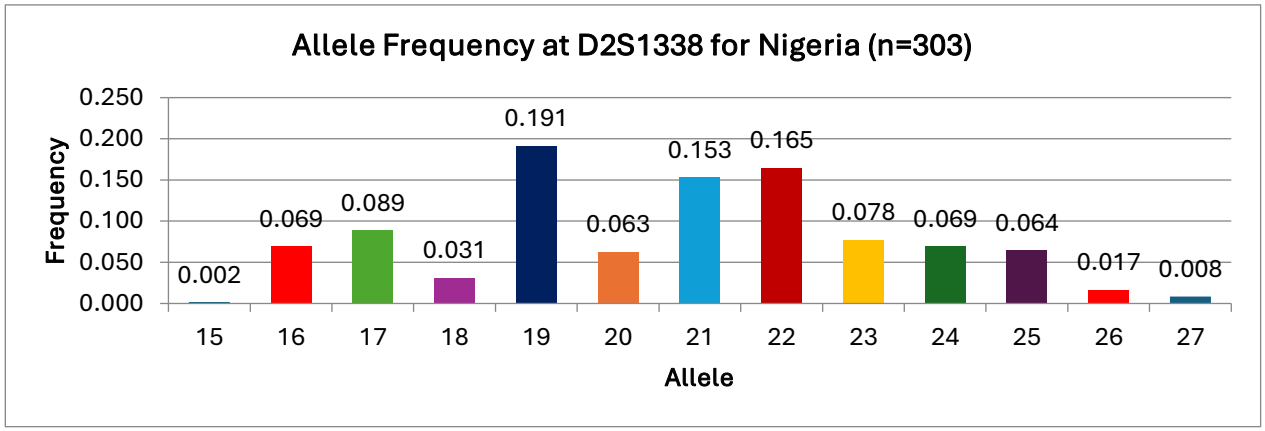



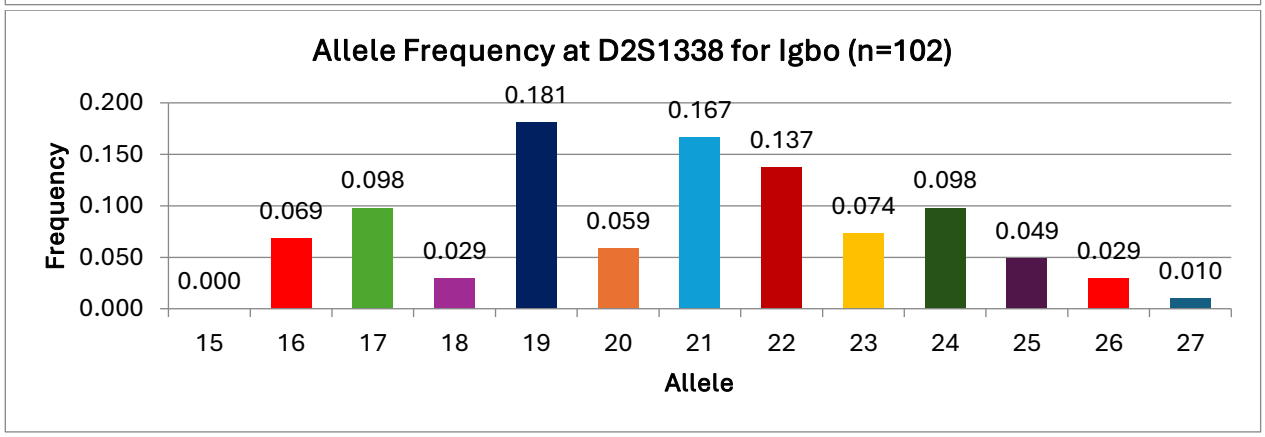



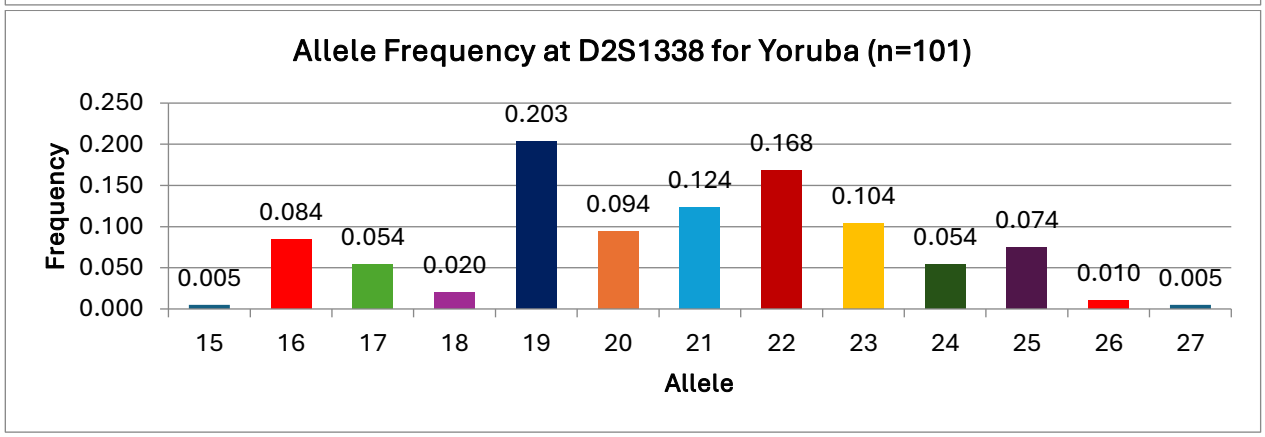



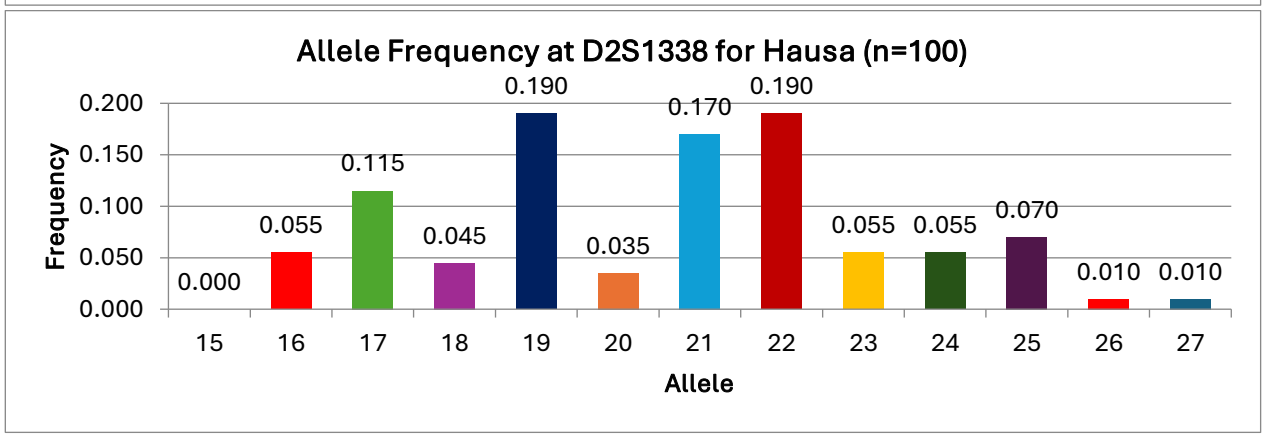
## APPENDIX 47

## ALLELE FREQUENCY DISTRIBUTION AT D2S1338 BY POPULATION USING GLOBALFILER™ EXPRESS KIT



## APPENDIX 48

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM CHI-SQUARE ANALYSES FOR THE NIGERIAN POPULATION USING THE GLOBALFILER™ EXPRESS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Nigerians | D3S1358 | 303 | 9 | 7.28673 | 0.994001 | 1 | False | False |
| Nigerians | vWA | 303 | 10 | 30.54965 | 0.90222 | 1 | False | False |
| Nigerians | D16S539 | 303 | 7 | 8.713353 | 0.981804 | 1 | False | False |
| Nigerians | CSF1PO | 303 | 14 | 28.97867 | 0.960008 | 1 | False | False |
| Nigerians | TPOX | 303 | 9 | 40.34105 | 0.720056 | 1 | False | False |
| Nigerians | D8S1179 | 303 | 8 | 7.397778 | 0.995601 | 1 | False | False |
| Nigerians | D21S11 | 303 | 17 | 111.6167 | 0.071986 | 1 | False | False |
| Nigerians | D18S51 | 303 | 18 | 78.44886 | 0.84903 | 1 | False | False |
| Nigerians | D2S441 | 303 | 9 | 17.78658 | 0.877624 | 1 | False | False |
| Nigerians | D19S433 | 303 | 14 | 48.23306 | 0.443111 | 1 | False | False |
| Nigerians | TH01 | 303 | 6 | 12.9297 | 0.977205 | 1 | False | False |
| Nigerians | FGA | 303 | 19 | 75.43546 | 0.4975 | 1 | False | False |
| Nigerians | D22S1045 | 303 | 10 | 9.33361 | 0.79904 | 1 | False | False |
| Nigerians | D5S818 | 303 | 9 | 29.91231 | 0.4989 | 1 | False | False |
| Nigerians | D13S317 | 303 | 11 | 139.3614 | 0.705859 | 1 | False | False |
| Nigerians | D7S820 | 303 | 8 | 327.4664 | 0.310138 | 1 | False | False |
| Nigerians | SE33 | 303 | 34 | 307.5308 | 0.346131 | 1 | False | False |
| Nigerians | D10S1248 | 303 | 12 | 114.7878 | 0.729454 | 1 | False | False |
| Nigerians | D1S1656 | 303 | 16 | 45.42333 | 0.330534 | 1 | False | False |
| Nigerians | D12S391 | 303 | 14 | 35.41532 | 0.896421 | 1 | False | False |
| Nigerians | D2S1338 | 303 | 13 | 61.24565 | 0.110178 | 1 | False | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 49

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM CHI-SQUARE ANALYSES FOR THE IGBO SUBPOPULATION USING THE GLOBALFILER™ EXPRESS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Igbo | D3S1358 | 102 | 7 | 9.178982 | 0.876425 | 1 | False | False |
| Igbo | vWA | 102 | 9 | 15.0321 | 0.291742 | 1 | False | False |
| Igbo | D16S539 | 102 | 7 | 8.410514 | 0.85163 | 1 | False | False |
| Igbo | CSF1PO | 102 | 10 | 10.15917 | 0.945411 | 1 | False | False |
| Igbo | TPOX | 102 | 8 | 27.48747 | 0.609078 | 1 | False | False |
| Igbo | D8S1179 | 102 | 8 | 12.29534 | 0.675265 | 1 | False | False |
| Igbo | D21S11 | 102 | 14 | 84.04559 | 0.054189 | 0.867027 | False | False |
| Igbo | D18S51 | 102 | 16 | 25.65916 | 0.138572 | 1 | False | False |
| Igbo | D2S441 | 102 | 8 | 19.58584 | 0.403519 | 1 | False | False |
| Igbo | D19S433 | 102 | 11 | 8.336365 | 0.993401 | 1 | False | False |
| Igbo | TH01 | 102 | 6 | 12.103 | 0.69966 | 1 | False | False |
| Igbo | FGA | 102 | 14 | 38.54851 | 0.169366 | 1 | False | False |
| Igbo | D22S1045 | 102 | 9 | 9.864757 | 0.611078 | 1 | False | False |
| Igbo | D5S818 | 102 | 8 | 7.711801 | 0.868026 | 1 | False | False |
| Igbo | D13S317 | 102 | 10 | 206.6647 | 0.229954 | 1 | False | False |
| Igbo | D7S820 | 102 | 6 | 21.31307 | 0.275745 | 1 | False | False |
| Igbo | SE33 | 102 | 24 | 125.3425 | 0.34893 | 1 | False | False |
| Igbo | D10S1248 | 102 | 11 | 81.37624 | 0.088782 | 1 | False | False |
| Igbo | D1S1656 | 102 | 13 | 42.11756 | 0.164167 | 1 | False | False |
| Igbo | D12S391 | 102 | 10 | 23.63335 | 0.69846 | 1 | False | False |
| Igbo | D2S1338 | 102 | 12 | 55.31386 | 0.083383 | 1 | False | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 50

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM CHI-SQUARE ANALYSES FOR THE YORUBA SUBPOPULATION USING THE GLOBALFILER™ EXPRESS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Yoruba | D3S1358 | 101 | 7 | 11.72803 | 0.654869 | 1 | False | False |
| Yoruba | vWA | 101 | 10 | 37.38235 | 0.269346 | 1 | False | False |
| Yoruba | D16S539 | 101 | 7 | 12.3066 | 0.486103 | 1 | False | False |
| Yoruba | CSF1PO | 101 | 10 | 23.26806 | 0.187962 | 1 | False | False |
| Yoruba | TPOX | 101 | 7 | 18.27621 | 0.543491 | 1 | False | False |
| Yoruba | D8S1179 | 101 | 8 | 8.853617 | 0.927614 | 1 | False | False |
| Yoruba | D21S11 | 101 | 15 | 117.0436 | 0.007998 | 0.167966 | True | False |
| Yoruba | D18S51 | 101 | 13 | 30.90531 | 0.904619 | 1 | False | False |
| Yoruba | D2S441 | 101 | 8 | 6.553908 | 0.922216 | 1 | False | False |
| Yoruba | D19S433 | 101 | 12 | 31.98178 | 0.423715 | 1 | False | False |
| Yoruba | TH01 | 101 | 5 | 5.382998 | 0.876625 | 1 | False | False |
| Yoruba | FGA | 101 | 17 | 32.15183 | 0.607279 | 1 | False | False |
| Yoruba | D22S1045 | 101 | 10 | 11.7162 | 0.10198 | 1 | False | False |
| Yoruba | D5S818 | 101 | 8 | 13.08741 | 0.536893 | 1 | False | False |
| Yoruba | D13S317 | 101 | 9 | 103.4509 | 0.566687 | 1 | False | False |
| Yoruba | D7S820 | 101 | 7 | 8.373012 | 0.720056 | 1 | False | False |
| Yoruba | SE33 | 101 | 27 | 121.3148 | 0.312737 | 1 | False | False |
| Yoruba | D10S1248 | 101 | 8 | 8.798508 | 0.885023 | 1 | False | False |
| Yoruba | D1S1656 | 101 | 15 | 37.42969 | 0.128174 | 1 | False | False |
| Yoruba | D12S391 | 101 | 11 | 22.22012 | 0.59948 | 1 | False | False |
| Yoruba | D2S1338 | 101 | 13 | 33.26308 | 0.410918 | 1 | False | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 51

## SUMMARY OF HARDY-WEINBERG EQUILIBRIUM CHI-SQUARE ANALYSES FOR THE HAUSA-FULANI SUBPOPULATION USING THE GLOBALFILER™ EXPRESS KIT

| Population | Locus | Number of Individuals | Number of alleles observed | Chi-square statistic | Monte-Carlo p-value | Bonferroni-corrected p-value | Significance at α = 0.05 (uncorrected) | Significance after Bonferroni correction (α = 0.05) |
|---|---|---|---|---|---|---|---|---|
| Hausa-Fulani | D3S1358 | 100 | 8 | 6.010559 | 0.560088 | 1 | False | False |
| Hausa-Fulani | vWA | 100 | 9 | 18.66617 | 0.35213 | 1 | False | False |
| Hausa-Fulani | D16S539 | 100 | 7 | 7.835599 | 0.929014 | 1 | False | False |
| Hausa-Fulani | CSF1PO | 100 | 10 | 23.09603 | 0.580684 | 1 | False | False |
| Hausa-Fulani | TPOX | 100 | 8 | 14.27765 | 0.182763 | 1 | False | False |
| Hausa-Fulani | D8S1179 | 100 | 8 | 5.151131 | 0.983203 | 1 | False | False |
| Hausa-Fulani | D21S11 | 100 | 17 | 108.5235 | 0.009198 | 0.193161 | True | False |
| Hausa-Fulani | D18S51 | 100 | 15 | 60.97538 | 0.728454 | 1 | False | False |
| Hausa-Fulani | D2S441 | 100 | 8 | 7.382894 | 0.935413 | 1 | False | False |
| Hausa-Fulani | D19S433 | 100 | 8 | 32.47814 | 0.176941 | 1 | False | False |
| Hausa-Fulani | TH01 | 100 | 5 | 2.156385 | 0.994801 | 1 | False | False |
| Hausa-Fulani | FGA | 100 | 15 | 58.96937 | 0.660868 | 1 | False | False |
| Hausa-Fulani | D22S1045 | 100 | 9 | 13.57248 | 0.538492 | 1 | False | False |
| Hausa-Fulani | D5S818 | 100 | 8 | 108.5186 | 0.062787 | 1 | False | False |
| Hausa-Fulani | D13S317 | 100 | 7 | 0.709464 | 0.994601 | 1 | False | False |
| Hausa-Fulani | D7S820 | 100 | 7 | 109.8896 | 0.029594 | 0.621476 | True | False |
| Hausa-Fulani | SE33 | 100 | 28 | 175.1946 | 0.955009 | 1 | False | False |
| Hausa-Fulani | D10S1248 | 100 | 9 | 6.258621 | 0.884223 | 1 | False | False |
| Hausa-Fulani | D1S1656 | 100 | 16 | 59.87861 | 0.302939 | 1 | False | False |
| Hausa-Fulani | D12S391 | 100 | 12 | 15.42037 | 0.868626 | 1 | False | False |
| Hausa-Fulani | D2S1338 | 100 | 12 | 94.52539 | 0.033989 | 0.543819 | True | False |

**Significance at α = 0.05 (uncorrected)**

- **True** → Significant deviation from HWE at α = 0.05 (uncorrected)
- **False** → No significant deviation (uncorrected)

**Significance after Bonferroni correction (α = 0.05)**

- **True** → Significant deviation from HWE *after* Bonferroni correction
- **False** → No significant deviation after correction

## APPENDIX 52

## NEI'S GENETIC DISTANCE PAIRWISE POPULATION MATRIX FOR THE GLOBALFILER™ EXPRESS KIT

| Sub-Population | Igbo | Yoruba | Hausa-Fulani | Egypt |
|---|---|---|---|---|
| **Igbo** | 0.000 | | | |
| **Yoruba** | 0.020 | 0.000 | | |
| **Hausa-Fulani** | 0.025 | 0.023 | 0.000 | |
| **Egypt** | 0.088 | 0.089 | 0.086 | 0.000 |

***Egypt is included as an outgroup.**

## APPENDIX 53

## NEI'S GENETIC IDENTITY PAIRWISE POPULATION MATRIX FOR THE GLOBALFILER™ EXPRESS KIT

| Sub-Population | Igbo | Yoruba | Hausa-Fulani | Egypt |
|---|---|---|---|---|
| **Igbo** | 1.000 | | | |
| **Yoruba** | 0.980 | 1.000 | | |
| **Hausa-Fulani** | 0.975 | 0.977 | 1.000 | |
| **Egypt** | 0.912 | 0.913 | 0.914 | 0.000 |

***Egypt is included as an outgroup.**

## APPENDIX 54

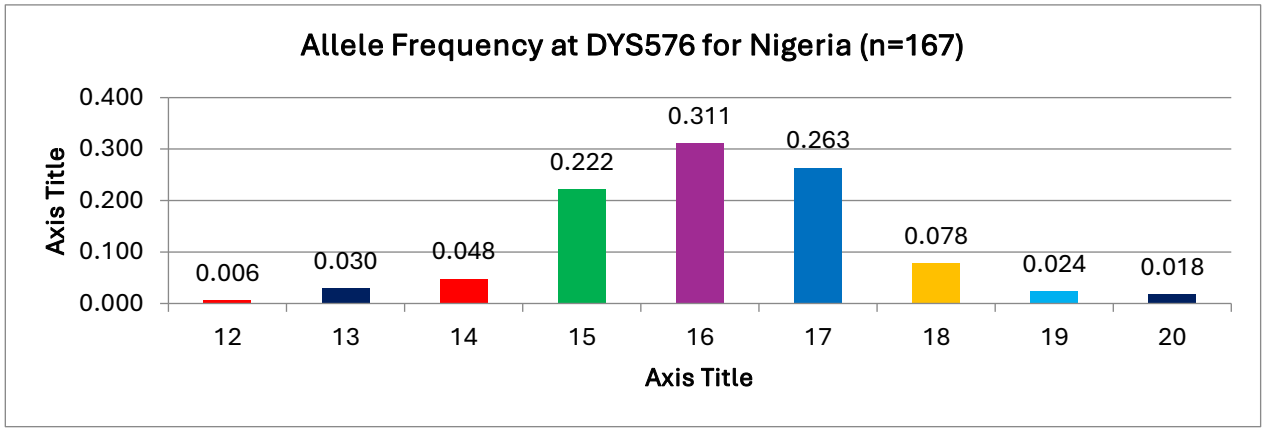
## MEAN LnP(K)±SD AND DELTA *K* VALUES FOR THE GLOBALFILER™ EXPRESS KIT

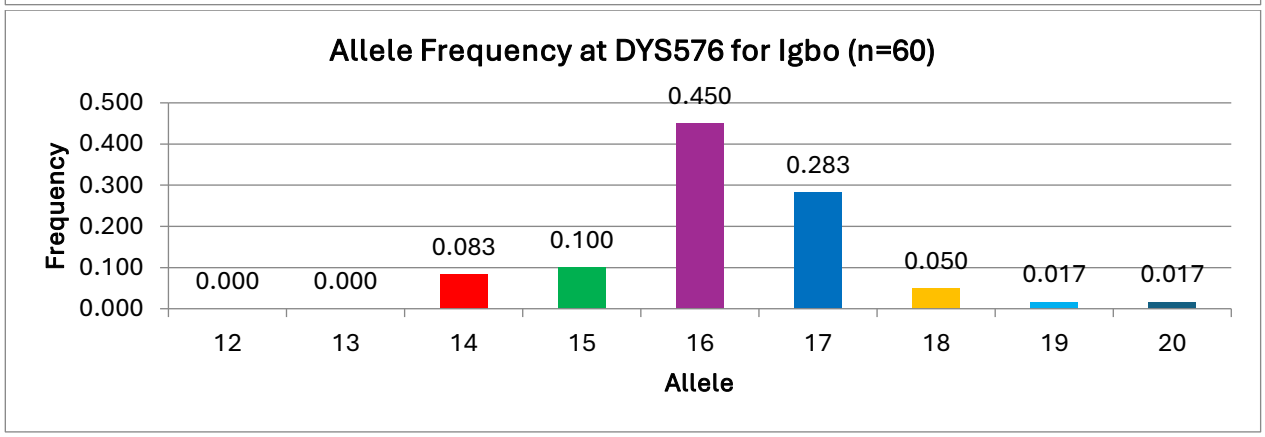
| *K* | Reps | Mean LnP(*K*) | SD LnP(*K*) | Ln'(*K*) | \|Ln''(K)\| | Delta *K* |
|---|---|---|---|---|---|---|
| 1 | 5 | -24131.4 | 0.96799 | NA | NA | NA |
| 2 | 5 | -24183.7 | 14.41676 | -52.24 | 22.18 | 1.53849 |
| 3 | 5 | -24258.1 | 57.65264 | -74.42 | 122.46 | 2.1241 |
| 4 | 5 | -24455 | 154.7978 | -196.88 | 51.1 | 0.33011 |
| 5 | 5 | -24702.9 | 160.6426 | -247.98 | 166.02 | 1.03347 |
| 6 | 5 | -24784.9 | 442.1469 | -81.96 | 54.28 | 0.12276 |
| 7 | 5 | -24812.6 | 315.2258 | -27.68 | 53.96 | 0.17118 |
| 8 | 5 | -24894.2 | 438.8419 | -81.64 | 64.42 | 0.1468 |
| 9 | 5 | -24911.4 | 513.501 | -17.22 | 151.6 | 0.29523 |
| 10 | 5 | -25080.3 | 668.7386 | -168.82 | NA | NA |

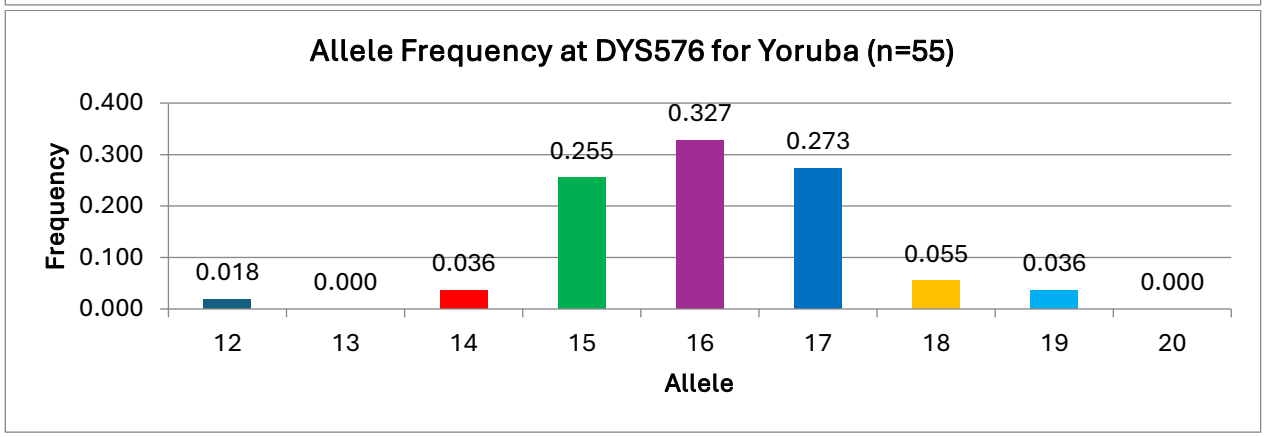
The peak of Delta *K* (*K* = 3), indicating the optimal number of clusters, is highlighted in red.

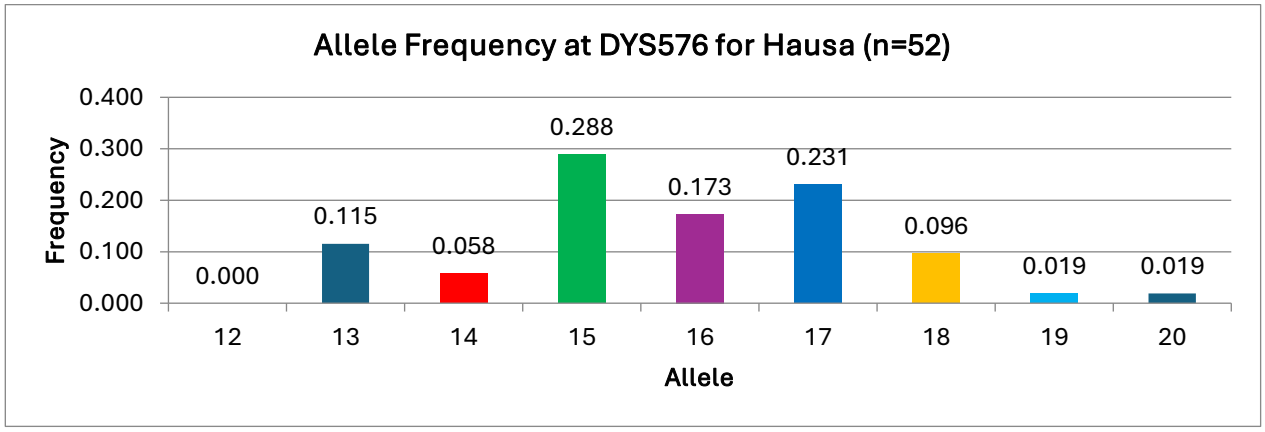
## APPENDIX 55

## ALLELE FREQUENCIES AT DYS576 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



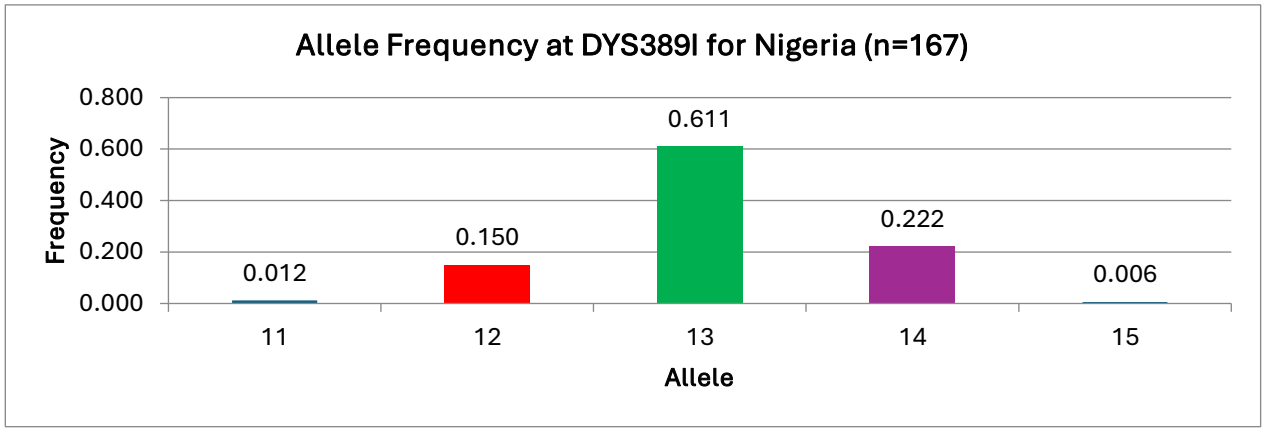



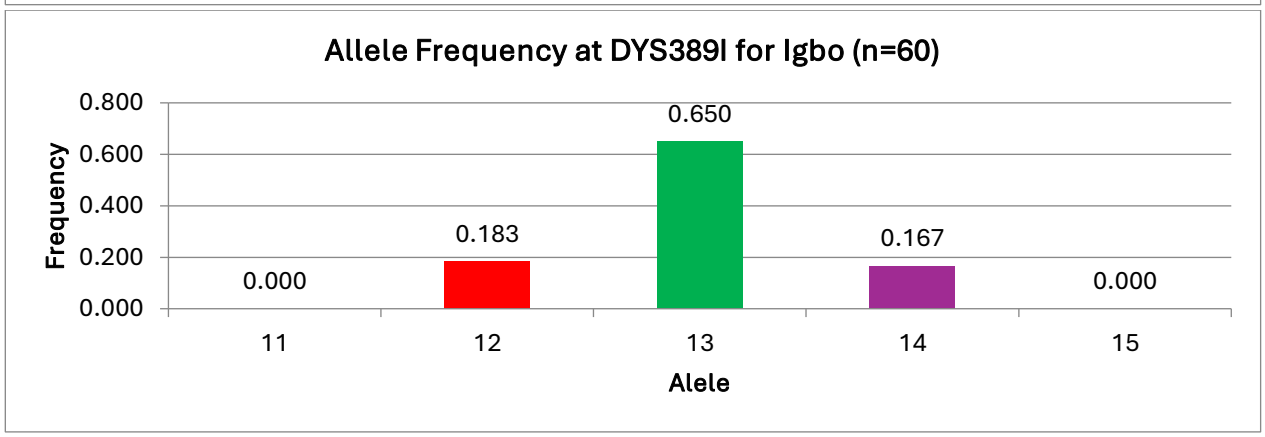



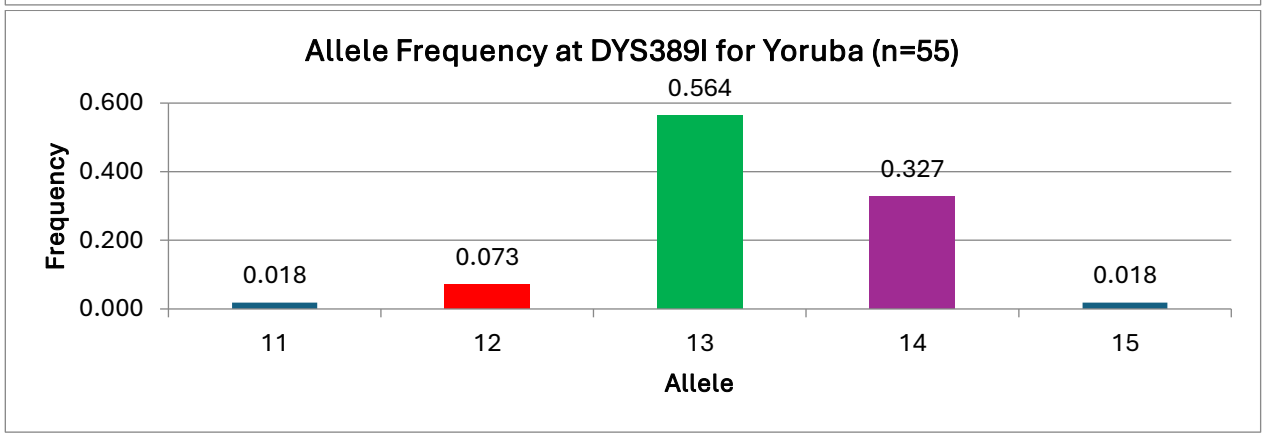



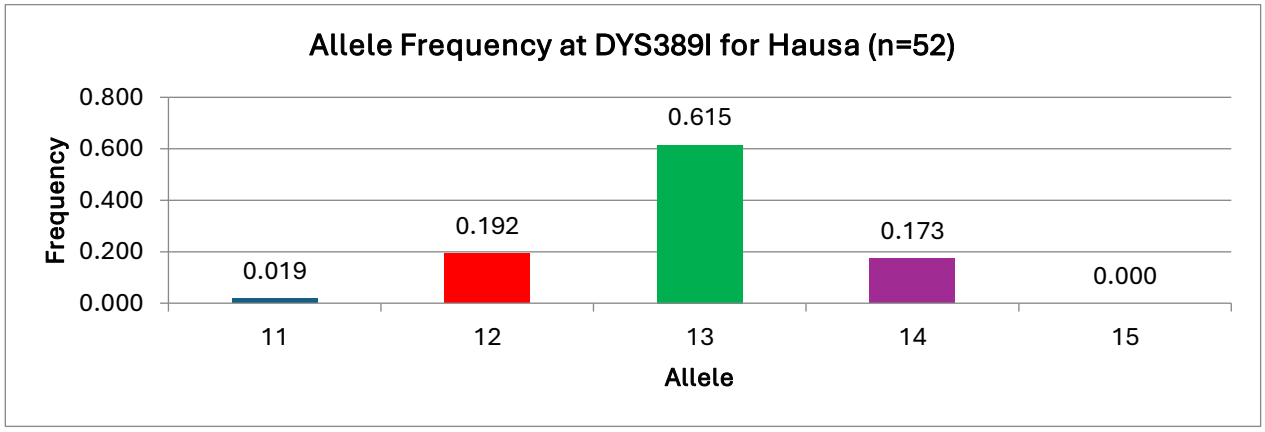
## APPENDIX 56

## ALLELE FREQUENCIES AT DYS389I WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



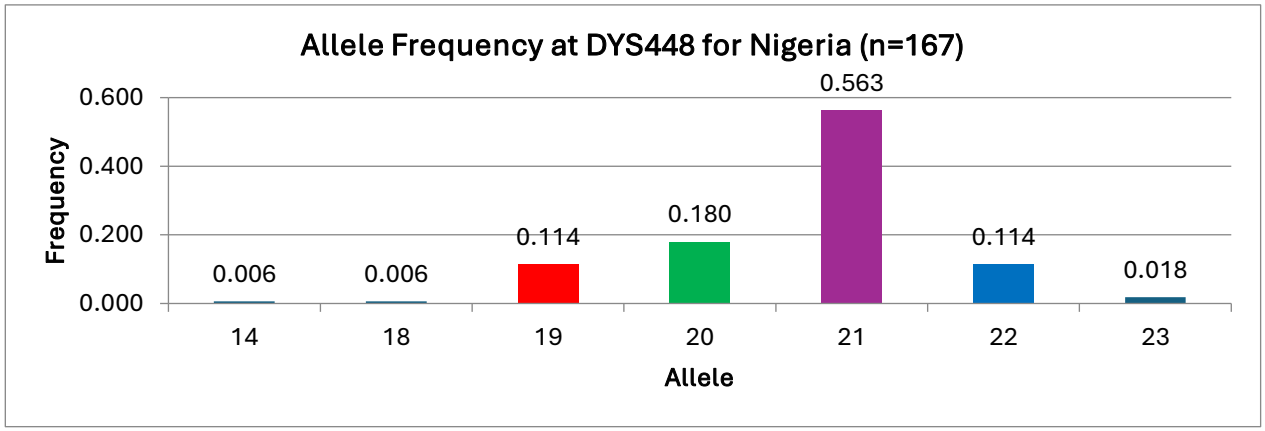



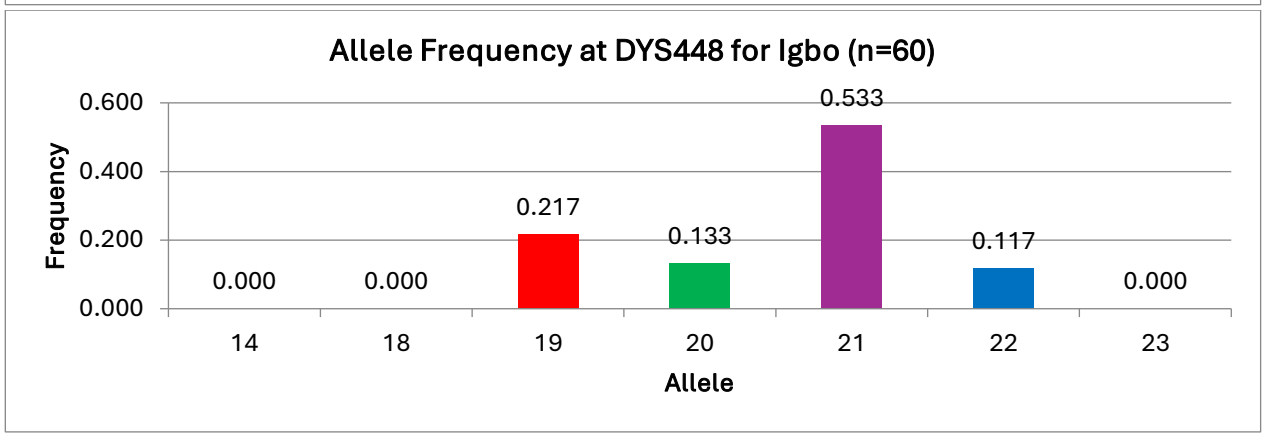



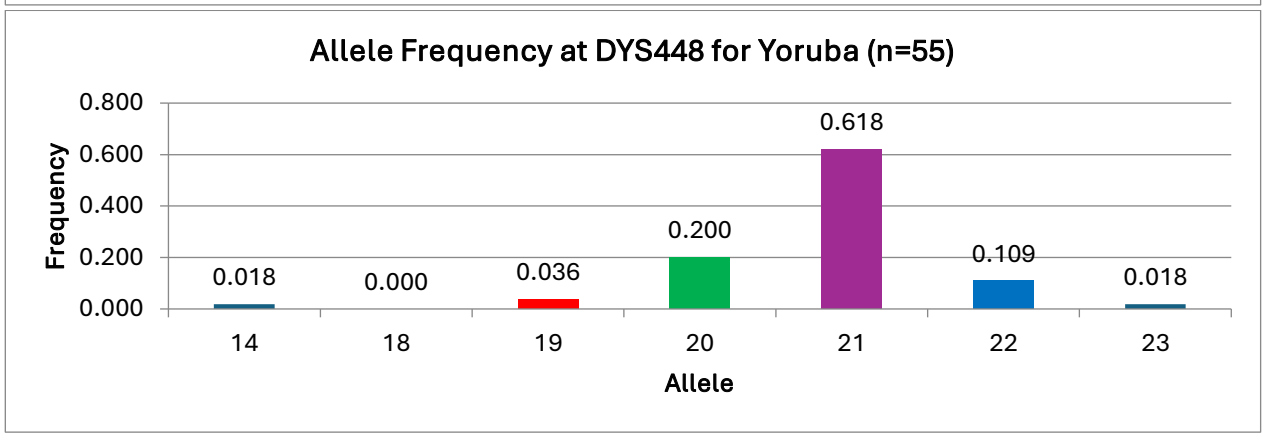



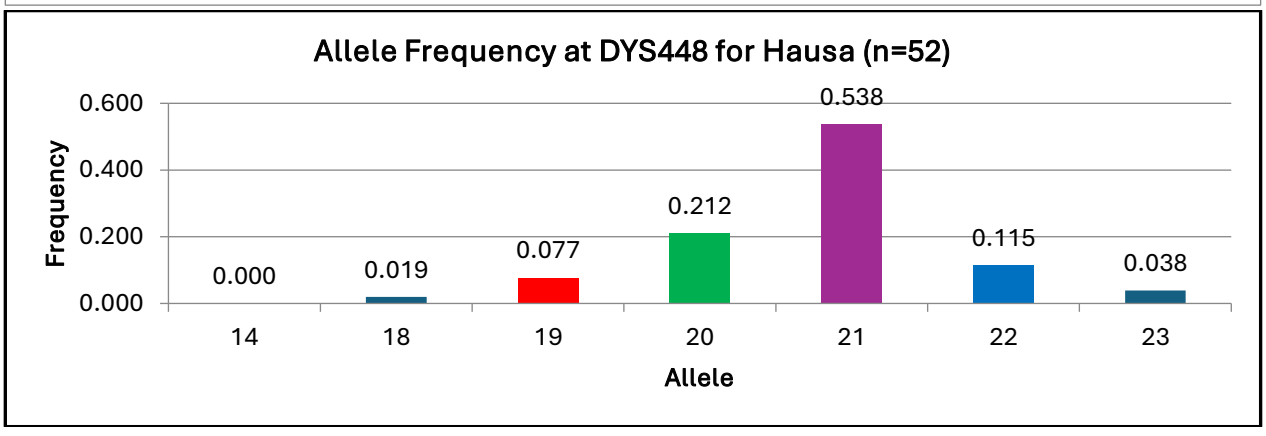
## APPENDIX 57

## ALLELE FREQUENCIES AT DYS448 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



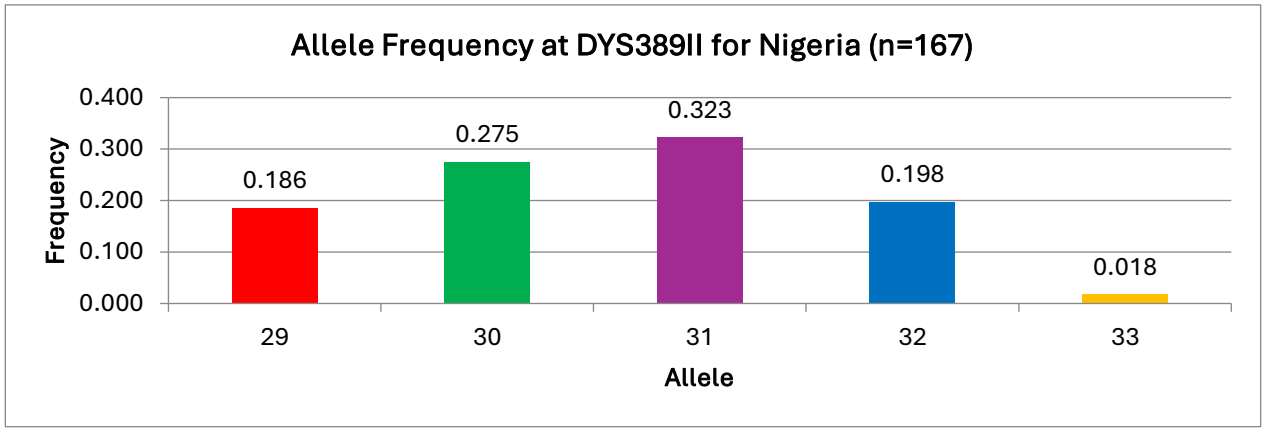



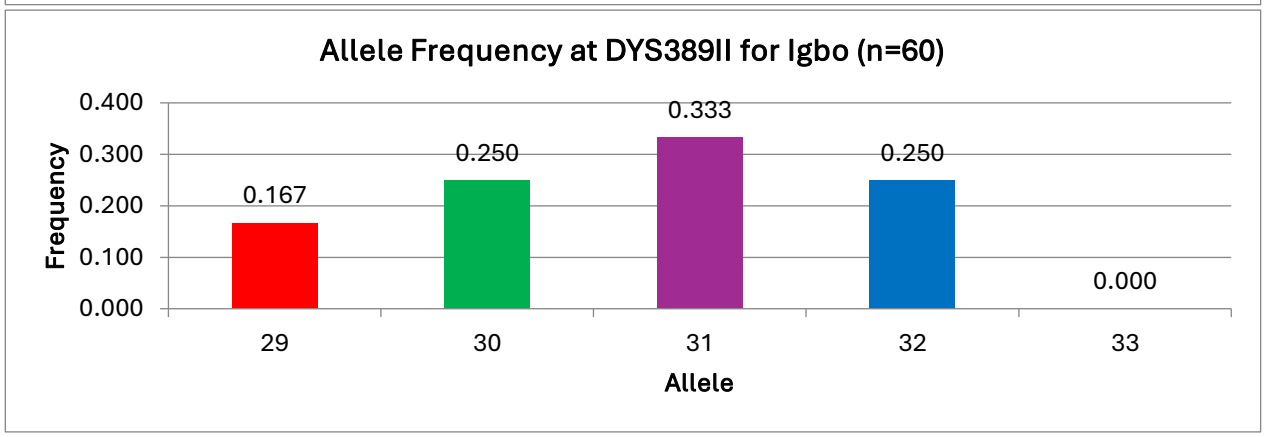



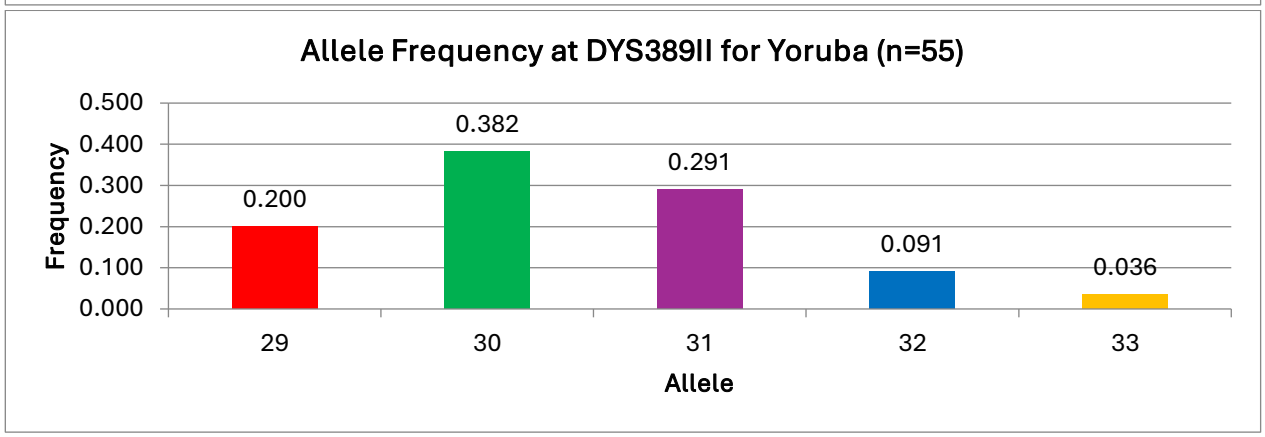



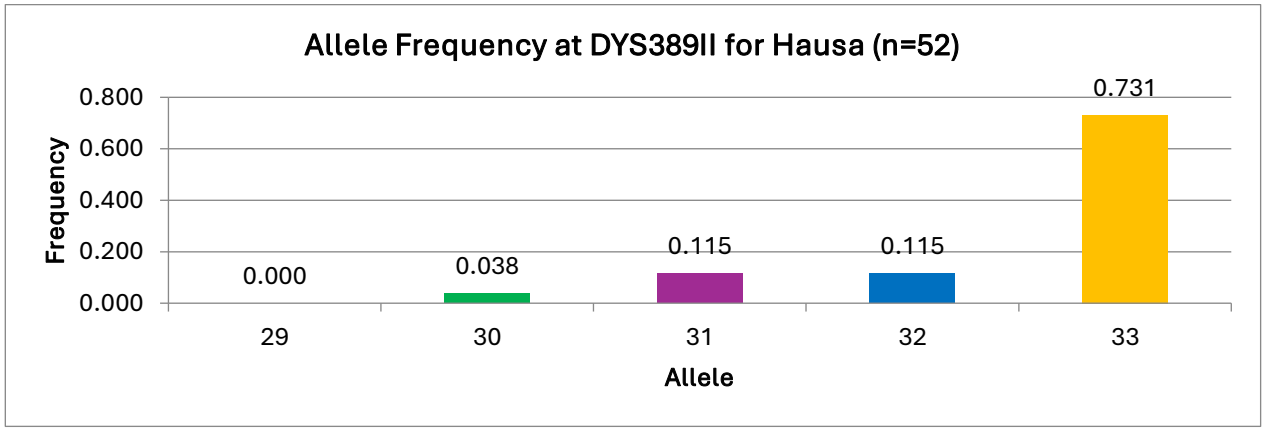
## APPENDIX 58

## ALLELE FREQUENCIES AT DYS389II WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



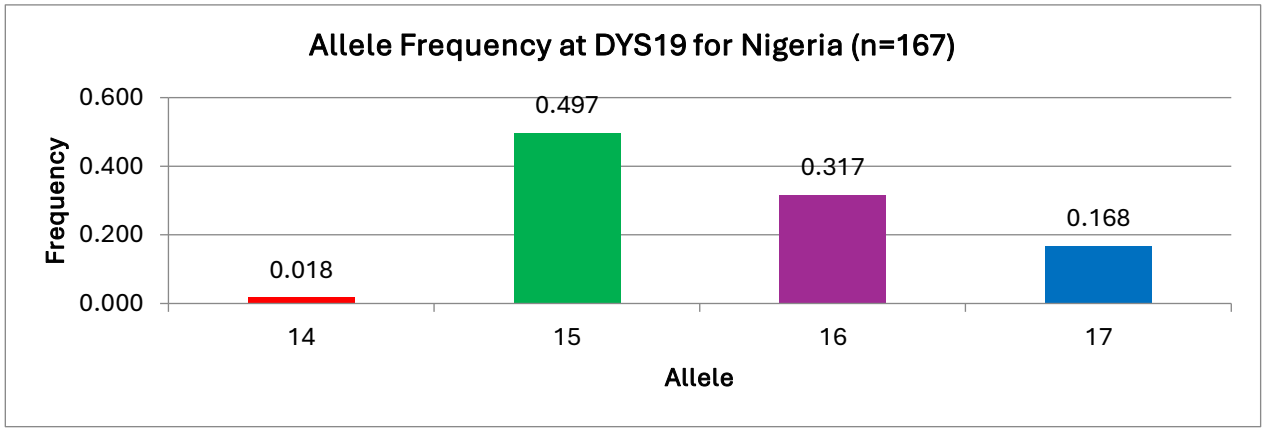



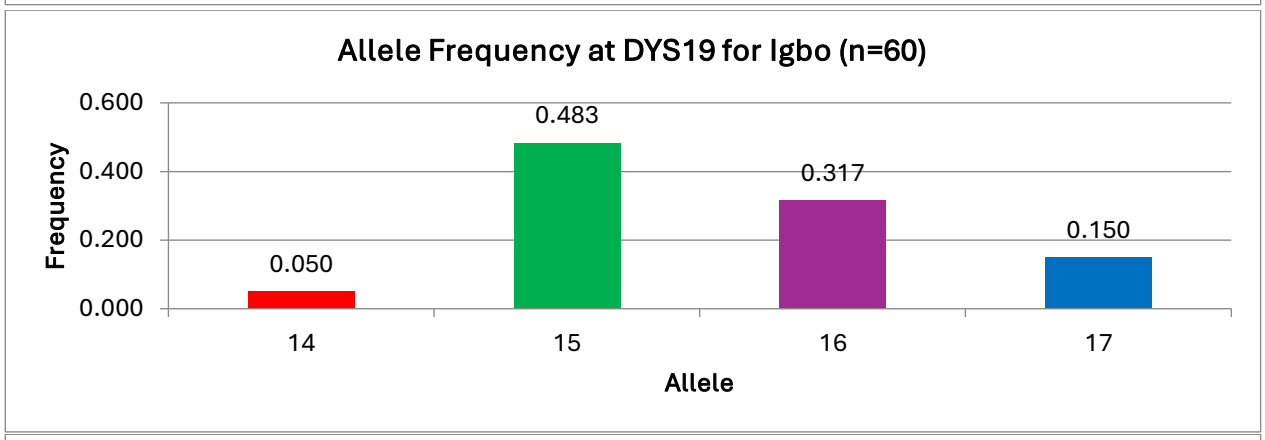



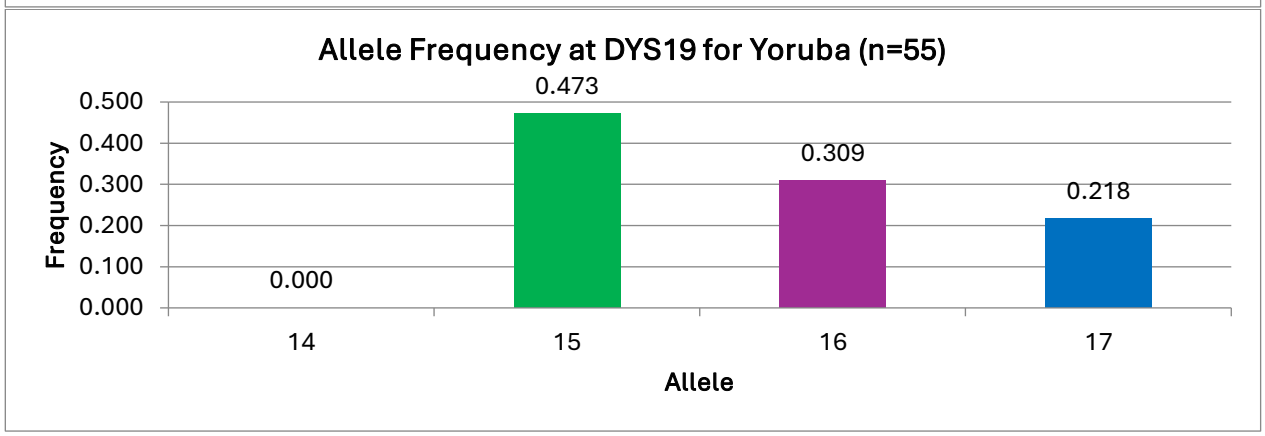



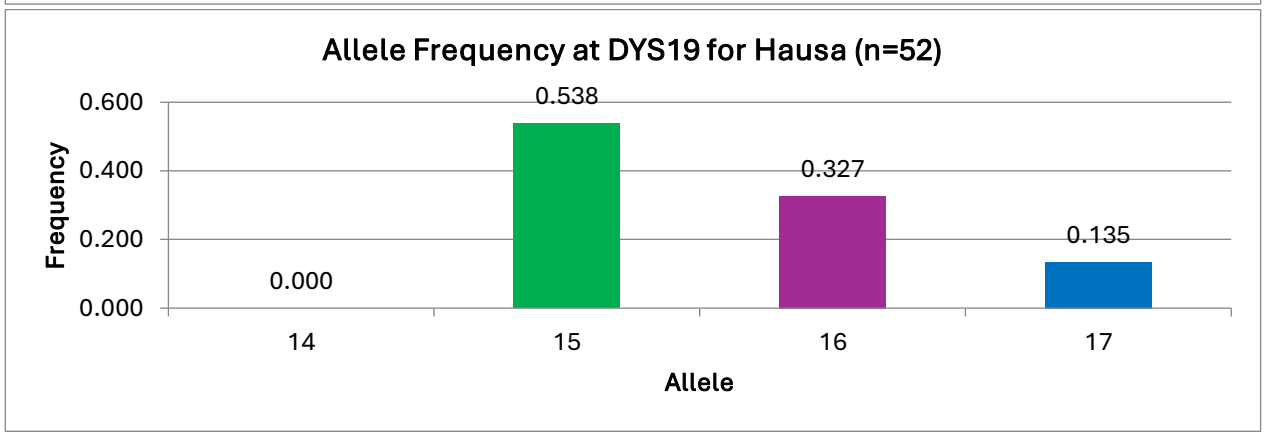
## APPENDIX 59

## ALLELE FREQUENCIES AT DYS19 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



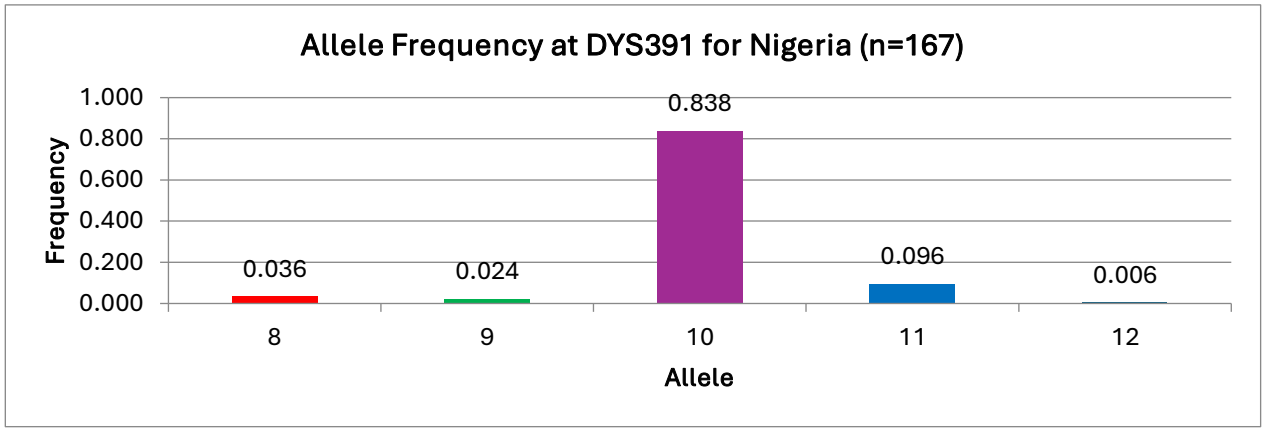



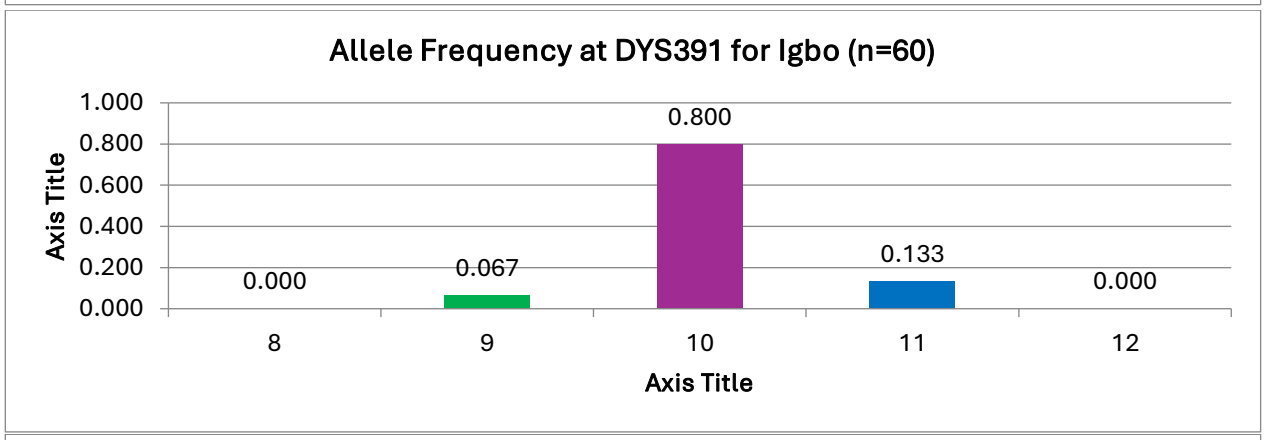



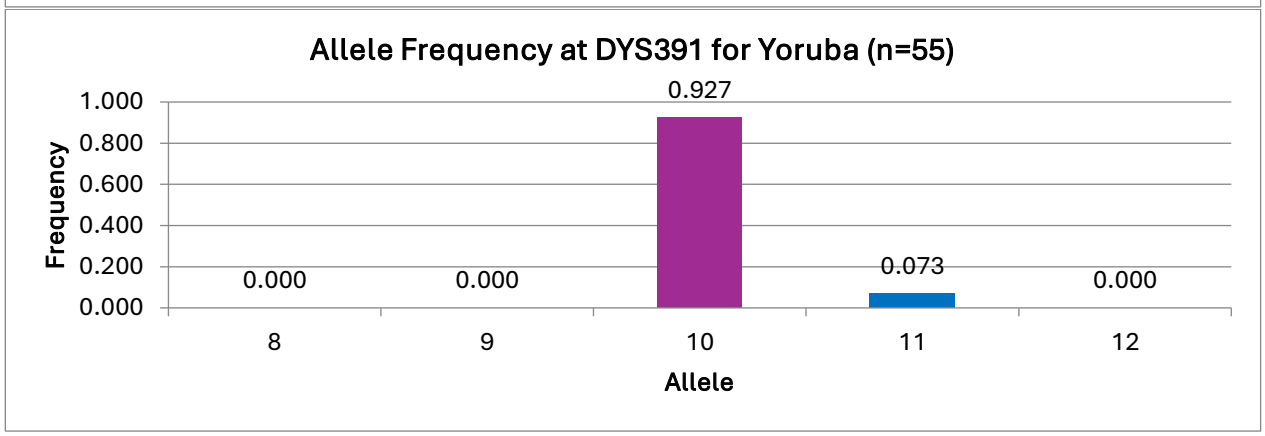



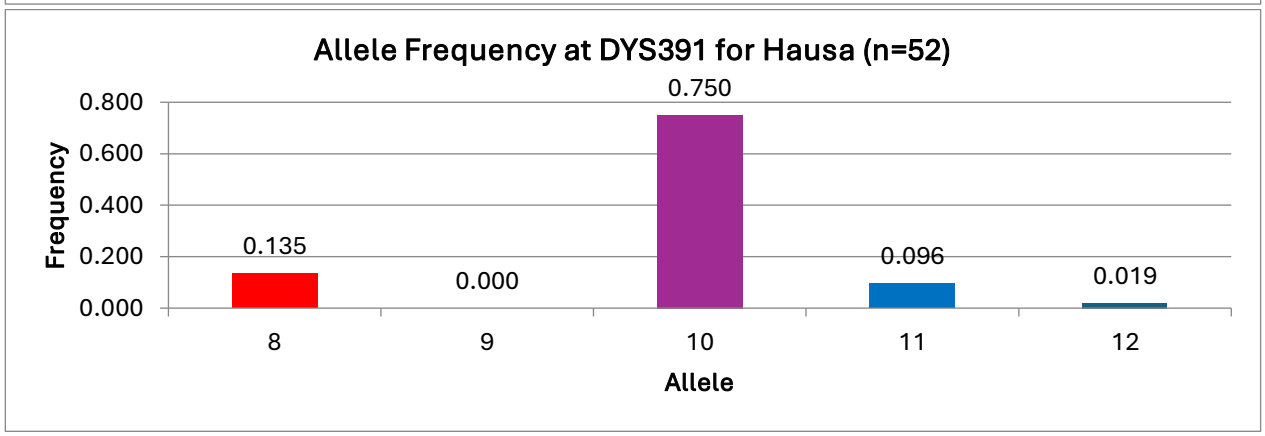
## APPENDIX 60

## ALLELE FREQUENCIES AT DYS391 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



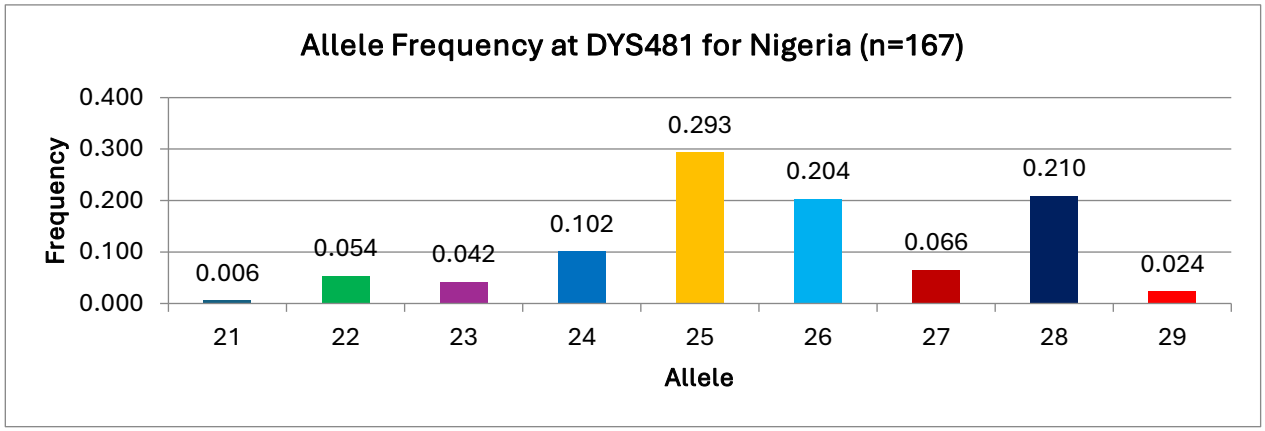



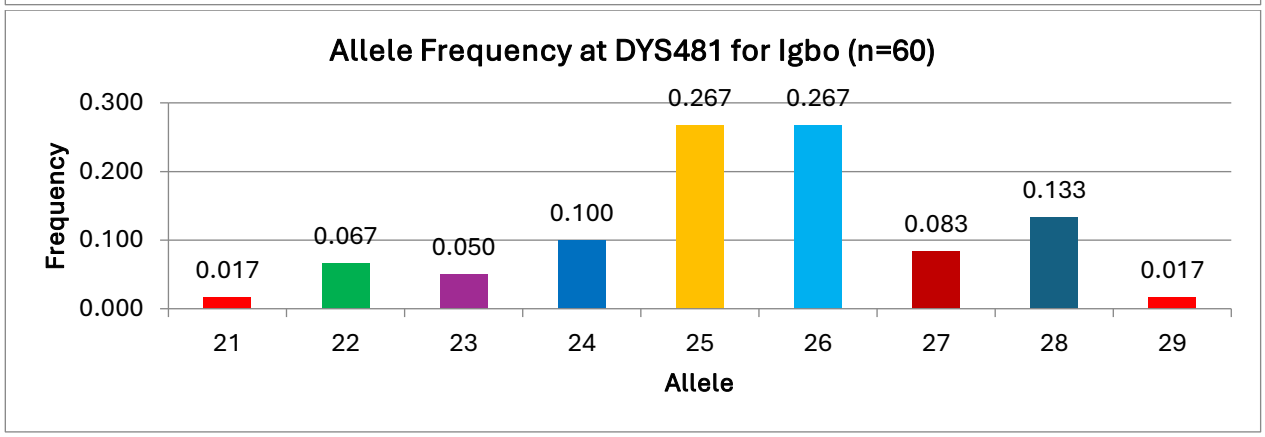



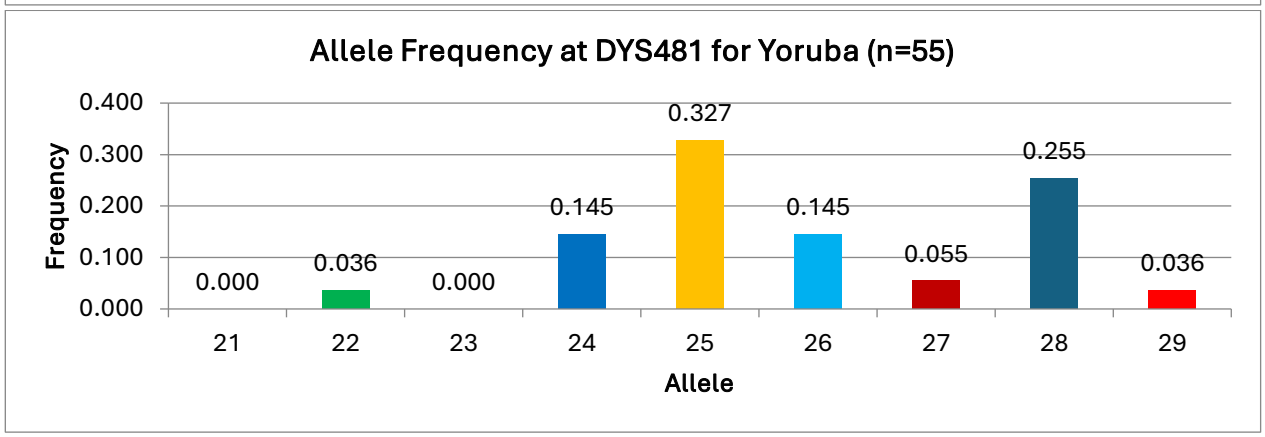



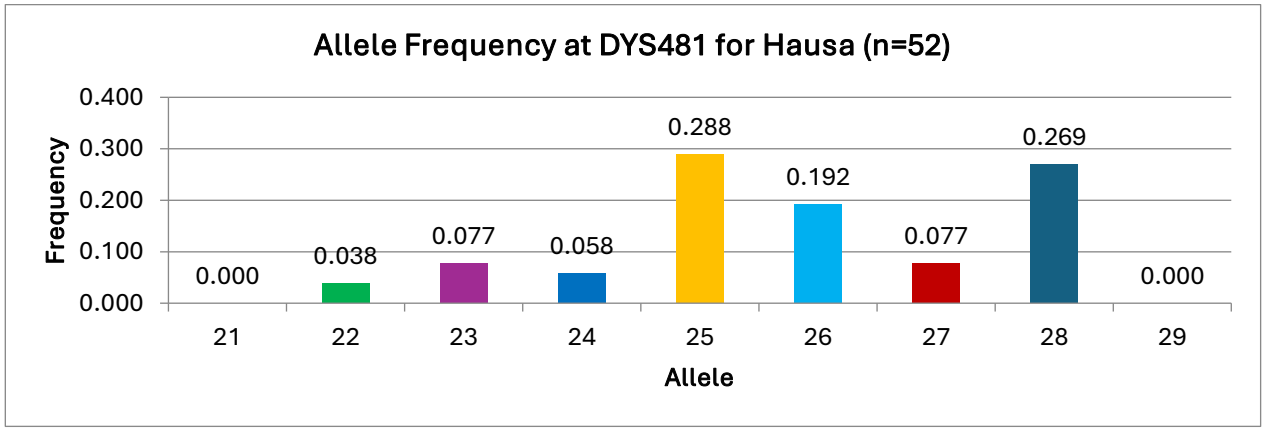
## APPENDIX 61

## ALLELE FREQUENCIES AT DYS481 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



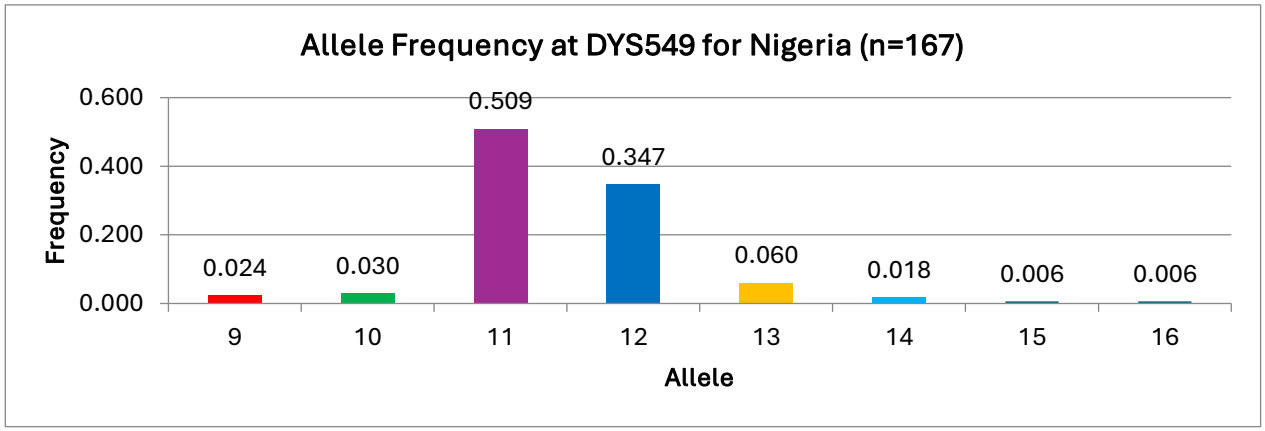



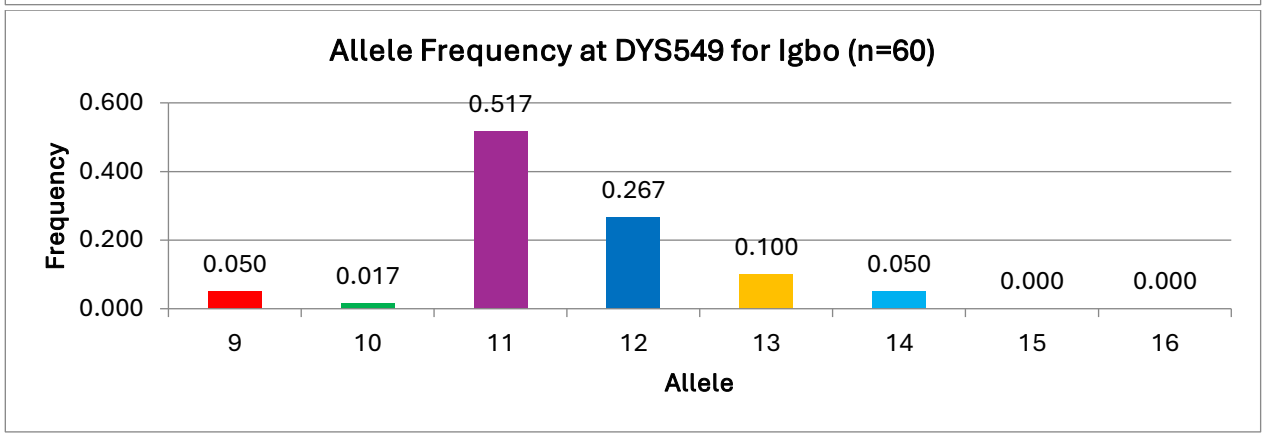



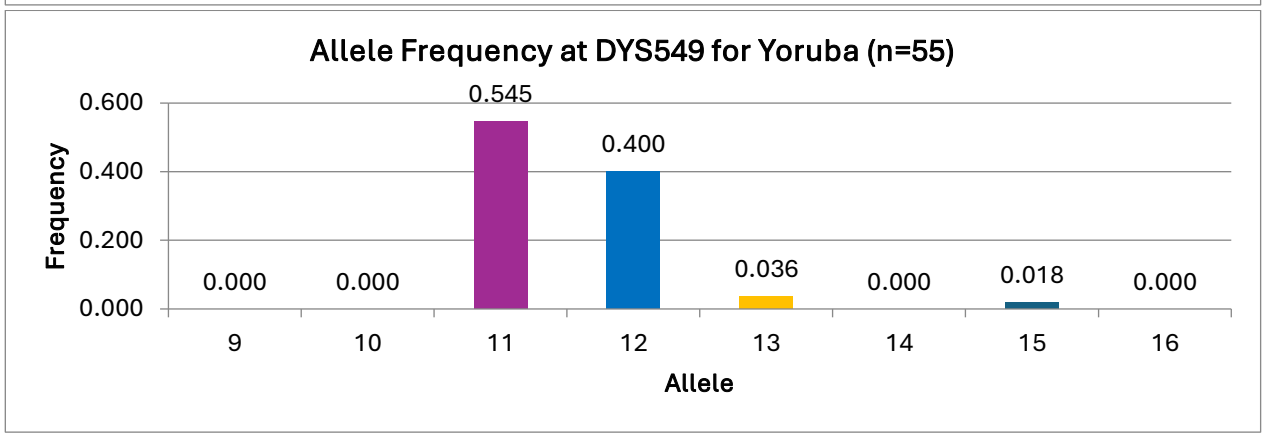



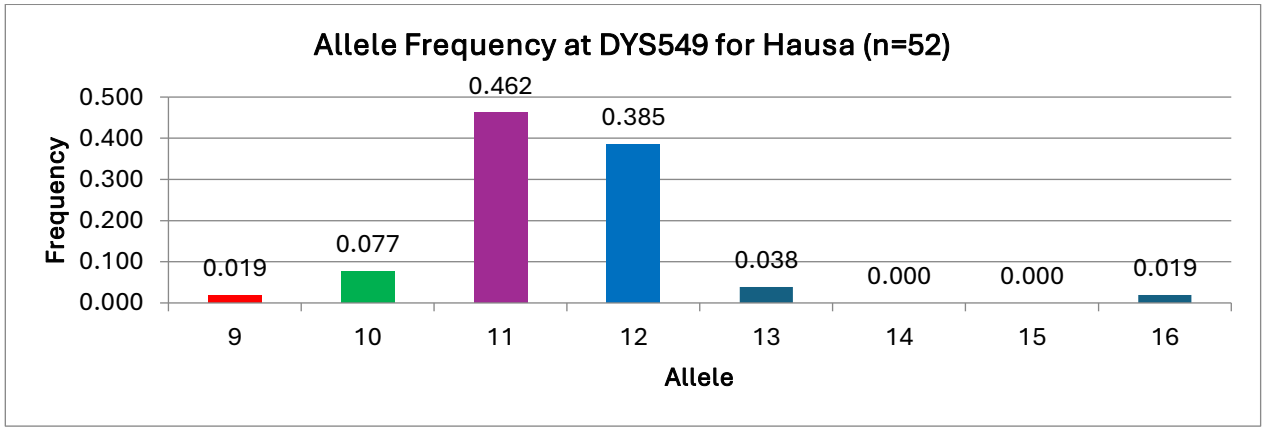
## APPENDIX 62

## ALLELE FREQUENCIES AT DYS549 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



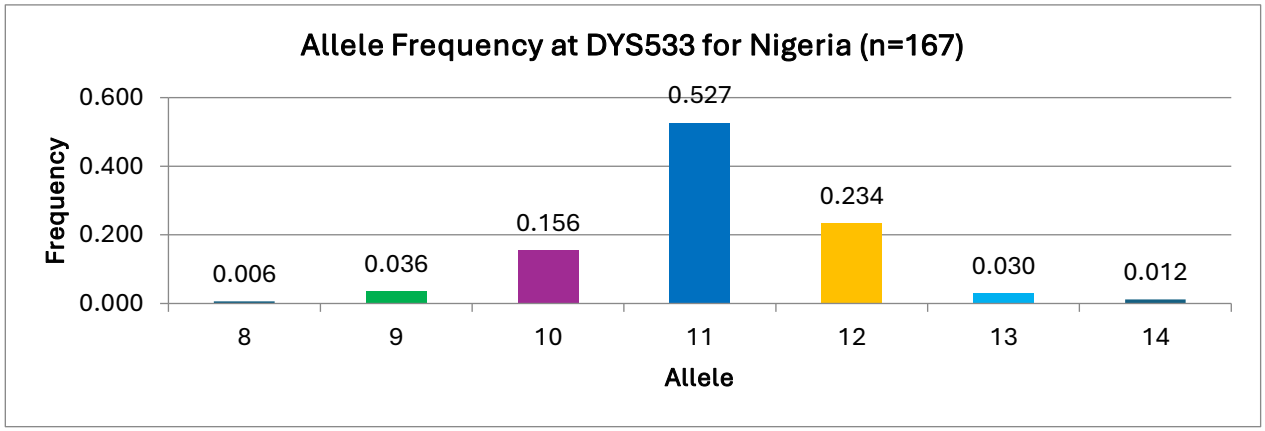



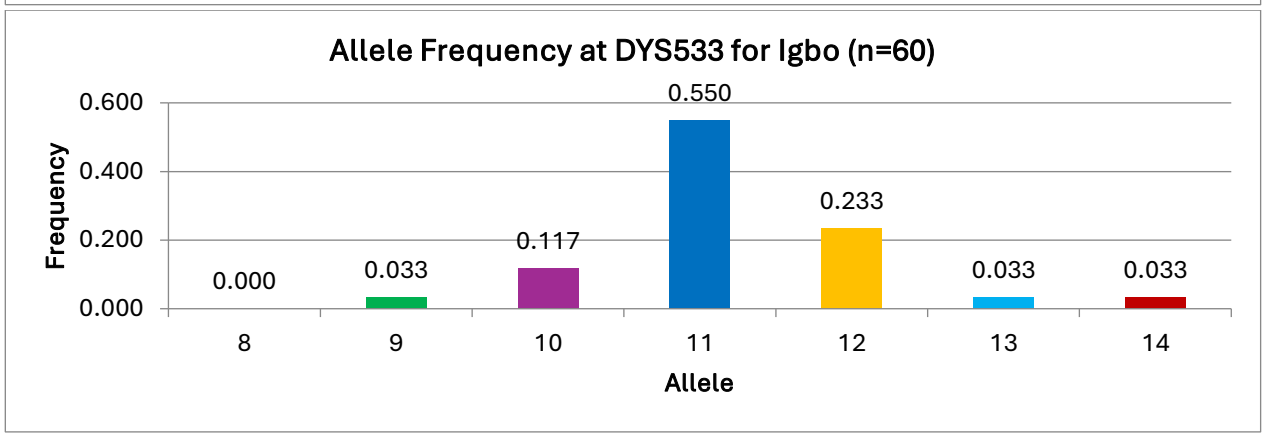



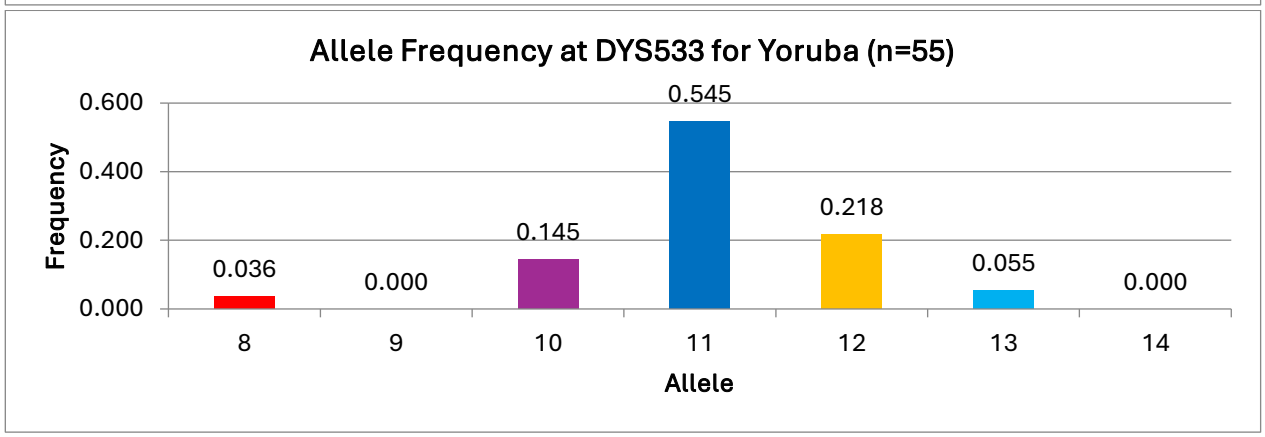



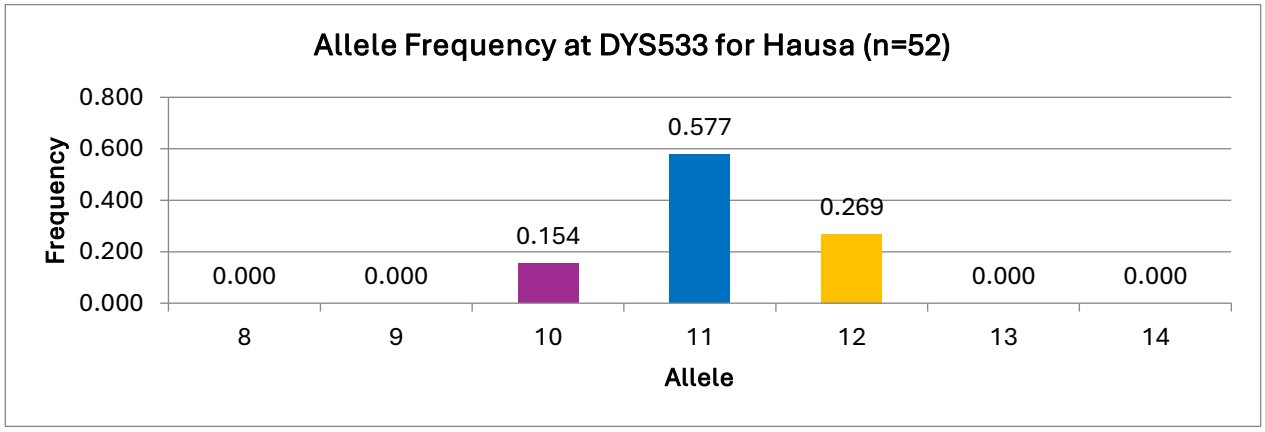
## APPENDIX 63

## ALLELE FREQUENCIES AT DYS533 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



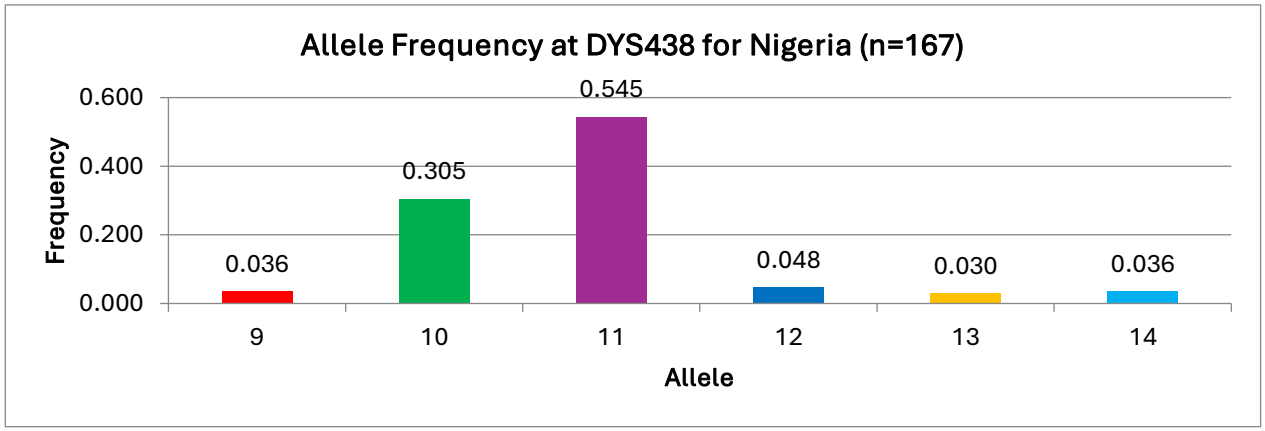



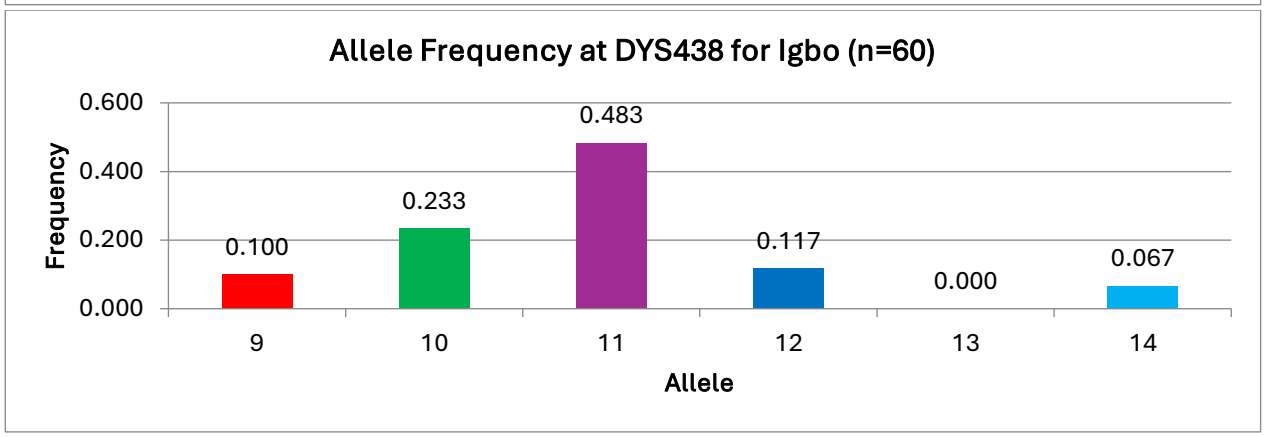



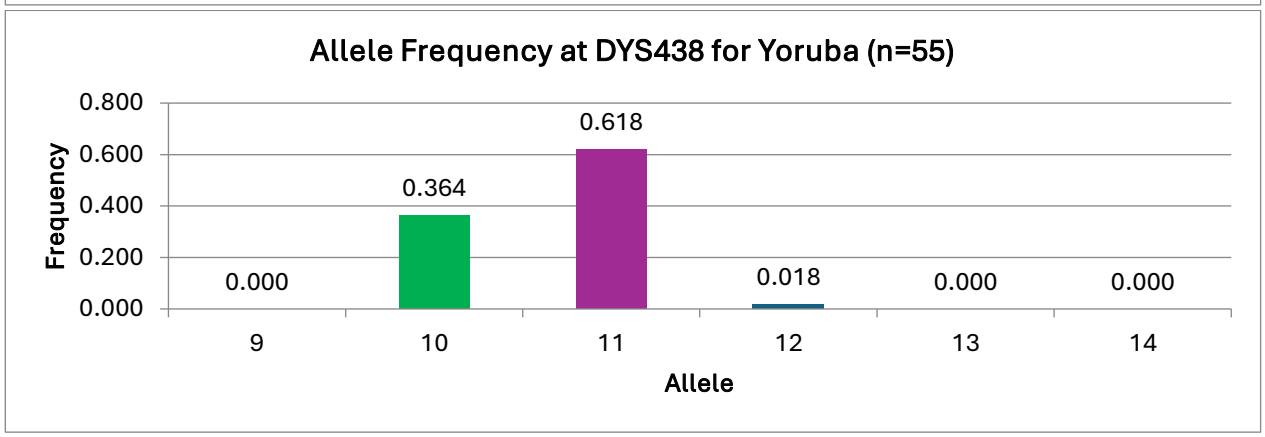



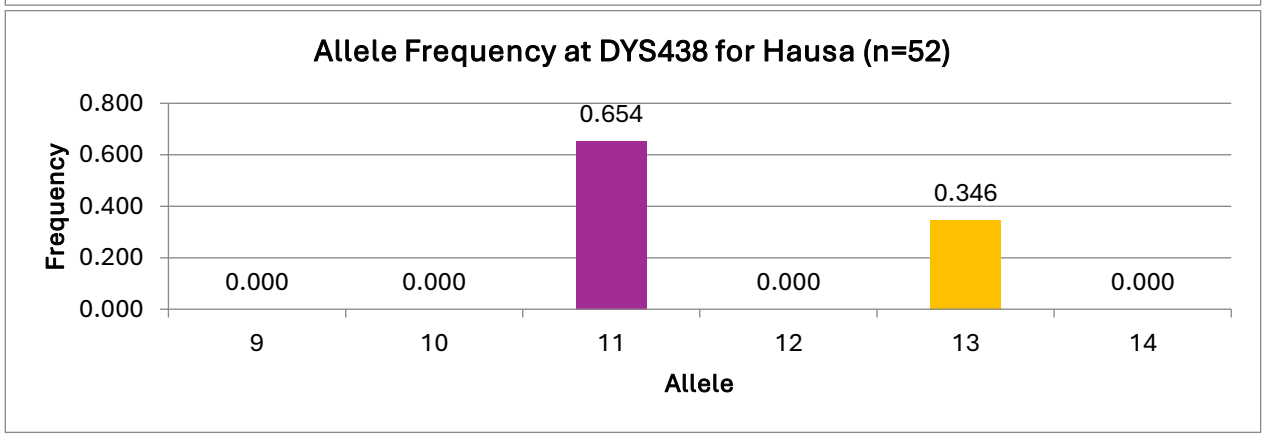
## APPENDIX 64

## ALLELE FREQUENCIES AT DYS438 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



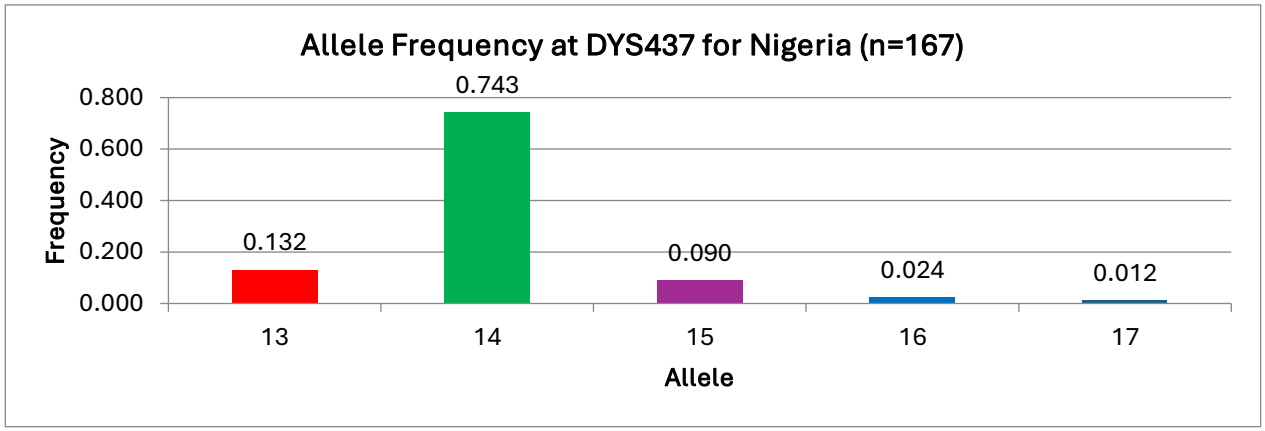



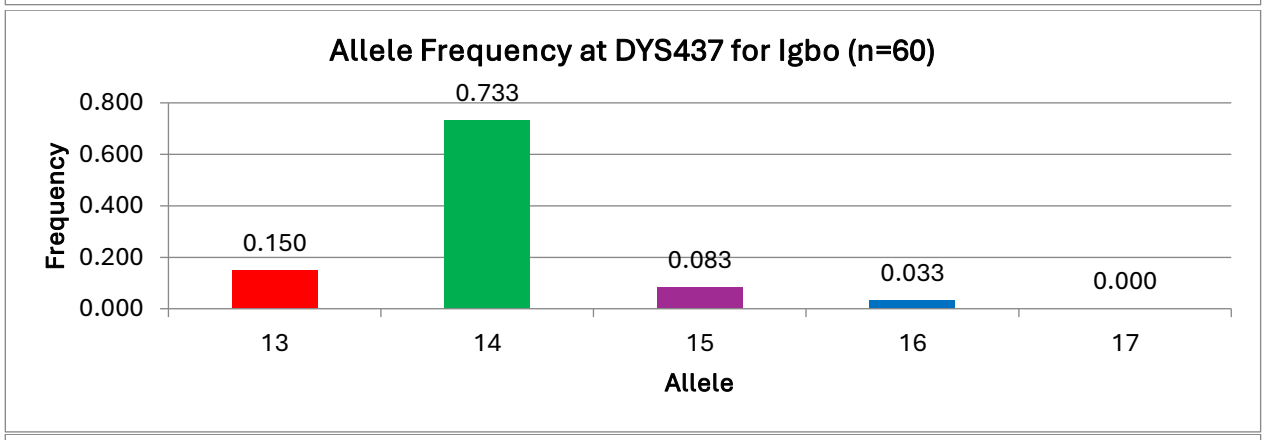



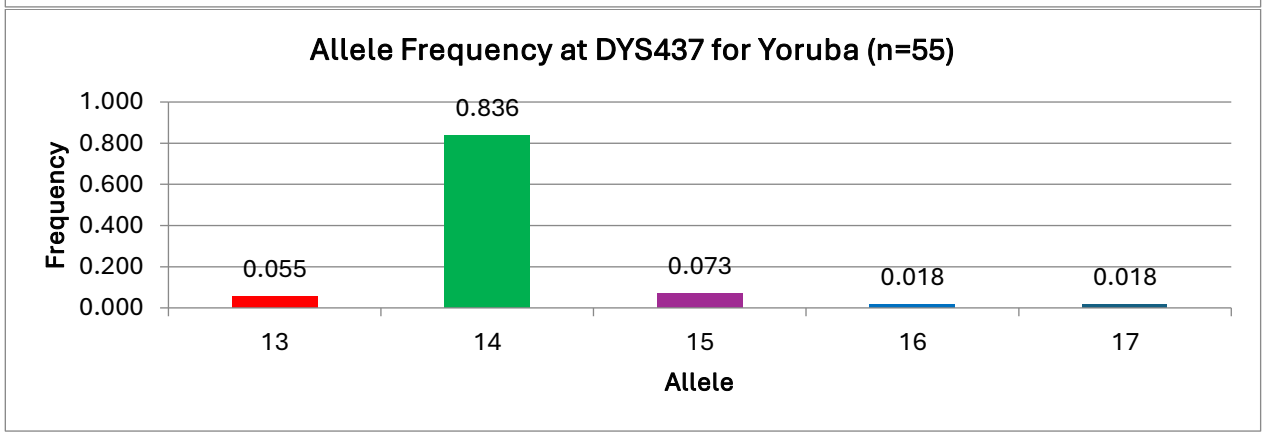



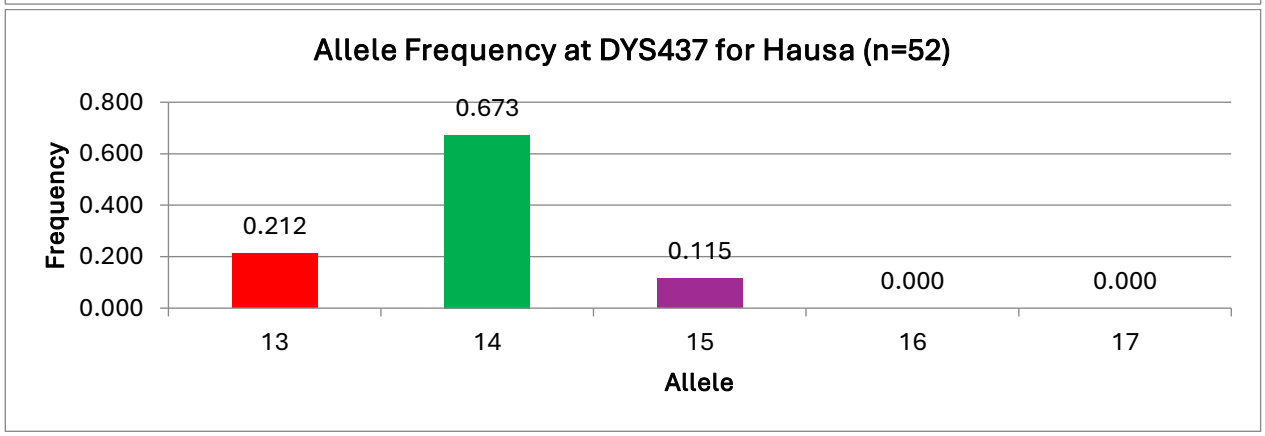
## APPENDIX 65

## ALLELE FREQUENCIES AT DYS437 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



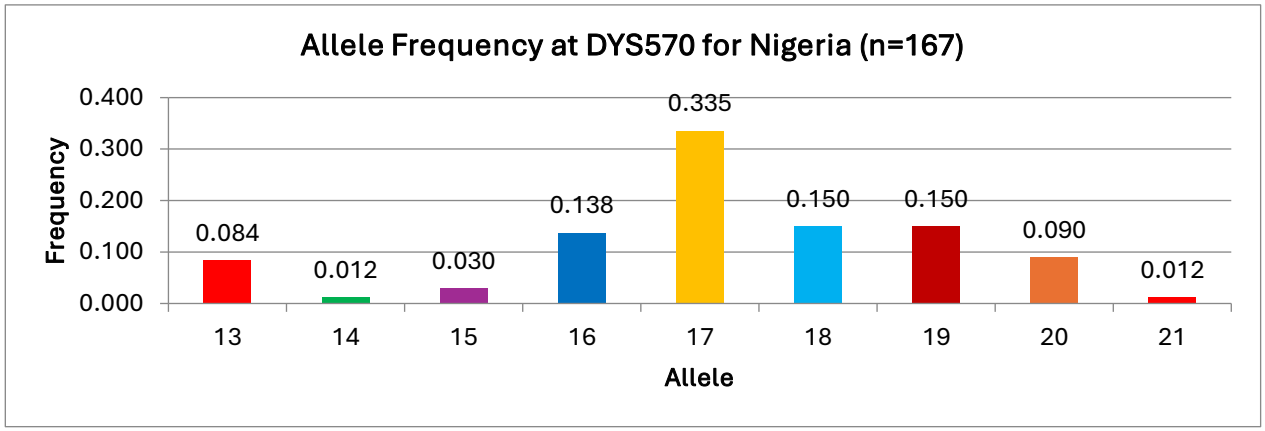



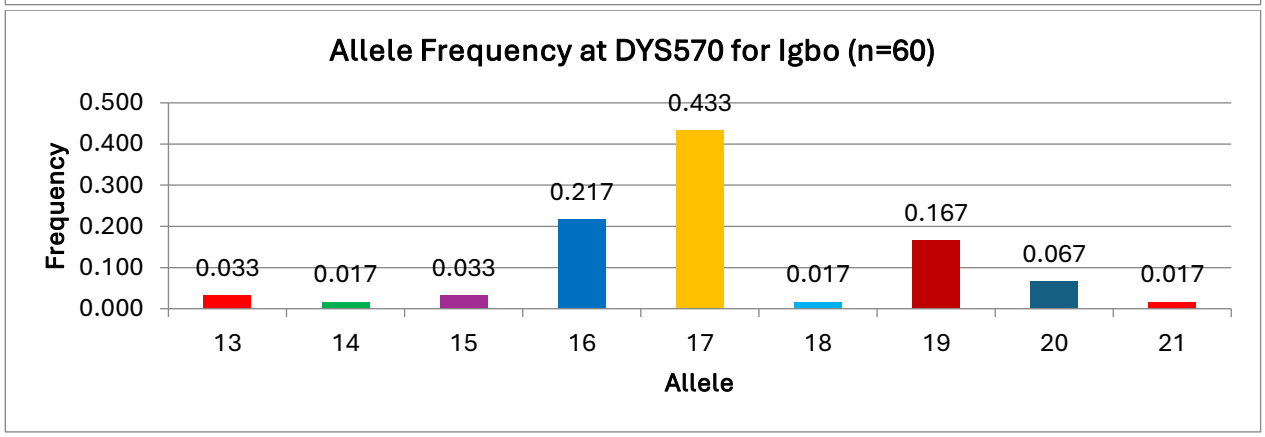



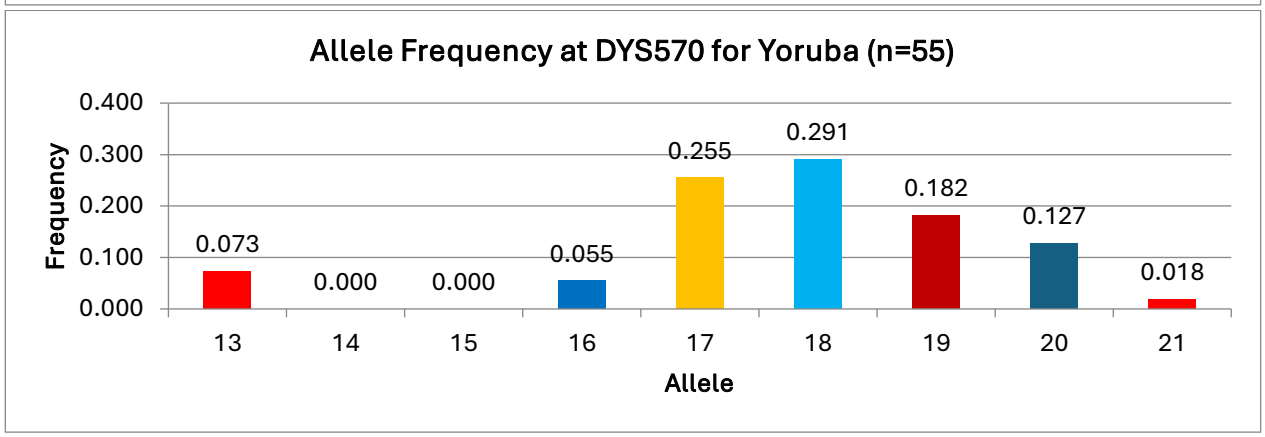



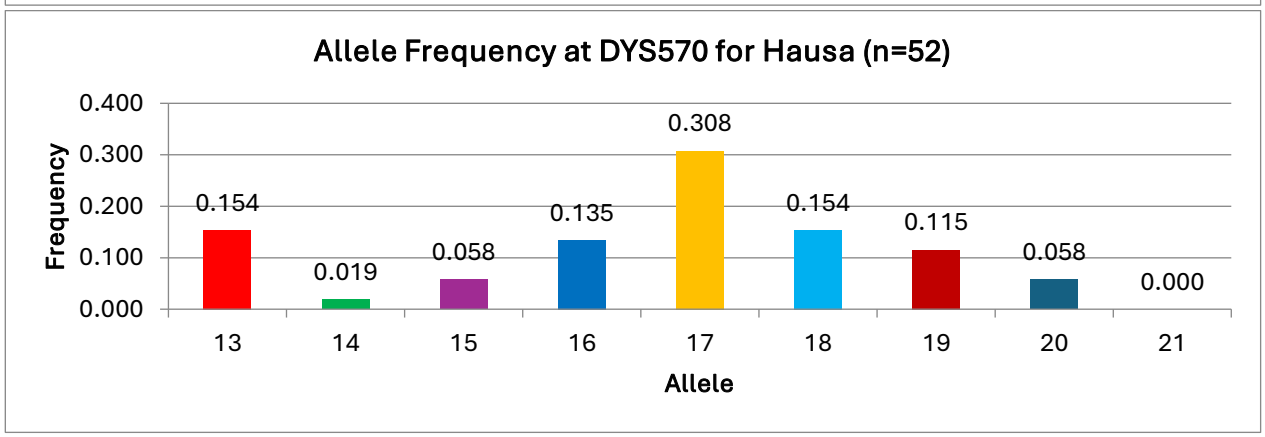
## APPENDIX 66

## ALLELE FREQUENCIES AT DYS570 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



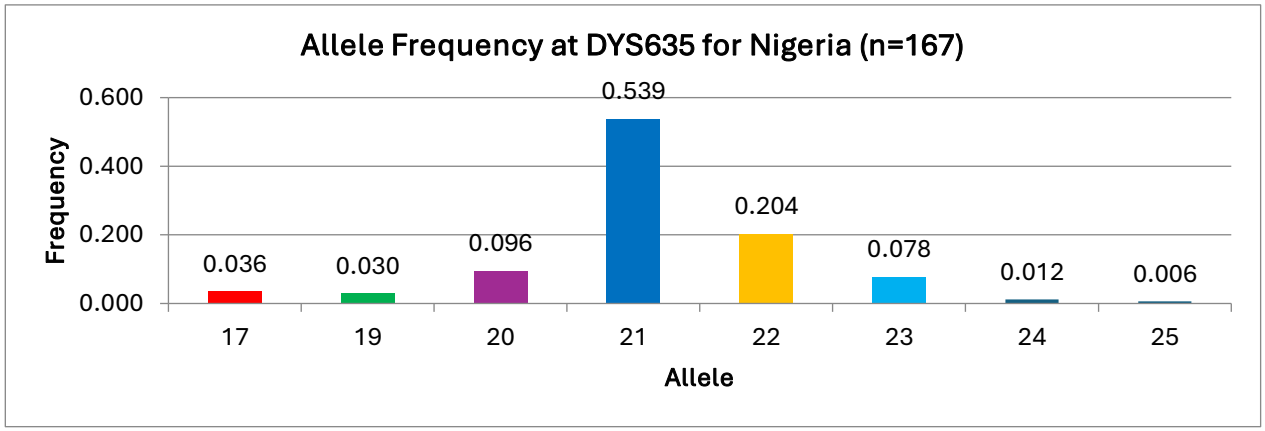



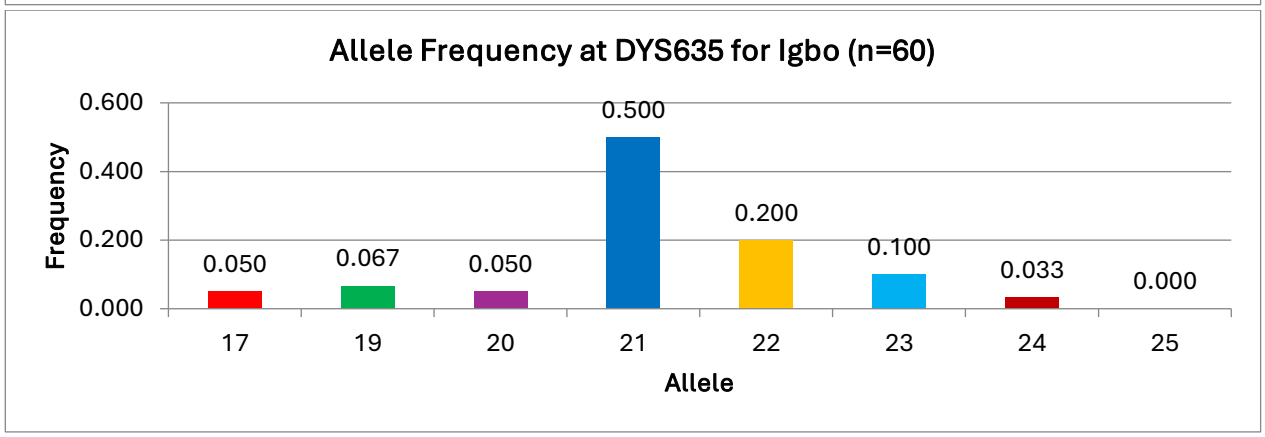



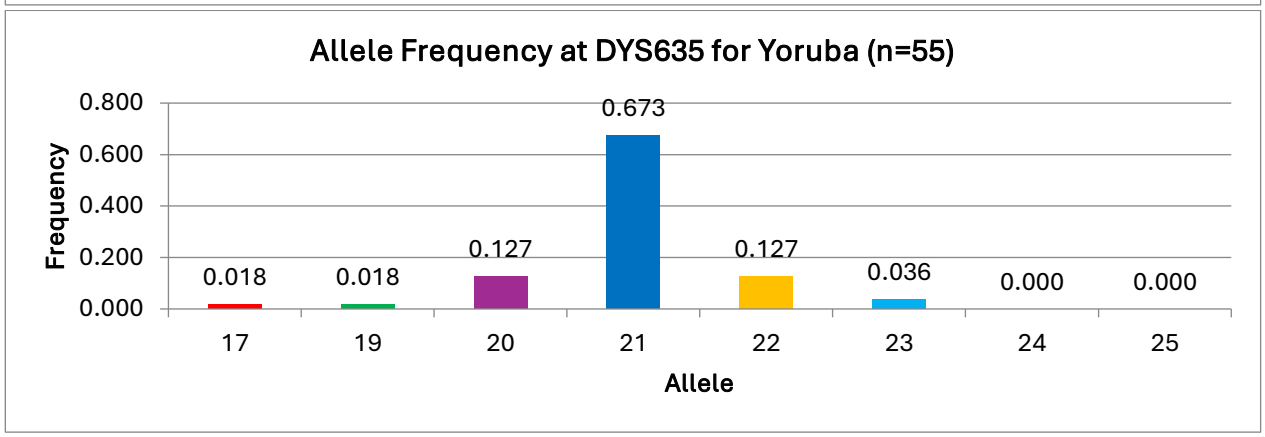



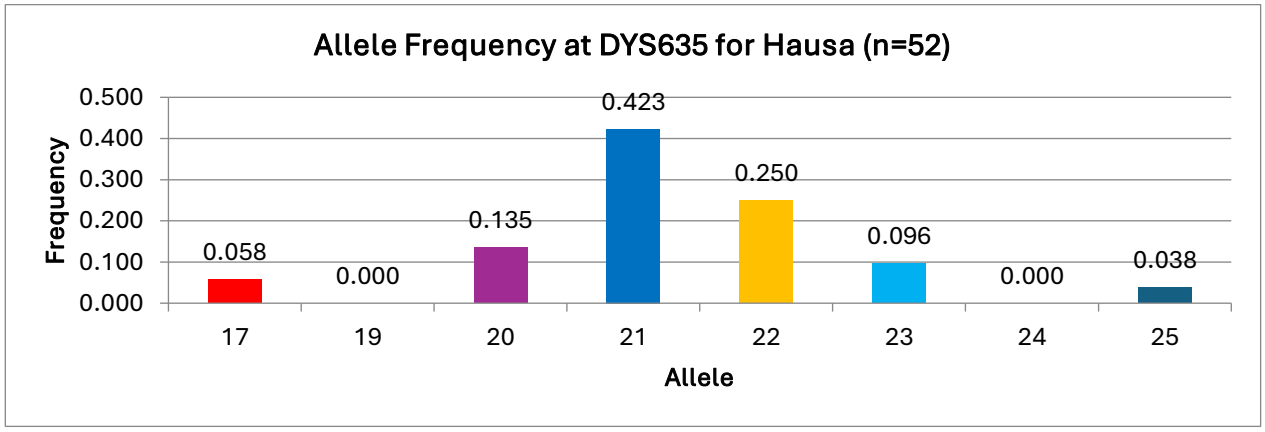
## APPENDIX 67

## ALLELE FREQUENCIES AT DYS635 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



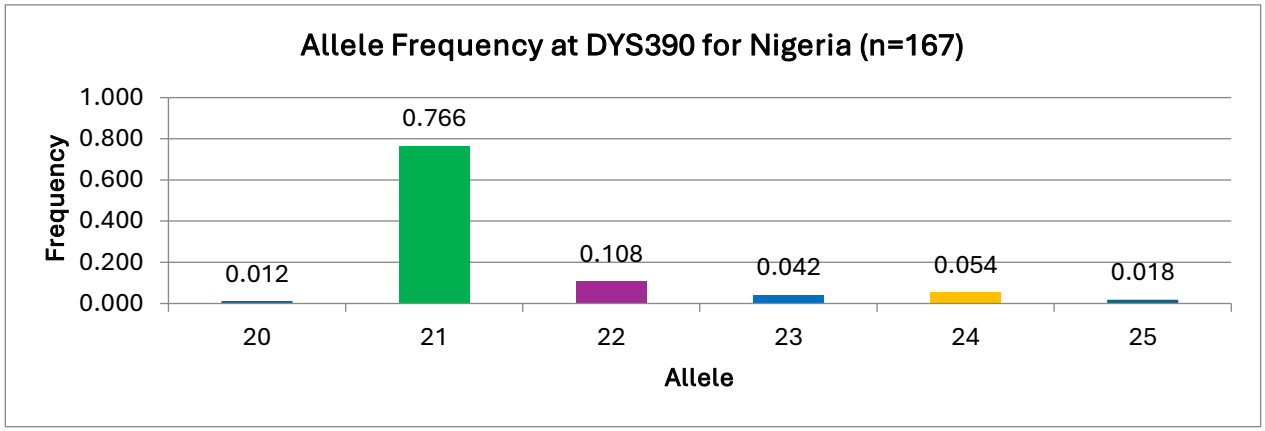



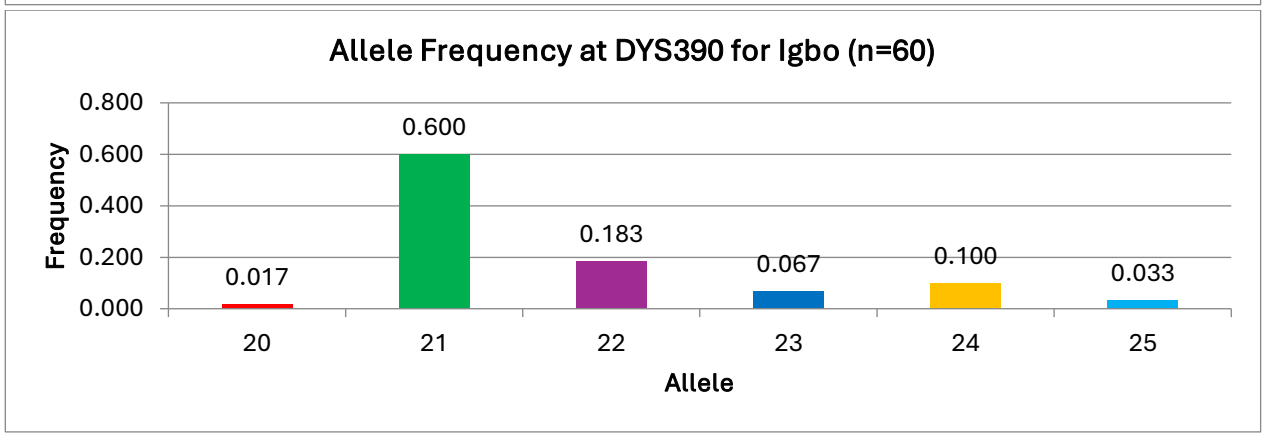



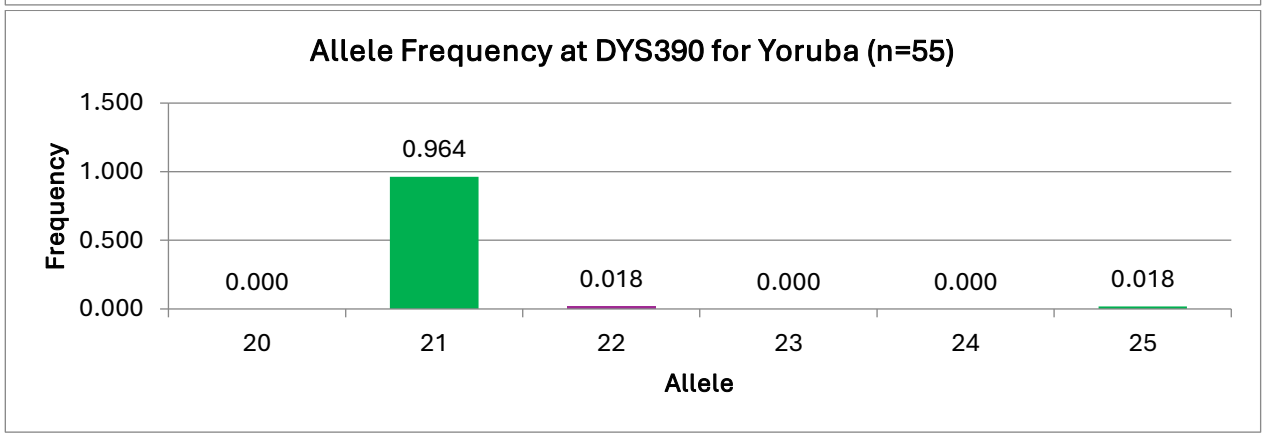



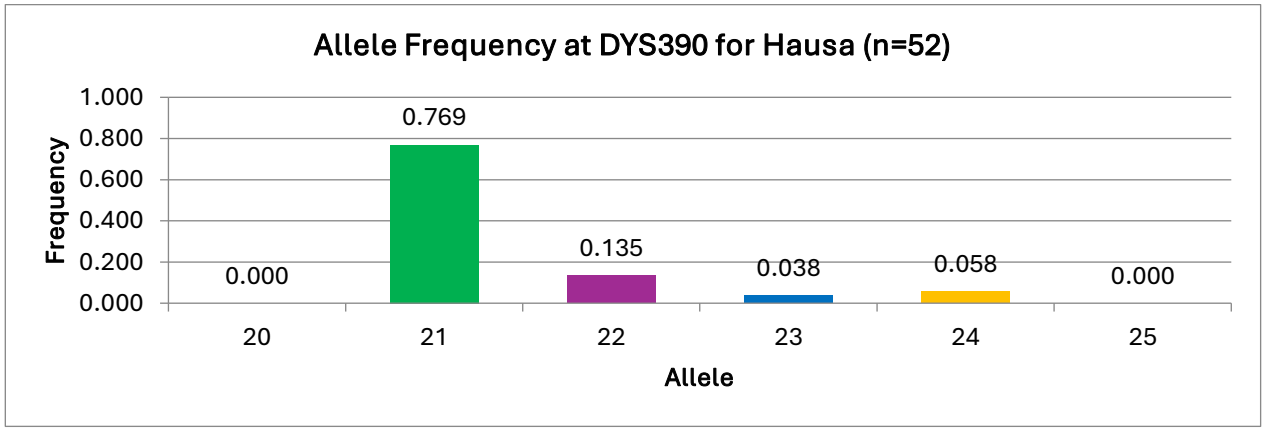
## APPENDIX 68

## ALLELE FREQUENCIES AT DYS390 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



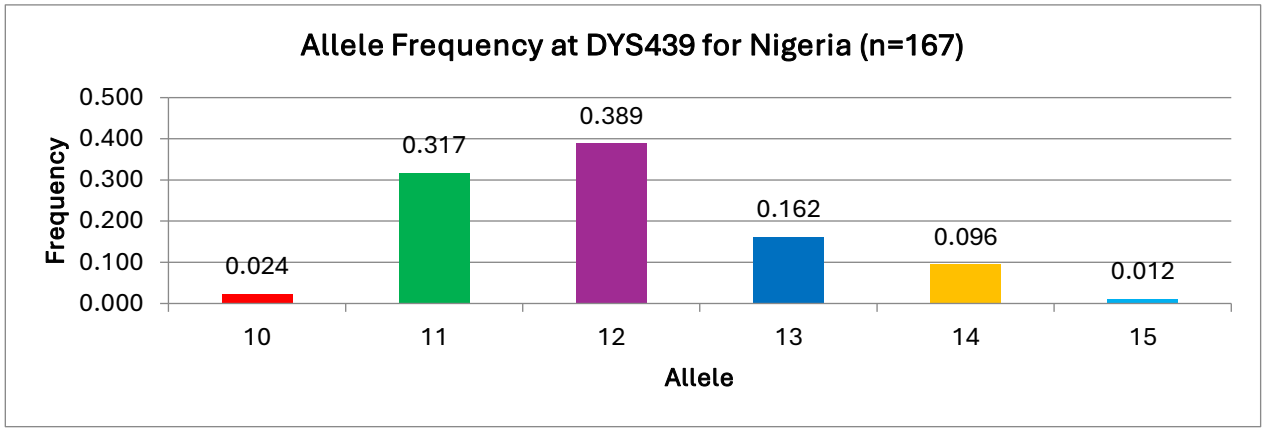



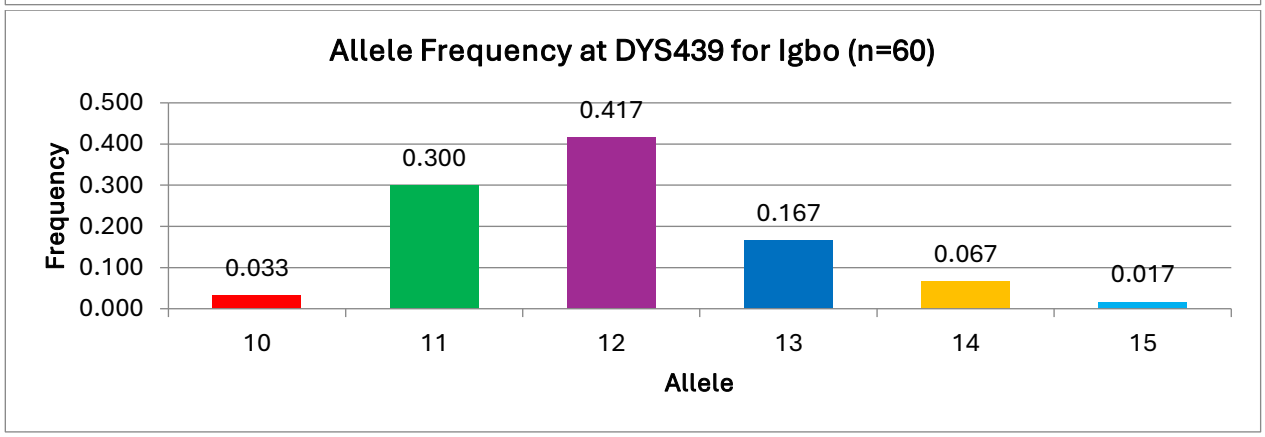



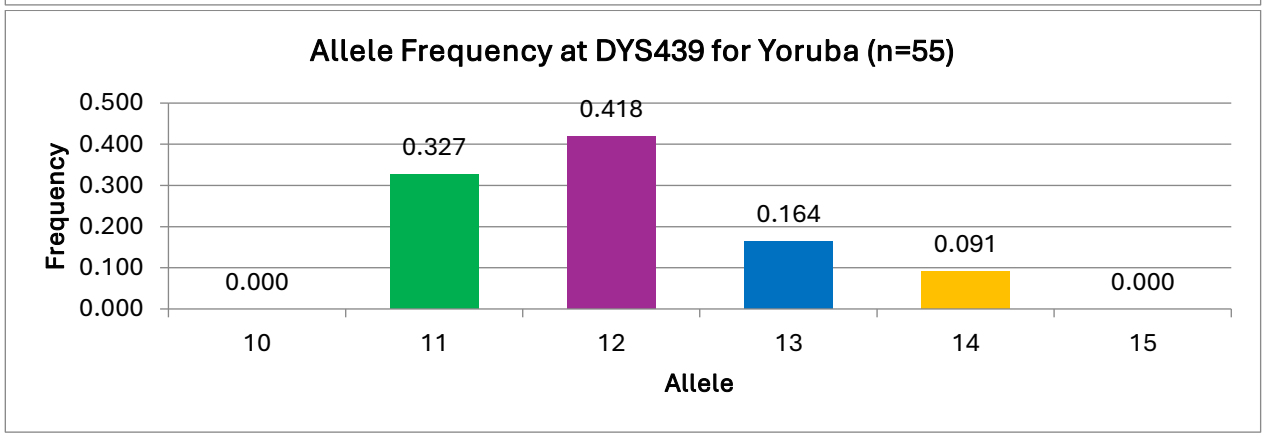



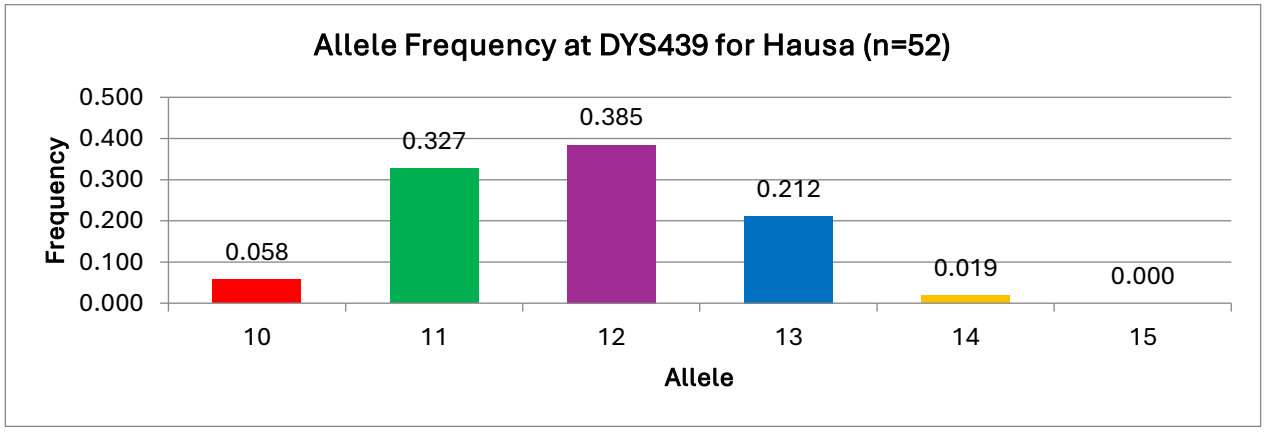
## APPENDIX 69

## ALLELE FREQUENCIES AT DYS439 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



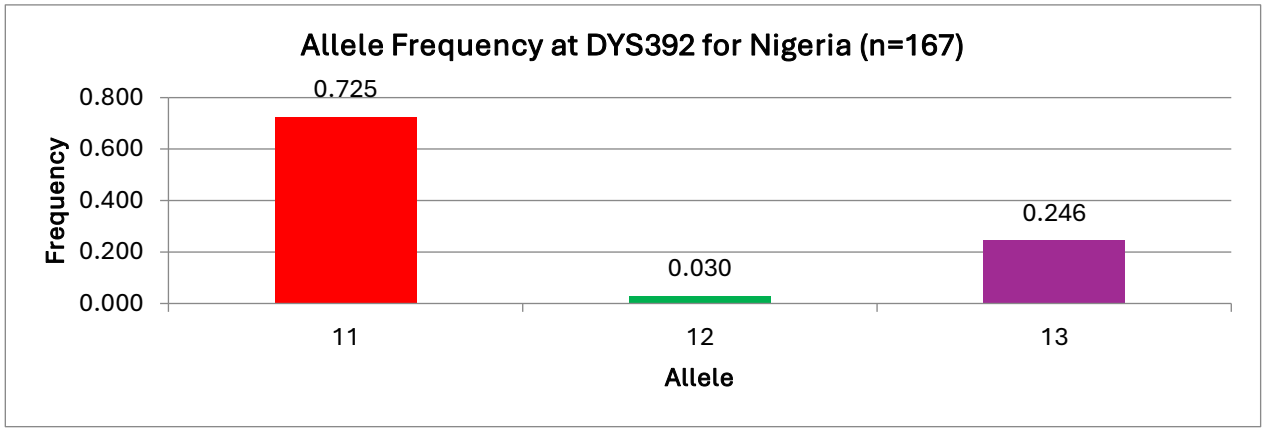



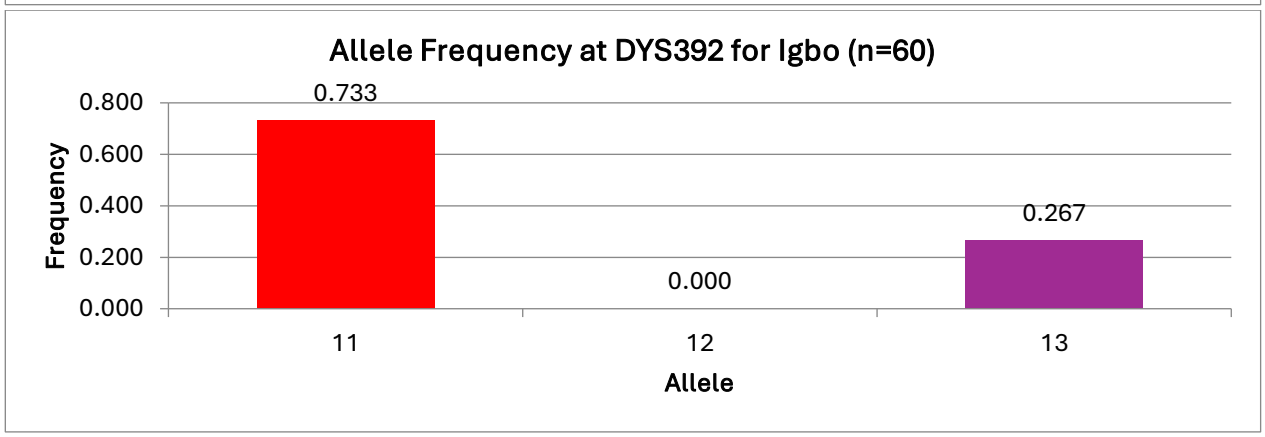



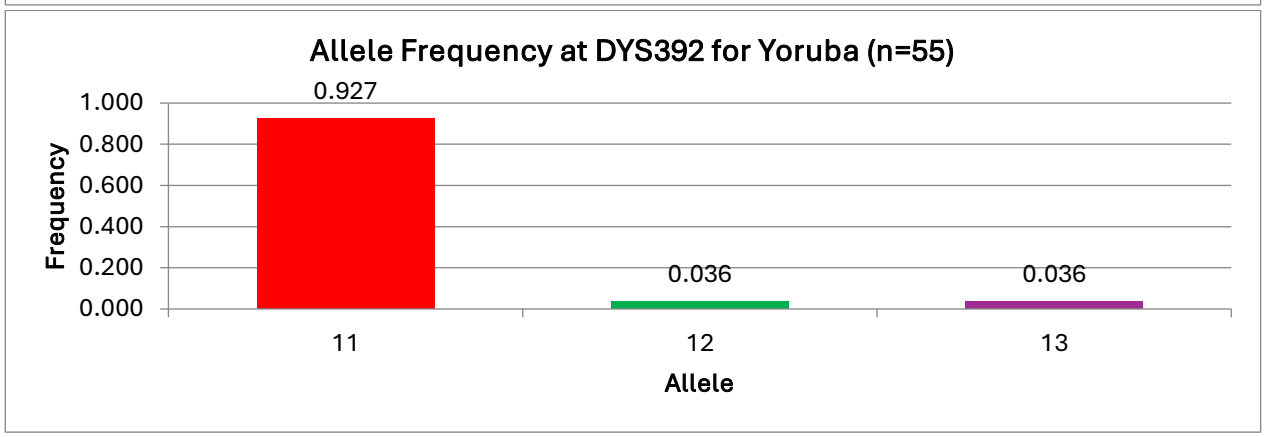



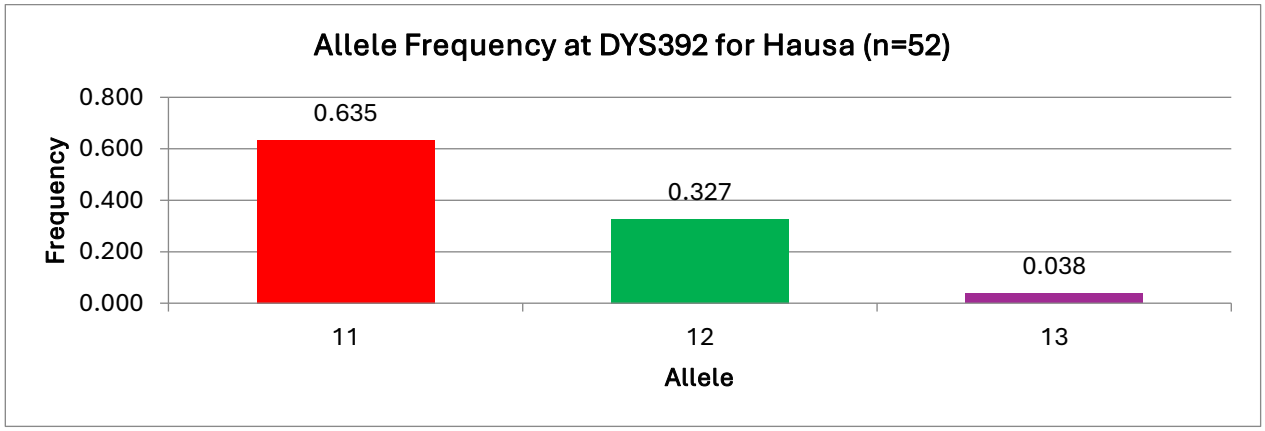
## APPENDIX 70

## ALLELE FREQUENCIES AT DYS392 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



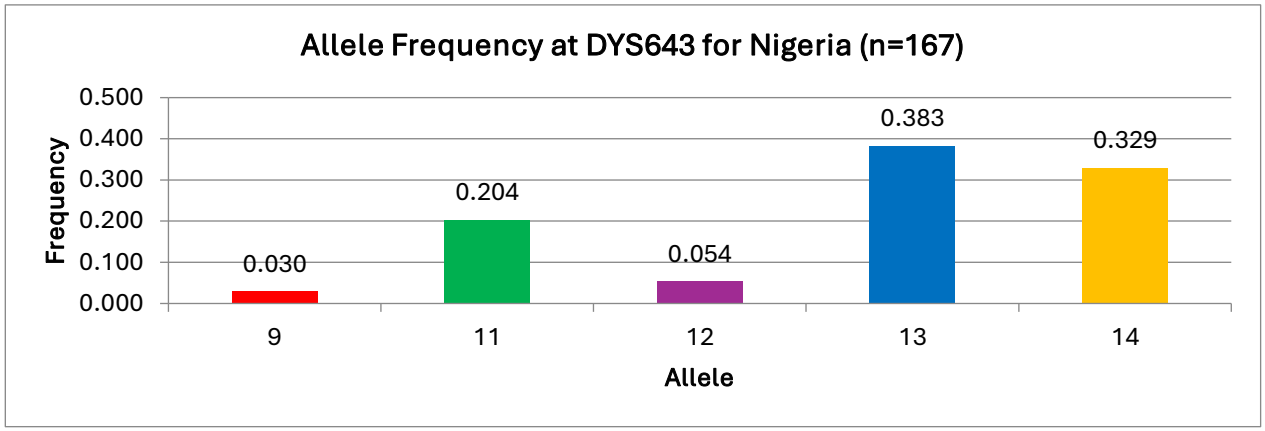



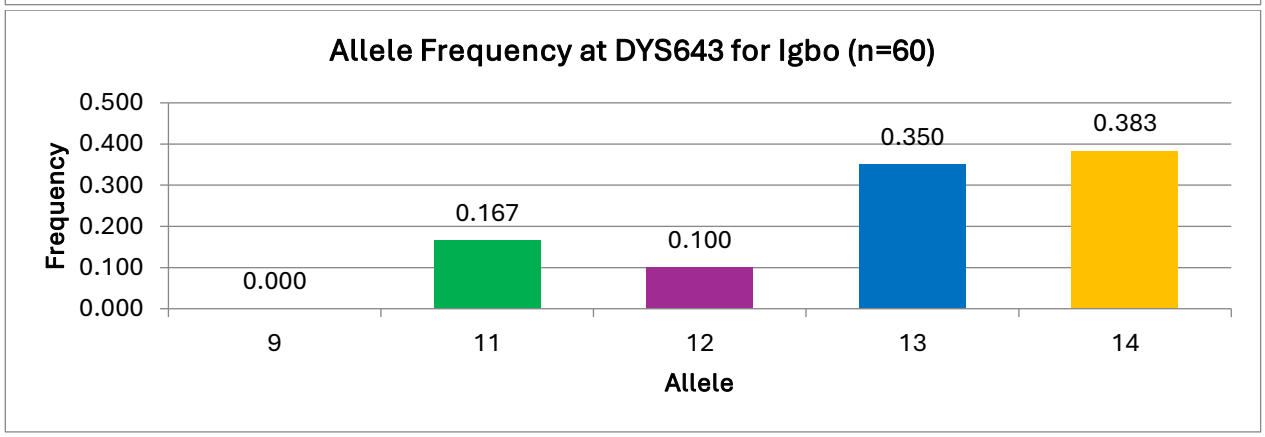



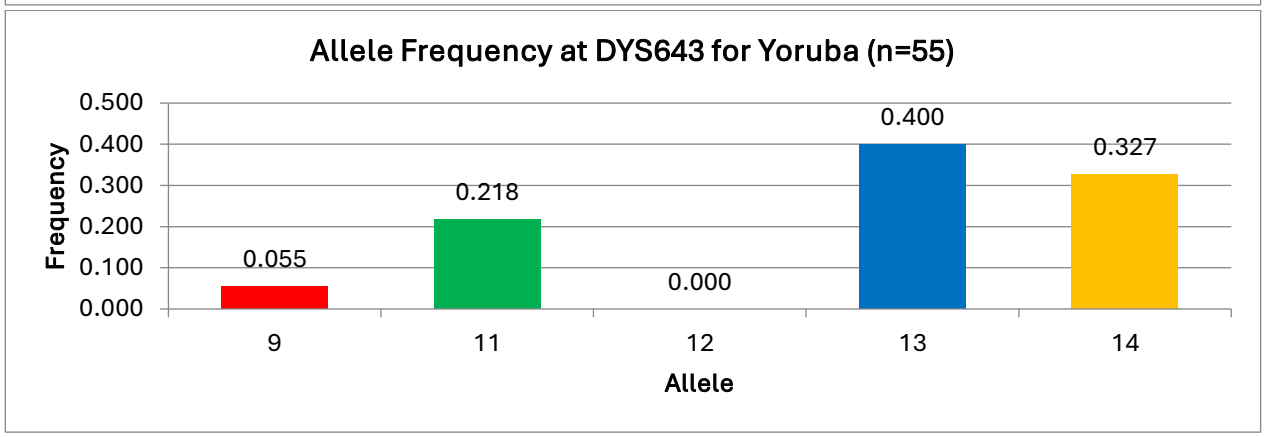



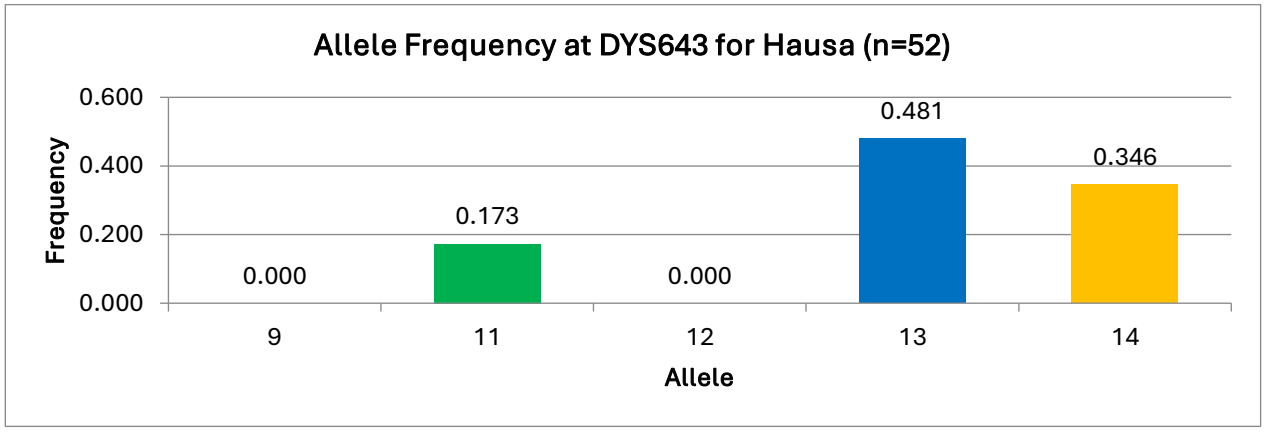
## APPENDIX 71

## ALLELE FREQUENCIES AT DYS643 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



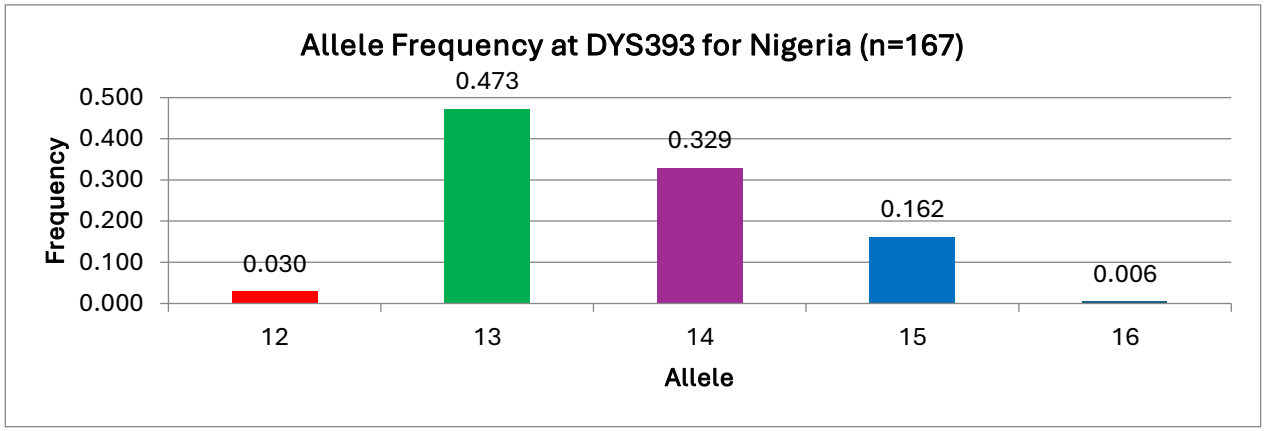



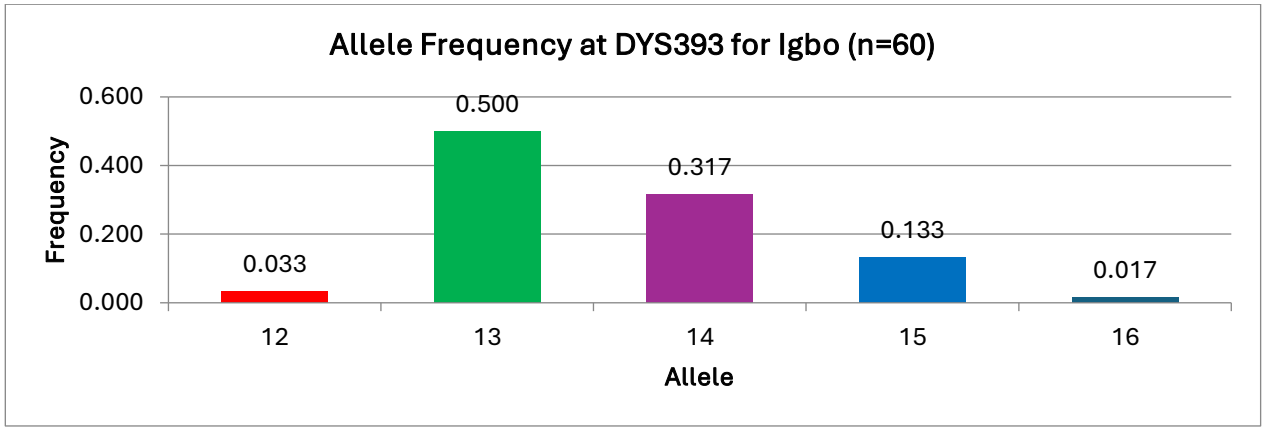



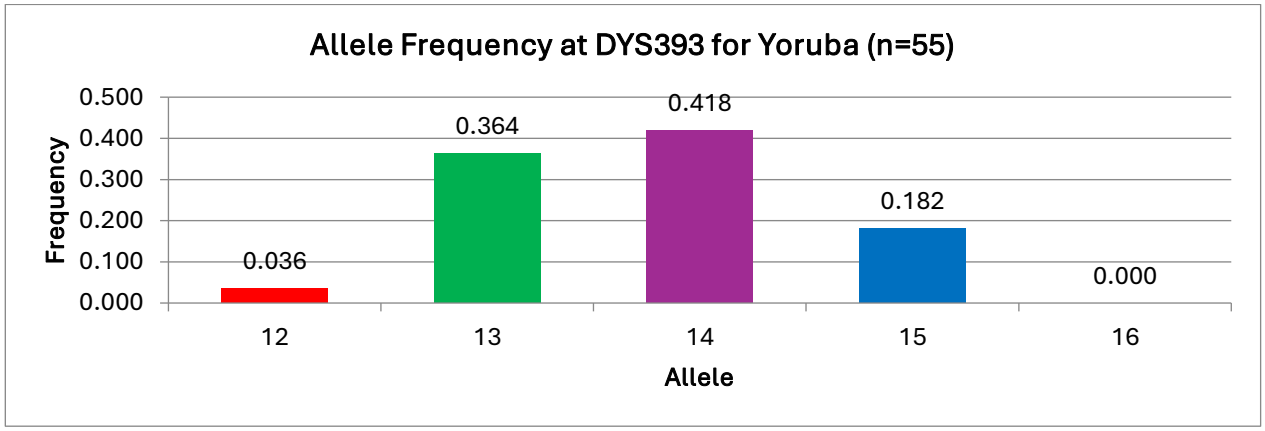



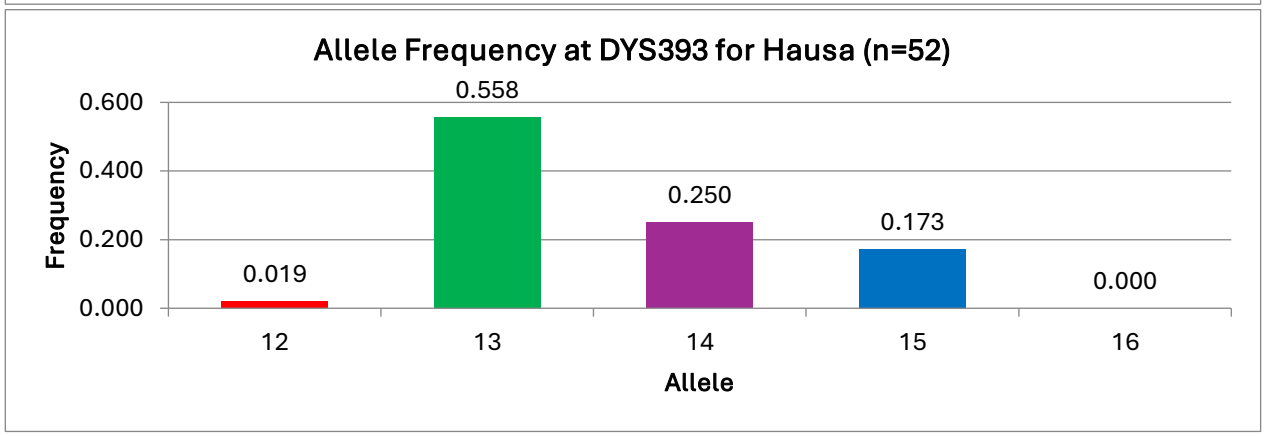
## APPENDIX 72

## ALLELE FREQUENCIES AT DYS393 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



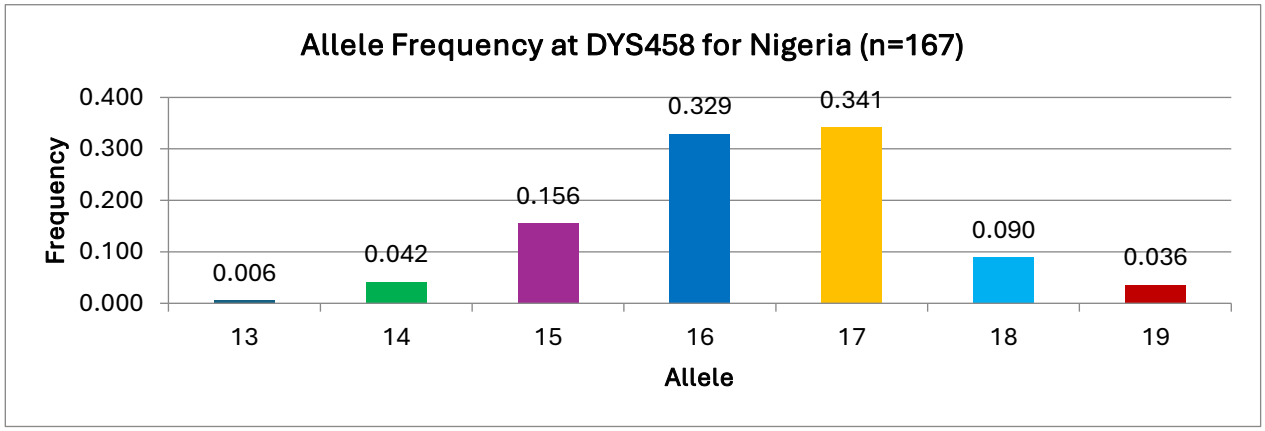



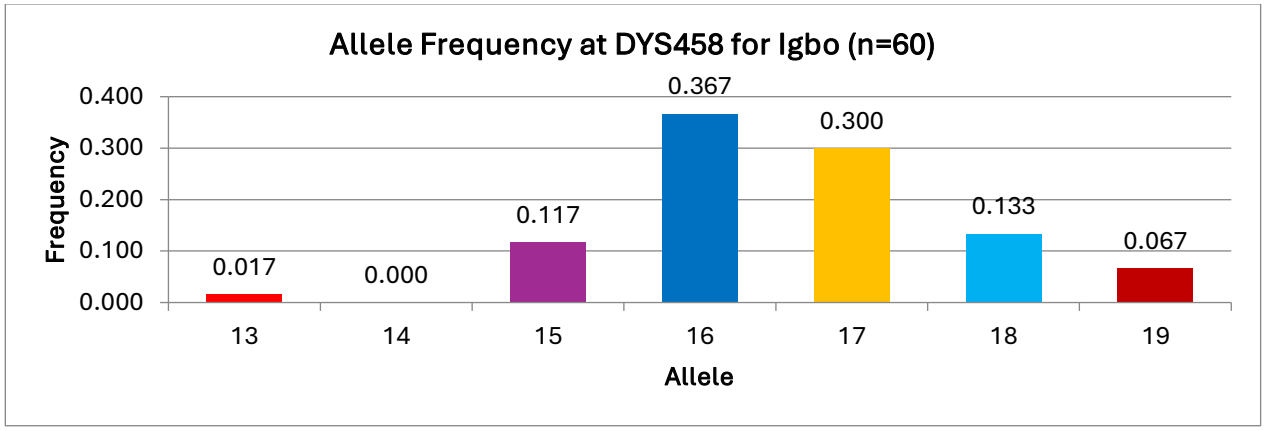



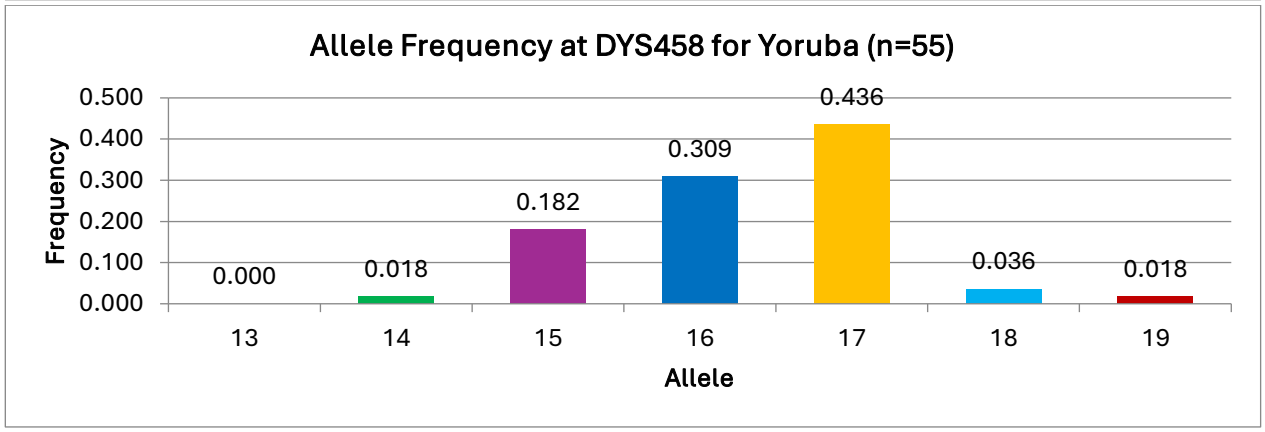



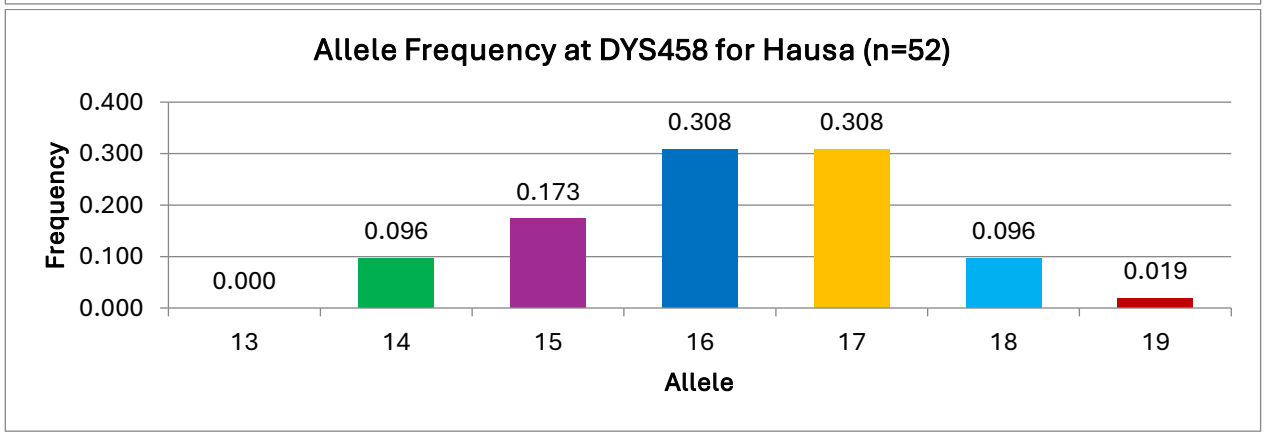
## APPENDIX 73

## ALLELE FREQUENCIES AT DYS458 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



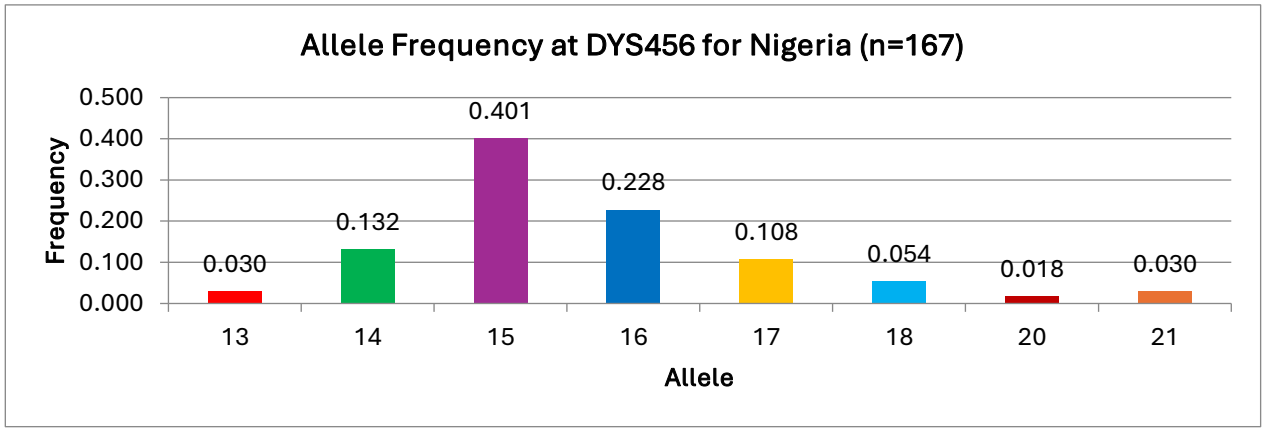



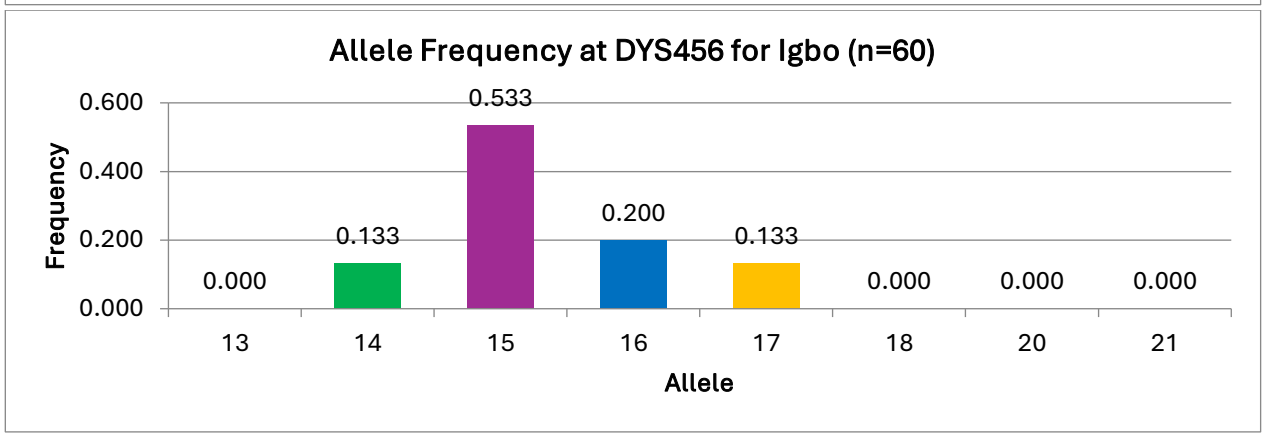



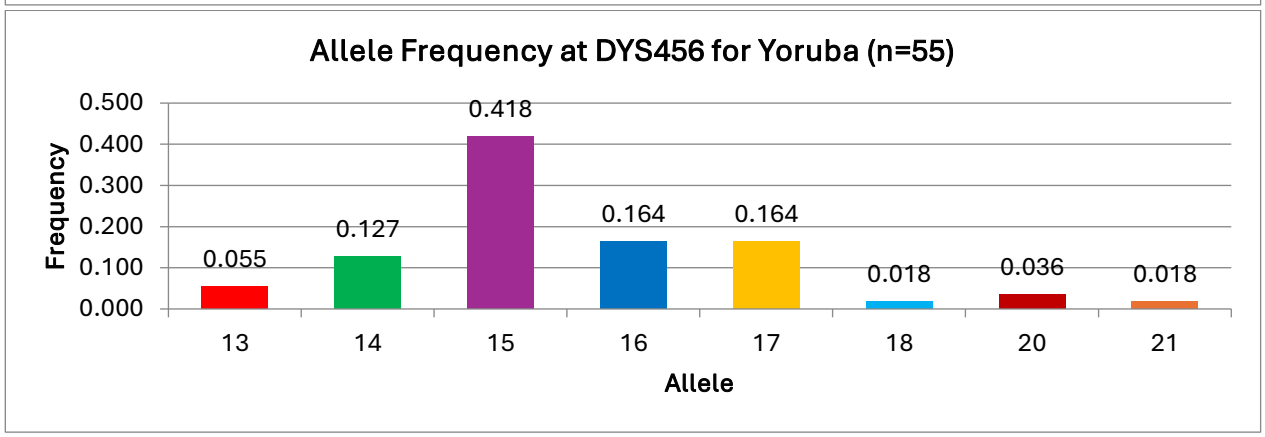



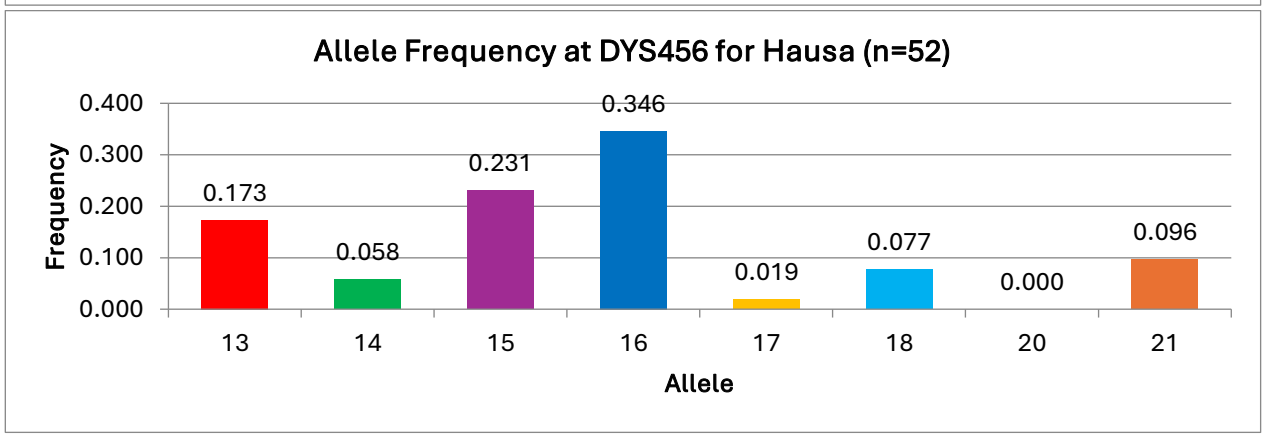
## APPENDIX 74

## ALLELE FREQUENCIES AT DYS456 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



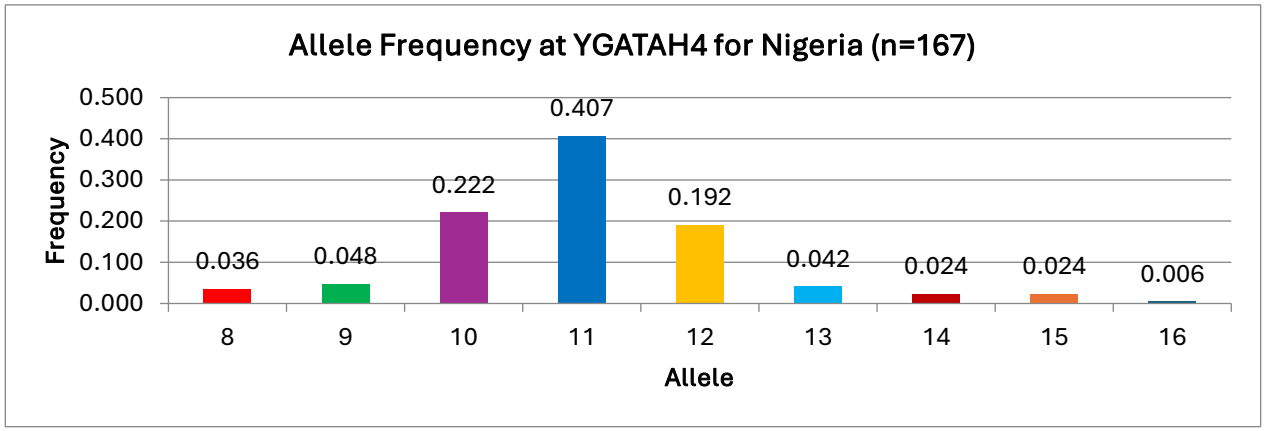



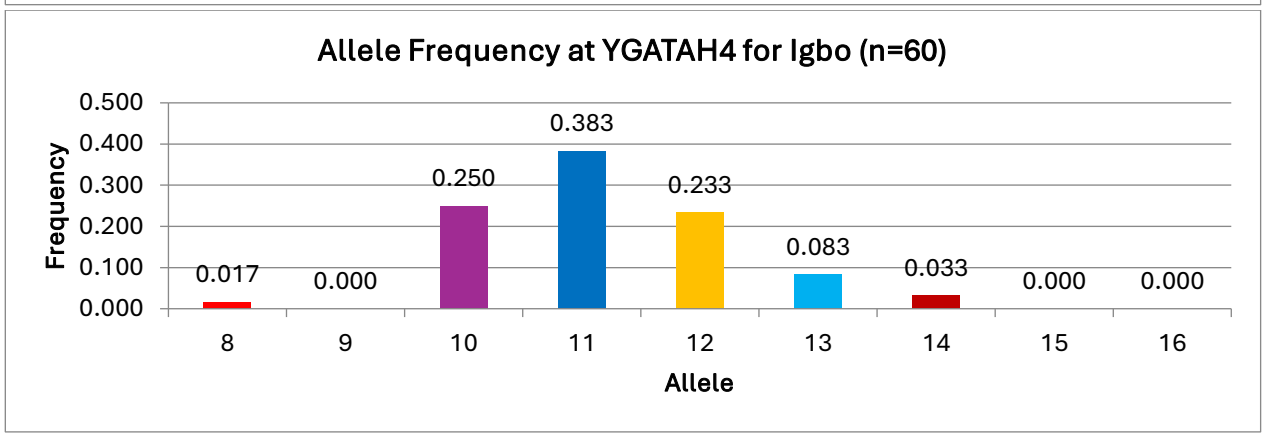



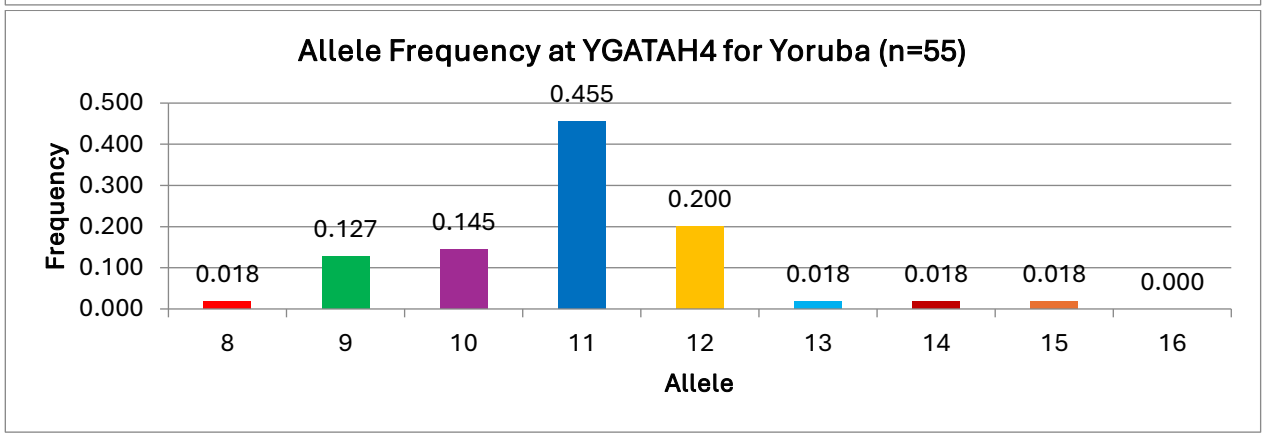



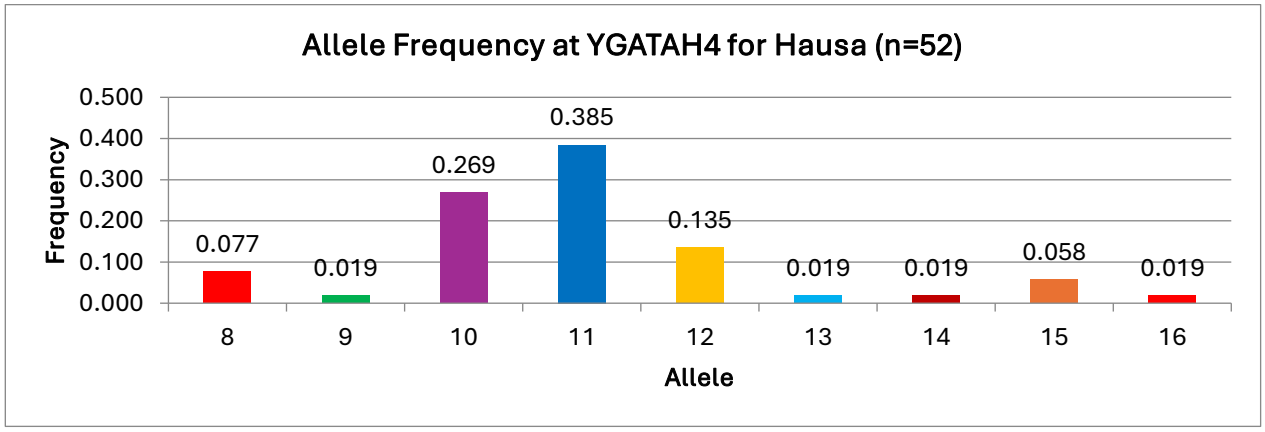
## APPENDIX 75

## ALLELE FREQUENCIES AT YGATAH4 WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



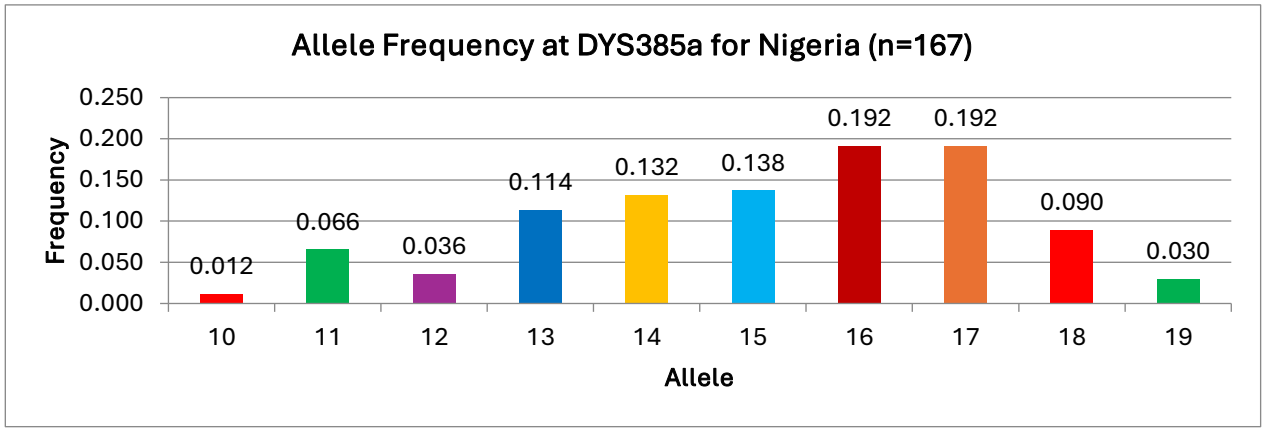



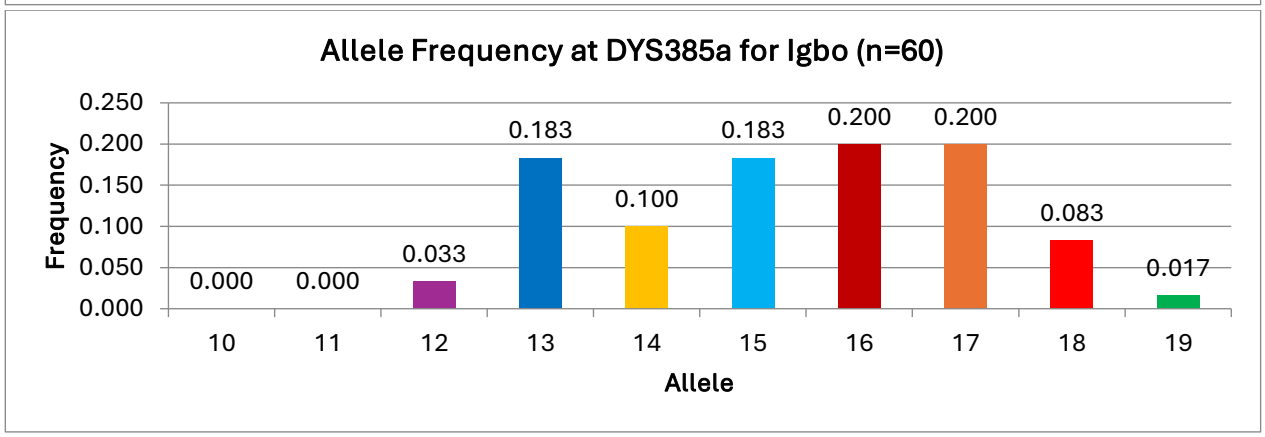



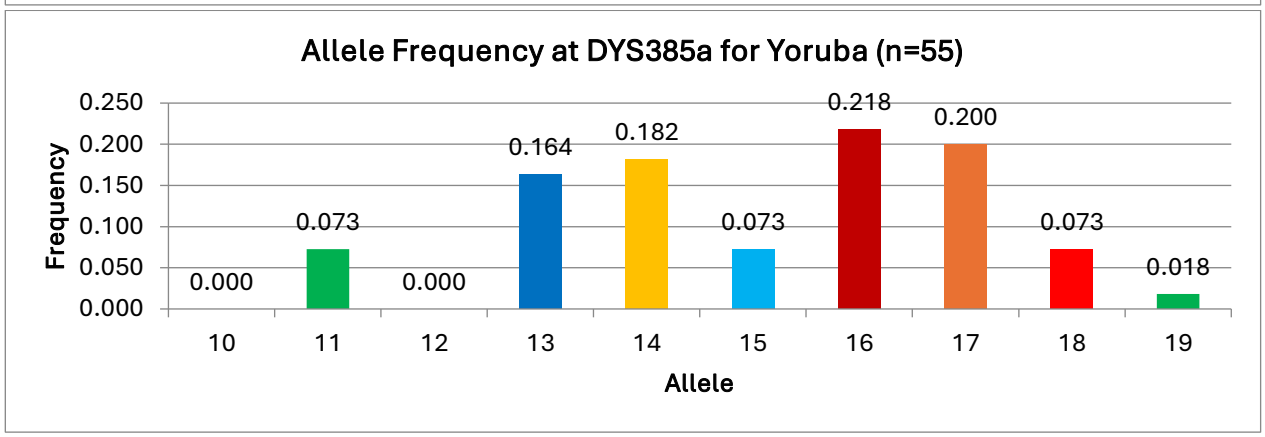



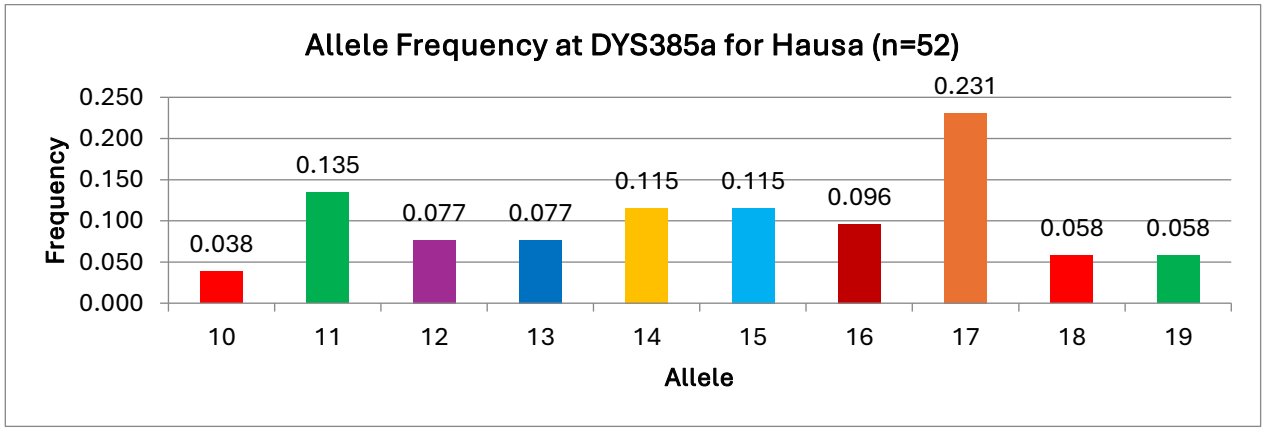
## APPENDIX 76

## ALLELE FREQUENCIES AT DYS385a WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



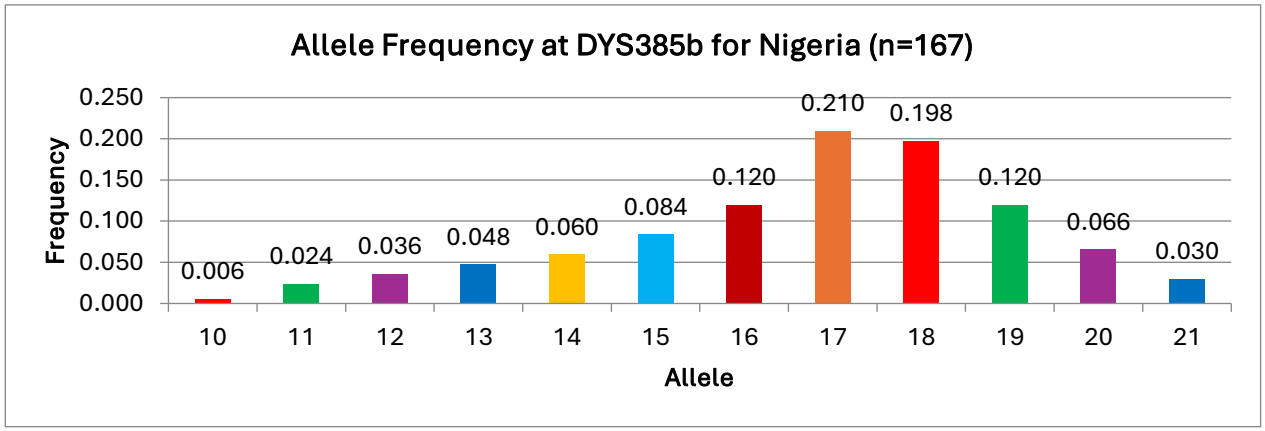



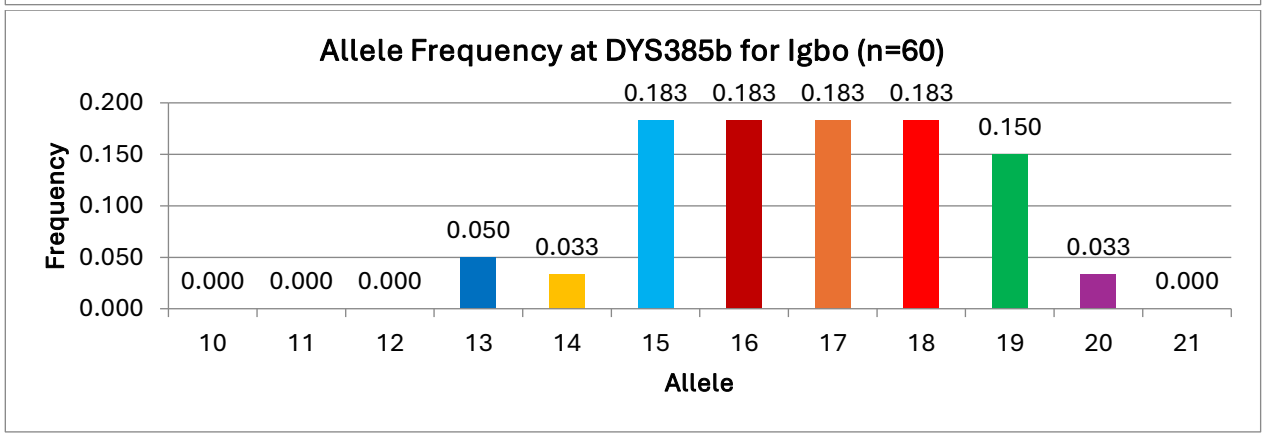



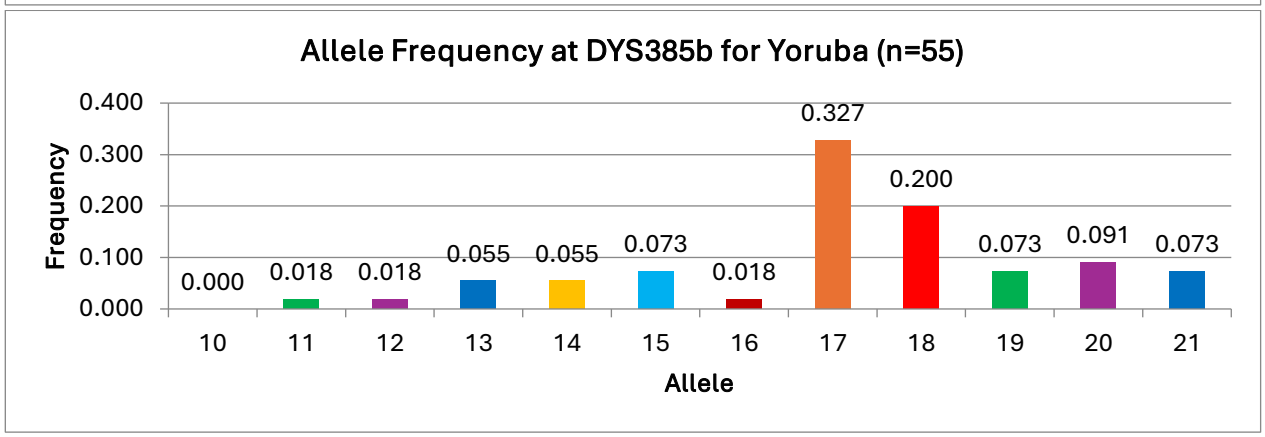



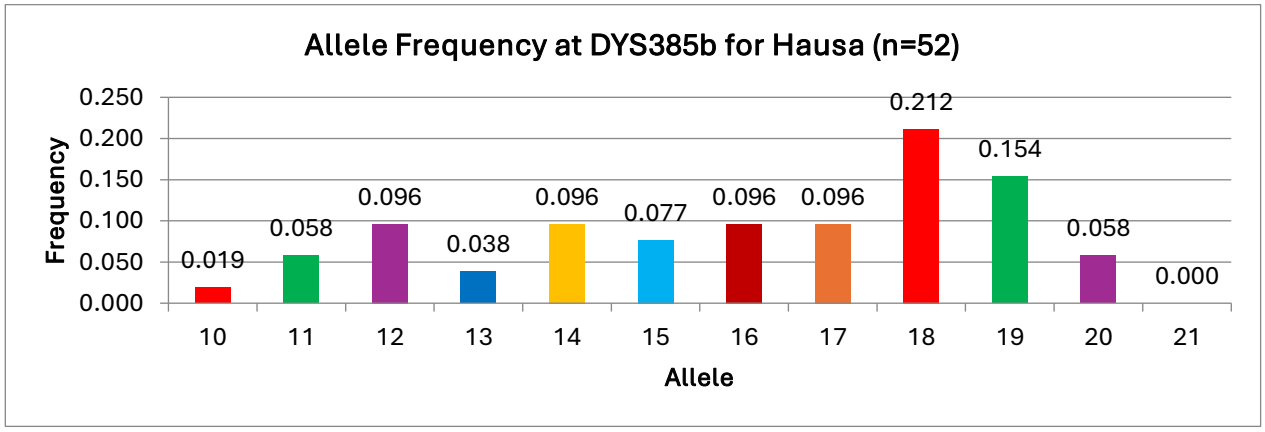
## APPENDIX 77

## ALLELE FREQUENCIES AT DYS385b WITH GRAPHS BY POPULATION FOR POWERPLEX® Y23 SYSTEM KIT



## APPENDIX 78

## NEI'S GENETIC DISTANCE PAIRWISE POPULATION MATRIX FOR THE POWERPLEX® Y23 SYSTEM KIT

| Sub-Population | Igbo | Yoruba | Hausa-Fulani | Algeria |
|---|---|---|---|---|
| **Igbo** | 0.000 | | | |
| **Yoruba** | 0.051 | 0.000 | | |
| **Hausa-Fulani** | 0.100 | 0.154 | 0.000 | |
| **Algeria** | 0.332 | 0.306 | 0.349 | 0.000 |

***Algeria is included as an outgroup.**

## APPENDIX 79

## NEI'S GENETIC IDENTITY PAIRWISE POPULATION MATRIX FOR THE POWERPLEX® Y23 SYSTEM KIT

| Sub-Population | Igbo | Yoruba | Hausa-Fulani | Algeria |
|---|---|---|---|---|
| **Igbo** | 1.000 | | | |
| **Yoruba** | 0.949 | 1.000 | | |
| **Hausa-Fulani** | 0.900 | 0.846 | 1.000 | |
| **Algeria** | 0.668 | 0.694 | 0.651 | 1.000 |

***Algeria is included as an outgroup.**

## APPENDIX 80

## ALLELIC DIVERSITY PER LOCUS AND POPULATION BASED ON THE POWERPLEX® Y23 SYSTEM KIT

| Locus\Population | Igbo | Yoruba | Hausa | Overall |
|---|---|---|---|---|
| **DYS576** | 7 | 7 | 8 | 9 |
| **DYS389I** | 3 | 5 | 5 | 6 |
| **DYS448** | 4 | 6 | 5 | 6 |
| **DYS389II** | 4 | 5 | 4 | 5 |
| **DYS19** | 4 | 3 | 3 | 4 |
| **DYS391** | 3 | 2 | 4 | 5 |
| **DYS481** | 9 | 7 | 7 | 9 |
| **DYS549** | 6 | 3 | 4 | 7 |
| **DYS533** | 8 | 5 | 3 | 9 |
| **DYS438** | 5 | 3 | 2 | 6 |
| **DYS437** | 4 | 5 | 3 | 5 |
| **DYS570** | 9 | 7 | 8 | 10 |
| **DYS635** | 8 | 6 | 6 | 9 |
| **DYS390** | 6 | 3 | 4 | 6 |
| **DYS439** | 6 | 4 | 5 | 6 |
| **DYS392** | 2 | 3 | 3 | 3 |
| **DYS643** | 4 | 4 | 3 | 5 |
| **DYS393** | 4 | 4 | 4 | 4 |
| **DYS458** | 6 | 6 | 7 | 8 |
| **DYS456** | 4 | 8 | 8 | 9 |
| **YGATAH4** | 6 | 6 | 3 | 9 |
| **DYS385a** | 8 | 9 | 10 | 11 |
| **DYS385b** | 9 | 12 | 11 | 14 |

## APPENDIX 81

## PAIRWISE $F_{ST}$ VALUES FOR GENETIC DISTANCE MATRIX BASED ON 13 AUTOSOMAL STR LOCI FROM 23 GLOBAL POPULATION GROUPS.

| | NIG | GHA | MOR | EGY | MOZ | SOM | U-AFM | U-CAU | U-HIS | U-ASI | BRA | MEX | HON | COL | MAL | CHI | IND | BAH | LAO | S-ARA | THA | ITA | SPA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NIG** | 0.000 | | | | | | | | | | | | | | | | | | | | | | |
| **GHA** | 0.002 | 0.000 | | | | | | | | | | | | | | | | | | | | | |
| **MOR** | 0.013 | 0.016 | 0.000 | | | | | | | | | | | | | | | | | | | | |
| **EGY** | 0.012 | 0.015 | 0.003 | 0.000 | | | | | | | | | | | | | | | | | | | |
| **MOZ** | 0.005 | 0.005 | 0.013 | 0.012 | 0.000 | | | | | | | | | | | | | | | | | | |
| **SOM** | 0.008 | 0.010 | 0.011 | 0.008 | 0.011 | 0.000 | | | | | | | | | | | | | | | | | |
| **U-AFM** | 0.002 | 0.003 | 0.009 | 0.007 | 0.003 | 0.006 | 0.000 | | | | | | | | | | | | | | | | |
| **U-CAU** | 0.018 | 0.021 | 0.007 | 0.007 | 0.018 | 0.014 | 0.010 | 0.000 | | | | | | | | | | | | | | | |
| **U-HIS** | 0.015 | 0.017 | 0.006 | 0.005 | 0.014 | 0.010 | 0.009 | 0.004 | 0.000 | | | | | | | | | | | | | | |
| **U-ASI** | 0.025 | 0.028 | 0.012 | 0.010 | 0.022 | 0.021 | 0.018 | 0.015 | 0.011 | 0.000 | | | | | | | | | | | | | |
| **BRA** | 0.011 | 0.012 | 0.004 | 0.003 | 0.009 | 0.009 | 0.005 | 0.003 | 0.003 | 0.012 | 0.000 | | | | | | | | | | | | |
| **MEX** | 0.025 | 0.027 | 0.013 | 0.014 | 0.024 | 0.018 | 0.018 | 0.011 | 0.004 | 0.017 | 0.010 | 0.000 | | | | | | | | | | | |
| **HON** | 0.006 | 0.008 | 0.012 | 0.012 | 0.006 | 0.012 | 0.005 | 0.015 | 0.012 | 0.022 | 0.009 | 0.019 | 0.000 | | | | | | | | | | |
| **COL** | 0.019 | 0.021 | 0.008 | 0.007 | 0.018 | 0.014 | 0.012 | 0.006 | 0.002 | 0.013 | 0.005 | 0.002 | 0.014 | 0.000 | | | | | | | | | |
| **MAL** | 0.022 | 0.024 | 0.011 | 0.008 | 0.020 | 0.018 | 0.015 | 0.012 | 0.009 | 0.005 | 0.009 | 0.015 | 0.021 | 0.012 | 0.000 | | | | | | | | |
| **CHI** | 0.026 | 0.028 | 0.011 | 0.010 | 0.023 | 0.021 | 0.019 | 0.016 | 0.012 | 0.004 | 0.012 | 0.019 | 0.023 | 0.015 | 0.003 | 0.000 | | | | | | | |
| **IND** | 0.020 | 0.021 | 0.007 | 0.008 | 0.016 | 0.015 | 0.014 | 0.012 | 0.009 | 0.012 | 0.007 | 0.016 | 0.016 | 0.011 | 0.009 | 0.011 | 0.000 | | | | | | |
| **BAH** | 0.011 | 0.013 | 0.004 | 0.003 | 0.011 | 0.009 | 0.006 | 0.005 | 0.004 | 0.012 | 0.002 | 0.012 | 0.011 | 0.006 | 0.008 | 0.011 | 0.007 | 0.000 | | | | | |
| **LAO** | 0.027 | 0.029 | 0.013 | 0.011 | 0.026 | 0.022 | 0.020 | 0.014 | 0.012 | 0.007 | 0.012 | 0.018 | 0.026 | 0.015 | 0.002 | 0.003 | 0.012 | 0.012 | 0.000 | | | | |
| **S-ARA** | 0.014 | 0.016 | 0.005 | 0.003 | 0.014 | 0.009 | 0.008 | 0.007 | 0.006 | 0.013 | 0.004 | 0.014 | 0.013 | 0.008 | 0.011 | 0.013 | 0.009 | 0.002 | 0.014 | 0.000 | | | |
| **THA** | 0.023 | 0.024 | 0.010 | 0.008 | 0.021 | 0.018 | 0.016 | 0.012 | 0.010 | 0.006 | 0.009 | 0.017 | 0.022 | 0.013 | 0.002 | 0.004 | 0.008 | 0.009 | 0.002 | 0.011 | 0.000 | | |
| **ITA** | 0.016 | 0.019 | 0.005 | 0.004 | 0.015 | 0.012 | 0.009 | 0.002 | 0.004 | 0.012 | 0.002 | 0.012 | 0.014 | 0.006 | 0.009 | 0.012 | 0.009 | 0.002 | 0.012 | 0.003 | 0.010 | 0.000 | |
| **SPA** | 0.017 | 0.019 | 0.006 | 0.005 | 0.017 | 0.012 | 0.010 | 0.002 | 0.003 | 0.014 | 0.002 | 0.011 | 0.015 | 0.005 | 0.010 | 0.014 | 0.009 | 0.003 | 0.013 | 0.005 | 0.011 | 0.001 | 0.000 |

**Populations:**

Nigeria (NIG), Ghana (GHA), Morocco (MOR), Egypt (EGY), Mozambique (MOZ), Somalia (SOM), US African American (U-AFM), US Caucasian (U-CAU), US Hispanic (U-HIS), US Asian (U-ASI), Brazil (BRA), Mexico (MEX), Honduras (HON), Colombia (COL), Malaysia (MAL), China (CHI), India (IND), Bahrain (BAH), Laos (LAO), Saudi Arabia (S-ARA), Thailand (THAI), Italy (ITA), Spain (SPA).

## APPENDIX 82

## MATRIX OF PAIRWISE UNBIASED NEI GENETIC DISTANCES ACROSS 23 GLOBAL POPULATIONS USING 13 AUTOSOMAL STR MARKERS

| | NIG | GHA | MOR | EGY | MOZ | SOM | U-AFM | U-CAU | U-HIS | U-ASI | BRA | MEX | HON | COL | MAL | CHI | IND | BAH | LAO | S-ARA | THA | ITA | SPA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NIG** | 0.000 | | | | | | | | | | | | | | | | | | | | | | |
| **GHA** | 0.004 | 0.000 | | | | | | | | | | | | | | | | | | | | | |
| **MOR** | 0.103 | 0.114 | 0.000 | | | | | | | | | | | | | | | | | | | | |
| **EGY** | 0.088 | 0.098 | 0.018 | 0.000 | | | | | | | | | | | | | | | | | | | |
| **MOZ** | 0.025 | 0.022 | 0.097 | 0.082 | 0.000 | | | | | | | | | | | | | | | | | | |
| **SOM** | 0.055 | 0.058 | 0.083 | 0.056 | 0.075 | 0.000 | | | | | | | | | | | | | | | | | |
| **U-AFM** | 0.011 | 0.011 | 0.064 | 0.044 | 0.016 | 0.039 | 0.000 | | | | | | | | | | | | | | | | |
| **U-CAU** | 0.136 | 0.146 | 0.048 | 0.048 | 0.129 | 0.097 | 0.074 | 0.000 | | | | | | | | | | | | | | | |
| **U-HIS** | 0.110 | 0.115 | 0.039 | 0.032 | 0.102 | 0.071 | 0.059 | 0.020 | 0.000 | | | | | | | | | | | | | | |
| **U-ASI** | 0.183 | 0.194 | 0.079 | 0.059 | 0.160 | 0.143 | 0.128 | 0.104 | 0.070 | 0.000 | | | | | | | | | | | | | |
| **BRA** | 0.077 | 0.080 | 0.026 | 0.021 | 0.067 | 0.062 | 0.033 | 0.015 | 0.013 | 0.079 | 0.000 | | | | | | | | | | | | |
| **MEX** | 0.183 | 0.192 | 0.088 | 0.092 | 0.173 | 0.127 | 0.128 | 0.071 | 0.020 | 0.113 | 0.066 | 0.000 | | | | | | | | | | | |
| **HON** | 0.040 | 0.048 | 0.092 | 0.087 | 0.038 | 0.083 | 0.035 | 0.113 | 0.085 | 0.156 | 0.068 | 0.136 | 0.000 | | | | | | | | | | |
| **COL** | 0.141 | 0.152 | 0.052 | 0.050 | 0.129 | 0.098 | 0.088 | 0.038 | 0.007 | 0.084 | 0.031 | 0.012 | 0.106 | 0.000 | | | | | | | | | |
| **MAL** | 0.165 | 0.170 | 0.072 | 0.052 | 0.148 | 0.125 | 0.109 | 0.082 | 0.056 | 0.023 | 0.058 | 0.101 | 0.162 | 0.081 | 0.000 | | | | | | | | |
| **CHI** | 0.193 | 0.205 | 0.077 | 0.064 | 0.168 | 0.152 | 0.140 | 0.111 | 0.080 | 0.011 | 0.085 | 0.128 | 0.177 | 0.104 | 0.012 | 0.000 | | | | | | | |
| **IND** | 0.152 | 0.155 | 0.051 | 0.052 | 0.117 | 0.111 | 0.104 | 0.082 | 0.064 | 0.070 | 0.050 | 0.110 | 0.124 | 0.076 | 0.058 | 0.066 | 0.000 | | | | | | |
| **BAH** | 0.083 | 0.089 | 0.029 | 0.016 | 0.079 | 0.061 | 0.041 | 0.031 | 0.025 | 0.078 | 0.012 | 0.080 | 0.084 | 0.040 | 0.056 | 0.077 | 0.049 | 0.000 | | | | | |
| **LAO** | 0.209 | 0.213 | 0.092 | 0.076 | 0.198 | 0.160 | 0.149 | 0.102 | 0.083 | 0.035 | 0.087 | 0.131 | 0.210 | 0.112 | 0.007 | 0.014 | 0.082 | 0.089 | 0.000 | | | | |
| **S-ARA** | 0.102 | 0.112 | 0.034 | 0.016 | 0.104 | 0.066 | 0.060 | 0.044 | 0.040 | 0.086 | 0.030 | 0.094 | 0.099 | 0.051 | 0.073 | 0.089 | 0.062 | 0.009 | 0.105 | 0.000 | | | |
| **THA** | 0.178 | 0.177 | 0.071 | 0.053 | 0.159 | 0.130 | 0.117 | 0.086 | 0.068 | 0.031 | 0.064 | 0.120 | 0.175 | 0.097 | 0.007 | 0.018 | 0.054 | 0.063 | 0.011 | 0.078 | 0.000 | | |
| **ITA** | 0.119 | 0.132 | 0.033 | 0.021 | 0.110 | 0.084 | 0.062 | 0.012 | 0.022 | 0.079 | 0.010 | 0.079 | 0.108 | 0.037 | 0.060 | 0.086 | 0.057 | 0.011 | 0.087 | 0.021 | 0.068 | 0.000 | |
| **SPA** | 0.127 | 0.139 | 0.036 | 0.032 | 0.124 | 0.087 | 0.071 | 0.008 | 0.020 | 0.098 | 0.011 | 0.076 | 0.111 | 0.037 | 0.072 | 0.099 | 0.062 | 0.020 | 0.095 | 0.031 | 0.077 | 0.007 | 0.000 |

**Populations:**

Nigeria (NIG), Ghana (GHA), Morocco (MOR), Egypt (EGY), Mozambique (MOZ), Somalia (SOM), US African American (U-AFM), US Caucasian (U-CAU), US Hispanic (U-HIS), US Asian (U-ASI), Brazil (BRA), Mexico (MEX), Honduras (HON), Colombia (COL), Malaysia (MAL), China (CHI), India (IND), Bahrain (BAH), Laos (LAO), Saudi Arabia (S-ARA), Thailand (THAI), Italy (ITA), Spain (SPA).

## APPENDIX 83

## MATRIX OF PAIRWISE UNBIASED NEI GENETIC IDENTITY ACROSS 23 GLOBAL POPULATIONS USING 13 AUTOSOMAL STR MARKERS

| | NIG | GHA | MOR | EGY | MOZ | SOM | U-AFM | U-CAU | U-HIS | U-ASI | BRA | MEX | HON | COL | MAL | CHI | IND | BAH | LAO | S-ARA | THA | ITA | SPA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NIG** | 1.000 | | | | | | | | | | | | | | | | | | | | | | |
| **GHA** | 0.996 | 1.000 | | | | | | | | | | | | | | | | | | | | | |
| **MOR** | 0.902 | 0.892 | 1.000 | | | | | | | | | | | | | | | | | | | | |
| **EGY** | 0.916 | 0.907 | 0.982 | 1.000 | | | | | | | | | | | | | | | | | | | |
| **MOZ** | 0.975 | 0.978 | 0.908 | 0.921 | 1.000 | | | | | | | | | | | | | | | | | | |
| **SOM** | 0.947 | 0.944 | 0.920 | 0.945 | 0.928 | 1.000 | | | | | | | | | | | | | | | | | |
| **U-AFM** | 0.989 | 0.989 | 0.938 | 0.956 | 0.984 | 0.961 | 1.000 | | | | | | | | | | | | | | | | |
| **U-CAU** | 0.873 | 0.864 | 0.953 | 0.953 | 0.879 | 0.908 | 0.928 | 1.000 | | | | | | | | | | | | | | | |
| **U-HIS** | 0.896 | 0.891 | 0.962 | 0.969 | 0.903 | 0.931 | 0.942 | 0.980 | 1.000 | | | | | | | | | | | | | | |
| **U-ASI** | 0.833 | 0.824 | 0.924 | 0.943 | 0.852 | 0.866 | 0.880 | 0.902 | 0.932 | 1.000 | | | | | | | | | | | | | |
| **BRA** | 0.926 | 0.923 | 0.974 | 0.979 | 0.936 | 0.940 | 0.967 | 0.985 | 0.987 | 0.924 | 1.000 | | | | | | | | | | | | |
| **MEX** | 0.833 | 0.825 | 0.916 | 0.912 | 0.841 | 0.881 | 0.880 | 0.931 | 0.981 | 0.893 | 0.936 | 1.000 | | | | | | | | | | | |
| **HON** | 0.960 | 0.954 | 0.912 | 0.917 | 0.963 | 0.920 | 0.966 | 0.893 | 0.918 | 0.856 | 0.934 | 0.873 | 1.000 | | | | | | | | | | |
| **COL** | 0.868 | 0.859 | 0.949 | 0.952 | 0.879 | 0.907 | 0.916 | 0.963 | 0.993 | 0.919 | 0.969 | 0.988 | 0.899 | 1.000 | | | | | | | | | |
| **MAL** | 0.848 | 0.843 | 0.930 | 0.949 | 0.863 | 0.882 | 0.897 | 0.921 | 0.945 | 0.977 | 0.943 | 0.904 | 0.850 | 0.922 | 1.000 | | | | | | | | |
| **CHI** | 0.824 | 0.814 | 0.926 | 0.938 | 0.845 | 0.859 | 0.869 | 0.895 | 0.923 | 0.989 | 0.918 | 0.880 | 0.838 | 0.901 | 0.988 | 1.000 | | | | | | | |
| **IND** | 0.859 | 0.857 | 0.950 | 0.949 | 0.890 | 0.895 | 0.901 | 0.922 | 0.938 | 0.932 | 0.951 | 0.896 | 0.883 | 0.927 | 0.944 | 0.936 | 1.000 | | | | | | |
| **BAH** | 0.921 | 0.915 | 0.972 | 0.984 | 0.924 | 0.941 | 0.960 | 0.969 | 0.975 | 0.925 | 0.989 | 0.923 | 0.919 | 0.961 | 0.945 | 0.925 | 0.953 | 1.000 | | | | | |
| **LAO** | 0.812 | 0.808 | 0.912 | 0.927 | 0.820 | 0.852 | 0.861 | 0.903 | 0.920 | 0.966 | 0.917 | 0.878 | 0.810 | 0.894 | 0.993 | 0.986 | 0.921 | 0.915 | 1.000 | | | | |
| **S-ARA** | 0.903 | 0.894 | 0.966 | 0.985 | 0.901 | 0.936 | 0.942 | 0.957 | 0.961 | 0.918 | 0.971 | 0.910 | 0.905 | 0.950 | 0.929 | 0.915 | 0.940 | 0.991 | 0.900 | 1.000 | | | |
| **THA** | 0.837 | 0.837 | 0.931 | 0.948 | 0.853 | 0.878 | 0.890 | 0.917 | 0.935 | 0.970 | 0.938 | 0.887 | 0.840 | 0.907 | 0.993 | 0.982 | 0.947 | 0.939 | 0.989 | 0.925 | 1.000 | | |
| **ITA** | 0.887 | 0.876 | 0.967 | 0.979 | 0.896 | 0.919 | 0.940 | 0.988 | 0.978 | 0.924 | 0.990 | 0.924 | 0.897 | 0.963 | 0.942 | 0.918 | 0.944 | 0.989 | 0.917 | 0.979 | 0.934 | 1.000 | |
| **SPA** | 0.880 | 0.870 | 0.964 | 0.968 | 0.883 | 0.916 | 0.931 | 0.992 | 0.980 | 0.907 | 0.989 | 0.927 | 0.895 | 0.964 | 0.930 | 0.906 | 0.940 | 0.980 | 0.910 | 0.969 | 0.926 | 0.994 | 1.000 |

**Populations:**

Nigeria (NIG), Ghana (GHA), Morocco (MOR), Egypt (EGY), Mozambique (MOZ), Somalia (SOM), US African American (U-AFM), US Caucasian (U-CAU), US Hispanic (U-HIS), US Asian (U-ASI), Brazil (BRA), Mexico (MEX), Honduras (HON), Colombia (COL), Malaysia (MAL), China (CHI), India (IND), Bahrain (BAH), Laos (LAO), Saudi Arabia (S-ARA), Thailand (THAI), Italy (ITA), Spain (SPA).

## APPENDIX 84

## PAIRWISE $R_{ST}$ VALUES FOR GENETIC DISTANCE MATRIX BASED ON 12 Y-STR LOCI FROM 50 GLOBAL POPULATION GROUPS.

| | NIG-IB | NIG-YO | NIG-HF | ALG-AA | ANG-B | BEN-AC | BEN-M | BOT-B | BF-M | BF-G | CAM-P | CAM-B | CAR-P | DRC-B | DRC-P | ETH-AA | GAB-B | GAB-P | IVC-AC | IVC-M | KEN-B | KEN-NS | LIB-AA | NAM-K | SEN-M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NIG-IB** | 0 | | | | | | | | | | | | | | | | | | | | | | | | |
| **NIG-YO** | 0.051 | 0 | | | | | | | | | | | | | | | | | | | | | | | |
| **NIG-HF** | 0.100 | 0.154 | 0 | | | | | | | | | | | | | | | | | | | | | | |
| **ALG-AA** | 0.332 | 0.306 | 0.349 | 0 | | | | | | | | | | | | | | | | | | | | | |
| **ANG-B** | 0.071 | 0.095 | 0.171 | 0.338 | 0 | | | | | | | | | | | | | | | | | | | | |
| **BEN-AC** | 0.027 | 0.023 | 0.134 | 0.356 | 0.063 | 0 | | | | | | | | | | | | | | | | | | | |
| **BEN-M** | 0.031 | 0.038 | 0.126 | 0.365 | 0.048 | 0.014 | 0 | | | | | | | | | | | | | | | | | | |
| **BOT-B** | 0.066 | 0.087 | 0.164 | 0.235 | 0.050 | 0.077 | 0.080 | 0 | | | | | | | | | | | | | | | | | |
| **BF-M** | 0.084 | 0.071 | 0.149 | 0.317 | 0.089 | 0.066 | 0.088 | 0.061 | 0 | | | | | | | | | | | | | | | | |
| **BF-G** | 0.034 | 0.030 | 0.122 | 0.322 | 0.034 | 0.011 | 0.016 | 0.053 | 0.044 | 0 | | | | | | | | | | | | | | | |
| **CAM-P** | 0.074 | 0.082 | 0.140 | 0.199 | 0.100 | 0.075 | 0.083 | 0.073 | 0.107 | 0.067 | 0 | | | | | | | | | | | | | | |
| **CAM-B** | 0.053 | 0.048 | 0.162 | 0.303 | 0.105 | 0.038 | 0.066 | 0.056 | 0.076 | 0.046 | 0.056 | 0 | | | | | | | | | | | | | |
| **CAR-P** | 0.082 | 0.085 | 0.131 | 0.221 | 0.100 | 0.074 | 0.086 | 0.087 | 0.082 | 0.067 | 0.035 | 0.068 | 0 | | | | | | | | | | | | |
| **DRC-B** | 0.052 | 0.081 | 0.148 | 0.365 | 0.045 | 0.046 | 0.045 | 0.063 | 0.075 | 0.028 | 0.106 | 0.098 | 0.084 | 0 | | | | | | | | | | | |
| **DRC-P** | 0.138 | 0.132 | 0.213 | 0.142 | 0.164 | 0.149 | 0.161 | 0.079 | 0.091 | 0.121 | 0.096 | 0.106 | 0.104 | 0.147 | 0 | | | | | | | | | | |
| **ETH-AA** | 0.183 | 0.216 | 0.307 | 0.244 | 0.227 | 0.202 | 0.208 | 0.182 | 0.238 | 0.196 | 0.215 | 0.205 | 0.268 | 0.230 | 0.128 | 0 | | | | | | | | | |
| **GAB-B** | 0.029 | 0.041 | 0.130 | 0.328 | 0.040 | 0.020 | 0.020 | 0.049 | 0.077 | 0.015 | 0.062 | 0.046 | 0.071 | 0.034 | 0.140 | 0.197 | 0 | | | | | | | | |
| **GAB-P** | 0.202 | 0.211 | 0.228 | 0.166 | 0.218 | 0.219 | 0.241 | 0.171 | 0.144 | 0.191 | 0.113 | 0.180 | 0.128 | 0.238 | 0.122 | 0.303 | 0.220 | 0 | | | | | | | |
| **IVC-AC** | 0.053 | 0.071 | 0.130 | 0.414 | 0.090 | 0.038 | 0.030 | 0.127 | 0.134 | 0.054 | 0.123 | 0.101 | 0.113 | 0.066 | 0.210 | 0.229 | 0.043 | 0.286 | 0 | | | | | | |
| **IVC-M** | 0.055 | 0.088 | 0.119 | 0.406 | 0.098 | 0.059 | 0.048 | 0.106 | 0.107 | 0.043 | 0.110 | 0.106 | 0.125 | 0.108 | 0.185 | 0.251 | 0.071 | 0.251 | 0.092 | 0 | | | | | |
| **KEN-B** | 0.043 | 0.063 | 0.157 | 0.287 | 0.068 | 0.043 | 0.053 | 0.038 | 0.067 | 0.030 | 0.058 | 0.046 | 0.071 | 0.056 | 0.081 | 0.142 | 0.045 | 0.180 | 0.095 | 0.069 | 0 | | | | |
| **KEN-NS** | 0.086 | 0.106 | 0.221 | 0.261 | 0.099 | 0.093 | 0.108 | 0.053 | 0.124 | 0.083 | 0.105 | 0.076 | 0.152 | 0.127 | 0.098 | 0.076 | 0.081 | 0.227 | 0.149 | 0.138 | 0.056 | 0 | | | |
| **LIB-AA** | 0.277 | 0.294 | 0.333 | 0.103 | 0.292 | 0.317 | 0.325 | 0.230 | 0.257 | 0.282 | 0.205 | 0.286 | 0.233 | 0.321 | 0.104 | 0.183 | 0.301 | 0.154 | 0.382 | 0.340 | 0.229 | 0.212 | 0 | | |
| **NAM-K** | 0.264 | 0.260 | 0.288 | 0.296 | 0.268 | 0.274 | 0.296 | 0.247 | 0.203 | 0.255 | 0.252 | 0.265 | 0.252 | 0.316 | 0.248 | 0.370 | 0.288 | 0.166 | 0.345 | 0.287 | 0.274 | 0.292 | 0.196 | 0 | |
| **SEN-M** | 0.044 | 0.024 | 0.152 | 0.367 | 0.053 | 0.013 | 0.030 | 0.071 | 0.072 | 0.021 | 0.083 | 0.048 | 0.093 | 0.067 | 0.157 | 0.212 | 0.028 | 0.212 | 0.069 | 0.085 | 0.072 | 0.084 | 0.321 | 0.262 | 0 |

**Populations:**

Nigeria-Igbo (NIG-IB), Nigeria-Yoruba (NIG-YO), Nigeria-Hausa-Fulani (NIG-HF), Algeria-Afroasiatic (ALG-AA), Angola-Bantu West (ANG-B), Benin-Niger Congo-Atlantic (BEN-AC), Benin-Niger Congo-Mende (BEN-M), Botswana-Bantu East (BOT-B), Burkina Faso-Niger Congo-Mende (BF-M), Burkina Faso-Niger Congo-Gur (BF-G), Cameroon-Niger Cong-Pygmy (CAM-P), Cameroon-Bantu West (CAM-B), Central African Republic-Pygmy (CAR-P), Democratic Republic Congo-Bantu West (DRC-B), Democratic Republic Congo-Nilotic-Pygmy (DRC-P), Ethiopia-Afroasiatic (ETH-AA), Gabon-Bantu West (GAB-B), Gabon-Pygmy West (GAB-P), Ivory Coast-Niger Congo-Atlantic (IVC-AC), Ivory Coast-Niger Congo-Mende (IVC-M), Kenya-Bantu West (KEN-B), Kenya-Nilotic (KEN-NS), Libya-Afroasiatic (LIB-AA), Namibia-Khoisan (NAM-K), Senegal-Niger Congo-Mende (SEN-M), South Africa-Bantu East (SA-B), Tanzania-Afroasiatic (TZ-AA), Tanzania-Nilotic (TZ-NS), Tanzania-Khoisan (TZ-K), Tanzania-Bantu West (TZ-B), Uganda-Nilotic (UGA-NS), Zambia-Bantu West (ZAW-B), Zambia-Bantu East (ZAE-B), China-Sino Tibetan (CHI), India-Dravidian (IND), Iraq-Afroasiatic-Semitic (IRQ), Israel-Afroasiatic Semitic (ISR), Kuwait-Afroasiatic-Semitic (KUW), Lebanon- Afroasiatic-Semitic (LEB), Philippine- Afroasiatic-Semitic (PHP), United Arab Emirates- Afroasiatic-Semitic (UAE), Germany-Indo-European-Germany (GER), Italy-Indo-European-Italian (ITA), Spain-Indo-European-Spanish (SPA), United Kingdom-British African (UK-BA), United Kingdom-Indo-European-English (UK-BW), United State-African American (US-AA), Brazil- Indo-European-Portuguese (BRA), Jamaica-Jamaica-Patois (JAM), Peru- Indo-European-Spanish (PER).

## APPENDIX 84 (Continued)

## PAIRWISE $R_{ST}$ VALUES FOR GENETIC DISTANCE MATRIX BASED ON 12 Y-STR LOCI FROM 50 GLOBAL POPULATION GROUPS.

| | NIG-IB | NIG-YO | NIG-HF | ALG-AA | ANG-B | BEN-AC | BEN-M | BOT-B | BF-M | BF-G | CAM-P | CAM-B | CAR-P | DRC-B | DRC-P | ETH-AA | GAB-B | GAB-P | IVC-AC | IVC-M | KEN-B | KEN-NS | LIB-AA | NAM-K | SEN-M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SA-B** | 0.123 | 0.126 | 0.171 | 0.343 | 0.128 | 0.117 | 0.114 | 0.113 | 0.123 | 0.114 | 0.145 | 0.128 | 0.150 | 0.123 | 0.170 | 0.246 | 0.109 | 0.231 | 0.150 | 0.158 | 0.123 | 0.167 | 0.308 | 0.282 | 0.129 |
| **TZ-AA** | 0.100 | 0.091 | 0.172 | 0.154 | 0.108 | 0.101 | 0.114 | 0.036 | 0.069 | 0.082 | 0.048 | 0.063 | 0.065 | 0.116 | 0.032 | 0.165 | 0.084 | 0.115 | 0.163 | 0.127 | 0.049 | 0.074 | 0.146 | 0.214 | 0.117 |
| **TZ-NS** | 0.200 | 0.207 | 0.303 | 0.198 | 0.187 | 0.214 | 0.226 | 0.132 | 0.144 | 0.180 | 0.181 | 0.186 | 0.187 | 0.208 | 0.079 | 0.116 | 0.195 | 0.212 | 0.279 | 0.266 | 0.139 | 0.089 | 0.132 | 0.282 | 0.202 |
| **TZ-K** | 0.102 | 0.106 | 0.162 | 0.175 | 0.121 | 0.121 | 0.136 | 0.068 | 0.045 | 0.089 | 0.086 | 0.103 | 0.088 | 0.132 | 0.058 | 0.180 | 0.118 | 0.085 | 0.196 | 0.132 | 0.072 | 0.105 | 0.151 | 0.198 | 0.117 |
| **TZ-B** | 0.078 | 0.080 | 0.161 | 0.230 | 0.078 | 0.076 | 0.091 | 0.020 | 0.055 | 0.058 | 0.053 | 0.040 | 0.061 | 0.086 | 0.062 | 0.196 | 0.057 | 0.138 | 0.135 | 0.115 | 0.048 | 0.073 | 0.213 | 0.232 | 0.082 |
| **UGA-NS** | 0.220 | 0.225 | 0.265 | 0.208 | 0.225 | 0.233 | 0.250 | 0.155 | 0.190 | 0.217 | 0.119 | 0.179 | 0.156 | 0.257 | 0.105 | 0.273 | 0.230 | 0.065 | 0.300 | 0.270 | 0.176 | 0.195 | 0.194 | 0.215 | 0.226 |
| **ZAW-B** | 0.055 | 0.072 | 0.152 | 0.349 | 0.006 | 0.044 | 0.032 | 0.046 | 0.075 | 0.025 | 0.090 | 0.089 | 0.087 | 0.030 | 0.157 | 0.223 | 0.024 | 0.225 | 0.069 | 0.087 | 0.063 | 0.099 | 0.307 | 0.284 | 0.039 |
| **ZAE-B** | 0.047 | 0.055 | 0.146 | 0.368 | 0.027 | 0.029 | 0.019 | 0.066 | 0.080 | 0.017 | 0.090 | 0.085 | 0.086 | 0.024 | 0.148 | 0.214 | 0.016 | 0.232 | 0.046 | 0.090 | 0.053 | 0.102 | 0.317 | 0.302 | 0.041 |
| **UK-BA** | 0.022 | 0.049 | 0.121 | 0.297 | 0.054 | 0.028 | 0.041 | 0.039 | 0.039 | 0.021 | 0.066 | 0.047 | 0.068 | 0.040 | 0.101 | 0.198 | 0.026 | 0.163 | 0.076 | 0.064 | 0.034 | 0.080 | 0.243 | 0.233 | 0.039 |
| **US-AA** | 0.058 | 0.098 | 0.125 | 0.226 | 0.073 | 0.077 | 0.092 | 0.041 | 0.051 | 0.060 | 0.064 | 0.067 | 0.075 | 0.082 | 0.082 | 0.189 | 0.065 | 0.110 | 0.130 | 0.086 | 0.064 | 0.087 | 0.179 | 0.183 | 0.077 |
| **CHI** | 0.282 | 0.319 | 0.346 | 0.250 | 0.324 | 0.321 | 0.335 | 0.244 | 0.227 | 0.293 | 0.256 | 0.298 | 0.266 | 0.324 | 0.149 | 0.273 | 0.327 | 0.160 | 0.379 | 0.349 | 0.264 | 0.260 | 0.164 | 0.227 | 0.320 |
| **IND** | 0.226 | 0.247 | 0.287 | 0.157 | 0.214 | 0.254 | 0.272 | 0.160 | 0.185 | 0.221 | 0.165 | 0.211 | 0.205 | 0.258 | 0.094 | 0.212 | 0.241 | 0.097 | 0.324 | 0.274 | 0.192 | 0.180 | 0.072 | 0.159 | 0.240 |
| **IRQ** | 0.238 | 0.262 | 0.298 | 0.129 | 0.242 | 0.268 | 0.281 | 0.168 | 0.196 | 0.236 | 0.178 | 0.232 | 0.201 | 0.267 | 0.075 | 0.176 | 0.259 | 0.117 | 0.335 | 0.284 | 0.194 | 0.172 | 0.044 | 0.185 | 0.265 |
| **ISR** | 0.228 | 0.246 | 0.309 | 0.151 | 0.226 | 0.250 | 0.263 | 0.158 | 0.198 | 0.226 | 0.190 | 0.225 | 0.219 | 0.249 | 0.087 | 0.130 | 0.233 | 0.166 | 0.316 | 0.286 | 0.186 | 0.140 | 0.048 | 0.199 | 0.238 |
| **KUW** | 0.234 | 0.257 | 0.311 | 0.156 | 0.217 | 0.271 | 0.272 | 0.154 | 0.231 | 0.241 | 0.184 | 0.234 | 0.223 | 0.255 | 0.095 | 0.159 | 0.235 | 0.153 | 0.323 | 0.302 | 0.196 | 0.154 | 0.050 | 0.168 | 0.256 |
| **LEB** | 0.249 | 0.275 | 0.316 | 0.152 | 0.263 | 0.280 | 0.296 | 0.188 | 0.210 | 0.253 | 0.198 | 0.250 | 0.225 | 0.285 | 0.089 | 0.183 | 0.272 | 0.139 | 0.354 | 0.304 | 0.208 | 0.183 | 0.042 | 0.170 | 0.275 |
| **PHP** | 0.323 | 0.375 | 0.364 | 0.287 | 0.397 | 0.377 | 0.397 | 0.286 | 0.268 | 0.341 | 0.289 | 0.335 | 0.290 | 0.388 | 0.167 | 0.317 | 0.384 | 0.193 | 0.440 | 0.385 | 0.295 | 0.319 | 0.221 | 0.297 | 0.394 |
| **UAE** | 0.242 | 0.278 | 0.328 | 0.170 | 0.249 | 0.292 | 0.290 | 0.191 | 0.261 | 0.265 | 0.203 | 0.265 | 0.243 | 0.286 | 0.109 | 0.173 | 0.260 | 0.195 | 0.345 | 0.328 | 0.214 | 0.180 | 0.047 | 0.221 | 0.286 |
| **BRA** | 0.269 | 0.302 | 0.303 | 0.149 | 0.271 | 0.290 | 0.306 | 0.197 | 0.240 | 0.264 | 0.186 | 0.250 | 0.204 | 0.293 | 0.119 | 0.245 | 0.281 | 0.122 | 0.342 | 0.309 | 0.242 | 0.215 | 0.126 | 0.214 | 0.295 |
| **JAM** | 0.035 | 0.081 | 0.123 | 0.273 | 0.067 | 0.057 | 0.065 | 0.044 | 0.056 | 0.041 | 0.064 | 0.068 | 0.069 | 0.055 | 0.096 | 0.200 | 0.043 | 0.152 | 0.098 | 0.079 | 0.050 | 0.091 | 0.216 | 0.228 | 0.065 |
| **PER** | 0.258 | 0.282 | 0.324 | 0.178 | 0.283 | 0.268 | 0.289 | 0.206 | 0.270 | 0.256 | 0.194 | 0.238 | 0.230 | 0.286 | 0.150 | 0.211 | 0.263 | 0.229 | 0.326 | 0.282 | 0.222 | 0.187 | 0.157 | 0.278 | 0.290 |
| **GER** | 0.305 | 0.349 | 0.320 | 0.215 | 0.285 | 0.331 | 0.343 | 0.225 | 0.259 | 0.303 | 0.214 | 0.293 | 0.220 | 0.318 | 0.175 | 0.328 | 0.322 | 0.097 | 0.371 | 0.364 | 0.291 | 0.279 | 0.196 | 0.222 | 0.332 |
| **ITA** | 0.257 | 0.277 | 0.304 | 0.118 | 0.281 | 0.279 | 0.299 | 0.196 | 0.217 | 0.258 | 0.185 | 0.241 | 0.212 | 0.298 | 0.092 | 0.203 | 0.281 | 0.109 | 0.347 | 0.294 | 0.223 | 0.199 | 0.073 | 0.179 | 0.284 |
| **SPA** | 0.350 | 0.383 | 0.371 | 0.210 | 0.351 | 0.361 | 0.377 | 0.274 | 0.326 | 0.340 | 0.253 | 0.322 | 0.272 | 0.365 | 0.186 | 0.305 | 0.351 | 0.167 | 0.405 | 0.388 | 0.323 | 0.286 | 0.205 | 0.278 | 0.372 |
| **UK-BW** | 0.362 | 0.415 | 0.366 | 0.280 | 0.358 | 0.383 | 0.402 | 0.286 | 0.329 | 0.358 | 0.276 | 0.343 | 0.287 | 0.371 | 0.205 | 0.344 | 0.369 | 0.180 | 0.418 | 0.412 | 0.340 | 0.310 | 0.259 | 0.291 | 0.386 |

**Populations:**

Nigeria-Igbo (NIG-IB), Nigeria-Yoruba (NIG-YO), Nigeria-Hausa-Fulani (NIG-HF), Algeria-Afroasiatic (ALG-AA), Angola-Bantu West (ANG-B), Benin-Niger Congo-Atlantic (BEN-AC), Benin-Niger Congo-Mende (BEN-M), Botswana-Bantu East (BOT-B), Burkina Faso-Niger Congo-Mende (BF-M), Burkina Faso-Niger Congo-Gur (BF-G), Cameroon-Niger Cong-Pygmy (CAM-P), Cameroon-Bantu West (CAM-B), Central African Republic-Pygmy (CAR-P), Democratic Republic Congo-Bantu West (DRC-B), Democratic Republic Congo-Nilotic-Pygmy (DRC-P), Ethiopia-Afroasiatic (ETH-AA), Gabon-Bantu West (GAB-B), Gabon-Pygmy West (GAB-P), Ivory Coast-Niger Congo-Atlantic (IVC-AC), Ivory Coast-Niger Congo-Mende (IVC-M), Kenya-Bantu West (KEN-B), Kenya-Nilotic (KEN-NS), Libya-Afroasiatic (LIB-AA), Namibia-Khoisan (NAM-K), Senegal-Niger Congo-Mende (SEN-M), South Africa-Bantu East (SA-B), Tanzania-Afroasiatic (TZ-AA), Tanzania-Nilotic (TZ-NS), Tanzania-Khoisan (TZ-K), Tanzania-Bantu West (TZ-B), Uganda-Nilotic (UGA-NS), Zambia-Bantu West (ZAW-B), Zambia-Bantu East (ZAE-B), China-Sino Tibetan (CHI), India-Dravidian (IND), Iraq-Afroasiatic-Semitic (IRQ), Israel-Afroasiatic Semitic (ISR), Kuwait-Afroasiatic-Semitic (KUW), Lebanon- Afroasiatic-Semitic (LEB), Philippine- Afroasiatic-Semitic (PHP), United Arab Emirates- Afroasiatic-Semitic (UAE), Germany-Indo-European-Germany (GER), Italy-Indo-European-Italian (ITA), Spain-Indo-European-Spanish (SPA), United Kingdom-British African (UK-BA), United Kingdom-Indo-European-English (UK-BW), United State-African American (US-AA), Brazil- Indo-European-Portuguese (BRA), Jamaica-Jamaica-Patois (JAM), Peru- Indo-European-Spanish (PER).

## APPENDIX 84 (Continued)

## PAIRWISE $R_{ST}$ VALUES FOR GENETIC DISTANCE MATRIX BASED ON 12 Y-STR LOCI FROM 50 GLOBAL POPULATION GROUPS.

| | SA-B | TZ-AA | TZ-NS | TZ-K | TZ-B | UGA-NS | ZAW-B | ZAE-B | UK-BA | US-AA | CHI | IND | IRQ | ISR | KUW | LEB | PHP | UAE | BRA | JAM | PER | GER | ITA | SPA | UK-BW |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SA-B** | 0 | | | | | | | | | | | | | | | | | | | | | | | | |
| **TZ-AA** | 0.137 | 0 | | | | | | | | | | | | | | | | | | | | | | | |
| **TZ-NS** | 0.232 | 0.100 | 0 | | | | | | | | | | | | | | | | | | | | | | |
| **TZ-K** | 0.143 | 0.042 | 0.101 | 0 | | | | | | | | | | | | | | | | | | | | | |
| **TZ-B** | 0.127 | 0.017 | 0.129 | 0.052 | 0 | | | | | | | | | | | | | | | | | | | | |
| **UGA-NS** | 0.270 | 0.101 | 0.214 | 0.106 | 0.120 | 0 | | | | | | | | | | | | | | | | | | | |
| **ZAW-B** | 0.114 | 0.107 | 0.193 | 0.118 | 0.072 | 0.233 | 0 | | | | | | | | | | | | | | | | | | |
| **ZAE-B** | 0.122 | 0.106 | 0.199 | 0.125 | 0.076 | 0.241 | 0.018 | 0 | | | | | | | | | | | | | | | | | |
| **UK-BA** | 0.103 | 0.055 | 0.169 | 0.071 | 0.035 | 0.185 | 0.039 | 0.035 | 0 | | | | | | | | | | | | | | | | |
| **US-AA** | 0.129 | 0.047 | 0.143 | 0.049 | 0.031 | 0.136 | 0.070 | 0.081 | 0.023 | 0 | | | | | | | | | | | | | | | |
| **CHI** | 0.301 | 0.210 | 0.213 | 0.185 | 0.249 | 0.221 | 0.324 | 0.331 | 0.247 | 0.180 | 0 | | | | | | | | | | | | | | |
| **IND** | 0.267 | 0.127 | 0.163 | 0.130 | 0.151 | 0.131 | 0.232 | 0.253 | 0.179 | 0.109 | 0.083 | 0 | | | | | | | | | | | | | |
| **IRQ** | 0.268 | 0.121 | 0.134 | 0.118 | 0.156 | 0.138 | 0.254 | 0.268 | 0.190 | 0.121 | 0.086 | 0.019 | 0 | | | | | | | | | | | | |
| **ISR** | 0.246 | 0.119 | 0.110 | 0.118 | 0.151 | 0.185 | 0.230 | 0.249 | 0.176 | 0.119 | 0.144 | 0.064 | 0.030 | 0 | | | | | | | | | | | |
| **KUW** | 0.264 | 0.128 | 0.138 | 0.143 | 0.164 | 0.155 | 0.229 | 0.253 | 0.196 | 0.134 | 0.134 | 0.038 | 0.025 | 0.015 | 0 | | | | | | | | | | |
| **LEB** | 0.280 | 0.133 | 0.137 | 0.128 | 0.171 | 0.157 | 0.274 | 0.286 | 0.201 | 0.129 | 0.109 | 0.040 | 0.012 | 0.017 | 0.018 | 0 | | | | | | | | | |
| **PHP** | 0.355 | 0.237 | 0.241 | 0.206 | 0.288 | 0.255 | 0.395 | 0.402 | 0.300 | 0.218 | 0.037 | 0.140 | 0.137 | 0.213 | 0.209 | 0.171 | 0 | | | | | | | | |
| **UAE** | 0.280 | 0.142 | 0.148 | 0.164 | 0.187 | 0.206 | 0.261 | 0.276 | 0.212 | 0.153 | 0.171 | 0.072 | 0.044 | 0.025 | 0.013 | 0.031 | 0.237 | 0 | | | | | | | |
| **BRA** | 0.299 | 0.140 | 0.201 | 0.152 | 0.169 | 0.145 | 0.282 | 0.299 | 0.209 | 0.121 | 0.115 | 0.082 | 0.066 | 0.107 | 0.108 | 0.083 | 0.157 | 0.141 | 0 | | | | | | |
| **JAM** | 0.114 | 0.054 | 0.166 | 0.069 | 0.037 | 0.172 | 0.055 | 0.052 | 0.009 | 0.009 | 0.220 | 0.151 | 0.161 | 0.153 | 0.170 | 0.170 | 0.269 | 0.179 | 0.165 | 0 | | | | | |
| **PER** | 0.308 | 0.150 | 0.221 | 0.197 | 0.188 | 0.246 | 0.285 | 0.298 | 0.210 | 0.154 | 0.157 | 0.156 | 0.134 | 0.144 | 0.164 | 0.148 | 0.185 | 0.182 | 0.073 | 0.178 | 0 | | | | |
| **GER** | 0.331 | 0.192 | 0.253 | 0.198 | 0.216 | 0.137 | 0.300 | 0.329 | 0.250 | 0.157 | 0.123 | 0.086 | 0.102 | 0.179 | 0.163 | 0.144 | 0.179 | 0.219 | 0.057 | 0.212 | 0.188 | 0 | | | |
| **ITA** | 0.292 | 0.132 | 0.175 | 0.130 | 0.167 | 0.126 | 0.291 | 0.299 | 0.204 | 0.127 | 0.096 | 0.044 | 0.021 | 0.059 | 0.058 | 0.031 | 0.141 | 0.092 | 0.029 | 0.170 | 0.105 | 0.073 | 0 | | |
| **SPA** | 0.376 | 0.218 | 0.284 | 0.238 | 0.245 | 0.193 | 0.359 | 0.373 | 0.286 | 0.195 | 0.146 | 0.137 | 0.122 | 0.172 | 0.168 | 0.149 | 0.193 | 0.224 | 0.029 | 0.241 | 0.109 | 0.061 | 0.069 | 0 | |
| **UK-BW** | 0.398 | 0.235 | 0.302 | 0.253 | 0.256 | 0.186 | 0.365 | 0.385 | 0.301 | 0.201 | 0.169 | 0.170 | 0.157 | 0.211 | 0.196 | 0.184 | 0.215 | 0.265 | 0.048 | 0.251 | 0.152 | 0.058 | 0.104 | 0.018 | 0 |

**Populations:**

Nigeria-Igbo (NIG-IB), Nigeria-Yoruba (NIG-YO), Nigeria-Hausa-Fulani (NIG-HF), Algeria-Afroasiatic (ALG-AA), Angola-Bantu West (ANG-B), Benin-Niger Congo-Atlantic (BEN-AC), Benin-Niger Congo-Mende (BEN-M), Botswana-Bantu East (BOT-B), Burkina Faso-Niger Congo-Mende (BF-M), Burkina Faso-Niger Congo-Gur (BF-G), Cameroon-Niger Cong-Pygmy (CAM-P), Cameroon-Bantu West (CAM-B), Central African Republic- Pygmy (CAR-P), Democratic Republic Congo-Bantu West (DRC-B), Democratic Republic Congo-Nilotic-Pygmy (DRC-P), Ethiopia-Afroasiatic (ETH-AA), Gabon-Bantu West (GAB-B), Gabon-Pygmy West (GAB-P), Ivory Coast-Niger Congo-Atlantic (IVC-AC), Ivory Coast-Niger Congo-Mende (IVC-M), Kenya-Bantu West (KEN-B), Kenya-Nilotic (KEN-NS), Libya-Afroasiatic (LIB-AA), Namibia-Khoisan (NAM-K), Senegal-Niger Congo-Mende (SEN-M), South Africa-Bantu East (SA-B), Tanzania-Afroasiatic (TZ-AA), Tanzania-Nilotic (TZ-NS), Tanzania-Khoisan (TZ-K), Tanzania-Bantu West (TZ-B), Uganda-Nilotic (UGA-NS), Zambia-Bantu West (ZAW-B), Zambia-Bantu East (ZAE-B), China-Sino Tibetan (CHI), India-Dravidian (IND), Iraq-Afroasiatic-Semitic (IRQ), Israel-Afroasiatic Semitic (ISR), Kuwait-Afroasiatic-Semitic (KUW), Lebanon- Afroasiatic-Semitic (LEB), Philippine- Afroasiatic-Semitic (PHP), United Arab Emirates- Afroasiatic-Semitic (UAE), Germany-Indo-European-Germany (GER), Italy-Indo-European-Italian (ITA), Spain-Indo-European-Spanish (SPA), United Kingdom-British African (UK-BA), United Kingdom-Indo-European-English (UK-BW), United State-African American (US-AA), Brazil- Indo-European-Portuguese (BRA), Jamaica-Jamaica-Patois (JAM), Peru- Indo-European-Spanish (PER).

## APPENDIX 85

## MEAN LnP(K)±SD AND DELTA *K* VALUES FOR 13 AUTOSOMAL STR LOCI FROM 23 GLOBAL POPULATION GROUPS.

| *K* | Reps | Mean LnP(K) | SD LnP(K) | Ln'(K) | \|Ln''(K)\| | Delta *K* |
|---|---|---|---|---|---|---|
| **1** | 10 | -236573 | 0.27 | | | |
| **2** | 10 | -233975 | 12.4 | 2597.32 | | |
| **3** | 10 | -232782 | 25.38 | 1193.01 | 1404.31 | 55.33136 |
| **4** | 10 | -232519 | 51.83 | 262.9 | 930.11 | 17.9454 |
| **5** | 10 | -232491 | 24.61 | 28.66 | 234.24 | 9.518082 |
| **6** | 10 | -233328 | 2262.03 | -836.84 | 865.5 | 0.382621 |
| **7** | 10 | -235435 | 5630.55 | -2107.85 | 1271.01 | 0.225735 |
| **8** | 10 | -238899 | 4670.61 | -3463.72 | 1355.87 | 0.290298 |
| **9** | 10 | -234064 | 478.03 | 4834.88 | 8298.6 | 17.36 |
| **10** | 10 | -233830 | 280.52 | 233.74 | 4601.14 | 16.40218 |
| **11** | 10 | -234532 | 1332.19 | -701.36 | 935.1 | 0.701927 |
| **12** | 10 | -252082 | 23195.17 | -17550.3 | 16848.94 | 0.726399 |
| **13** | 10 | -247270 | 16738.81 | 4811.81 | 22362.11 | 1.335944 |
| **14** | 10 | -254667 | 22924.99 | -7396.61 | 12208.42 | 0.532538 |
| **15** | 10 | -253913 | 22300.44 | 753.89 | 8150.5 | 0.365486 |
| **16** | 10 | -260303 | 19578.87 | -6390.27 | 7144.16 | 0.364891 |
| **17** | 10 | -242271 | 5701.84 | 18032.06 | 24422.33 | 4.283237 |
| **18** | 10 | -268551 | 28498.45 | -26279.5 | 44311.55 | 1.554876 |
| **19** | 10 | -270430 | 48949.19 | -1879.39 | 24400.1 | 0.498478 |
| **20** | 10 | -268536 | 25157.61 | 1894.48 | 3773.87 | 0.150009 |
| **21** | 10 | -260534 | 15312.82 | 8001.67 | 6107.19 | 0.398829 |
| **22** | 10 | -275218 | 30504.52 | -14683.5 | 22685.2 | 0.743667 |
| **23** | 10 | -263747 | 30718.56 | 11470.25 | 26153.78 | 0.8514 |
| **24** | 10 | -273546 | 18060.93 | -9798.98 | 21269.23 | 1.177638 |
| **25** | 10 | -297185 | 44290.65 | -23638.8 | 13839.78 | 0.312476 |