Genetic insight into Nigerian population groups using an X chromosome decaplex system

Catarina Isabel Ferreira Gomes
MSc in Forensic Genetics
Faculty of Sciences
University of Porto

2018/2019

Supervisor

Dr. Iva Gomes

Instituto de Investigação e Inovação em Saúde (i3S), University of Porto, Portugal. Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Portugal

Co-supervisor

Dr. Leonor Gusmão

DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE INVESTIGAÇÃO
E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

25 ANOS
A DESCOBRIR O CANCRO
ipatimup

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, 16 / 12 / 2019

Acknowledgements

Em primeiro lugar, quero agradecer à minha orientadora Iva. Obrigada por ter aceitado a orientar-me nesta dissertação, obrigada por toda a paciência e disponibilidade para me esclarecer todas as dúvidas, por todas as oportunidades que me proporcionou e obrigada por todos os conhecimentos transmitidos ao longo deste ano. Foi um prazer trabalhar com uma pessoa que admiro cada vez mais.
Quero agradecer, igualmente, à minha co-orientadora Leonor, que apesar da distância esteve sempre disponível e foi uma ajuda crucial. Obrigada pelas opiniões, atenção e paciência.

Agradecer, também, ao professor Amorim pela oportunidade de integrar o seu grupo e por tudo o que me ensinou ao longo dos últimos 2 anos. E obrigada ao grupo de Genética Populacional do IPATIMUP por me ter recebido, a todos os elementos pela ajuda e disponibilidade.

Tenho de agradecer às companheiras de laboratório mais 'problemáticas' de sempre, Raquel e Sara! Agradecer por todos os momentos passados nos últimos meses, aqueles almoços prolongados, aquelas conversas que faziam o dia e o apoio recorrente que fizeram os dias longos parecerem mais curtos.
Ao Sr. Adão que teve sempre paciência para ouvir as minhas crises existências e lamentações e, além disso, teve sempre bons conselhos para dar. Obrigada por conseguires tornar esta caminhada menos difícil e mais divertida. E ao Rui também tenho de agradecer as palavras de apoio incondicional, mesmo nos piores momentos estava lá para me ouvir e distrair, um grande obrigada.
Um agradecimento especial às minhas miúdas Lima, Débora, Telma, Laura e Mariana, que apesar da distância, sempre tiveram uma palavra de força a dar! We'll always be there for each other!

À minha irmã, que sem ela, muito provavelmente, não teria sido capaz de chegar até aqui. Não só este ano, mas ao longo da vida: obrigada por me ajudares a tomar as decisões difíceis, por teres sempre um conselho para dar, por tornares os momentos difíceis um bocadinho mais fáceis e por seres "mais que minha mãe", obrigada por tudo!

Por último, mas definitivamente não menos importante, um agradecimento especial aos meus pais, que possibilitaram essa experiência com seu grande apoio e ajuda. Estarei sempre agradecida por tudo que fizeram por mim. Sem a ajuda deles e apoio contínuo, esta experiência não teria sido possível. Obrigada!

Resumo

Na genética forense a determinação do perfil genético de um indivíduo é utilizada principalmente na identificação individual, como por exemplo, em investigações de ordem criminal e na investigação de parentesco biológico. A probabilidade de identidade é calculada usando as frequências alélicas de cada marcador de um determinado perfil genético. Para a determinação desse perfil genético são utilizados marcadores moleculares, mais frequentemente do tipo microssatélites, denominados STRs (Short Tandem Repeats), que se encontram dispersos por todo o genoma humano. A escolha deste tipo de marcadores advém de inúmeros fatores, nomeadamente pelo facto de possuírem um elevado poder de discriminação e de serem facilmente amplificados através da reação de PCR. Neste caso, foram utilizados STRs presentes no cromossoma X, pois o cromossoma X exibe um modo distinto de transmissão. Nos indivíduos masculinos apenas uma cópia é transmitida inteiramente às descendentes do sexo feminino, permitindo a reconstrução direta do haplótipo; enquanto que nos indivíduos femininos as informações fornecidas pela recombinação no cromossoma X são uma visão adicional da história das populações humanas, alterando a variação genética em cada geração.

Neste estudo foi feita a caracterização genética de 10 X-STRs (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 e DXS6789) e avaliada a sua eficiência em aplicações forenses na população da Nigéria, um país situado no Oeste de África, no Golfo da Guiné. A amostra foi composta por 230 indivíduos masculinos não aparentados, pertencentes a três dos maiores grupos étnicos da Nigéria, nomeadamente Hausa (n= 90), Yoruba (n= 44) e Igbo (n= 96).

A análise de associação significativa entre pares de loci, com base no teste exato de desiquilíbrio de ligação (LD), não forneceu evidências de LD estatisticamente significante (p≥ 0,0011) entre os marcadores estudados na presente amostra.

A afinidade genética entre os três grupos étnicos estudados foi analisada através de distâncias genéticas (F_{ST}) e análise de variância molecular (AMOVA). Com base nos resultados não foram detetadas diferenças significativas, o que apoia a homogeneidade entre os três grupos estudados para os X-STRs utilizados neste trabalho.

Foram, também, comparadas as frequências alélicas entre as populações da Nigéria, Angola, Moçambique, Argélia, Costa do Marfim, Egipto, Gana, Guiné-Bissau, Marrocos, Tunísia e Uganda. Não foram detetadas diferenças genéticas significativas de F_{ST} entre os três grupos da Nigéria e as populações Subsarianas de Angola, Moçambique, Costa do Marfim, Gana, Guiné-Bissau e Uganda. Por conseguinte, as diferenças significativas

encontradas resultaram da comparação com os grupos Norte Africanos, nomeadamente Argélia, Egípto, Marrocos e Tunísia.

Na caracterização da população da Nigéria observou-se que o locus DXS6809 apresentou maior grau de polimorfismo (0,852), sendo também o locus com maior poder de discriminação (0,979) em mulheres e o maior poder de exclusão (0,881) em trios com filha. Por outro lado, o locus menos polimórfico foi o DXS7133 com um valor de diversidade genética igual a 0,564, sendo também este o que apresentou menor poder de discriminação em mulheres (0,561) e a menor probabilidade de exclusão em trios com filha (0,375). Os elevados valores obtidos de poder de discriminação acumulado comprovam o potencial do sistema decaplex desenvolvido em estudos de identificação, e as elevadas probabilidades de exclusão obtidas confirmam a utilidade destes marcadores em testes de parentesco, nomeadamente em testes de paternidade envolvendo filhas quando analisados trios completos ou em duos pai/filha.

Pode, assim, concluir-se que o decaplex X-STRs preenche alguns dos requisitos necessários para uso em identificação genética na população da Nigéria, sendo o seu uso uma mais-valia nesta área, uma vez que permite aumentar o poder de discriminação e de exclusão em casos de parentesco complexos em que os marcadores autossómicos são pouco informativos.

Palavras-chave:

Cromossoma X, X-STRs, marcadores do cromossoma X, STRs, Nigéria, Yoruba, Igbo, Hausa, África, Genética Forense, Genética de Populações

Abstract

In forensic genetics the determination of an individual's genetic profile is mainly used in particular cases where human individual identification is required and in kinship-based investigations. The probability of identity is calculated using the allelic frequencies of each marker of a given genetic profile. To determine this genetic profile, molecular markers, most often microsatellite type, called STRs (Short Tandem Repeats) are used, which are dispersed throughout the human genome. The choice of this type of markers comes from a number of factors, namely because they have a high discrimination power and are easily amplified through PCR reaction. In this case, were used STRs present on the X chromosome because it exhibits a distinct mode of transmission. In male individuals, only one copy is transmitted entirely to female offspring, allowing for direct reconstruction of the haplotype; whereas in female individuals the information provided by X chromosome recombination is an additional insight into the history of human populations, altering genetic variation in each generation.

In this study, 10 X-STRs were genetically characterized (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789) and their efficiency in forensic applications in the population of Nigeria, a country located in West Africa in the Gulf of Guinea. The sample consisted of 230 unrelated male individuals from three of Nigeria's largest ethnic groups, namely Hausa (n= 90), Yoruba (n= 44) and Igbo (n= 96).

The analysis of significant association between loci pairs, based on the exact linkage disequilibrium (LD) test, did not provide evidence of statistically significant LD ($p \geq 0.0011$) between the markers studied in the present sample.

Genetic affinities among the three ethnic groups studied was analyzed by genetic distances (F_{ST}) and molecular variance analysis (AMOVA). Based on the results, no significant differences were detected, which supports the homogeneity between the three groups of Nigeria studied for the X-STRs used.

Allelic frequencies were also compared between the populations of Nigeria, Angola, Mozambique, Algeria, Ivory Coast, Egypt, Ghana, Guinea-Bissau, Morocco, Tunisia and Uganda. No significant genetic differences of *Fst* were detected between the three groups in Nigeria and the sub-Saharan populations of Angola, Mozambique, Ivory Coast, Ghana, Guinea-Bissau and Uganda. Therefore, the significant differences found resulted from the comparison with the North African groups, namely Algeria, Egypt, Morocco and Tunisia.

In the characterization of the population of Nigeria, it was observed that the DXS6809 locus showed the highest degree of polymorphism (0.852), being also the locus with the

highest discrimination power (0.979) in female and the highest exclusion power (0.881) in trios involving daughters. On the other hand, the least polymorphic locus was the DXS7133 with a genetic diversity value of 0.564, which also had the lowest discriminating power in female (0.561) and the lowest probability of exclusion in trios involving daughters (0.375). The high values of accumulated discrimination power prove the potential of the decaplex system developed in identification studies, and the high probability of exclusion obtained confirm the usefulness of these markers in kinship tests, particularly in paternity tests involving daughters when complete trios are analyzed or in father / daughter duos.

Therefore, can be concluded that the decaplex X-STRs fulfills some of the necessary requirements for use in genetic identification in the population of Nigeria and its use is an added advantage in this area as it allows for greater discrimination and exclusion in cases of complex kinship where autosomal markers are poorly informative.

Keywords:

X chromosome, X-STRs, X chromosomal markers, Short Tandem Repeats, STRs, Nigeria, Yoruba, Igbo, Hausa, Africa, Forensic Genetics, Population Genetics

Table of Contents

List of Tables

List of Figures

List of Abbreviations

µL	Microliter
AMOVA	Analysis of molecular variance
AS	Autosomal
bp	Base pair
CE	Capillary electrophoresis
ChrX	X chromosome
ChrY	Y chromosome
ddNTPs	Dideoxyribonucleotide triphosphates
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
F_{ST}	F-statistics test
FTA	Fitzco/Flinder Technology Agreement
GD	Gene diversities
GHEP-ISFG	Spanish and Portuguese Group of the International Society of Forensic Genetics
He	Expected heterozygosity
HGDP	Human genome diversity project
INDEL	Insertion and deletion
kya	Thousand years ago
LD	Linkage disequilibrium
LDD	DNA Diagnostic Laboratory
Mb	Megabasepairs
ml	Milliliters
mtDNA	Mitochondrial DNA
MYA	Million years ago
PAGE	Polyacrylamide gel electrophoresis

PAR	Pseudoautosomal Region
PCR	Polymerase chain reaction
PD_F	Power of discrimination in females
PDHA1	Pyruvate dehydrogenase E1
PD_M	Power of discrimination in males
PE_D	Power of exclusion in duos
PE_{HS}	Power of exclusion in half-sisters
PE_T	Power of exclusion in trios
***p*-value**	Probability value
RNA	Ribonucleic acid
s.e.	Standard error
SDS	Sodium dodecyl sulfate
SMM	Stepwise *mutation* model
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
Tris	Tris(hydroxymethyl)aminomethane
UERJ	State University of Rio de Janeiro
VNTR	Variable Number Tandem Repeat
XAR	X-added region
XCI	X-chromosome inactivation
XCR	X-conservative region

Introduction

DNA structure and organization

The human genome holds a complete set of our genetic information, which is organized into deoxyribonucleic acid (DNA) molecules localized in the nucleus of the cells. The DNA is a double stranded molecule arranged in helical form, disclosed in 1953 by Watson and Crick (Watson and Crick, 1953). It is formed by nucleotide units, which comprises a triphosphate group, a deoxyribose sugar and a nitrogenous base – adenine (A), cytosine (C), thymine (T) or guanine (G).

There are two types of DNA: 1) nuclear, which is present in the nucleus cell organized in chromosomes, and 2) mitochondrial DNA (mtDNA) a small portion of extranuclear DNA named as such due to being present in the mitochondria (Goodwin et al., 2007). The DNA material in chromosomes can be divided into coding and non-coding regions. The coding regions are called genes and contain all the information needed for encoding and regulation of proteins synthesis. The function of non-coding DNA regions is still unclear; it does not encode proteins, but contains numerous elements that are involved in gene regulation (Butler, 2005, Goodwin et al., 2007). These organizational structures are contained in the nucleus of the cells and the entire genetic information of a cell is called the genome. The human genome was initially studied through the Human Genome Project, that sequenced 99% of the euchromatic human DNA from 2001 to 2004 (International Human Genome Sequencing, 2004).

Concerning human beings, in some regions, the DNA sequence is the same in all of the individuals of the specie and, in other regions, different. Some of these differences are, for example, responsible for the distinct physical features of each individual. For this reason, DNA typing has undergone numerous changes and has become an extremely important tool in human identification in the field of forensic sciences.

DNA in forensics

DNA markers are currently used in forensic genetics to conduct individual identification in criminal investigations, such as in the identification or confirmation of suspects, in cases of missing persons and in the identification of corpses and skeletal remains or in the investigation of biological kinship (especially paternity testing). The type of DNA used (nuclear or mitochondrial) depends on many factors (like sample type, i.e., bone, or on the state of degradation), but whenever possible nuclear DNA is preferred as it

has more power of discrimination and allows for individual identification (Jobling and Gill, 2004, Goodwin et al., 2007).

One of the main characteristics of human populations is their great genetic variability, responsible for inter-individual differences. This genetic variability is created by mutation and is influenced by evolutionary forces such as selection, genetic drift and recombination. All individuals are expected to have a different DNA sequence, except for monozygotic twins, and therefore the probability that two individuals share the same DNA profile is “virtually zero” for recombination markers (Butler, 2009, Amorim and Budowle, 2016).

DNA can be collected from any biological material, like for example, blood, semen, hair, bones, skin and saliva, being the same regardless of the cell types that are used. There are some rare exceptions, such as individuals called chimaeras (Castella et al., 2009). Genetic variation can be found in the form of sequence or length polymorphisms. Due to the research done on human DNA markers, many specific polymorphisms are well known and the most suitable are chosen for the above mentioned identification type of cases.

Types of genetic polymorphisms

The countless small and large differences in nucleotide sequence between individuals are known as DNA polymorphisms. A DNA polymorphism corresponds to the simultaneous existence of more than one allele (with frequencies higher than 1%) at a given locus in a given population. These polymorphisms can be a single-base sequence variation between individuals at a particular point in the genome called single nucleotide polymorphisms (SNPs) or an entire repetitive region of DNA with a tandem pattern, which called DNA satellites (Houck, 2015). DNA satellites range in size from 5 to 500 bp and cover approximately 3% of the genome. The regions of satellite DNA include Variable Number Tandem Repeats (VNTRs) or minisatellites, which are regions with sequences of 8 to 100 bp, and Short Tandem Repeats (STRs) or microsatellites, containing repetitive sequences with 1 to 7 bp. In the forensic field, degraded DNA is common, thus the use of STR markers are preferred because of their small size alleles (~100 bp to 400 bp) compared to minisatellite VNTR alleles (~400 bp to 1000 bp) (Butler, 2011, Jobling et al., 2014). The polymorphic DNA markers used for forensic purposes are required to be located, mainly, in non-coding regions (Schneider, 1997).

SNPs and INDELs

About 85% of the variations at the DNA sequence level correspond to differences in a single nucleotide (substitutions, insertions or deletions) and are designated SNPs or INDELs (insertion or deletion). SNPs are single-base variations in the genetic code that occur about every 1000 bases along the 3 billion bases of the human genome (Brenner et al., 2002, Budowle and van Daal, 2008). However, SNPs are so abundant throughout the genome that it is theoretically possible to type hundreds of them. This will make the combined power discrimination very high (Goodwin et al., 2007). Thus, it is to be expected that the distribution of SNPs in the genome is not homogeneous. In fact, there are regions in which the density of this type of polymorphism is greatest (for example, near the telomeres).

An INDEL can be the insertion or deletion of a segment of DNA ranging from one nucleotide to hundreds of nucleotides. The two alleles for biallelic INDELs can simply be classified as “short” and “long”. Most biallelic INDELs exhibit allele-length differences of only a few nucleotides (Butler, 2011). Despite the fact that the use of INDELs is still recent, there is a clear increase in the number of studies based on this type of polymorphisms that have been published in recent years, with several different purposes, being one of the main ancestral affiliations. Importantly, INDELs are very informative to deal with the study of the genetic structure of human populations (Ibarra et al., 2014, Fan et al., 2016). Recently, Pereira R. and co-authors (Pereira et al., 2012) described an X-INDEL multiplex system designed to amplify 32 biallelic markers in a single PCR. This multiplex includes X-INDELs described with a high degree of polymorphism in the major human population groups: Africa, Europe and Asia.

Despite the well proven utility of SNP and INDEL markers in forensic genetics, STRs are the markers of choice in human identification cases.

STRs: excellence markers in forensic genetics

The STRs have become the most common markers in the genetic field of human identification and criminal investigation (Hammond et al., 1994, Chakraborty et al., 1999, Butler et al., 2003, Pinto et al., 2011), and the main reasons are due to the following characteristics (Goodwin et al., 2007, Butler, 2011):

- Abundance in all over the genome and high degree of polymorphism. This high degree of polymorphism may be due to the number of replicates, size or composition of the repeat unit;

- Ability to perform amplification by PCR technique and amenable to automation processes involving fluorescence detection, reproducing robust results even when multiplexed with other markers;
- High discriminating power with observed heterozygosity of 70%;
- Predicted length of alleles that fall in the range of 100 bp to 400 bp with smaller sizes better suited for analysis of degraded DNA samples;
- In order to take advantage of the product rule and be able to combine the genetic information across multiple loci, it is possible to choose STR markers to be used in forensic DNA typing from separated chromosomes or that are widely spaced on the same chromosome to avoid any problems with physical linkage between markers;
- Low stutter products. Stutter products are amplicons that have one or more repeats units smaller or larger than the true allele and that are formed by strand slippage by the DNA polymerase during the extension of the target DNA strand in PCR amplification. The variation of these products may depend on the locus and on the structure of the core repeat. Although even in good profiles there are some stutters peaks, these are easily recognized and do not interfere with interpretation.

Another phenomenon that can happen during the amplification of DNA fragments that contain STR repeat regions is “allele dropout”. Allele dropout may occur due to mutations (variants) at the flanking region within the repeat region or in the primer-binding region. If the mutation occurs in the DNA template at the primer binding region, the hybridization of the primer can be disrupted, leading to amplification failure and consequently to non-detection of an allele in the DNA template. In this case, it’s called null (or silent) allele, and this phenomenon is rather rare, since the flanking sequence around STR repeats are quite stable and consistent across samples (Butler, 2011).

The number of STR repeats are according to the number of the repeats in the amplified region, for example, a dinucleotide marker has two repetitive nucleotides and a tri, tetra and pentanucleotide marker have 3, 4 and 5 nucleotides repeating in tandem, respectively. These last two are the most used and preferred type of repeats in forensic genetics, of which, the most used are tetranucleotide markers (Ellegren, 2004, Rapley and Whitehouse, 2007).
The advantages of using tetranucleotide STR loci in forensic DNA typing over di- and trinucleotide repeat STRs, include a well-defined allele size range that permits

multiplexing and that reduces allelic dropout from preferential amplification of smaller alleles. In comparison to dinucleotide repeats, tetranucleotides exhibit a reduced stutter product formation, that benefit the interpretation of sample mixtures (Butler, 2011).

X chromosome

Origin and evolution

About 300 million years ago, the first event of X-Y differentiation occurred. A consensus model for these sex-chromosomes evolution has been developed, hypothesizing that, in mammals, the human X and Y chromosomes have evolved from an ordinary autosomal pair that took different evolutionary paths (Lahn et al., 2001, Schaffner, 2004, Malone and Oliver, 2008). One way in which this could have happened was through inversions (or other chromosomal rearrangements), that included the sex-determining locus and imposed a direct hindrance to crossing over between the proto-sex chromosomes during meiosis. Over time, and in the absence of recombination, the two chromosomes independently accumulated mutations that made them increasingly different from each other (Ellegren, 2011). The Y chromosome has lost most of its sequence and its genes developing a unique pattern of repeated sequences. On the other hand, the X chromosome has not lost its autosomal character, indeed, the X chromosome is still able to recombine in females, while the Y chromosome only recombines with the X in the two pseudoautosomal regions (PAR) (Schaffner, 2004). These pseudoautosomal regions – Xp and Xq – referred to as PAR1 and PAR2, located in the telomeres (Figure 1) of the human X and Y chromosomes, are two short homologous regions, which are autosomal-like and can pair in the male meiosis (Lahn et al., 2001, Bachtrog, 2006).

When observing the physical arrangement of the X chromosome, discrete clusters were found with respect to the degree of divergence, suggesting a historical series of progressive reductions in recombination (Figure 1). These clusters, called 'evolutionary strata', are ordered so that the cluster with the lowest degree of divergence between X–Y paralogues is located next to the pseudoautosomal region PAR 1, with the other three clusters showing increasing divergence with expanding distance from PAR. A fifth stratum has subsequently been suggested (Ellegren, 2011).

Representing ~5% of the human genome, the X chromosome is distinctly bigger than the Y (155 Mb compared to 60 Mb) and is richer in genes, including many involved in both sexes in reproduction and cognitive functions (Jobling et al., 2014).

To balance the gene dosage between males and females, one of a female's two X chromosomes, in somatic tissues, is randomly silenced during early embryogenesis. This random X-chromosome inactivation (XCI) results in cellular hemizygosity of expression of most X-linked genes and in balanced mosaicism at the level of the organism, wherein about half of the cells expressing the maternal X and the other half

the paternal, but maintained clonally thereafter in daughter cells (Morey and Avner, 2010, Jobling et al., 2014). This regulation of gene expression was first described by Lyon (Lyon, 1961).

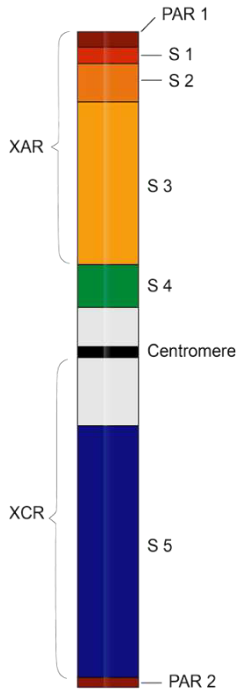


Figure 1. A schematic illustration of the organization of the human X chromosome. The evolutionary history of this chromosome includes a region defined as the X-conserved region (XCR) in blue and white, and the X-added region (XAR) in yellow, orange and red. The pseudoautosomal regions (PAR 1 and PAR 2) in brown, recombine with a corresponding region on the Y chromosome during male meiosis. S1–S5 (red, orange, yellow, green and blue, respectively) represent the five evolutionary strata that have been identified, each stratum tentatively represents the result of stepwise cessation of recombination between the proto-sex chromosomes. Figure modified from Ellegren, 2011

Main inheritance features of the X chromosome

Another interesting characteristic of the X chromosome is due to its special transmission and inheritance pattern, being inherited twice as frequently from females as from males. In a similar manner as for autosomes, women inherit one X chromosome from the mother and one from the father, which recombine only in female outside PARs, before being

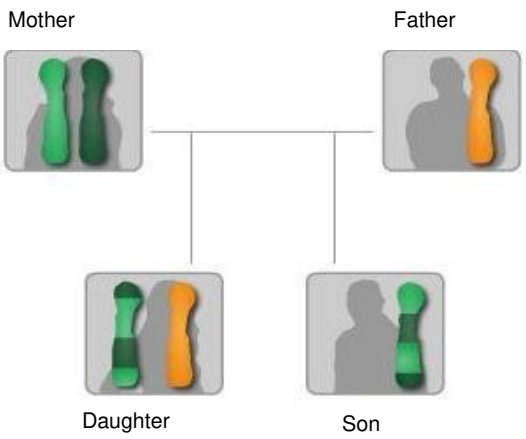


Figure 2. A schematic illustration of the mode of transmission of the X chromosome. A genealogy representative of the mode of transmission of the X chromosome where only two generations are represented; it is possible to verify that in the case of the mother the X chromosome undergoes recombination before being transmitted to the offspring contrary to what is observed in the father. Illustration modified from learn.genetics.utah.edu

transmitted to all their descendants, while the men receive only one X recombined chromosome from the mother that is then transmitted as a block to female offspring without recombination (except for PAR 1 and PAR2) (Figure 2) (Schaffner, 2004, learn.genetics.utah.edu, 2016).

In addition, due to the haploidy in males, this gender stays more exposed to the risk of X-linked diseases, since in females the second copy can mask the effect of mutation. For this reason, there has been strong sex bias ascertainment of severe diseases linked to the X, because although the X chromosome contains only 4% of the human genes, ~10% of the Mendelian diseases have been assigned to it (Jobling et al., 2014). Haploid state in males also has positive practical implications, such as, the single copy makes X-chromosomal DNA variation relatively easy to analyze in male DNA samples.

Population and forensic genetics

Population genetics is the study of inherited variation and its modulation in time and space (Suzuki and Griffiths, 1976). In other words, population genetics is a subdivision of the field of genetics that deals with variations in populations of organisms and provides a mathematical basis for evolutionary theory. For example, the examination and modelling of changes in the frequencies of alleles and genotypes in populations over relatively short time and (typically) approximately in a small number of generations. This type of evolutionary change is also known as microevolution (Jobling et al., 2004), and there are two ways of looking at genetic variation: variation within and among populations (Hamilton, 2009, Relethford, 2012). This variation allows to make inferences about the evolution of the human species (human evolution and history), performing lineage studies that permit to create reliable genealogical trees, or to simply determine degrees of kinship between individuals (individual's ancestry).

In forensic genetics the frequency of a given genetic profile, relevant to identify individuals, is calculated from the allelic frequencies of a given set of markers in the population to which it belongs, allowing to determine if it is unique or a common genetic profile (Jobling and Gill, 2004). This means that the fields of population genetics and forensic genetics are intrinsically linked: collecting population data in large scale is essential for the construction of reference databases and understanding the genetic variation of specific markers among worldwide populations prior to forensic applications.

Advantages of using X-STRs in population and forensic applications

The X chromosome has an important role in population and forensic genetics, mainly due to its characteristics of inheritance that have been mentioned above. In more recent years, it has been the subject of research in the fields of population genetics and forensics, mainly because it combines desirable features of other commonly used genetic markers (Chen et al., 2009, Gottipati et al., 2011, Arbiza et al., 2014).

X chromosome genetic variation studies have been used, mostly, in three categories: i) global population characteristics, that includes the occurrence of expansions or bottlenecks and the extent of subdivision within the population; ii) large-scale geographical history, for example, contributing significantly to the debate between the Out of Africa and Multiregional models of the expansion of modern humans (Schaffner, 2004, Athanasiadis et al., 2007) and; iii) the use of the X chromosome in forensic genetics and fine-scale geographical studies, with focus on smaller geographical and

ethnical levels, which continues to be overshadowed by the autosomes, Y chromosome and mtDNA.

The potential of genetic markers present on autosomes, Y chromosome and mtDNA has been extensively explored in the study of genetic diversity of human populations and individual identification (Melton et al., 1995, Poetsch et al., 2013b, Vullo et al., 2015). Therefore, a major contribution for the genetic characterization of populations and assessment of the forensic value in human identification using X chromosome markers is still missing, in particular for African populations (Athanasiadis et al., 2007, Gomes et al., 2007). Partly, this can be explained by the difficult access to samples from this continent and in addition because the bulk of work has been a phylogenetic effort done for the Y chromosome and mtDNA.

Interestingly, previous studies have used different regions of the X chromosome to infer past demographic events: from the Out of Africa migrations to regional patterns. Some of these studies are presented in Table 1. In population genetics, analyses often prefer the X regions with very low recombination to avoid genetic diversity from recombination events. Commonly, these regions are situated within segments with lengthy linkage disequilibrium (Nagaraja et al., 1997, Kaessmann et al., 1999, Garrigan et al., 2005, Laan et al., 2005). As previously mentioned, the X chromosome travels between both sexes in each generation, telling a different story of the one-parent genomes, only recombining in females, leading to a lower recombination rate compared to autosomes. Compared to the uniparental markers, the X chromosome has a superior effective population size and a weaker genetic drift. However, in comparison to autosomal chromosomes, the effective population size is lower, and the genetic drift is stronger, consequently, population structure is higher, which justifies its smaller global genetic diversity. If genetic drift is strong, as in the uniparental markers, it could lead to the loss of some alleles and a decrease of the genetic diversity. Thus, a good balance between the effective population size and the level of genetic drift is crucial to define a good population genetic marker (Schaffner, 2004). Furthermore, its homozygote status in males allows direct access to haplotypes. Also, the presence of a single copy of the X chromosome in males means that X-linked alleles are more exposed to natural selection, making this chromosome an attractive place to examine the role of selection in human history and an ideal source of information, including haplotyped-based phylogenetic studies often applied in human population genetics and anthropological research (e.g., (Schaffner, 2004, Szibor, 2007).

Table 1. Selection of relevant studies on the X chromosome to infer past demographic events

Samples	Marker	Conclusions	Reference
860 X chromosomes	chrX: dys44 (7622 bp) locus spanning exon 44 of the dystrophin gene	Modern human history reconstructed by the genetic diversity at the dys44 locus: ancestral population with Ne = 10,000 around 100-200 KYA which is divided into Africans and non-Africans with expansion outside the African continent	(Zietkiewicz et al., 1998)
35 males, 2 chimpanzees and 1 orangutan	chrX: PDHA1(4200 bp)	A striking haplotype differentiation between African and non-African samples was observed so a strong pattern of population differentiation was suggested.	(Harris and Hey, 1999)
336 X chromosomes	chrX: last intron (1089 bp) of the ZFX gene	Modern humans diverged from African ancestor 100 KYA with a subsequent range expansion	(Jaruzelska et al., 1999)
69 males distributed worldwide	chrX: Xq13.3 (10,200 bp)	The over-all pattern is most easily reconciled with an African origin for the sequence variation observed at Xq13.3	(Kaessmann et al., 1999)
139 X chromosomes	chrX: PDHA1(456 bp)	The differences between Africans and non-Africans in nucleotide frequencies and in the distribution of haplotypes do indicate substantial population subdivision between the two groups	(Yu and Li, 2000)
62 males distributed worldwide	chrX: 10,346 bp of non-coding sequence	Population expansion with significant excess of singletons	(Yu et al., 2002)
672 males from the HGDP panel	chrX: 10,084 bp and 2 STRs	Geographic and genetic patterns consistent with previous studies; One highly divergent haplotype	(Shimada et al., 2007)
667 males from the HGDP panel	chrX: 1 Alu element and 7 STRs	Increased diversity in African populations compared to non-African populations; Single origin in Africa and then migration to occupy the other continents by serial founder effect	(Santos-Lopes et al., 2007)
6092 X chromosomes	chrX: dys44 (7622 bp) locus spanning the exon 44 of the dystrophin gene	9% of Neanderthal-derived X chromosome segment in modern humans	(Yotova et al., 2011)

The study of autosomal and uniparental genetic markers is a widespread practice in forensic genetics, allowing to successfully solve the vast majority of questions. However, in particular cases, X chromosome specific markers are a powerful source of information and have the potential to efficiently complement the analysis of other genetic markers. X-STRs are very useful in investigating kinship relationships, especially in some cases of mixture analysis, namely in identifying the female component in male contaminated traces. X-STRs are more efficient than AS-STR (autosomal STR) in these particular scenarios because the female alleles can only be completely included in the male component if the female coincidentally happens to be homozygous at all loci. In addition, one of the main applications of X-STRs, and which autosomal STRs alone do not have, is the capacity of excluding the relationship between two putative paternal half-sisters. Males transmit their whole chromosome X to their daughters (unless mutation occurs) without recombination. Therefore, the paternal X chromosome haplotype can be detected in their daughters. It should be noted that several relationships that make up the so-called “deficient paternity testing” are also possible where the use of X chromosome polymorphisms may be of use. Some of these cases have been well addressed in regards to the paternity exclusion power in the study of Pinto et al. (2013). Resumed below are some of the potential applications of X-STRs in kinship analyses (Szibor et al., 2003b, Szibor, 2007, Krawczak, 2007, Pinto et al., 2013):

- Paternity testing in trios and duos

When paternity cases involve the common trio constellation of mother, daughter and alleged father and father/daughter relationships, X-STRs markers are more efficient than AS-STRs, because the average probability of exclusion of X-STR markers is higher than that of AS-STRs. However, this is not applied due to the common practice of autosomal markers in routine casework, i.e., it would not be routinely practical to use AS-STRs for one particular offspring gender (male) and X-STRs for the other particular descendant gender (female).

- Paternity and other kinship testing, involving relatives in various contexts, for example in incest and rape cases in incest

After criminal sexual assault by incest, pregnant women sometimes decide to voluntarily terminate the pregnancy. Investigating paternity in fetal material resulting from abortions proves to be a complex problem, in most cases it involves a mixture of fetal and mother DNA. If the fetus has male gender, paternity can still be tested efficiently using Y-STRs, if it is a female, the analysis can only include autosomal and X-STRs. Confirmation of

paternity can only be achieved with fetal alleles that are not shared by the mother. In incest cases in which the father is in fact the perpetrator of the crime, testing of the abortion material cannot demonstrate paternity since all fetal alleles would necessarily coincide with alleles of the daughter (and consequently with the father/grandfather). In these cases, X chromosome testing of mixed abortion material can only be used for supporting exclusion (and not positive paternity).

- Maternity tests

There are some situations in which mother/child testing is requested, for example, disagreement between biological and legal parents. Maternity can be demonstrated by sequencing mtDNA, however this procedure is expensive and the level of discrimination required in forensic cases is not always achieved. For testing mother–daughter relationships, X-STRs are equivalent to AS markers and do not provide any specific advantage, however, testing mother–son kinship is more efficient with X-STRs.

- “Deficient paternity” cases, when no sample is available from the putative father

When a biological sample from a putative father is not available, the DNA of paternal relatives must be analyzed. The mother of the unavailable father (i.e. the putative grandmother) is the key-figure since all X chromosome alleles from the presumed father can be determined by his mother's genotype. In the absence of the presumed grandmother, an investigation of two-sisters or half-sisters ChrX markers may exclude paternity, notably through the presence of four different alleles or haplotypes, even when neither parent is available for the test.
The putative grandmother's ChrX marker genotypes can also be reconstructed to some extent from their children. If she has several daughters, it is possible to determine the origin of the parents of most of her ChrX alleles and thus the grandmother's genotype. If brothers of the putative father are available, the data is even more informative. Obtaining X chromosome haplotypes from male individuals is very useful in kinship testing as it allows to compensate for information gaps over several generations, given the particularities of the X chromosome mode of transmission.

The human genome assembles a record of genetic variation, responsible for the increase of knowledge concerning our history and differences among populations, as well as its value in human identification studies. It is well established that polymorphisms spread throughout the entire human DNA allow us to understand our origin and evolutionary route representing a powerful tool in population genetic studies and on the other hand, due to their high individualization power and laboratorial easiness, are also the most common used markers in forensic studies. Autosomal, Y chromosomal and mtDNA polymorphisms have exceeded the use of X chromosome genetic markers in population genetic studies and in forensic applications, i.e., X-STRs are rarely employed in forensic practice. However, the formal genetic peculiarities make them particularly suitable for kinship testing and therefore they are mostly useful in particular scenarios where AS-STRs cannot offer the full answer and need complemental data from X-STRs. In the last few years, X linked loci have the deserved attention, particularly in the field of forensic genetics. A vast number of X chromosome STR markers have been reported and studied for forensic purposes as well as used to characterize several populations worldwide. Figure 3 illustrates the main in use and described X-STRs in forensic and population genetics along the ideogram. The X chromosome markers used in this work, DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789 (Figure 3), have been widely studied (e.g., (Gomes et al., 2007, Zeng et al., 2009, Poetsch et al., 2013a, Zidkova et al., 2014).

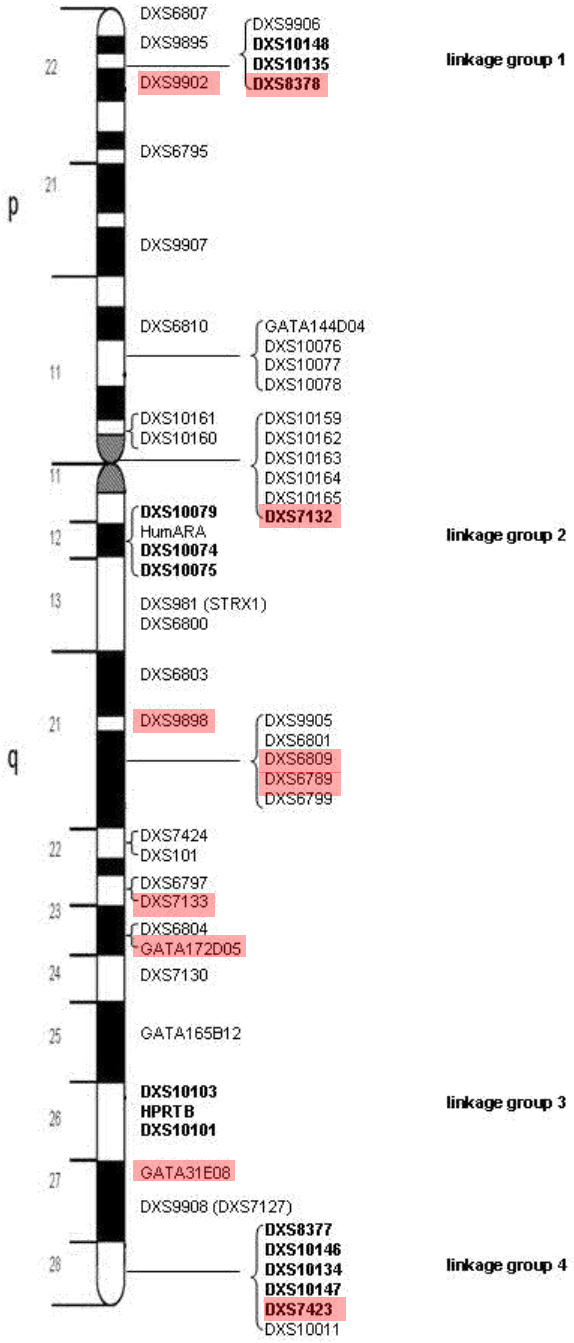


Figure 3. Map of commonly used X-chromosomal markers in forensics as well as four main recognized linkage groups (Szibor et al., 2006). In red are highlighted the 10 X-STRs used in this present study. Figure adapted from chrx-str.org

Parameters in population and forensic genetics

In order to evaluate the usefulness of a given locus in human identification, forensic parameters must be calculated in each population by being genetically characterized for this locus. The genetic diversity of a population is reflected in the number of alleles of a particular genetic marker. The variability is greater when the number of alleles of a marker is greater too. Population studies are part of the procedures needed to perform internal validation, as this data will enable the study of allele frequencies. Consequently, this data will be used in reporting population statistics and calculations regarding the most part of forensic parameters (Butler, 2005).

Some of the genetic parameters in population and forensic genetic studies commonly estimated are: observed and expected heterozygosities (or gene diversities, GD), powers of discrimination (PD) and powers of exclusion (PE) in particular kinship constellations. In addition, population comparisons to assess population structure (if, for example, several subgroups/regions compose the studied population) is also a common analysis and usually done through genetic distances estimation using F_{ST}'s.

Gene diversity is a measure of genetic variation within a population and is defined as the probability that two randomly chosen alleles from the population are different. The higher the heterozygosity the greater the genetic variation in the population.

Power of discrimination (PD) and power of exclusion (PE), first described by Fisher (Fisher, 1951), are parameters of high relevance for assessing forensic value of DNA markers, which are calculated according to kinship constellations. Power of discrimination indicates the probability that, in a population, two randomly selected individuals, unrelated, will have different genotypes for a given locus. Thus, the higher the number of genetic markers analyzed, the higher the capability of discriminating individuals in a population. Power of exclusion is the probability of a given locus to exclude a "random man" as a father of a child, in a typical paternity case. This is the proportion of individuals that have a DNA profile different from a randomly selected individual, permitting to exclude a putative father in a population (Evett et al., 1996) .

As for population comparisons using genetic distances such as F_{ST}, Sewall Wright (Wright, 1951) defined the F-statistics as the inbreeding coefficient to describe population structure, which measures the differences in the allele frequency distribution between two populations. The F_{ST} measure can range from 0 to 1, where 0 means complete sharing of genetic material and 1 means no sharing. The greater the genetic distance between populations, the less inbreeding among them is occurring and the more isolated they are from one another.

Africa and Nigeria: the study population

Today, it is unanimously accepted that Africa has in the past offered conditions favorable to human evolution. It is generally believed that three to two million years ago (MYA) several hominids coexisted in Africa, but only *Homo* was able to survive. There is little or no discussion regarding the origin of the *Homo* genus in East Africa, nor that *Homo erectus* was the first species to spread to Europe and Asia in the last two MYA (Jobling et al., 2004). There are two main hypotheses focus on the expansion of early modern humans from Africa (*Homo sapiens*). The most known hypotheses on this subject are: The Multiregional and the Out of Africa model (Figure 4). The Multiregional model proposes that the transition from *erectus* to *sapiens* took place in a number of places in the Old World, with the diverse modern human characteristics arising at different times in different places. Thus, according to this hypothesis, millions of individuals were ancestors of modern humanity. A direct consequence of this substantial exchange of genes would have been a huge genetic diversity, which however, does not occur in the present populations. In fact, genetic data point to the Out of Africa hypothesis, and only a few researchers still attribute any validity to the multiregional model.

The Out of Africa model proposes that the transition took place recently, about 200 thousand years ago (kya), in Africa, from a small number of ancestors, and that these humans replaced the hominids already present in other continents (Jobling et al., 2004, Gomez et al., 2014).

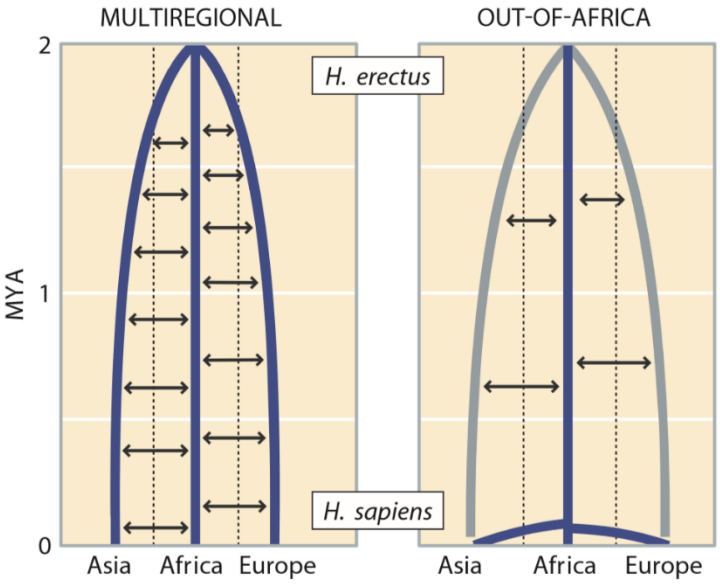


Figure 4. Both models begin with *H. erectus* after 2 MYA and lead to contemporary humans. Horizontal arrows indicate gene flow between populations on different continents. In the multiregional model, extensive gene flow is required; the Out of Africa model requires less. Blue lines: ancestors of modern humans. Gray lines: lineages that are not ancestors of modern humans. Illustration from Jobling et al, 2004

There are over 2000 distinct ethnolinguistic groups representing nearly one-third of the world's languages in Africa, and they are classified into four major linguistic families: Afroasiatic, Nilo-Saharan, Niger-Congo and Khoisan (Figure 5).

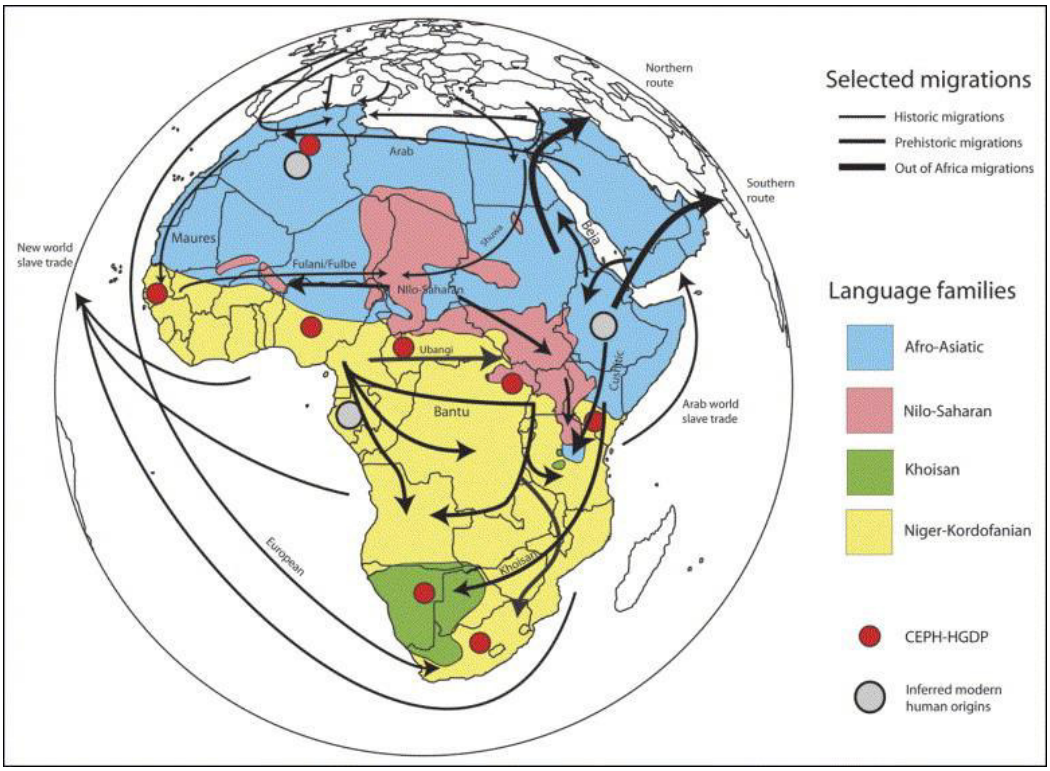


Figure 5. A map of selected migrations and major linguistic family's distribution in Africa. Languages change very rapidly over time, and a shared language family indicates some degree of recent shared cultural history. Currently, more than 2000 distinct languages exist in Africa, representing about a third of the world's languages, which can be classified into four major linguistic families – Afroasiatic, Nilo-Saharan, Khoisan and Niger-Kordofanian. This map also shows a number of key migration events, most notably geographic expansion of Bantu Niger-Kordofanian-speakers across Africa from a homeland near the Nigeria/Cameroon border (prehistoric migrations are represented by medium arrows), as well as the general geographic regions, where enslaved Africans were transported from Africa to the New World based on historical records (more recent migrations in historical times are represented by thin arrows). And hypothetical ancient migrations out of Africa are represented by the thickest arrows. The grey circles indicated the regions of origin of modern humans. Figure adapted from Reed and A Tishkoff, 2007

Afroasiatic languages are mainly spoken by agro-pastoralist and agriculturalist populations in northern and eastern Africa. Nilo-Saharan languages are spoken mainly by pastoralists in central and eastern Africa. The Niger-Congo, with 1436 languages, is the largest language phylum in the world. The Bantu languages, which are a subfamily of the Niger-Congo, are a collection of around 500 closely related languages and are spoken by at least 200 million people due to the migration within the last four thousand years of Bantu-speaking people across eastern and southern sub-Saharan Africa – the Bantu expansion. Finally, the Khoisan language, characterized by click consonants, is the smallest of the four languages in Africa. Thus, populations that speak languages belonging to the same linguistic family, tend to have high levels of genetic relatedness (Reed and A Tishkoff, 2007, Campbell et al., 2014, Fan et al., 2019).

The Bantu expansion occurred about 5,000 years ago, where Bantu-speaking people left their homeland in northwest Cameroon / southern Nigeria and spread throughout sub-Saharan Africa. This expansion is described as one of the major and fastest population movement in the recent history of Africa, involving the transmission of culture, language, technology and most of the genes.

In Figure 5 is shown the distribution of spoken language families in Africa and it also represents that the expansion did not follow a single continuous migration route, but involved at least two large dispersions with different expansion centers (one in the west and one in the east), different geographic constraints and at different times. Studies of autosomal and Y-chromosomal loci have reported a relatively high level of shared variation among Western Bantu Niger-Kordofanian speakers, as well as the presence of Bantu Niger-Kordofanian descendants in many eastern and southern African populations, consistent with widespread migration in Africa. Also, in mtDNA studies it was possible to confirm this migration, therefore, some mtDNA lines were postulated as genetic footprints of Bantu expansion (Tishkoff and Verrelli, 2003, Jobling et al., 2004, Berniell-Lee et al., 2009, Tishkoff et al., 2009, Campbell et al., 2014)

In West Africa lies the most populated country of the continent, Nigeria, bordered to the north by Niger and Chad, the western by Benin and in the eastern border by Cameroon and the coastline lies on the Gulf of Guinea in the south (Figure 6) (Levy, 2004). As said, Nigeria is the most populous country in Africa: in 2016 the population was over 180 million people (thecommonwealth.org). Most Nigerians live in rural areas; however, the number of people who come to cities for a better life has increased. There are many ethnic groups and therefore different languages, but English has become the official language of the country since 1960 (Levy, 2004, Visentini, 2011).

Nigeria has a rich and multifaceted culture consisting of traditions from more than 250 ethnic groups. Of these 250 groups, based on their subpopulation size, there are three major ethnic groups, namely the Hausa, the Igbo and the Yoruba, which make up 29%, 21% and 18% of the population, respectively (Visentini, 2011).

The Hausa are located in the northern savannas, while the Yoruba are located in the southwestern part of the country and the Igbo in the southeast (Figure 6). Nigerians belong to many different religions as well. The Muslims who make up about 50% of the population are concentrated in the northern region of the country - Hausa - and Christians who make up about 40% of the population are concentrated in the southern region – Igbo and Yoruba (Falola and Heaton, 2008, Okolie et al., 2018, nigeria.gov.ng, 2019).

Figure 6. Major cities and ethnic groups in present-day Nigeria. In red boxes are highlighted the three ethnic groups of Hausa, Yoruba and Igbo. Illustration adapted from Levy, 2004

The culture in Nigeria is still influenced by its indigenous traditions and also by more recent lifestyles that were incorporated from the West. For example, the traditional reliance on extended family and kinship networks remains strong throughout Nigeria, but a growing on smaller, nuclear families and on individual achievement is perceptible, particularly in urban areas. Further, while polygamy is still a common practice in the country, monogamous marriage is also common, particularly among Christians and the educated elite (Falola and Heaton, 2008). Regarding the ethnic relations, more than 90%

of marriages were within rather than between ethnic units, or at least within identical region and language groups. Marriages between subgroups of Igbo, Yoruba, Hausa, Fulani, or Kanuri occurred without stigma and for many decades. But in the south, Yoruba - Igbo unions were uncommon, and north-south marriages were even rarer (Metz and Division, 1992).

Over the 1450-1850 period slavery became an increasingly ingrained institution in many states of Nigeria region. The Atlantic slave trade was the largest intercontinental migration in world history before the nineteenth century. In Nigeria there were three main centers for slave trade, one in the north and two in the south (Figure 7). In the north, Borno and the Hausa states had continued a trade in slaves with north Africa through the trans-Saharan trade routes, even before 1500. In the south, on the coast, the trade was in Bight of Benin in the west and the Bight of Biafra in the east (Falola and Heaton, 2008).

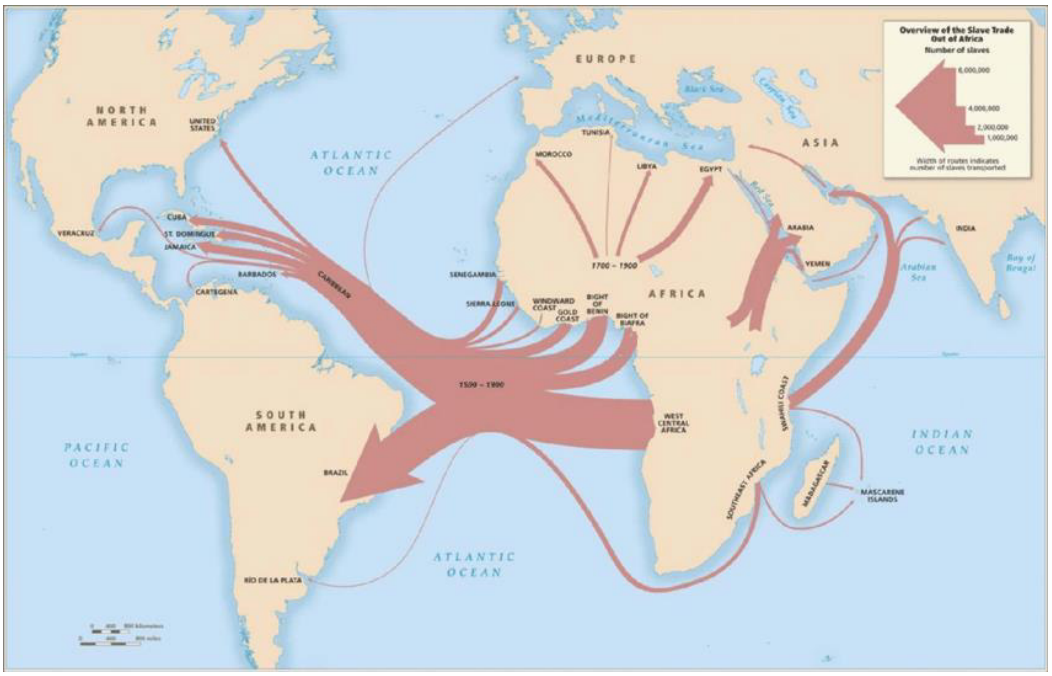


Figure 7. Overview of the slave trade out of Africa, 1500–1900. Adapted from (Busey and Cruz, 2015)

In the 19th century, 1807 to be more precisely, the Britain declared that slavery was wrong and illegal, and ceased trade with the slave dealers who had for many years benefitted from a lucrative relationship with the Britain. Slavery was slowly eradicated, and it was gradually replaced by other commodities, especially palm oil (Metz and Division, 1992, Levy, 2004). In 1914, Britain established the borders of Nigeria to form a unified colonial state, although geographical and social barriers still divided the north from the east and the west. Nigeria today is marked by the emergence of various periods of civilizations,

kingdoms, states and empires, as well as colonial rule, before the founding of the Nigerian Nation State in 1914 and its later independence in 1960 (Falola and Heaton, 2008).

Regarding the three main ethnic groups of the country, they have some differences, such as geographical, religious and linguistic, but also social, cultural and political. The Hausa community the transition girl to wife and mother may be sudden and marriage before age 16 is common (pbs.org, 2007, Odimegwu and Somefun, 2017). The cultural borders of Hausa have been constantly expanding, from a fund of centralized governments of the emirates, extended families and markets, also to modern communications, transport and the fast necessity of a lingua franca. Hausa was rapidly becoming either the first or second language of the entire northern area of the country, belonging to Afro-Asiatic language family (Metz and Division, 1992). In relation to other ethnics, the Yoruba had a strong sense of ethnic identity and of region, history and leadership among Nigeria's people. Yoruba are one of Nigeria's most urban ethnic groups. For example, Yoruba and Igbo girls tend to marry in the third decade of life. Compared to the other ethnic groups, the Igbos value Western education and they are family-oriented, possessing a strong kinship system. Unlike neighboring Hausa and Yoruba cultures, Igbo society was traditionally decentralized and non-hierarchical. For Igbo as an ethnic group, personal advancement and participation in local affairs were matters of individual initiative and skill. A council of the most respected elders of the locality ran villages. Their language tied them, historically, to regions east and south of their contemporary locations, belonging to the Niger-Kordofanian language family, like the Yoruba people (Metz and Division, 1992, pbs.org, 2007, Odimegwu and Somefun, 2017).

After the independence of the country, in 1960, the relations between the ethnic groups remained a major problem for such a large and diverse society. However, a conjunction of locations, language, religion, and common and differentiating customs created a strong sense of destiny shared among the co-ethnics and formed a constant basis for organizing ethnically related groups in political circles. Thus, when political parties emerged, they represented the Hausa, the Yoruba and the Igbo (Metz and Division, 1992).

Aims

The analyses of genetic polymorphisms, in particular short tandem repeats, have uncovered important and useful applications in many different fields with special focuses in population and forensic genetic studies. Although Africa is considered the source of all modern humans, the pattern of broad genetic diversity in geographically and ethnically diverse, African populations is largely uncharacterized. In all continent, the population shows a wide variety of languages, cultures and phenotypes, so characterizing the pattern of genetic variation among ethnically diverse African populations is fundamental for elucidating fine-scale population structure and demographic patterns in these populations. On the other hand, collecting population data in large scale is essential for the construction of reference databases, and to understand the genetic variation of specific markers prior to forensic applications. The work described in this dissertation centered on the genetic characterization of genetic markers located on the X chromosome in male samples from Nigerian ethnic groups. With this main aim underlined specific objectives were established and different points of study were investigated:

1) To amplify a decaplex system of X-STRs (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789) in one single PCR following the procedure described in Gusmão et al. (2009);

2) To genetically characterize a representative number of samples from the three ethnic groups of Nigeria, namely Hausa, Igbo and Yoruba by genotyping the 10 X-STRs using the approach described in aim 1);

3) To calculate population and forensic genetic parameters of relevance, that include:
   (a) Construction of allelic frequency databases;
   (b) Determination of gene diversity levels;
   (c) Assessment of linkage disequilibrium/association status among markers;
   (d) Calculation of genetic distances for the assessment of population genetic affinities between the three Nigerian groups and other African countries, namely, Angola, Mozambique, Algeria, Ivory Coast, Egypt, Ghana, Guinea-Bissau, Morocco, Tunisia and Uganda.

   e) Evaluation of the applicability of the selected X chromosome polymorphisms in human identification (forensic and kinship cases) by assessing discrimination capacity parameters for the different populations, in both males and females,

and the exclusion capacity parameters for specific family constellations such as father/mother/daughter trios, father/daughter (or mother/son) duos and half-sisters.

Material & Methods

Population samples

From all voluntary donors, appropriate informed consent and information about the birthplace of all their known ancestors, up to the third generation, were obtained in personal inquiries under strictly confidential circumstances. The sample set of Nigeria in this study is divided into three ethnic groups: Hausa (n= 90, Afroasiatic language family), Igbo and Yoruba (n= 96 and n=44, respectively, both groups are Niger-Congo speakers). All samples were obtained from unrelated male individuals.

DNA extraction

Total DNA was obtained in this work from blood samples that had been previously extracted, collected in FTA cards, by Phenol-chloroform and/or Chelex methods, in the DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro - UERJ.

FTA paper is an absorbent cellulose-based paper that contains four chemical substances to protect DNA molecules from nuclease degradation and preserve the paper from bacterial growth (Burgoyne and Hieftje, 1996). As a result, DNA on FTA paper is stable at room temperature over a period of several years. The use of FTA paper simply involves adding a spot of blood to the paper and allowing the stain to dry (Butler, 2011).

Phenol chloroform extraction involves the serial addition of several chemicals. First sodium dodecyl sulfate (SDS) and proteinase K are added to break open the cell membranes and to break down the proteins that protect the DNA molecules while they are in the chromosomes. Next a phenol/chloroform mixture is added to separate the proteins from the DNA into hydrophobic and aqueous phases. Finally, ethanol precipitation passing through a membrane filter concentrates the DNA in the sample. While the organic extraction method works well for recovery of high molecular weight DNA, it is time-consuming, involves the use of hazardous chemicals, and requires the sample to be transferred between multiple tubes, which increases the risk of error or contamination (Butler, 2011, Barnett and Larson, 2012).

An alternative and inexpensive procedure for DNA extraction, that has become popular among forensic scientists, is the use of a chelating-resin suspension that can be added directly to the sample (e.g., blood, bloodstain, or semen). Introduced in 1991 to the forensic DNA community, Chelex 100 is an ion-exchange resin that is added as a suspension to the samples (Walsh et al., 1991). The Chelex method of DNA extraction is more rapid than the organic extraction method and involves fewer steps and thus fewer

opportunities for sample contamination. Chelex denatures double-stranded DNA and yields single-stranded DNA from the extraction process. Thus, it can only be followed by PCR-based analyses but is an advantage for PCR-based typing methods because it removes inhibitors of PCR and uses only a single tube for the DNA extraction, which reduces the potential for laboratory-induced contamination. The addition of too much whole blood or too large of a bloodstain to the Chelex extraction solution can result in some PCR inhibition (Butler, 2011).

Amplification of X-STRs

The PCR amplification of the X chromosomal STR markers was performed in a multiplex system, amplifying in one single-PCR reaction the 10 X-STR loci represented in Figure 8: DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789. The features of the primers of the respective loci are in Table 2. The mix of reagents (reaction mix) and thermocycling conditions used are described in Table 3 and 4, respectively. The reaction was carried out in a final volume of 5 µL, the amount of input H_2O and DNA differ as shown in Table 3, depending on the quantity of DNA of the sample. Quantification values were not available, however, when low peaks and/or allelic dropout were observed PCR was repeated by increasing the input of DNA quantity. In each PCR run, a negative and positive control sample were included. In negative controls, the template DNA was substituted with water and were used in order to screen for any contamination that could occur. As positive controls, samples with a known genetic profile were used, namely the reference DNA 9948 (Figure 9) (Promega, Madison, WI, USA; originally established by Coriell Institute as NA9948) (Szibor et al., 2003a), which genetic profile is represented in Table 5, and an internal DNA sample (2089), which X-STR profile was established by using fragment sizing and the reference 9948 (Figure 9 and Table 5).

Figure 8. Multiplex used for the 10 X-STR, with refence to the interval sizes expected for each marker in base pairs (bp). Also, are represented the fluorescent colors attributed to each marker (blue: 6-FAM; green: VIC; yellow: NED)

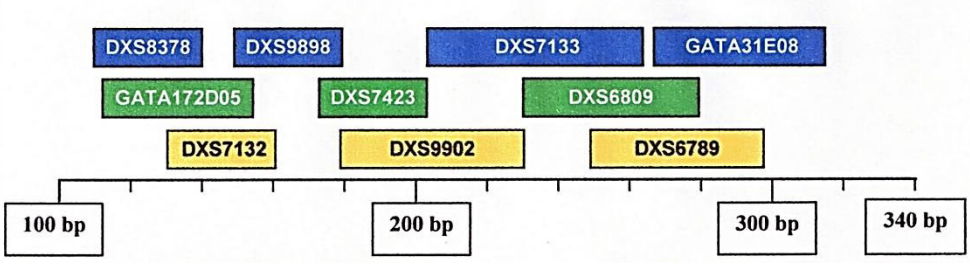


Table 2. Primer sequences and dye labels, as describe in Gusmão et al. (2009)

Locus	Primer sequence (5'-3')	
	Forward primer	Reverse primer
DXS8378	**6-FAM**_TTAGGCAACCCGGTGGTCC	ACAAGAACGAAACTCCAACTC
DXS9898	**6-FAM**_CGAGCACACCTACAAAAGCTG	TAGGCTCACCTCACTGAGCA
DXS7133	**6-FAM**_CACTTCCAAAAGGGGAAAAA	ACTTGTACTTGGTGGGAGGAA
GATA31E08	**6-FAM**_GCAAGGGGAGAAGGCTAGAA	TCAGCTGACAGAGCACAGAGA
GATA172D05	**VIC**_TAGTGGTGATGGTTGCACAG	ATAATTGAAAGCCCGGATTC
DXS7423	**VIC**_GTCTTCCTGTCATCTCCCAAC	TAGCTTAGCGCCTGGCACATA
DXS6809	**VIC**_TCCATCTTTCTCTGAACCTTCC	TGCTTTAGGCTGATGTGAGG
DXS7132	**NED**_TCCCCTCTCATCTATCTGACTG	CACTCCTGGTGCCAAACTCT
DXS9902	**NED**_CTGGGTGAAGAGAAGCAGGA	GGCAATACACATTCATATCAGGA
DXS6789	**NED**_CTTCATTATGTGCTGGGGTAAA	ACCTCGTGATCATGTAAGTTGG

Table 3. Set up of the mix used in the DNA amplification in a total final volume of 5 µL using the Qiagen Multiplex PCR kit

Reaction mix	Volume per reaction (µL)
QIAGEN Multiplex PCR master mix	2.5
Primer mix	0.5
H_2O PCR-grade kit QIAGEN	1 - 1.5
Template DNA	0.5 - 1

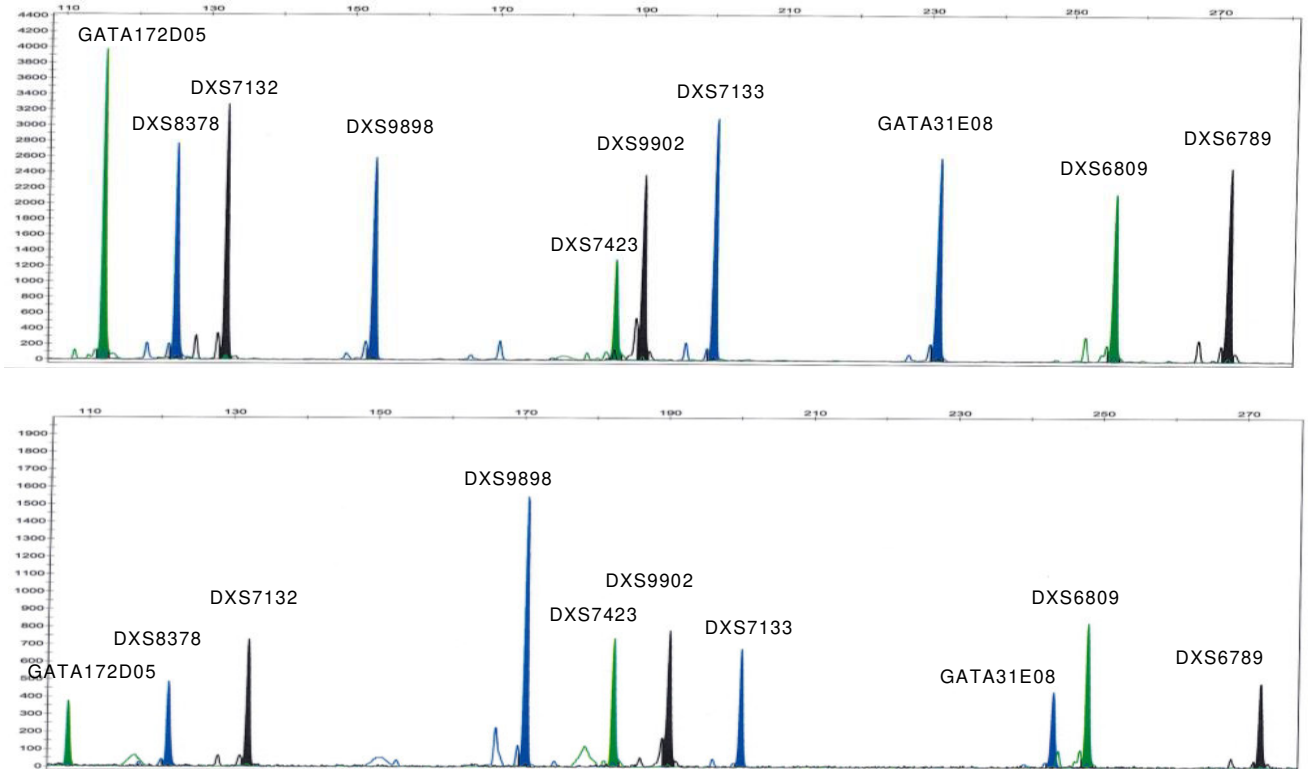
Table 4. Thermocycling conditions used for the DNA amplification

		Temperature (°C)	Time
Incubation		95	15 min
10 cycles	Denaturation	94	30 sec
	Annealing	60	90 sec
	Extension	72	60 sec
20 cycles	Denaturation	94	30 sec
	Annealing	58	90 sec
	Extension	72	60 sec
Final extension		60	75 min
Hold		12	∞

Table 5. X-STR profiles of standard DNA samples used as references in this work

	2089	9948
DXS8378	12	11
DXS9898	8.3	13
DXS7133	11	11
GATA31E08	9	12
GATA172D05	8	6
DXS7423	15	14
DXS6809	33	31
DXS7132	13	13
DXS9902	13	13
DXS6789	20	20

Figure 9. Electropherogram obtained by the ABI PRISM 3500 Genetic Analyzer of the control samples 2089 (above) and 9948 (below) for the X-STR decaplex



Detection and genotyping of PCR products

Polyacrylamide gel electrophoresis (PAGE)

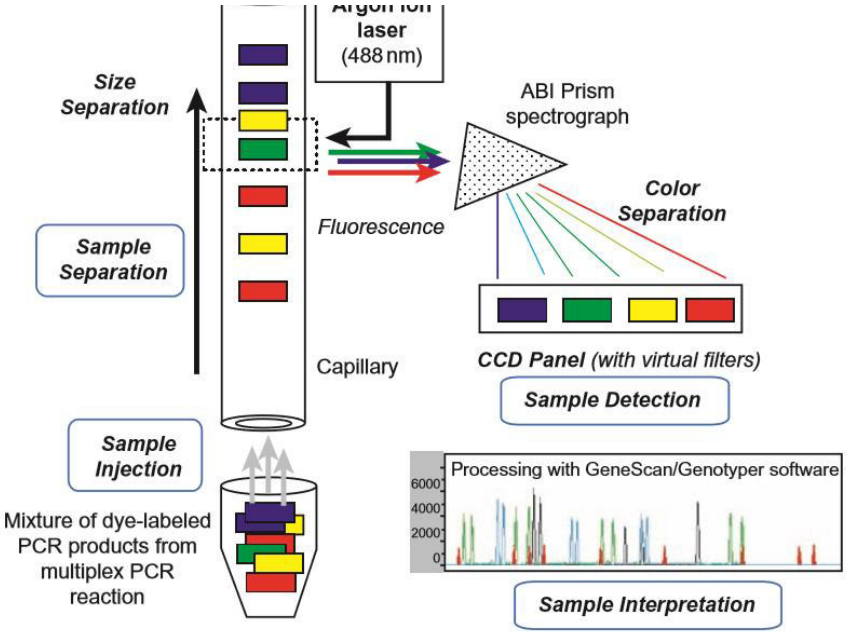
When needed, in particular prior to sequencing, some amplified samples were pre-examined by polyacrylamide gel electrophoresis (PAGE) in order to confirm the success of PCR amplification. The PCR amplified STR alleles, ranged in size from ~100 bp to ~400 bp, are better separated with polyacrylamide gels than with agarose gels, since the polyacrylamide gels have much smaller pore size (~100 - 200 Å) than agarose gels (~1500 - 2000 Å). Therefore, the PCR products were separated by horizontal polyacrylamide gel (acrylamide:bisacrylamide 19:1) electrophoresis using a Multiphor II Electrophoresis System (GE Healthcare), with MultiTemp III Thermostatic Circulatior (Amersham Biosciences) and Consort EV243 power supply. The electrophoresis run was performed at 180V with 100bp ZR DNA Ladder (Zymo Research). Thereafter, the results were visualized by silver staining method, with the following sequential steps: 10 minutes in ethanol (10%), 5 minutes in nitric acid (1%), two washings with distilled water, 20 minutes in silver nitrate, two washings with distilled water, DNA visualization with solution made with 3 g sodium carbonate (0.28M) + 1 ml formaldehyde (0,02%) + 100 ml water, reaction termination with acid acetic (10%) and final washing with distilled water.

Capillary electrophoresis and genotyping

The principle of capillary electrophoresis (CE) is based on the same principle of electrophoresis in “normal” gel, however the gel and sample are transported inside a capillary and the fragments are identified by a detector. Thus, separation and detection of the amplicons were done by capillary electrophoresis (Figure 10). First, using a 4-capillary electrophoresis system, the ABI PRISM 3130 Genetic Analyzer (Applied Biosystems), then later, an 8-capillary electrophoresis system, the ABI PRISM 3500 Genetic Analyzer (Applied Biosystems).

To each PCR products obtained, 9.5 µl of Hi-Di formamide (Applied Biosystems) and 1.5 µl of internal size standard GS500 LIZ (Applied Biosystems) were added. Fragment sizes were automatically determined using GeneMapper v5 software (Applied Biosystems) and genotyping performed through comparison with DNA control reference samples and a panel of bins created for the decaplex, using the same software.

Figure 10. Schematic illustration of the separation and detection of STR alleles with a single capillary in an ABI 3130 or another multi-capillary instrument. Illustration modified from Butler, 2011



Sequencing

In order to confirm the accuracy of allelic assignment of some markers, some samples were sequenced. The selection criterion included potentially rare, non-consensus and/or off-ladder alleles and, in addition, standard DNA 2089 was also sequenced to support genotype data. The chosen markers for sequencing as well as and the primer sequences used are summarized in Table 6.

Prior to the sequencing reaction, the samples were amplified in singleplex and PCR amplification was confirmed through polyacrylamide gel electrophoresis. Post-PCR enzymatic purification consisted in adding to each 2 µl PCR product, 1 µl of ExoSAP-IT (Applied Biosystems), and then the solution was submitted to the thermocycling conditions described in Table 7, first the enzyme will remove the primers and dNTPs that weren't used, and then the inactivation of the enzyme. After enzymatic purification, sequencing reaction was performed using ABI BigDye Terminator v3.1 (Applied Biosystems). The mix of reagents (reaction mix) and thermocycling conditions used are described in Table 8 and 9, respectively.

Table 6. Primer sequences used for sequencing according to (Gomes et al., 2009a, Gomes et al., 2009c)

Locus	Primer sequence (5'-3')	
	Forward primer	Reverse primer
DXS9898	CGAGCACACCTACAAAAGCTG	TGCTCAGTGAGGTGAGCCTA
DXS6809	TCCATCTTTCTCTGAACCTTCC	CCTCACATCAGCCTAAAGCA
DXS7132	TCCCCTCTCATCTATCTGACTGT	AGAGTTTGGCACCAGGAGTG

Table 7. Thermocycling conditions used for the purification step

Temperature (°C)	Time
37	15 min
85	15 min

Table 8. Set up of the mix used in the actual sequencing reaction

Reaction mix	Volume per reaction (µL)	
	Example Forward	Example Reverse
BigDye™ Terminator 3.1 Ready Reaction Mix	1	
BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer	1	
Forward primer	0.5	-
Reverse primer	-	0.5
Deionized water	1.5 - 2	
Template DNA	0.5 - 1	

Table 9. Thermocycling conditions for the sequencing

		Temperature (°C)	Time
Incubation		96	1 min
25 - 35 cycles	Denaturation	96	10 sec
	Annealing	55	5 sec
	Extension	60	2 min
Hold		4	∞

Final sequenced products were purified with 5% Sephadex using in-house filtration columns, following the protocol:

-Put columns in 2 mL collection microtubes;
-Put 750 µl of Sephadex in the columns;
-Centrifuge at 4400 rpm for 4 minutes;
-Transfer the columns to 1.5 mL eppendorf's;

-Pipette the total amplification product into the center of the column;

-Centrifuge at 4400 rpm for 4 minutes;

-Discard the column and store the eppendorf containing the purified sample.

Finally, 10 µl of Hi-Di formamide (Applied Biosystems) were added to each purified sample that were then separated and detected on the ABI PRISM 3500 Analyzer Electrophoresis System (Applied Biosystems) and the results were analyzed using the software Sequencing Analysis (Applied Biosystems).

All this process is based on the traditional Sanger method for DNA sequencing, which was first described over 40 years ago (Sanger et al., 1977). In theory, the process involves the polymerase incorporation of dideoxyribonucleotide triphosphates (ddNTPs) as chain terminators followed by a separation step capable of single nucleotide resolution. There is no hydroxyl group at the 3'-end of the DNA nucleotide with a ddNTP and therefore chain growth terminates when the polymerase incorporates a ddNTP into the synthesized strand. Extendable dNTPs and ddNTP terminators are both present in the reaction mix so that some portions of the DNA molecules are extended. At the end of the sequencing reaction, a series of molecules are present that differ by one base from one another (Butler, 2011).

Statistical analysis

The information from the genotyping of STR markers was analyzed with Arlequin v.3.5.2 software (EXCOFFIER and LISCHER, 2010) in order to estimate statistical parameters, such as allele frequencies, gene diversities (GD), exact test of population differentiation based on population pairwise genetic distances, F_{ST} (for intra and inter-population variation), analysis of molecular variance (AMOVA) and pairwise exact test of linkage disequilibrium (LD). Statistics for forensic efficiency evaluation of each locus, namely power of exclusion in trios involving daughters (PE_T) as well as in father/daughter duos (PE_D), and power of discrimination in females (PD_F) and in males (PD_M) were calculated using the formulae according to Desmarais et al. (1998). In addition, power of exclusion in half-sisters (PE_{HS}) were calculated using the formula described in Gomes et al. (2009b).

Results and Discussion

To genetically characterize the three ethnic groups of Nigeria, ten X-STR markers (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789) were used. These X-linked STR polymorphisms were successfully amplified in one single-PCR multiplex, following the above reported conditions (see Material and Methods). Despite the need for some protocol adjustments and a second amplification on a small number of samples, a set of 230 well-amplified samples were genotyped. After genotyping the ten X-STR markers, a statistical analysis of the genetic data obtained proceeded, and is presented in this chapter.

X-STRs alleles: genetic characterizarion of Nigerian population

Genetic variation

Ten X-chromosome-specific STR markers were used to genetically characterize the Nigerian population. The genetic characterization of this population is necessary to infer the usefulness of these markers for forensic analysis, namely for human identification cases in Nigeria. This requires the construction of an allelic frequency database, gene diversity determination, linkage disequilibrium analyses between markers, and forensic parameter estimations such as Power of Exclusion (PE) and Power of Discrimination (PD).

The complete genotypes obtained for the 10 X-STR markers in the 230 samples from the three Nigerian population groups are presented in Appendix A. Some rare (new) and null alleles, not yet described in the literature, were found throughout the genotyping of the samples. These unusual alleles were sequenced and will be presented later. Allele frequencies and gene diversities were estimated for each locus in each ethnic group and are included in Table 10.

Table 10. Allele frequencies and gene diversities (GD) for the 10 X-STRs of the three ethnic groups of Nigeria: Igbo n=96, Hausa n=90 and Yoruba n=44

DXS8378			
	Igbo	**Hausa**	**Yoruba**
0	-	0.011	0.023
10	0.250	0.244	0.273
11	0.469	0.378	0.250
12	0.250	0.300	0.409
13	0.031	0.056	0.023
14	-	0.011	0.023
GD	**0.661**	**0.706**	**0.697**

DXS9898			
	Igbo	**Hausa**	**Yoruba**
7	0.021	-	0.023
8	-	0.011	0.023
8.3	0.031	0.033	0.023
9	0.010	0.011	-
10	0.146	0.111	0.182
11	0.188	0.111	0.205
12	0.354	0.400	0.386
12.3	0.010	-	-
13	0.115	0.200	0.091
13.3	0.052	0.089	-
14	0.063	0.033	0.068
14.3	0.010	-	-
GD	**0.805**	**0.774**	**0.779**
DXS7133			
	Igbo	**Hausa**	**Yoruba**
8	0.010	0.022	0.023
9	0.021	0.056	0.045
10	0.146	0.222	0.182
11	0.677	0.589	0.591
12	0.115	0.078	0.136
13	0.021	0.033	0.023
14	0.010	-	-
GD	**0.511**	**0.600**	**0.610**
GATA31E08			
	Igbo	**Hausa**	**Yoruba**
7	0.031	0.033	0.023
8	0.010	0.022	0.023
9	0.177	0.222	0.091
10	0.135	0.133	0.227
11	0.010	0.067	0.045
12	0.396	0.244	0.318
13	0.177	0.200	0.227
14	0.052	0.078	0.045
15	0.010	-	-
GD	**0.766**	**0.830**	**0.800**
GATA172D05			
	Igbo	**Hausa**	**Yoruba**
6	0.250	0.300	0.295
7	0.031	0.067	0.159
8	0.229	0.144	0.045
9	0.323	0.367	0.364
10	0.073	0.078	0.091
11	0.083	0.011	0.045
12	0.010	0.033	-
GD	**0.775**	**0.751**	**0.760**
DXS7423			
	Igbo	**Hausa**	**Yoruba**
8	0.010	-	-
12	0.031	0.022	-
13	0.063	0.078	0.136
14	0.552	0.511	0.409
15	0.250	0.289	0.341
16	0.083	0.056	0.091
17	0.010	0.044	0.023
GD	**0.627**	**0.651**	**0.705**
DXS6809			
	Igbo	**Hausa**	**Yoruba**
27	0.010	-	-
28	0.010	-	-

29	0.052	-	-
30	0.177	0.111	0.068
31	0.094	0.067	0.227
31.1	-	-	0.045
32	0.125	0.200	0.114
32.1	-	0.011	-
33	0.188	0.267	0.159
34	0.188	0.211	0.227
35	0.104	0.089	0.136
36	0.021	0.033	-
37	0.021	-	0.023
38	0.010	0.011	-
GD	**0.868**	**0.827**	**0.852**
DXS7132			
	Igbo	**Hausa**	**Yoruba**
0	-	-	0.023
10	-	0.011	-
11	0.031	0.022	0.023
12	0.073	0.144	0.091
13	0.281	0.278	0.205
14	0.417	0.322	0.273
15	0.156	0.133	0.250
16	0.031	0.078	0.114
17	-	0.011	0.023
18	0.010	-	-
GD	**0.723**	**0.782**	**0.808**
DXS9902			
	Igbo	**Hausa**	**Yoruba**
0	-	-	0.023
8	-	0.022	-
9	0.083	0.067	0.023
10	0.073	0.067	0.091
11	0.365	0.322	0.295
12	0.313	0.400	0.341
13	0.146	0.122	0.136
14	0.021	-	0.068
15	-	-	0.023
GD	**0.743**	**0.720**	**0.771**
DXS6789			
	Igbo	**Hausa**	**Yoruba**
0	-	-	0.023
15	0.281	0.289	0.227
16	0.083	0.111	0.091
17	-	0.011	0.023
18	0.031	-	0.023
19	0.073	0.078	0.068
20	0.198	0.189	0.182
21	0.146	0.244	0.273
22	0.167	0.067	0.045
23	0.021	0.011	0.023
24	-	-	0.023
GD	**0.828**	**0.807**	**0.835**

"0"= null alleles (silent alleles)

As for the genetic diversity of the X-STR markers in the three ethnic groups of Nigeria, the least polymorphic marker was DXS7133, with an average diversity of 0.574 across the three groups. The highest genetic diversity was observed for the DXS6809 locus, with an average diversity of 0.850, being, therefore, the most discriminating marker in

the Nigerian ethnic groups. Furthermore, all the selected X-STRs in the present Nigerian population sample (excluding the DXS7133, the less polymorphic marker) have high genetic diversities ranging from 63% (DXS7423) to 87% (DXS6809), similar to what was observed in other African populations (e.g., (Gomes et al., 2009b, Laouina et al., 2013, Messoussi et al., 2019)). Their high genetic variability shows that these X chromosome markers are highly informative and, therefore, useful in forensic applications such as individual identification studies and kinship tests.

Detection of uncommon rare (new) and null (silent) alleles

During genotyping of samples from the three ethnic groups of Nigeria, unusual alleles appeared in some markers as well as some inconsistencies such as “out of bin” or “out of marker range” or even allele dropouts in the electropherograms that resulted from the capillary electrophoresis runs. These markers will get some attention during this subchapter, as well as the alleles that undergone sequencing to clarify some of the doubts that appeared during genotyping of samples.

DXS8378

A loss of an allele at locus DXS8378 (silent allele) was observed in one sample in the Nigerian population (Yoruba, sample A145), after repetition of multiplex and singleplex amplifications. The presence of polymorphisms in the primer annealing site of STR loci, can lead to serious problems, such as false exclusions in paternity tests (Gusmão et al., 1996, Dakin and Avise, 2004, Robino et al., 2008). Therefore, several attempts were performed to sequence the sample with another in-house existing reverse primer; however, no acceptable results were obtained that could allow sequencing, as several, intense, unspecific bands were observed in the polyacrylamide gels. Therefore, a possible solution to detect the mutation responsible for the allele dropout (potentially located at a primer-binding site) would be to re-design a new set of forward and reverse primers away from the previous flanking regions for PCR amplification and sequencing. However, unfortunately, DNA sample quantity was already scarce and after the several experiment attempts, no sample was left available.

DXS9898

The tetranucleotide DXS9898 has been genetically characterized in different populations exhibiting a simple structure with non-consensus repeat. In this work, two alleles not previously described were found in the DXS9898 locus with a size corresponding to 8 and 12.3 repeats. The non-consensual (12.3) allele was sequenced, and the result is

represented in Table 11. According to our knowledge, it has not been reported and the sequence results confirmed the presence of an additional trinucleotide (ATC) (Table 11). The non-consensus nature of this STR is further supported by an 8.3 allele that is present in high frequencies in most of the studied populations (also sequenced in this study for reference sample 2089, Table 11). Regarding allele 8, which according to our literature research has also not been reported until date, was genotyped twice throughout the work in the different ethnic groups of Hausa and Yoruba. However, also the absence of the required amount of sample, did not allow for sequencing of the respective samples to confirm the size of the amplicon and identify of the allele structure.

Table 11. STR and allele sequence structure for DXS9898. Allele structure variation obtained in two samples according to the adopted allele nomenclature (Hering and Szibor, 2000). In the allele structure, lower case **atc** is not considered for allele nomenclature. In STR sequence structure, uppercase bold letters represent primer hybridization sites

cGAGCACACCTACAAAAGCTGagatatatacctctg $(TATC)_2$ atc$(TATC)_{1-5}$ $(ATC)_{0-1}$ $(TATC)_{5-13}$ tctatttatcaatctattgatctgtctatctggaTGCTCAGTGAGGTGAGCCTA			
Allele	**Allele structure: $(TATC)_2$ atc$(TATC)_{1-5}$ $(ATC)_{0-1}$ $(TATC)_{5-13}$**	**n**	**Sequenced in**
8.3	**$(TATC)_2$ atc(TATC) (ATC) $(TATC)_5$**	1	Reference 2089
12.3	**$(TATC)_2$ atc$(TATC)_5$ (ATC) $(TATC)_5$**	1	Igbo

DXS6809

The locus DXS6809 has a complex and compound allele repeat structure composed of three different motifs (CTAT)n (ATCT)3 N9 (TATC)n (ATCT)n N10 (ATCT)n. In the present work, this complex and compound repeat sequence structure was also observed in all three ethnic groups. All the intermediate alleles were sequenced (Table 19) and all revealed a thymine insertion -T- in the repeat region (31.1 and 32.1). In addition, allele 31.1 presented a T->C transition mutation at the 3rd base nucleotide at the tenth-base nucleotide stretch attatctatc (that is not considered for allele nomenclature by the described structure) and does not interfere with fragment size. In the study by Gomes and collaborators (Gomes et al., 2009a) all the intermediate alleles reported at DXS6809 were also from African populations, which is supported as well by the findings in this present work. In this locus two other substitution mutations, in different samples, at different sites, were found: 1) allele 32 exhibits a C->A transversion mutation at the 1st base nucleotide of the CTAT 3rd repeat motif; 2) in allele 37 a T->C transition mutation at the 7th base nucleotide, downstream from the ATCT last repeat motif, at the flanking region was found (Table 12).

Table 12. STR and allele sequence structure for DXS6809. Allele structure variation obtained in 9 samples according to the adopted allele nomenclature (Edelmann et al., 2003). In the allele structure, italic bold black letters represent the insertion mutation and the italic underline bold letters represent the substitution mutation. In STR sequence structure, uppercase bold letters represent primer hybridization sites. N9 – atcatctat, N10 – attatctatc

TCCATcTTTCTCTGAACCTTCCtagctcaggaatactgagggcatgactagattatgtaggaatttggg (CTAT)$_n$ (ATCT)$_3$ N$_9$ (TATC)$_n$ (ATCT)$_n$ N$_{10}$ (ATCT)$_n$ atcctctatctcttCCTCACATCAGCCTAAAGCA			
Allele	**Allele structure: (CTAT)$_n$ (ATCT)$_3$ N$_9$ (TATC)$_n$ (ATCT)$_n$ N$_{10}$ (ATCT)$_n$**	**n**	**Sequenced in**
31.1	(CTAT)$_8$ (ATCT)$_3$ N$_9$ (TATC)$_4$ (ATCT)$_6$ at***c***atctatc ***T*** (ATCT)$_{10}$	1	Yoruba
	(CTAT)$_9$ (ATCT)$_3$ N$_9$ (TATC)$_6$ ***T*** N$_{10}$ (ATCT)$_{13}$	1	Yoruba
32	(CTAT)$_2$ ***A***TAT (CTAT)$_3$ (ATCT)$_3$ N$_9$ (TATC)$_4$ (ATCT)$_5$ N$_{10}$ (ATCT)$_{14}$	1	Hausa
32	(CTAT)$_{11}$ (ATCT)$_3$ N$_9$ (TATC)$_3$ (ATCT)$_5$ N$_{10}$ (ATCT)$_{10}$	1	Hausa
32.1	(CTAT)$_{10}$ (ATCT)$_3$ N$_9$ (TATC)$_6$ ***T*** N$_{10}$ (ATCT)$_{13}$	1	Hausa
33	(CTAT)$_9$ (ATCT)$_3$ N$_9$ (TATC)$_3$ (ATCT)$_5$ N$_{10}$ (ATCT)$_{13}$	2	Reference 2089, Igbo
37	(CTAT)$_{10}$ (ATCT)$_3$ N$_9$ (TATC)$_3$ (ATCT)$_7$ N$_{10}$ (ATCT)$_{14}$ ATCCTC***C***	1	Igbo
38	(CTAT)$_{10}$ (ATCT)$_3$ N$_9$ (TATC)$_3$ (ATCT)$_5$ N$_{10}$ (ATCT)$_{17}$	1	Igbo

DXS7132

Two samples were sequenced, as exhibited in Table 13, to confirm the genotype data attributed to fragment sizing since there were some uncertainties.

Table 13. STR and allele sequence structure for DXS7132. Allele structure variation obtained in 2 samples according to the adopted allele nomenclature (Edelman et al., 2001). In STR sequence structure, uppercase bold letters represent primer hybridization sites

TCCCCTCTCATCTATCTGACTgT(CTAT)$_{9,11\text{-}17}$(CAT)$_{0\text{-}1}$(CTAT)$_{0\text{-}2}$ cctattggttctgtttctctggagaacgttgactaatAGAGTTTGGCACCAGGAGTG			
Allele	**Allele structure: (CTAT)$_{9,11\text{-}17}$ (CAT)$_{0\text{-}1}$ (CTAT)$_{0\text{-}2}$**	**n**	**Sequenced in**
13	(CTAT)$_{13}$	1	Reference 2089
18	(CTAT)$_{18}$	1	Igbo

In this locus, anomalous size mobilities were noticed that initially were thought to be a high number of off-ladder alleles equivalent to 15.1 and 16.1 alleles, which could lead to genotyping errors. A similar case was reported in a collaborative study by Gusmão et al. (2009), which was detected as a sizing problem when using GeneScan-500 LIZ size standard for fragments in the range of 139–150 bp. The solution to this problem was to

use GeneScan-600 LIZ to detect the real differences between alleles. Therefore, in this work we used the same procedure. Samples containing the unusual alleles and the control DNA sample used (2089) were re-run with the two different size standards, GeneScan-500 LIZ and GeneScan-600 LIZ (Applied Biosystems). From the results, it was possible to compare the sizes between the alleles (Table 14) and confirm genotypes. As expected, it was possible to verify that when using the GeneScan-500 LIZ the size difference is overestimated. Allele calling was confirmed in each sample at this locus. Caution must be taken when using GeneScan-500 LIZ. However, the best approach is to use a different internal size standard such as the GeneScan-600 LIZ to avoid wrong fragment sizing.

Table 14. Comparative results obtained when using GeneScan-500 LIZ and GeneScan-600 LIZ size standards (Applied Biosystems) in fragment size analysis of the DXS7132. ESD: Expected size differences between consecutive alleles, OSD: observed size differences between consecutive alleles

		LIZ500		LIZ600	
Alleles	**ESD**	**SIZE**	**OSD**	**SIZE**	**OSD**
13	4.00	132.26	4.07	130.70	3.90
14	4.00	136.33	4.53	134.60	4.06
15	4.00	140.86	4.75	138.66	4.10
16	4.00	145.61	4.06	142.76	3.85
17	4.00	149.67	5.13	146.61	4.02
18		154.80		150.63	

DXS9902

During the analysis it was found that at locus DXS9902, allele 15 was genotyped only once in 230 samples, belonging to a Yoruba individual. This allele has been described in several other populations, mainly in South American populations, however, always with low frequency (Gusmão et al., 2009) therefore it was not sequenced.

The increased attention given to X-STRs in forensics and population genetics strengthens the importance of a well-established nomenclature. In this work, we are towards the presence of new, rare and null alleles as well as anomalous mobility of sizes. All these characteristics need to be well established and clarified, so that there is a unification allowing the comparison of data from different laboratories and to minimize adverse consequences in forensic situations.

Linkage disequilibrium

Linkage disequilibrium (LD) is the allelic dependency between alleles at different loci, that can be the result of physical proximity between loci or can be due to population substructure (Schaffner, 2004). As such LD is an important aspect that should be taken into account in both population and forensic genetic studies, because if markers are linked they cannot be treated independently, but should be analyzed together (as haplotypes). Since this study comprehended the analysis of markers located in the same chromosome this is even more relevant. Therefore, especial attention needs to be paid to the level of linkage disequilibrium (LD) between markers.

In this case, the study comprised 10 X chromosome STRs and LD tests were performed for the presence of significant associations between pairs of loci, based on an exact test totalizing 45 comparisons. The significance level assumed was $p \leq 0.0011$, which was obtained after applying the Bonferroni's correction for multiple testing ($p \leq 0.05/45$). Results showed no significant association between alleles of any pair of loci (Table 15). This lack of LD may be due to a longer evolutionary history of Sub-Saharan populations that allowed divergence among these loci due to recombination (Tishkoff and Williams, 2002). In addition, it could also be due to the low power of the exact test of LD to detect allelic associations in small size population samples, as demonstrated by Kling et al. (2015). In this latter study, authors demonstrated that the power of estimates for an exact test of linkage disequilibrium increased when the number of profiles tested also increased. This could also be the case of the present sample set, as the total number of individuals studied in each ethnic group was quite small n<100 (Igbo n= 96, Hausa n=90 and Yoruba n=44).

In summary, linkage disequilibrium tests did not reveal consistent evidence of strong associations between the X-STR markers studied in the Nigerian population, and therefore these can be treated as independent loci for DNA testing and analysis.

Table 15. Test of linkage disequilibrium in the 230 male samples. After Bonferroni's correction (0.05/45), *p*-values lower than 0.0011 were considered statistically significant

Markers	*p*-value ± s.e.
(DXS8378, DXS9898)	*p*=0.361584 ± 0.00199256
(DXS8378, DXS7133)	*p*=0.0558416 ± 0.00207551
(DXS9898, DXS7133)	*p*=0.341782 ± 0.0024272
(DXS8378, GATA31E08)	*p*=0.703564 ± 0.0025009
(DXS9898, GATA31E08)	*p*=0.481584 ± 0.00123903
(DXS7133, GATA31E08)	*p*=0.417822 ± 0.00253308
(DXS8378, GATA172D05)	*p*=0.332079 ± 0.00354737
(DXS9898, GATA172D05)	*p*=0.958317 ± 0.00106521
(DXS7133, GATA172D05)	*p*=0.580495 ± 0.00198846
(GATA31E08, GATA172D05)	*p*=0.299802 ± 0.0026815
(DXS8378, DXS7423)	*p*=0.0613861 ± 0.00174069
(DXS9898, DXS7423)	*p*=0.389505 ± 0.00312216
(DXS7133, DXS7423)	*p*=0.795149 ± 0.00146717
(GATA31E08, DXS7423)	*p*=0.972673 ± 0.00117841
(GATA172D05, DXS7423)	*p*=0.548812 ± 0.00272608
(DXS8378, DXS6809)	*p*=0.603168 ± 0.00164524
(DXS9898, DXS6809)	*p*=0.226238 ± 0.00135985
(DXS7133, DXS6809)	*p*=0.0613861 ± 0.00106246
(GATA31E08, DXS6809)	*p*=0.391188 ± 0.0029869
(GATA172D05, DXS6809)	*p*=0.0942574 ± 0.00168567
(DXS7423, DXS6809)	*p*=0.731188 ± 0.00234796
(DXS8378, DXS7132)	*p*=0.934653 ± 0.00191925
(DXS9898, DXS7132)	*p*=0.125743 ± 0.00223618
(DXS7133, DXS7132)	*p*=0.558416 ± 0.00251739
(GATA31E08, DXS7132)	*p*=0.396139 ± 0.00201733
(GATA172D05, DXS7132)	*p*=0.0538614 ± 0.000882435
(DXS7423, DXS7132)	*p*=0.617228 ± 0.00177572
(DXS6809, DXS7132)	*p*=0.409604 ± 0.00208092
(DXS8378, DXS9902)	*p*=0.58297 ± 0.00270063
(DXS9898, DXS9902)	*p*=0.630297 ± 0.00217658
(DXS7133, DXS9902)	*p*=0.126634 ± 0.00144404
(GATA31E08, DXS9902)	*p*=0.408218 ± 0.00240825

(GATA172D05, DXS9902)	*p*=0.776337 ± 0.00201536
(DXS7423, DXS9902)	*p*=0.170891 ± 0.00278133
(DXS6809, DXS9902)	*p*=0.0845545 ± 0.00118929
(DXS7132, DXS9902)	*p*=0.535446 ± 0.00228515
(DXS8378, DXS6789)	*p*=0.485347 ± 0.00200495
(DXS9898, DXS6789)	*p*=0.590198 ± 0.00177948
(DXS7133, DXS6789)	*p*=0.632178 ± 0.00325007
(GATA31E08, DXS6789)	*p*=0.851089 ± 0.00334775
(GATA172D05, DXS6789)	*p*=0.620594 ± 0.00264731
(DXS7423, DXS6789)	*p*=0.768812 ± 0.00126164
(DXS6809, DXS6789)	*p*=0.613168 ± 0.00214466
(DXS7132, DXS6789)	*p*=0.10901 ± 0.0012859
(DXS9902, DXS6789)	*p*=0.89099 ± 0.0019372

Nigerian population comparisons

The allelic database to be constructed depends on its population structure, i.e., should a single allelic database comprise the three ethnic groups of Nigeria or are three different databases needed? In the absence of substructure, it is possible to consider a single database for this population. On the contrary, the presence of substructure determines the need to use specific population databases for each subpopulation group (Hausa, Igbo and Yoruba). Thus, following the genetic characterizing of the ten X-STR markers in each of the three ethnic groups of Nigeria, an analysis of its population structure was performed, which included comparisons between samples from each ethnicity and analysis of the distribution of the genetic variance (AMOVA) in the population.

After characterizing the 10 X-STR markers in all samples, an analysis of the genetic structure of the population of Nigeria was made in order to determine the need to use specific population databases for each ethnic subpopulation or a single database for the three groups. Thus, to evaluate the population structure of Nigeria, population comparisons were made between each ethnic group: considering all markers together (haplotype level) and at the single locus level. Comparative analyses were performed based on F_{ST} genetic distance measures. The F_{ST} values are calculated based on the number of different alleles between haplotypes, in contrast to the R_{ST} genetic distance that takes into account the *Stepwise Mutational Model and incorporates a measure of the evolutionary distance between alleles - estimated from size differences. Our choice for using F_{ST} rather than R_{ST} genetic distance was supported by the study of Caglia et al. (2003) on genetic distances in Sub-Saharan African populations. Based on the results of their study, they concluded that F_{ST} had a better performance in an African context, because of the important role that genetic drift had in shaping relationships among African populations (rather than mutation as the principal cause of population differentiation).

**Stepwise Mutational Model: The Stepwise Mutation Model (SMM) is the most consensual model in microsatellite evolution. It assumes that the length of a microsatellite changes by one or a few repetitive units at each mutational event.*

Genetic distances (F_{ST}s)

Genetic distances were calculated between the three groups and are presented in Table 16.

Table 16. Genetic distances (F_{ST}) between the three ethnic groups of Nigeria for the 10 X-STRs. After Bonferroni's correction (0.05/3), *p*-values lower than 0.017 were considered statistically significant

	F_{ST}	*p*-value ± s.e.
Igbo *vs.* Hausa	0.00096	0.32650± 0.0045
Igbo *vs.* Yoruba	0.00620	0.06544± 0.0023
Hausa *vs.* Yoruba	-0.00070	0.52817± 0.0052

The results obtained showed small F_{ST} genetic distances in the three pairwise comparisons and all *p*-values were higher than 0.06544, indicating that there are no statistically significant differences ($F_{ST} \leq$ 0.00620; $p \geq$ 0.06544). This result supports homogeneity among the groups of Hausa, Yoruba and Igbo for the studied 10 X-chromosomal markers. These results are rather unexpected since Hausa belong to Afroasiatic language family while Igbo and Yoruba are Niger-Congo language speakers, therefore it would be expected to detect some variation between Hausa and the other two groups (Table 16).

For single locus comparisons, F_{ST} values were also relatively low and statistically non-significant ($p \geq$ 0.03346) (Table 17), after applying Bonferroni's correction for multiple testing ($p \leq$ 0.017). These results further support the haplotype level comparisons where no differentiation was detected in the present sample set with the X-STRs studied.

Table 17. Genetic distances (F_{ST}) locus by locus between the three ethnic groups of Nigeria for 10 X-STRs. After Bonferroni's correction (0.05/3), values equal and less than 0.017 were considered to be statistically significant

Marker	Igbo *vs.* Hausa		Igbo *vs.* Yoruba		Hausa *vs.* Yoruba	
	F_{ST}	*p*-value ± s.e.	F_{ST}	*p*-value ± s.e.	F_{ST}	*p*-value ± s.e.
DXS8378	-0.00268	0.48451 ± 0.0049	0.03765	0.03346 ± 0.0021	0.00511	0.27067 ± 0.0043
DXS9898	0.00153	0.31838 ± 0.0046	-0.01243	0.96218 ± 0.0020	0.00614	0.21909 ± 0.0040
DXS7133	0.00418	0.22948 ± 0.0040	-0.00734	0.60984 ± 0.0047	-0.01270	0.88377 ± 0.0032
GATA31E08	0.00772	0.13355 ± 0.0035	0.00017	0.39065 ± 0.0040	0.00370	0.28482 ± 0.0046
GATA172D05	0.01420	0.32700 ± 0.0040	0.01918	0.07098 ± 0.0025	-0.00312	0.50015± 0.0049
DXS7423	-0.00647	0.76240 ± 0.0037	0.01053	0.17830 ± 0.0035	-0.0029	0.45005 ± 0.0044
DXS6809	0.00185	0.30957 ± 0.0050	0.00586	0.21582 ± 0.0041	0.01466	0.08752 ± 0.0029
DS7132	0.00066	0.36086 ± 0.0050	0.01131	0.15107 ± 0.0034	-0.00077	0.42026 ± 0.0046
DXS9902	-0.00309	0.55410 ± 0.0051	-0.00881	0.73973 ± 0.0043	-0.00926	0.74141 ± 0.0041
DXS6789	0.00259	0.27126 ± 0.0041	0.0048	0.25799 ± 0.0044	-0.01334	0.95792 ± 0.0019

In summary, considering the locus by locus results as well as the haplotype level analyses, no significant differences were observed between the three ethnic groups of Nigeria, for all markers. These results point also to a genetic homogeneity among these groups for the 10 X-STRs studied in this work.

Analysis of Molecular Variance (AMOVA)

The distribution of genetic variability of X chromosome markers among the three ethnic groups of Nigeria was further studied by AMOVA (Table 18). When considering all X-STRs, the overall value obtained among the populations of Nigeria was low (F_{ST}= 0.0018; *p*= 0.18964). Nearly all of the existing genetic variability (99.98%) is due to differences between individuals in the same population, and only 0.02% can be attributed to differences among the population groups of Igbo, Hausa and Yoruba.

Table 18. AMOVA results for X-STRs studied in the ethnic groups of Nigeria

Groupings and source of variation	Percentage of variation
Yoruba *vs.* Hausa *vs.* Igbo	
Among populations	0.02
Within populations	99.98
Yoruba & Igbo *vs.* Hausa	
Among populations	-0.09
Within populations	100.09

AMOVA was also performed for the ethnic groups of Nigeria clustered in a different way, namely Yoruba & Igbo *vs.* Hausa. This grouping was performed because of their linguistic affinity, since both belong to the Niger-Congo language and on the other hand, the Hausa ethnicity belongs to the Afroasiatic language, as previously mentioned (Metz and Division, 1992). Supporting the results obtained before in the comparisons between the three groups (Table 16), the AMOVA results also show that most of the genetic variation is within populations and not among populations. When the Yoruba & Igbo were grouped, the variation within populations (Yoruba & Igbo *vs.* Hausa) further supports the lack of heterogeneity between the groups.

This work began by genetically characterizing the ten X-STR markers in each of the three ethnic groups of Nigeria, followed by an analysis of population genetic structure, which include comparisons between samples from the three groups using several approaches. Comparisons between the three ethnic groups on a locus by locus and as well as on a haplotype level did not indicate a sub-structuring of the Nigeria population, and furthermore, AMOVA results confirmed that there was no significant genetic differentiation between the three groups.
In summary, results obtained in the present work by population comparisons and by AMOVA generally point to the absence of population sub-structuring in Nigeria, for the studied markers. Therefore, a global genetic database could be constructed and used for identification purposes for the 10 X-STRs comprising the three ethnic groups addressed in this study (Igbo, Hausa and Yoruba) (Table 19). However, more population data from the three ethnic groups of Nigeria involving larger sample sizes will allow increasing databases and a more complete analysis of the structure of the Nigeria population as well as confirmation of the results obtained in this work.

Furthermore, it should be highlighted that the results obtained for the Nigerian population group comparisons were presented as a poster communication format (Appendix B) at the 28th Congress of the International Society for Forensic Genetics (9-13th of September, 2019, Prague, Czech Republic). Also, a conference proceeding resulted from the same congress and has been submitted for publication in the special issue edition of the Forensic Science International: Genetics Supplement Series (Appendix C).

Table 19. Allele frequencies and heterozygosity for 10 X-STRs in Nigeria

	DXS8378	DXS9898	DXS7133	GATA31E08	GATA172D05	DXS7423	DXS6809	DXS7132	DXS9902	DXS6789
0	0.009							0.004	0.004	0.004
6					0.278					
7		0.013		0.030	0.070					
8		0.009	0.017	0.017	0.161	0.004			0.009	
8.3		0.030								
9		0.009	0.039	0.178	0.348				0.065	
10	0.252	0.139	0.183	0.152	0.078			0.004	0.074	
11	0.391	0.161	0.626	0.039	0.048			0.026	0.335	
12	0.300	0.378	0.104	0.322	0.017	0.022		0.104	0.352	
12.3		0.004								
13	0.039	0.143	0.026	0.196		0.083		0.265	0.135	
13.3		0.057	0.004							
14	0.009	0.052	0.039	0.061		0.509		0.352	0.022	
14.3		0.004								
15				0.004		0.283		0.165	0.004	0.274
16						0.074		0.065		0.096
17						0.026		0.009		0.009
18								0.004		0.017
19										0.074
20										0.191
21										0.209
22										0.104
23										0.017
24										0.004
27							0.004			
28							0.004			
29							0.022			
30							0.130			
31							0.109			
31.1							0.009			
32							0.152			
32.1							0.004			
33							0.213			
34							0.204			
35							0.104			
36							0.022			
37							0.013			
38							0.009			
He	**0.689**	**0.787**	**0.564**	**0.800**	**0.765**	**0.651**	**0.852**	**0.764**	**0.738**	**0.821**

Interpopulational comparisons

To allow for a more specific comparison between populations, both on a geographical and ethno-linguistically level, genetic distances were analyzed between the three ethnic groups of Nigeria (Igbo, Hausa and Yoruba) and other African population samples for which frequency data are published. For the studies with data regarding markers overlapping with this work (Appendix D), the following populations were selected, namely, Angola and Mozambique (Gomes et al., 2007), Algeria (Bekada et al., 2010), Ivory Coast (Pasino et al., 2011), Egypt (Fukuta et al., 2019), Ghana (Poetsch et al., 2009), Guinea-Bissau (Gomes et al., 2017), Morocco (Laouina et al., 2013), Tunisia (Messoussi et al., 2019) and Uganda (Gomes et al., 2009b). Results obtained in the comparisons of other African population countries with Igbo, Hausa and Yoruba were made with estimations using single locus genetic distances, namely the F_{ST} measure and applying the Bonferroni's correction of the significance level (Tables 20a-c).

Igbo *vs.* other African populations

Significant differences for the population comparisons were observed (Table 20a) between Igbo *vs.* Morocco and between Igbo *vs.* Tunisia for six X-STRs (DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423 and DXS6789). Genetic distances varied between 1.7% (p=0.00535) for DXS6789 (Morocco *vs.* Igbo) and up to 10.1% for DXS7133 ($p \leq 10^{-5}$) (Morocco *vs.* Igbo and Tunisia *vs.* Igbo). In addition, significant differentiation was also observed between Igbo and the population of Egypt for seven out of the eight analyzed loci, namely DXS7133, GATA31E08, GATA172D05, DXS7423, DXS7132, DXS9902 and DXS6789 (Table 20a), with F_{ST} values ranging from 3.9% for DXS9902 (p=0.00515) and 18.5% for DXS7133 ($p \leq 10^{-5}$). Finally, Igbo also demonstrated significant genetic differentiation ($F_{ST} \geq 2.7\%; p \leq 0.00059$) when compared to Algeria for half of the tested loci (GATA172D05, DXS7423 and DXS6789) that had data available for this population (Table 20a).
For the remaining populations compared with Igbo, specifically Angola, Mozambique, Ivory Coast, Ghana, Guinea Bissau and Uganda no significant genetic distances were detected for any of the studied loci (Table 20a).

Hausa *vs.* other African populations

Similar to results obtained between the population comparisons performed between Igbo and the other African populations here studied, Hausa also presented significant differentiation between Morocco, Tunisia, Egypt and Algeria. For Morocco and Tunisia,

differences were observed at four loci, namely DXS7133, GATA172D05, DXS7423 and DXS6789 when compared to the Hausa group (Table20b). Values of genetic distances with significant *p*-values were between 2.2% (p=0.00178) for DXS6789 (Morocco *vs*. Hausa) and 9.3%; ($p \leq 10^{-5}$) for GATA172D05 (Tunisia *vs*. Hausa). When comparing Egypt with Hausa, the highest number of differences were observed between these two populations, i.e., seven out of eight analyzed X-STRs showed significant genetic distances ($F_{ST} \geq 3.5\%$; $p \leq 0.00158$). As observed when comparing Igbo with Algeria, similar results were also obtained between Hausa and Algeria (Table 20b) since genetic differentiation was detected for the same three loci (GATA172D05, DXS7423 and DXS6789) ($F_{ST} \geq 4.4\%$; $p \leq 0.00109$).

As well, the populations of Angola, Mozambique, Ivory Coast, Ghana, Guinea Bissau and Uganda presented no significant genetic distances for any of the studied loci when compared to the Nigerian group of Hausa (Table 20b) similar to the comparisons of Igbo with these other African countries.

Yoruba *vs*. other African populations

When analyzing the genetic distances obtained in comparisons between Yoruba and the other studied populations, less significant differences were found compared to Igbo and Hausa (Table 20c). Nevertheless, significant differentiation was observed between Yoruba *vs*. Morocco ($F_{ST} \geq 4.2\%$; $p \leq 0.00178$) and between Yoruba *vs*. Tunisia ($F_{ST} \geq 4.9\%$; $p \leq 0.00248$) at three loci, namely DXS7133, GATA31E08 and GATA172D05. Again, Egypt presented the most significant genetic distances when compared to Yoruba with differences observed at DXS7133, GATA31E08, GATA172D05, DXS9902 and DXS6789 loci ($F_{ST} \geq 4.2\%$; $p \leq 0.00545$) (Table 20c). When compared Yoruba with Algeria, only at one locus (out of six) presented significant differences, namely at GATA172D05 (F_{ST}=8.8%; $p \leq 10^{-5}$).

Yoruba presented no significant differentiation with the African population groups of Angola, Mozambique, Ivory Coast, Ghana, Guinea Bissau and Uganda as observed with Igbo and Hausa (Table 20b).

As a final observation of the results obtained from the interpopulational comparisons of the Nigerian populations with the African populations here studied, all of the significant differences observed were with the Northern African populations of Morocco, Tunisia, Egypt and Algeria, which are populations that belong to the Afroasiatic linguistic family. Although the Hausa population also belongs to the Afroasiatic language family, there were no significant affinities between Hausa and the latter four populations. These

outcomes support the results obtained previously in the population comparisons among the three ethnic groups of Nigeria (subsection *Genetic distances ($F_{ST's}$)*) where homogeneity among the groups for the studied ten X-chromosomal markers was observed.

On the other hand, genetic affinities were observed with the Sub-Saharan populations of Angola, Mozambique, Ivory Coast, Ghana, Guinea Bissau and Uganda, which are populations belonging to the Niger-Congo language family such as the Igbo and Yoruba populations.

However, more population studies involving larger sample sizes will allow increasing databases and a more complete analysis of the structure of African populations.

Tables 20 a-c. Genetic distances between the three populations of Nigeria and other African populations. (1st line: F_{ST}, 2nd line: *p*-value of genetic distance and respective standard errors (s.e.)). Statistically significant values after Bonferroni's correction are in bold and the respective *p*-values shown under X-STR name. Ang: Angola, Moz: Mozambique, Alg: Algeria, CI: Ivory Coast, Egy: Egypt, Gha: Ghana, GB: Guinea-Bissau, Mor: Morocco, Tun: Tunisia, Uga: Uganda

a)

Nigeria_Igbo	Ang	Moz	Alg	IC	Egy	Gha	GB	Mor	Tun	Uga
DXS8378 *p*≤0.00556	0.00504 0.22978± 0.0036	0.00629 0.18008± 0.0042	0.00951 0.09425± 0.0027	0.01451 0.06257± 0.0019	-	0.00250 0.26027± 0.0045	0.00741 0.11910± 0.0035	0.02228 0.01356± 0.0010	0.01826 0.04079± 0.0019	-0.00471 0.52708± 0.0049
DXS9898 *p*≤0.00833	-0.00773 0.88031± 0.0030	0.00687 0.12662± 0.0037	-	-	-	-0.00063 0.45164± 0.0051	-	**0.02345 0.00327± 0.0005**	**0.02999 0.00059± 0.0002**	-0.00539 0.66578± 0.0046
DXS7133 *p*≤0.001	-	-	-	-	**0.18490 0.00000± 0.0000**	0.01189 0.06039± 0.0025	-	**0.10189 0.00000± 0.0000**	**0.10140 0.00000± 0.0000**	-0.00972 0.81249± 0.0036
GATA31E08 *p*≤0.0125	-	-	-	-	**0.13051 0.00000± 0.0000**	-	-	**0.04107 0.00000± 0.0000**	**0.04546 0.00000± 0.0000**	-0.00107 0.44085± 0.0046
GATA172D05 *p*≤0.00625	0.01019 0.11326± 0.0036	-0.00481 0.74438± 0.0044	**0.05291 0.00000± 0.0000**	0.01275 0.05405± 0.0024	**0.07483 0.00000± 0.0000**	-	-	**0.05941 0.00000± 0.0000**	**0.06371 0.00000± 0.0000**	-0.01400 0.99515± 0.0008
DXS7423 *p*≤0.005	-0.00312 0.50391± 0.0051	-0.00418 0.59103± 0.0052	**0.05745 0.00000± 0.0000**	0.00162 0.29066± 0.0039	**0.04912 0.00040± 0.0002**	0.00029 0.35620± 0.0046	-0.00266 0.54826± 0.0049	**0.05492 0.00000± 0.0000**	**0.04658 0.00139± 0.0004**	0.00989 0.16187± 0.0038
DXS6809 *p*≤0.00625	-0.00048 0.45055± 0.0048	0.00420 0.17493± 0.0035	-0.00036 0.44956± 0.0052	0.00695 0.10138± 0.0029	0.00830 0.03891± 0.0020	-	-	0.00625 0.07920± 0.0031	0.00637 0.11138± 0.0028	0.00816 0.14771± 0.0030
DXS7132 *p*≤0.005	-0.00555 0.66756± 0.0055	0.00383 0.22473± 0.0044	-0.00525 0.86467± 0.0037	-0.00158 0.47015± 0.0050	**0.07206 0.00000± 0.0000**	-0.00259 0.60152± 0.0049	-0.00417 0.79715± 0.0046	0.00130 0.29492± 0.0046	-0.00464 0.72894± 0.0043	0.01087 0.14533± 0.0032
DXS9902 *p*≤0.0125	-	-	-	-	**0.03890 0.00515± 0.0006**	-	-	0.00563 0.17830± 0.0039	0.00700 0.13504± 0.0030	-0.00441 0.56846± 0.0043
DXS6789 *p*≤0.00625	0.01100 0.07653± 0.0030	0.00554 0.14642± 0.0038	**0.02721 0.00059± 0.0002**	-0.00581 0.89090± 0.0034	**0.04904 0.00000± 0.0000**	-	-	**0.01715 0.00535± 0.0007**	**0.03723 0.00010± 0.0001**	-0.00120 0.46728± 0.0044

b)

Nigeria_Hausa	Ang	Moz	Alg	IC	Egy	Gha	GB	Mor	Tun	Uga
DXS8378 *p*≤0.00556	-0.00770 0.75606± 0.0044	-0.00650 0.75794± 0.0040	-0.00025 0.38570± 0.0048	-0.00162 0.46005± 0.0051	-	-0.00699 0.97159± 0.0015	-0.00303 0.59529± 0.0044	0.00111 0.30690± 0.0044	0.00237 0.28878± 0.0044	-0.01205 0.89357± 0.0031
DXS9898 *p*≤0.00833	-0.00267 0.53034± 0.0044	0.00817 0.10643± 0.0031	-	-	-	0.00367 0.18770± 0.0036	-	0.01875 0.01267± 0.0011	**0.03659 0.00079± 0.0003**	-0.00160 0.46589+- 0.0047
DXS7133 *p*≤0.001	-	-	-	-	**0.14643 0.00000± 0.0000**	-0.00538 0.83675± 0.0034	-	**0.05698 0.00000± 0.0000**	**0.06767 0.00000± 0.0000**	0.00679 0.20750± 0.0042
GATA31E08 *p*≤0.0125	-	-	-	-	**0.08112 0.00000± 0.0000**	-	-	0.00742 0.08207± 0.0026	0.00778 0.09851± 0.0030	-0.00956 0.86764± 0.0036
GATA172D05 *p*≤0.00625	-0.00626 0.72696± 0.0047	-0.00430 0.68459± 0.0046	**0.08083 0.00000± 0.0000**	-0.00047 0.40491± 0.0044	**0.09457 0.00000± 0.0000**	-	-	**0.07634 0.00000± 0.0000**	**0.09258 0.00000± 0.0000**	-0.00002 0.39560± 0.0050
DXS7423 *p*≤0.005	-0.00766 0.78012± 0.0040	-0.00741 0.85397± 0.0037	**0.04320 0.00109± 0.0003**	-0.00549 0.72448± 0.0043	**0.03514 0.00158± 0.0004**	-0.00284 0.57123± 0.0054	0.00184 0.26017± 0.0040	**0.03798 0.00158± 0.0004**	**0.03596 0.00386± 0.0006**	-0.00290 0.45956± 0.0053
DXS6809 *p*≤0.00625	0.00934 0.10643± 0.0029	0.00962 0.05514± 0.0021	0.00007 0.40956± 0.0048	0.00733 0.10583± 0.0028	-0.00000 0.40679± 0.0047	-	-	0.00499 0.12454± 0.0035	-0.00650 0.93466± 0.0025	0.02041 0.02901± 0.0015
DXS7132 *p*≤0.005	0.00231 0.29175± 0.0052	0.01007 0.09207± 0.0030	-0.00319 0.64241± 0.0046	0.00594 0.17454± 0.0039	**0.05995 0.00000± 0.0000**	-0.00271 0.61816± 0.0047	-0.00239 0.59410± 0.0045	-0.00113 0.46966± 0.0051	0.00012 0.38422± 0.0050	-0.00254 0.48886± 0.0053
DXS9902 *p*≤0.0125	-	-	-	-	**0.05519 0.00010± 0.0001**	-	-	0.00170 0.28542± 0.0044	0.00085 0.33848± 0.0053	-0.01175 0.90941± 0.0030
DXS6789 *p*≤0.00625	0.00975 0.10583± 0.0031	0.01300 0.03713± 0.0020	**0.04360 0.00000± 0.0000**	-0.00352 0.67548± 0.0047	**0.06128 0.00000± 0.0000**	-	-	**0.02225 0.00178± 0.0004**	**0.04831 0.00000± 0.0000**	-0.00207 0.50064± 0.0053

c)

Nigeria_Yoruba	Ang	Moz	Alg	IC	Egy	Gha	GB	Mor	Tun	Uga
DXS8378 *p*≤0.00556	0.00914 0.19572± 0.0038	-0.00228 0.43184± 0.0053	0.00846 0.19889± 0.0036	0.00311 0.27522± 0.0043	-	0.00508 0.25146± 0.0043	0.00387 0.26611± 0.0043	-0.00849 0.72221± 0.0036	-0.00115 0.40580± 0.0051	0.00743 0.25285± 0.0044
DXS9898 *p*≤0.00833	-0.01330 0.92050± 0.0026	0.00907 0.16068± 0.0035	-	-	-	-0.00628 0.72894± 0.0041	-	0.02702 0.02020± 0.0016	0.03427 0.00871± 0.0008	-0.00944 0.71805± 0.0044
DXS7133 *p*≤0.001	-	-	-	-	**0.14481 0.00000± 0.0000**	-0.00818 0.74735± 0.0048	-	**0.05725 0.00178± 0.0004**	**0.06604 0.00248± 0.0005**	-0.00507 0.49490± 0.0048
GATA31E08 *p*≤0.0125	-	-	-	-	**0.10774 0.00000± 0.0000**	-	-	**0.04251 0.00158± 0.0004**	**0.04868 0.00119± 0.0003**	-0.00075 0.41768± 0.0047
GATA172D05 *p*≤0.00625	-0.00631 0.61479± 0.0050	0.00084 0.37016± 0.0049	**0.08826 0.00000± 0.0000**	0.00053 0.37036± 0.0046	**0.09256 0.00000± 0.0000**	-	-	**0.07905 0.00000± 0.0000**	**0.09247 0.00000± 0.0000**	0.01714 0.11553± 0.0034
DXS7423 *p*≤0.005	-0.01095 0.75369± 0.0041	-0.00187 0.41501± 0.0048	0.00997 0.16167± 0.0039	-0.00798 0.66132± 0.0046	0.00815 0.18563± 0.0037	-0.00433 0.53816± 0.0044	0.02805 0.03742± 0.0018	0.00647 0.20543± 0.0039	0.00710 0.21077± 0.0043	-0.01851 0.95733± 0.0019
DXS6809 *p*≤0.00625	0.00509 0.25770± 0.0047	0.00380 0.25750± 0.0047	0.00249 0.29740± 0.0049	-0.00884 0.86803± 0.0029	0.00703 0.15593± 0.0040	-	-	-0.00129 0.47520± 0.0046	0.00944 0.13721± 0.0033	-0.00304 0.52282± 0.0050
DXS7132 *p*≤0.005	-0.00386 0.51985± 0.0045	-0.00336 0.51203± 0.0047	0.01315 0.10049± 0.0029	-0.00113 0.41699± 0.0048	0.00793 0.16256± 0.0040	0.00009 0.38432± 0.0048	0.01169 0.10949± 0.0030	-0.00268 0.50520± 0.0051	-0.00037 0.39758± 0.0057	-0.00957 0.70498± 0.0043
DXS9902 *p*≤0.0125	-	-	-	-	**0.04247 0.00545± 0.0007**	-	-	-0.00233 0.47282± 0.0052	-0.00378 0.51955± 0.0054	-0.01603 0.92219± 0.0030
DXS6789 *p*≤0.00625	0.00214 0.34234± 0.0048	0.00761 0.16939± 0.0040	0.03242 0.00554± 0.0008	-0.00471 0.64033± 0.0049	**0.04950 0.00069± 0.0003**	-	-	0.00973 0.10999± 0.0031	0.03415 0.00792± 0.0008	-0.00133 0.44976± 0.0043

Forensic Parameters

The statistical parameters for evaluating the forensic efficiency of the present X-STR markers in the population of Nigeria are shown in Table 21. All markers selected for this study were highly polymorphic, thus confirming their usefulness for forensic purposes.

Table 21. Forensic efficiency statistics for each X-STR locus and combined values for decaplex system: Power of exclusion in trios involving daughters (PE_T) and power of exclusion in father/daughter duos (PE_D), power of discrimination in males (PD_M), power of discrimination in females (PD_F) and probability of exclusion in half-sisters (PE_{HS})

	PE_T	PE_D	PE_{HS}	PD_M	PD_F
DXS8378	0.633	0.487	0.262	0.692	0.846
DXS9898	0.759	0.632	0.420	0.784	0.929
DXS7133	0.523	0.375	0.203	0.561	0.770
GATA31E08	0.769	0.644	0.428	0.797	0.931
GATA172D05	0.727	0.593	0.372	0.762	0.908
DXS7423	0.597	0.451	0.240	0.648	0.825
DXS6809	0.881	0.798	0.531	0.890	0.979
DXS7132	0.727	0.593	0.370	0.762	0.908
DXS9902	0.694	0.555	0.334	0.735	0.888
DXS6789	0.795	0.676	0.468	0.819	0.943
Combined values	**99.9997805382327**	**99.9888689015234**	**99.0249852287118**	**99.9999375932436**	**99.9999999965716**

Regarding forensic statistical parameters, the power of discrimination (PD) presented values ranging from 56.1% (DXS7133) to 89% (DXS6809) in males and 77% (DXS7133) and 98% (DXS6809) in females. The marker DXS6809 gives the highest probability that two randomly chosen individuals in the population can be discriminated against each other, contrary to the marker DXS7133. This supports that discriminating individual capacity of these loci, for example, the locus DXS6809 can discriminate 89% and 98% of samples with male and female Nigerian origin, respectively. With regard to the power of exclusion in trios involving daughters the values range were between 52.3% (DXS7133) and 88.1% (DXS6809), concerning to power of exclusion in mother/son duos the values range were between 37.5% (DXS7133) and 79.8% (DXS6809). This parameter is defined as the probability that a specific genetic system may contribute to the exclusion of a particular individual, i.e., for the DXS6809 marker, in the Nigerian population, the average probability of excluding a random man as the father in paternity analysis, when both mother and daughter are tested, just for this locus alone is 88%.
One of the most important features of X-STRs is the potential to exclude the relationship between two putative half-sisters, which is not possible with autosomal STRs. Therefore, the expected probability of exclusion in half-sisters (PE_{HS}) has been estimated according

to Gomes et al. (2009b) as being an example of an important scenario of application of X-STRs.

Regarding overall forensic efficiency in kinship analysis and other identification cases where X-STRs can complement other genetic information, the developed decaplex system revealed high combined power of exclusion values, in both trios and duos type of scenarios, as well power of discrimination values for male and females, for the ten X-STRs. In conclusion, the decaplex X-STR system represent a potential tool in kinship testing as well as identification forensic scenarios in the population of Nigeria.

Conclusion

After the end of this work, which aimed to genetically characterize a Nigerian population sample using a decaplex system composed by 10 X-STRs (already used to characterize many other different populations), it is possible to draw some conclusions:

- The decaplex system of X-STR markers allowed genotyping of three main ethnic groups of Nigeria, namely Yoruba, Igbo and Hausa, in a fast, simple and efficient way;
- This system, as shown in other studies, remains a robust tool for straightforward and simple X-STR genotyping using common forensic routine instrumentation;
- The X-STRs selected for this study were found to be highly polymorphic, with high genetic diversities in the populations of Nigeria here studied;
- Regarding the association between markers in the studied population, LD analysis revealed no statistically significant associations; thus, markers can be treated as independent loci in statistical calculations in identification cases, namely in forensic or kinship testing, in the present population groups;
- The analysis of the population structure of Nigeria showed population homogeneity for the studied 10 X chromosome markers among Yoruba, Igbo and Hausa groups, allowing for the construction of a single allelic frequency database for the Nigerian population;
- The high probabilities of exclusion obtained in trios and duos prove the usefulness of this system in specific cases of biological kinship investigation, especially in paternity cases where the disputed child is a female (or in maternity cases with male offspring) or in half-paternal sister scenarios, in Nigeria;
- High probabilities of discrimination were also obtained allowing for the use of the decaplex system in human individual identification in Nigeria.

Thus, it can be concluded that the X-chromosomal decaplex system here studied is a robust tool and proved to be useful for potential human identification analyses such as in forensic or kinship scenarios in the population of Nigeria (Igbo, Hausa and Yoruba).

References

AMORIM, A. & BUDOWLE, B. 2016. *Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation*, World Scientific.

ARBIZA, L., GOTTIPATI, S., SIEPEL, A. & KEINAN, A. 2014. Contrasting X-linked and autosomal diversity across 14 human populations. *Am J Hum Genet,* 94**,** 827-44.

ATHANASIADIS, G., ESTEBAN, E., VIA, M., DUGOUJON, J. M., MOSCHONAS, N., CHAABANI, H. & MORAL, P. 2007. The X chromosome Alu insertions as a tool for human population genetics: data from European and African human groups. *Eur J Hum Genet,* 15**,** 578-83.

BACHTROG, D. 2006. A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev,* 16**,** 578-85.

BARNETT, R. & LARSON, G. 2012. A Phenol–Chloroform Protocol for Extracting DNA from Ancient Samples. *In:* SHAPIRO, B. & HOFREITER, M. (eds.) *Ancient DNA: Methods and Protocols.* Totowa, NJ: Humana Press.

BEKADA, A., BENHAMAMOUCH, S., BOUDJEMA, A., FODIL, M., MENEGON, S., TORRE, C. & ROBINO, C. 2010. Analysis of 21 X-chromosomal STRs in an Algerian population sample. *Int J Legal Med,* 124**,** 287-94.

BERNIELL-LEE, G., CALAFELL, F., BOSCH, E., HEYER, E., SICA, L., MOUGUIAMA-DAOUDA, P., VAN DER VEEN, L., HOMBERT, J. M., QUINTANA-MURCI, L. & COMAS, D. 2009. Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol,* 26**,** 1581-9.

BRENNER, S., MILLER, J. H. & BROUGHTON, W. J. 2002. *Encyclopedia of Genetics*, Academic Press.

BUDOWLE, B. & VAN DAAL, A. 2008. Forensically relevant SNP classes. *BioTechniques,* 44**,** 603-610.

BURGOYNE, T. W. & HIEFTJE, G. M. 1996. An introduction to ion optics for the mass spectrograph. *Mass Spectrometry Reviews,* 15**,** 241-259.

BUSEY, C. & CRUZ, B. 2015. A Shared Heritage: Afro-Latin@s and Black History. *The Social Studies,* 106**,** 293-300.

BUTLER, J. M. 2005. *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*, Elsevier Science.

BUTLER, J. M. 2009. *Fundamentals of Forensic DNA Typing*, Elsevier Science.

BUTLER, J. M. 2011. *Advanced Topics in Forensic DNA Typing: Methodology*, Elsevier Science.

BUTLER, J. M., SHEN, Y. & MCCORD, B. R. 2003. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci,* 48**,** 1054-64.

CAGLIA, A., TOFANELLI, S., COIA, V., BOSCHI, I., PESCARMONA, M., SPEDINI, G., PASCALI, V., PAOLI, G. & DESTRO-BISOL, G. 2003. A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F(ST) and R(ST) genetic distances. *Hum Biol,* 75**,** 313-30.

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. *Curr Opin Genet Dev,* 29**,** 120-32.

CASTELLA, V., LESTA, M. & MANGIN, P. 2009. One person with two DNA profiles: A(nother) case of mosaicism or chimerism. *International journal of legal medicine,* 123**,** 427-30.

CHAKRABORTY, R., STIVERS, D. N., SU, B., ZHONG, Y. & BUDOWLE, B. 1999. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis,* 20**,** 1682-96.

CHEN, D. P., TSENG, C. P., TSAI, S. H., WANG, M. C., LU, S. C., WU, T. L., CHANG, P. Y. & SUN, C. F. 2009. Use of X-linked short tandem repeats loci to confirm mutations in parentage caseworks. *Clin Chim Acta,* 408**,** 29-33.

DAKIN, E. E. & AVISE, J. C. 2004. Microsatellite null alleles in parentage analysis. *Heredity,* 93**,** 504-509.

DESMARAIS, D., ZHONG, Y., CHAKRABORTY, R., PERREAULT, C. & BUSQUE, L. 1998. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). *J Forensic Sci,* 43**,** 1046-9.

EDELMAN, J., HERING, S., MICHAEL, M., LESSIG, R., DEISCHEL, D., MEIER-SUNDHAUSEN, G., ROEWER, L., PLATE, I. & SZIBOR, R. 2001. 16 X-chromosome STR loci frequency data from a German population. *Forensic Sci Int,* 124**,** 215-8.

EDELMANN, J., DEICHSEL, D., PLATE, I., KASER, M. & SZIBOR, R. 2003. Validation of the X-chromosomal STR DXS6809. *Int J Legal Med,* 117**,** 241-4.

ELLEGREN, H. 2004. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet,* 5**,** 435-45.

ELLEGREN, H. 2011. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nature Reviews Genetics,* 12**,** 157-166.

EVETT, I. W., LAMBERT, J. A., BUCKLETON, J. S. & WEIR, B. S. 1996. Statistical analysis of a large file of data from STR profiles of British Caucasians to support forensic casework. *Int J Legal Med,* 109**,** 173-7.

EXCOFFIER, L. & LISCHER, H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* 10**,** 564-567.

FALOLA, T. & HEATON, M. M. 2008. *A History of Nigeria*, Cambridge University Press.

FAN, G. Y., YE, Y. & HOU, Y. P. 2016. Detecting a hierarchical genetic population structure via Multi-InDel markers on the X chromosome. *Scientific reports,* 6**,** 32178-32178.

FAN, S., KELLY, D. E., BELTRAME, M. H., HANSEN, M. E. B., MALLICK, S., RANCIARO, A., HIRBO, J., THOMPSON, S., BEGGS, W., NYAMBO, T., OMAR, S. A., MESKEL, D. W., BELAY, G., FROMENT, A., PATTERSON, N., REICH, D. & TISHKOFF, S. A. 2019. African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol,* 20**,** 82.

FISHER, R. A. 1951. Standard calculations for evaluating a blood-group system. *Heredity (Edinb),* 5**,** 95-102.

FUKUTA, M., GABALLAH, M., TAKADA, K., MIYAZAKI, H., KATO, H., AOKI, Y., HAMED, S. S., ELMORSI, D. A. A. & ELDAKROORY, S. A. 2019. Genetic polymorphism of 27 X-chromosomal short tandem repeats in an Egyptian population. *Leg Med (Tokyo),* 37**,** 64-66.

GARRIGAN, D., MOBASHER, Z., KINGAN, S. B., WILDER, J. A. & HAMMER, M. F. 2005. Deep haplotype divergence and long-range linkage disequilibrium at xp21.1 provide evidence that humans descend from a structured ancestral population. *Genetics,* 170**,** 1849-56.

GOMES, I., ALVES, C., MAXZUD, K., PEREIRA, R., PRATA, M. J., SANCHEZ-DIZ, P., CARRACEDO, A., AMORIM, A. & GUSMAO, L. 2007. Analysis of 10 X-STRs in three African populations. *Forensic Sci Int Genet,* 1**,** 208-11.

GOMES, I., PEREIRA, P. J. P., HARMS, S., OLIVEIRA, A. M., SCHNEIDER, P. M. & BREHM, A. 2017. Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: Inferences from a multiethnic population. *Forensic Sci Int Genet,* 31**,** 89-94.

GOMES, I., PEREIRA, R., MAYR, W. R., AMORIM, A., CARRACEDO, A. & GUSMAO, L. 2009a. Evaluation of DXS9902, DXS7132, DXS6809, DXS7133, and DXS7423 in humans and chimpanzees: sequence variation, repeat structure, and nomenclature. *Int J Legal Med,* 123**,** 403-12.

GOMES, I., PEREIRA, V., GOMES, V., PRATA, M. J., PINTO, N., CARRACEDO, A., AMORIM, A. & GUSMAO, L. 2009b. The Karimojong from Uganda: genetic characterization using an X-STR decaplex system. *Forensic Sci Int Genet,* 3**,** e127-8.

GOMES, I., PRINZ, M., PEREIRA, R., BIESCHKE, E., MAYR, W. R., AMORIM, A., CARRACEDO, A. & GUSMAO, L. 2009c. X-chromosome STR sequence variation, repeat structure, and nomenclature in humans and chimpanzees. *Int J Legal Med,* 123**,** 143-9.

GOMEZ, F., HIRBO, J. & TISHKOFF, S. A. 2014. Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harb Perspect Biol,* 6**,** a008524.

GOODWIN, W., LINACRE, A. & HADI, S. 2007. *An Introduction to Forensic Genetics*, John Wiley & Sons.

GOTTIPATI, S., ARBIZA, L., SIEPEL, A., CLARK, A. G. & KEINAN, A. 2011. Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nat Genet,* 43**,** 741-3.

GUSMÃO, L., AMORIM, A., PRATA, M. J., PEREIRA, L., LAREU, M. V. & CARRACEDO, A. 1996. Failed PCR amplifications of MBP-STR alleles due to polymorphism in the primer annealing region. *Int J Legal Med,* 108**,** 313-5.

GUSMÃO, L., SANCHEZ-DIZ, P., ALVES, C., GOMES, I., ZARRABEITIA, M. T., ABOVICH, M., ATMETLLA, I., BOBILLO, C., BRAVO, L., BUILES, J., CAINE, L., CALVO, R., CARVALHO, E., CARVALHO, M., CICARELLI, R., CATELLI, L., CORACH, D., ESPINOZA, M., GARCIA, O., MALAGHINI, M., MARTINS, J., PINHEIRO, F., JOAO PORTO, M., RAIMONDI, E., RIANCHO, J. A., RODRIGUEZ, A., RODRIGUEZ, A., RODRIGUEZ CARDOZO, B., SCHNEIDER, V., SILVA, S., TAVARES, C., TOSCANINI, U., VULLO, C., WHITTLE, M., YURREBASO, I., CARRACEDO, A. & AMORIM, A. 2009. A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *Int J Legal Med,* 123**,** 227-34.

HAMILTON, M. B. 2009. *Population genetics,* Oxford, Wiley-Blackwell.

HAMMOND, H. A., JIN, L., ZHONG, Y., CASKEY, C. T. & CHAKRABORTY, R. 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am J Hum Genet,* 55**,** 175-89.

HARRIS, E. E. & HEY, J. 1999. X chromosome evidence for ancient human histories. *Proceedings of the National Academy of Sciences,* 96**,** 3320-3324.

HERING, S. & SZIBOR, R. 2000. Development of the X-linked tetrameric microsatellite marker DXS9898 for forensic purposes. *J Forensic Sci,* 45**,** 929-31.

HOUCK, M. M. 2015. *Forensic Biology*, Elsevier Science.

IBARRA, A., RESTREPO, T., ROJAS, W., CASTILLO, A., AMORIM, A., MARTÍNEZ, B., BURGOS, G., OSTOS, H., ÁLVAREZ, K., CAMACHO, M., SUAREZ, Z., PEREIRA, R. & GUSMÃO, L. 2014. Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion polymorphisms. *PloS one,* 9**,** e87202-e87202.

INTERNATIONAL HUMAN GENOME SEQUENCING, C. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-45.

JARUZELSKA, J., ZIETKIEWICZ, E., BATZER, M., COLE, D. E., MOISAN, J. P., SCOZZARI, R., TAVARÉ, S. & LABUDA, D. 1999. Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics,* 152**,** 1091-1101.

JOBLING, M. A. & GILL, P. 2004. Encoded evidence: DNA in forensic analysis. *Nat Rev Genet,* 5**,** 739-51.

JOBLING, M. A., HURLES, M. & TYLER-SMITH, C. 2004. *Human Evolutionary Genetics: Origins, Peoples & Disease*, Garland Science.

JOBLING, M. A., KIVISILD, T. & TYLER-SMITH, C. 2014. *Human Evolutionary Genetics*, Garland Science.

KAESSMANN, H., HEIßIG, F., HAESELER, A. V. & PÄÄBO, S. 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nature Genetics,* 22**,** 78-81.

KLING, D., DELL'AMICO, B. & TILLMAR, A. O. 2015. FamLinkX – implementation of a general model for likelihood computations for X-chromosomal marker data. *Forensic Science International: Genetics,* 17**,** 1-7.

KRAWCZAK, M. 2007. Kinship testing with X-chromosomal markers: mathematical and statistical issues. *Forensic Sci Int Genet,* 1**,** 111-4.

LAAN, M., WIEBE, V., KHUSNUTDINOVA, E., REMM, M. & PAABO, S. 2005. X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. *Eur J Hum Genet,* 13**,** 452-62.

LAHN, B. T., PEARSON, N. M. & JEGALIAN, K. 2001. The human Y chromosome, in the light of evolution. *Nat Rev Genet,* 2**,** 207-16.

LAOUINA, A., AZNAR, J. M., CHBEL, F. & DE PANCORBO, M. M. 2013. Genetic data of X-chromosomal STRs in a Moroccan population sample (Casablanca) using an X-STR decaplex. *Forensic Sci Int Genet,* 7**,** e90-2.

LEARN.GENETICS.UTAH.EDU. 2016. *Introduction to Molecular Genealogy* [Online]. Available: https://learn.genetics.utah.edu/content/basics/molgen/ [Accessed 5/9 2019].

LEVY, P. 2004. *Nigeria*, Benchmark Books/Marshall Cavendish.

LYON, M. F. 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature,* 190**,** 372-3.

MALONE, J. H. & OLIVER, B. 2008. The sex chromosome that refused to die. *Bioessays,* 30**,** 409-411.

MELTON, T., PETERSON, R., REDD, A. J., SAHA, N., SOFRO, A. S., MARTINSON, J. & STONEKING, M. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet,* 57**,** 403-14.

MESSOUSSI, M., PRIETO-FERNANDEZ, E., BAETA, M., NUNEZ, C., GAAIED, A. B. A., DE PANCORBO, M. M. & FADHLAOUI-ZID, K. 2019. Genetic variation of 17 X-chromosome STR loci in Tunisian population of Nabeul. *Int J Legal Med,* 133**,** 85-88.

METZ, H. C. & DIVISION, L. O. C. F. R. 1992. *Nigeria: A Country Study*, U.S. Government Printing Office.

MOREY, C. & AVNER, P. 2010. Genetics and epigenetics of the X chromosome. *Ann N Y Acad Sci,* 1214**,** E18-33.

NAGARAJA, R., MACMILLAN, S., KERE, J., JONES, C., GRIFFIN, S., SCHMATZ, M., TERRELL, J., SHOMAKER, M., JERMAK, C., HOTT, C., MASISI, M., MUMM, S., SRIVASTAVA, A., PILIA, G., FEATHERSTONE, T., MAZZARELLA, R., KESTERSON, S., MCCAULEY, B., RAILEY, B., BUROUGH, F., NOWOTNY, V., D'URSO, M., STATES, D., BROWNSTEIN, B. & SCHLESSINGER, D. 1997. X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. *Genome Res,* 7**,** 210-22.

NIGERIA.GOV.NG. 2019. *About Nigeria* [Online]. Available: http://www.nigeria.gov.ng/index.php/2016-04-06-08-38-30 [Accessed 5/9 2019].

ODIMEGWU, C. & SOMEFUN, O. D. 2017. Ethnicity, gender and risky sexual behaviour among Nigerian youth: an alternative explanation. *Reproductive Health,* 14**,** 16.

OKOLIE, V. O., CISANA, S., SCHANFIELD, M. S., AKANMU, A. S., ADEKOYA, K. O., OYEDEJI, O. A. & PODINI, D. 2018. Correction to: Population data of 21 autosomal STR loci in the Hausa, Igbo and Yoruba people of Nigeria. *Int J Legal Med,* 132**,** 739.

PASINO, S., CARATTI, S., DEL PERO, M., SANTOVITO, A., TORRE, C. & ROBINO, C. 2011. Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast. *Int J Legal Med,* 125**,** 749-52.

PBS.ORG. 2007. *Ethnicity in Nigeria* [Online]. Available: https://www.pbs.org/newshour/arts/africa-jan-june07-ethnic_04-05 [Accessed 5/9 2019].

PEREIRA, R., PEREIRA, V., GOMES, I., TOMAS, C., MORLING, N., AMORIM, A., PRATA, M. J., CARRACEDO, A. & GUSMAO, L. 2012. A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. *Int J Legal Med,* 126**,** 97-105.

PINTO, N., GUSMAO, L. & AMORIM, A. 2011. X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial. *Forensic Sci Int Genet,* 5**,** 27-32.

PINTO, N., GUSMÃO, L., EGELAND, T. & AMORIM, A. 2013. Paternity exclusion power: Comparative behaviour of autosomal and X-chromosomal markers in standard and deficient cases with inbreeding. *Forensic Science International: Genetics,* 7**,** 290-295.

POETSCH, M., EL-MOSTAQIM, D., TSCHENTSCHER, F., BROWNE, E. N., TIMMANN, C., HORSTMANN, R. D. & VON WURMB-SCHWARK, N. 2009. Allele frequencies of 11 X-chromosomal loci in a population sample from Ghana. *Int J Legal Med,* 123**,** 81-3.

POETSCH, M., WASSENAAR, D. & VON WURMB-SCHWARK, N. 2013a. Allele frequencies for 11 X chromosomal short tandem repeats in a population from Turkey. *Int J Legal Med,* 127**,** 913-4.

POETSCH, M., WIEGAND, A., HARDER, M., BLOHM, R., RAKOTOMAVO, N., FREITAG-WOLF, S. & VON WURMB-SCHWARK, N. 2013b. Determination of population origin: a comparison of autosomal SNPs, Y-chromosomal and mtDNA haplogroups using a Malagasy population as example. *Eur J Hum Genet,* 21**,** 1423-8.

RAPLEY, R. & WHITEHOUSE, D. 2007. *Molecular Forensics*, Wiley.

REED, F. & A TISHKOFF, S. 2007. African human diversity, origins and migrations. *Current opinion in genetics & development,* 16**,** 597-605.

RELETHFORD, J. H. 2012. *Human Population Genetics*, Wiley.

ROBINO, C., INTURRI, S., VARACALLI, S., PICCININI, A., GINO, S. & TORRE, C. 2008. Molecular characterization of a null allele at locus DXS8378. *Forensic Science International: Genetics Supplement Series,* 1**,** 160-161.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SANTOS-LOPES, S. S., PEREIRA, R. W., WILSON, I. J. & PENA, S. D. 2007. A worldwide phylogeography for the human X chromosome. *PLoS One,* 2**,** e557.

SCHAFFNER, S. F. 2004. The X chromosome in population genetics. *Nat Rev Genet,* 5**,** 43-51.

SCHNEIDER, P. M. 1997. Basic issues in forensic DNA typing. *Forensic Sci Int,* 88**,** 17-22.

SHIMADA, M. K., PANCHAPAKESAN, K., TISHKOFF, S. A., NATO, A. Q., JR. & HEY, J. 2007. Divergent haplotypes and human history as revealed in a worldwide survey of X-linked DNA sequence variation. *Mol Biol Evol,* 24**,** 687-98.

SUZUKI, D. T. & GRIFFITHS, A. J. F. U. H. B. G. P. B. I. J. U. 1976. *An Introduction to Genetic Analysis*, W.H. Freeman & Company.

SZIBOR, R. 2007. X-chromosomal markers: past, present and future. *Forensic Sci Int Genet,* 1**,** 93-9.

SZIBOR, R., EDELMANN, J., HERING, S., PLATE, I., WITTIG, H., ROEWER, L., WIEGAND, P., CALI, F., ROMANO, V. & MICHAEL, M. 2003a. Cell line DNA typing in forensic genetics--the necessity of reliable standards. *Forensic Sci Int,* 138**,** 37-43.

SZIBOR, R., HERING, S. & EDELMANN, J. 2006. A new Web site compiling forensic chromosome X research is now online. *Int J Legal Med,* 120**,** 252-4.

SZIBOR, R., KRAWCZAK, M., HERING, S., EDELMANN, J., KUHLISCH, E. & KRAUSE, D. 2003b. Use of X-linked markers for forensic purposes. *Int J Legal Med,* 117**,** 67-74.

THECOMMONWEALTH.ORG. *Niger: History* [Online]. Available: http://thecommonwealth.org/our-member-countries/nigeria [Accessed 5/9 2019].

TISHKOFF, S. A., REED, F. A., FRIEDLAENDER, F. R., EHRET, C., RANCIARO, A., FROMENT, A., HIRBO, J. B., AWOMOYI, A. A., BODO, J. M., DOUMBO, O., IBRAHIM, M., JUMA, A. T., KOTZE, M. J., LEMA, G., MOORE, J. H., MORTENSEN, H., NYAMBO, T. B., OMAR, S. A., POWELL, K., PRETORIUS, G. S., SMITH, M. W., THERA, M. A., WAMBEBE, C., WEBER, J. L. & WILLIAMS, S. M. 2009. The genetic structure and history of Africans and African Americans. *Science,* 324**,** 1035-44.

TISHKOFF, S. A. & VERRELLI, B. C. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet,* 4**,** 293-340.

TISHKOFF, S. A. & WILLIAMS, S. M. 2002. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet,* 3**,** 611-21.

VISENTINI, P. F. 2011. *O Livro na Rua Nigeria*.

VULLO, C., GOMES, V., ROMANINI, C., OLIVEIRA, A. M., ROCABADO, O., AQUINO, J., AMORIM, A. & GUSMAO, L. 2015. Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations. *Int J Legal Med,* 129**,** 673-80.

WALSH, P. S., METZGER, D. A. & HIGUCHI, R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques,* 10**,** 506-13.

WATSON, J. D. & CRICK, F. H. C. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature,* 171**,** 737-738.

WRIGHT, S. 1951. The genetical structure of populations. *Ann Eugen,* 15**,** 323-54.

YOTOVA, V., LEFEBVRE, J. F., MOREAU, C., GBEHA, E., HOVHANNESYAN, K., BOURGEOIS, S., BEDARIDA, S., AZEVEDO, L., AMORIM, A., SARKISIAN, T., AVOGBE, P. H., CHABI, N., DICKO, M. H., KOU' SANTA AMOUZOU, E. S., SANNI, A., ROBERTS-THOMSON, J., BOETTCHER, B., SCOTT, R. J. & LABUDA, D. 2011. An X-linked haplotype of Neandertal origin is present among all non-African populations. *Mol Biol Evol,* 28**,** 1957-62.

YU, N., FU, Y.-X. & LI, W.-H. 2002. DNA Polymorphism in a Worldwide Sample of Human X Chromosomes. *Molecular Biology and Evolution,* 19**,** 2131-2141.

YU, N. & LI, W. 2000. No fixed nucleotide difference between Africans and Non-Africans at the pyruvate dehydrogenase E1 alpha-subunit locus. *Genetics,* 155**,** 1481-1483.

ZENG, X., RAKHA, A. & LI, S. 2009. Genetic polymorphisms of 10 X-chromosome STR loci in Chinese Daur ethnic minority group. *Leg Med (Tokyo),* 11**,** 152-4.

ZIDKOVA, A., COUFALOVA, P. & CAPEK, P. 2014. X-STR decaplex study on the population of Czech Republic. *Int J Legal Med,* 128**,** 271-2.

ZIETKIEWICZ, E., YOTOVA, V., JARNIK, M., KORAB-LASKOWSKA, M., KIDD, K. K., MODIANO, D., SCOZZARI, R., STONEKING, M., TISHKOFF, S., BATZER, M. & LABUDA, D. 1998. Genetic structure of the ancestral population of modern humans. *J Mol Evol,* 47**,** 146-55.

Appendix

Appendix A. Genotypes obtained for the X-STR markers in the 230 samples studied from the Nigeria population

Samples	DXS8378	DXS9898	DXS7133	GATA31 E08	GATA172 D05	DXS7423	DXS6809	DXS7132	DXS9902	DXS6789
I1	10	12	11	8	8	14	34	13	12	21
I2	10	11	11	10	9	14	34	14	10	19
I3	11	13	10	13	8	15	34	15	11	20
I4	11	12	11	12	9	14	33	15	13	15
I5	11	13	12	12	6	16	34	13	11	15
I6	11	10	13	12	6	15	34	16	11	19
I7	12	12	12	10	9	16	33	15	13	15
I8	11	12	10	13	9	14	34	14	9	21
I9	11	12	11	12	8	15	34	15	11	20
I10	11	12	11	12	8	15	34	15	11	20
I11	11	10	11	12	8	16	33	14	13	22
I15	10	10	11	15	10	15	30	12	11	20
I16	12	12	11	12	10	15	31	13	11	15
I17	11	14	12	9	9	12	35	18	12	22
I18	12	11	11	10	6	15	34	13	12	15
I19	11	14	11	14	7	14	33	14	13	20
I20	12	7	11	12	6	14	33	12	11	15
I21	11	12	11	7	11	14	30	14	11	22
I22	11	13	11	13	7	14	31	14	11	21
I23	12	12	10	9	8	15	32	13	11	20
I24	11	14	11	13	6	15	35	15	13	15
I25	11	12	12	13	9	14	31	13	12	15
I26	12	11	11	14	11	14	35	14	12	15
I27	12	13	12	9	9	13	33	14	11	22
I28	12	13	11	9	9	13	29	13	10	22
I29	11	11	11	10	11	14	34	13	12	21
I30	10	13.3	11	12	6	13	32	15	12	15
I31	11	8.3	11	13	6	14	33	14	12	18
I32	11	13	11	12	9	14	32	14	12	20
I33	10	12	8	13	8	14	31	14	9	20
I34	12	12	11	14	9	13	33	13	11	19
I35	11	13.3	11	9	6	14	33	13	11	21
I36	11	12	11	12	9	14	32	15	11	20
I37	10	13	11	9	10	14	35	16	12	20
I38	11	11	11	12	6	14	37	14	12	19
I40	10	11	12	7	6	14	34	12	11	19
I41	11	12	12	10	6	16	35	13	12	21
I42	10	14	11	12	7	14	32	14	12	18
I43	11	11	11	12	10	14	30	13	11	22
I44	12	12	11	12	8	14	34	14	13	23
I45	11	12	11	9	11	14	37	14	12	15
I47	12	11	11	12	6	15	31	13	9	22
I48	10	11	11	13	9	16	30	14	12	20
I49	11	14	11	12	6	14	34	14	9	22
I50	11	12	11	7	6	14	30	14	12	20
I51	12	8.3	10	13	11	14	30	16	11	22
I52	11	10	10	12	8	14	30	14	13	15
I53	12	11	11	12	12	13	33	14	10	20
I54	13	10	11	13	8	15	33	14	12	16
I55	10	12	11	13	6	15	28	13	13	20
I56	12	11	11	9	11	13	29	14	11	20
I57	10	12.3	11	11	9	14	29	12	11	15
I58	11	10	11	12	6	15	38	13	9	21
I59	13	12	11	10	9	15	35	14	12	19

I60	12	12	11	12	9	14	36	14	12	15
I61	11	10	10	9	9	14	32	14	11	21
I62	12	13.3	11	9	9	15	30	15	11	15
I63	10	12	11	12	6	14	33	15	12	16
I64	11	10	11	13	8	14	33	14	12	19
I65	10	11	11	13	10	8	31	14	9	22
I66	11	10	11	12	6	15	33	15	11	22
I67	10	12	10	9	9	15	29	13	12	21
I68	10	8.3	9	12	8	14	27	14	11	16
I69	11	11	11	13	8	14	34	14	10	15
I70	12	13	11	10	6	14	30	14	13	21
I71	11	13	10	10	8	14	30	15	11	21
I72	12	12	11	14	8	14	32	13	12	21
I73	10	11	11	10	9	16	35	15	11	16
I74	11	12	12	13	6	14	31	11	11	15
I75	12	12	11	12	9	14	32	11	12	15
I76	12	12	11	10	6	14	35	13	12	20
I77	10	11	10	12	9	15	34	15	11	20
I78	10	12	11	14	9	15	30	15	12	22
I79	10	12	11	10	9	15	34	14	11	16
I80	12	13	12	12	11	14	35	14	13	18
I81	10	10	11	12	10	14	30	13	13	16
I82	13	9	14	13	6	14	32	13	14	22
I83	11	11	11	12	9	14	32	12	14	15
I84	10	12	10	10	9	14	30	13	10	22
I85	11	13.3	11	12	10	15	33	14	11	20
I86	11	13	9	12	8	14	33	14	13	20
I87	11	11	11	12	8	14	32	14	12	22
I88	12	10	11	9	9	14	31	13	10	21
I89	11	12	11	12	8	16	30	14	13	15
I90	11	10	13	12	8	15	33	14	13	23
I91	12	12	11	13	8	14	36	11	11	21
I92	10	12	10	9	9	16	30	13	11	15
I93	11	10	12	12	6	17	33	12	12	15
I94	12	13.3	10	12	8	15	29	13	12	15
I95	11	14	11	12	8	12	30	14	10	15
I96	10	10	10	10	9	14	32	14	11	22
I97	11	12	11	12	9	15	30	13	9	15
I98	11	7	10	9	9	14	34	13	11	16
I99	11	12	11	9	6	14	34	13	9	15
I100	11	11	11	9	11	12	35	12	11	15
I101	10	14.3	12	9	9	14	31	14	12	16
1H	11	12	11	9	10	15	34	15	10	15
2H	12	12	11	9	8	16	32	16	11	21
3H	12	14	11	13	6	14	32	13	13	20
4H	12	10	10	13	9	13	30	16	11	16
5H	12	12	10	13	6	14	31	13	12	15
6H	12	14	11	9	8	15	34	14	12	21
7H	13	13	13	9	8	15	30	14	12	15
8H	10	13	9	11	6	14	32	14	12	20
9H	10	12	10	12	6	15	34	13	12	19
10H	11	10	11	13	8	14	35	12	12	17
11H	12	10	11	12	9	14	30	15	10	15
12H	12	12	10	13	10	14	35	13	10	20
13H	12	12	12	13	10	14	31	13	13	15
14H	10	12	11	13	9	14	32	12	12	15
15H	10	12	10	8	6	14	35	13	11	15
16H	11	11	11	12	10	15	35	14	9	21
17H	11	13	10	7	9	15	34	14	9	20
18H	10	12	11	9	9	14	32	17	12	21
19H	11	8.3	11	12	6	14	34	13	11	19
20H	12	13.3	10	12	6	14	32	13	11	15
21H	12	8	12	7	6	14	32	12	11	15
22H	12	11	11	13	9	14	32	14	11	23

23H	11	12	12	11	9	14	33	15	11	20
24H	13	12	11	12	7	14	31	15	11	15
25H	12	13	11	13	6	13	35	15	11	20
26H	13	13	8	10	8	15	34	12	12	15
27H	12	10	11	12	9	12	32	14	12	16
28H	10	13	11	13	9	14	35	13	13	16
29H	12	12	11	14	9	14	34	14	9	20
30H	10	12	11	9	7	15	32	16	12	22
31H	13	12	8	12	7	15	35	15	10	16
32H	0	12	11	14	6	14	31	14	13	19
33H	10	13	11	10	8	16	33	14	11	15
34H	10	11	10	14	6	14	32	14	11	15
35H	12	13	10	10	9	14	33	13	12	20
36H	12	12	10	9	6	14	33	13	10	15
37H	12	13	11	12	9	14	34	14	12	20
38H	10	10	10	14	7	16	30	12	12	19
43H	14	13.3	11	12	6	13	34	13	12	15
44H	11	11	11	13	6	14	36	16	11	19
45H	11	12	13	9	6	14	33	13	11	15
46H	12	12	11	13	6	14	32	14	11	22
47H	10	11	11	13	6	15	34	14	12	21
48H	12	12	11	10	9	15	30	14	9	21
49H	13	10	11	14	6	14	35	12	13	21
50H	11	11	10	12	9	15	32	12	12	15
51H	11	13.3	11	12	9	14	36	15	12	20
52H	10	12	11	9	9	16	33	13	11	20
53H	11	10	10	10	6	14	32	14	11	15
54H	11	12	11	11	8	13	33	14	8	19
55H	11	12	10	10	6	15	34	15	12	16
56H	10	12	11	10	6	12	33	15	13	15
57H	11	13	9	9	6	14	34	13	12	15
58H	10	13	11	8	7	13	33	13	12	16
59H	11	12	9	14	12	14	32	13	12	20
60H	12	12	9	9	8	14	33	15	9	16
61H	11	12	10	10	9	14	32	16	12	21
62H	11	9	11	9	10	14	33	13	11	20
63H	11	13.3	11	9	9	15	33	14	12	16
64H	11	13	11	11	9	15	34	15	13	15
65H	10	13.3	11	12	9	17	34	14	12	21
66H	10	13.3	11	12	9	17	34	14	12	21
67H	10	8.3	13	12	9	14	33	14	12	21
68H	12	12	10	12	9	14	30	14	12	22
69H	12	12	11	9	7	15	32	14	8	15
70H	10	12	11	13	11	15	33	11	13	22
71H	12	13	11	9	9	14	34	12	11	16
72H	11	12	11	9	9	14	30	15	12	20
73H	11	11	9	9	9	14	33	16	12	16
74H	11	12	12	12	8	14	34	16	13	21
75H	11	12	12	11	12	14	33	14	11	15
76H	11	14	11	13	8	15	34	14	12	21
77H	11	10	11	7	9	15	33	13	12	22
78H	10	13	10	10	10	17	30	12	12	21
79H	12	13.3	11	10	9	15	33	14	12	21
80H	11	11	11	11	12	16	30	10	11	21
81H	10	12	10	12	9	14	33	11	12	21
82H	11	10	11	9	6	15	33	12	11	15
83H	12	12	10	12	6	15	33	12	11	15
84H	10	13	11	9	9	15	30	12	11	21
85H	11	12	11	13	8	15	36	12	11	21
86H	11	12	11	9	8	14	34	14	11	21
87H	12	10	11	12	8	13	32	14	9	20
88H	11	13	11	12	9	13	31	13	13	15
89H	10	11	11	13	9	17	38	13	11	20
90H	11	11	11	13	10	15	33	13	11	19

91H	11	13	12	10	6	14	33	13	10	20
92H	12	8.3	11	14	6	14	31	14	13	21
93H	11	13.3	12	12	9	15	32.1	13	11	21
94H	11	13	10	10	6	14	33	13	12	22
Y1	12	13	11	12	9	14	34	13	12	21
Y2	12	10	11	12	11	13	33	12	11	20
Y3	11	12	11	12	6	15	31	13	12	21
Y4	10	12	12	13	9	15	31	13	9	16
Y5	12	10	11	10	9	13	30	15	10	21
Y6	12	12	11	9	9	15	35	15	12	20
Y7	11	14	11	12	10	14	34	16	11	21
Y8	12	10	11	9	7	14	35	15	12	15
Y9	11	7	12	13	10	15	37	13	10	22
Y10	10	11	11	13	9	16	35	16	13	16
Y11	12	13	11	12	9	14	34	15	12	15
Y12	12	12	11	8	9	14	31	15	11	21
Y13	10	12	10	14	6	14	31	15	13	19
Y14	10	11	11	13	8	15	34	15	12	15
Y15	12	13	11	12	6	16	34	14	10	15
Y16	12	11	11	12	6	14	30	12	11	21
Y17	10	8	11	9	6	14	35	16	14	15
Y18	11	11	11	10	9	14	31.1	14	12	24
Y19	10	12	11	11	9	15	33	14	12	16
Y20	12	12	9	14	9	14	35	13	11	15
Y21	10	11	12	12	7	15	33	16	12	21
Y22	10	10	11	10	11	14	34	17	12	21
Y23	10	8.3	11	13	6	13	31	14	13	20
Y24	11	11	11	10	9	15	34	14	12	21
Y25	11	12	11	10	6	15	32	14	12	21
A020	11	13	11	13	6	15	32	16	12	16
A049	10	12	12	10	9	14	32	12	11	18
A074	11	12	12	13	7	14	31	11	13	20
A092	12	10	10	13	7	15	31.1	13	13	19
A094a	12	10	10	10	9	16	34	13	11	17
A094b	13	12	10	10	8	14	32	12	12	15
A127	11	14	9	12	6	14	33	13	11	19
A128	12	10	11	9	9	15	30	15	11	21
A134	10	12	10	12	6	13	33	14	14	20
A135	12	12	12	7	10	15	31	14	13	15
A137a	12	12	11	13	10	14	33	13	12	15
A137b	10	11	11	12	6	15	35	0	0	0
A140	11	12	13	10	7	13	33	15	10	20
A141	12	10	8	12	6	14	34	14	15	21
A142	12	11	10	13	7	17	31	14	11	15
A145	0	11	11	12	6	15	32	15	11	20
A146	14	14	10	12	9	13	34	14	11	20
A150	12	12	11	11	7	14	31	15	11	22
A153	11	12	10	10	9	16	31	14	14	23

Appendix B. Poster communication format (Appendix B) at the 28th Congress of the International Society for Forensic Genetics (9-13th of September, 2019, Prague, Czech)

26th World Congress of the International Society for Forensic Genetics

Genetic insight into Nigerian population groups using an X-chromosome decaplex system

Gomes C [1,2,3], Amorim A [1,2,3], Okolie VO [4], Keshinro SO [5], Starke A [6], Vullo C [7], Gusmão L [8], Gomes I [*1,2]

[1] i3S, Institute for Research and Innovation in Health, Porto, Portugal. [2] IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Portugal. [3] Department of Biology, Faculty of Sciences, University of Porto, Portugal. [4] Molecular Biology Research Laboratory, Lagos University Teaching Hospital (LUTH), Nigeria. [5] FCIID Annex, Nigeria Police Force, Lagos, Nigeria. [6] Institute of Medical Microbiology, Immunology and Parasitology, University Hospital of Bonn, Germany. [7] DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (EAAF), Córdoba, Argentina. [8] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil.

Figure 1. The three major ethnic groups in present-day Nigeria and the respective percentage that make up the population [1].

Introduction

Nigeria is one of the most populous countries in the world, situated in the Gulf of Guinea in West Africa. It is a highly diverse country with over 250 ethnic groups, of which the Hausa, the Yoruba and the Igbo make up 21%, 20% and 17% of the population, respectively [1]. However, their genetic diversity is still poorly understood as not many population genetic studies have been performed mainly due to the difficult access to samples.

In recent years the informative power of X chromosome short tandem repeats (X-STRs) has been used in human identification, paternity testing (mainly kinship testing) and in the establishment of databases in different populations around the world. X-STR polymorphisms have also complemented information of phylogeny-based studies that extensively use Y chromosome and mtDNA.

The X-chromosome displays a distinctive mode of transmission: in males only one copy is entirely transmitted to female descendants allowing direct haplotype reconstruction, while in females, recombination on the X chromosome shuffles genetic variation in each gametogenesis [2].

The aim of this work was to study a set of highly informative and well-characterized X-STRs in individuals from Nigeria and to construct a population allelic frequency database useful for genetic studies as well as for their application to forensic identification analysis.

Materials & Methods

The sample set of Nigeria (n= 218 male individuals) is divided into three ethnic groups: Hausa (n= 87; Afroasiatic language family), Igbo and Yoruba (n= 94 and n=37, respectively; both groups are Niger-Congo speakers).

- PCR: X-STR decaplex (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, DXS6789); conditions used according to [3]
- PCR products: ran on a polyacrylamide gel to confirm amplification of fragments
- Capillary electrophoresis of PCR products on an ABI3500 and results analyzed with GeneMapper v5 software (Applied Biosystems)
- Arlequin v3.5.2 software [4]: estimation of population genetic distances (F_{ST}) and exact test of linkage disequilibrium (LD)
- Statistic parameters for forensic efficiency evaluation: MEC in trios involving daughters (MEC_I), in father/daughter duos (MEC_{II}), power of discrimination in females (PD_F) and in males (PD_M) estimated according to Desmarais et al. [5]

Results & Discussion

Genetic distances between the three groups were computed using F_{ST} rather then R_{ST} (this approach is according to Caglià *et al.* [6], who compared F_{ST} and R_{ST} genetic distances in Sub-Saharan African populations). Based on their results and on the theoretical argument of the important role that genetic drift had in shaping relationships among African populations (rather than mutation as the principal cause of population differentiation) F_{ST} have a better performance in an African context.

Population differentiation between the three groups showed no statistically significant differences (significance level of 0.017, after Bonferroni's correction for multiple testing; $F_{ST} \leq 0.0069$; $p \geq 0.045$). This supports homogeneity among the groups of Hausa, Yoruba and Igbo for the studied ten X-chromosomal markers. Despite the fact that these results suggest a common database for Nigeria comprising the three ethnic groups, the number of individuals of each group is relatively small in the present work and **more samples and analyses are required to confirm the preliminary data here presented**. Therefore, parameters of forensic interest were estimated for the three studied ethnic groups of Nigeria for each locus (Table 1).

Linkage disequilibrium (LD) testing revealed no significant association between alleles of any pair of loci was found ($p \geq 0.011$). This lack of LD may be due to a longer evolutionary history of Sub-Saharan populations that allowed extensive recombination (Tishkoff *et al.* [7]) or merely to the low power of detecting LD with small sample sizes. Thus, caution on inferring on the association status of the present loci should be taken.

Table 1. Forensic efficiency statistics for each X-STR locus and combined values for decaplex system: MEC_I and MEC_{II}; PD_M and PD_F; and expected heterozygosity (He).

	MEC_I			MEC_{II}			PD			PD			He		
Markers	Igbo	Hausa	Yoruba	Igbo	Hausa	Yoruba	Igbo	Hausa	Yoruba	Igbo	Hausa	Yoruba	Igbo	Hausa	Yoruba
DXS8378	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS9898	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS7133	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
GATA31E08	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
GATA172D05	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS7423	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS6809	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS7132	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS9902	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS6789	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DXS6789	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Note: **The highest gene diversities were observed for loci DXS6809 (0.860) and DXS6789 (0.830) in all ethnic groups. The less discriminating locus was DXS7133 (0.298).**

At present, most of the allelic and haplotypic diversity distribution studies in sub-Saharan Africa are demonstrated by autosomal, Y-chromosomal and mtDNA markers, as the number of studies describing X-STRs diversity is very limited.

The present results on the X-STR genetic characterization of the population of Nigeria support the use of the decaplex system as a powerful tool in human identification in the establishment of complex kinship testing especially in cases where the disputed child is a female.

References

[1] T. Falola and M. M. Heaton, *A History of Nigeria*. Cambridge University Press, 2008. [2] S. F. Schaffner (2004). The X chromosome in population genetics, *Nat. Rev. Genet.* 5: 43–51. [3] L. Gusmão *et al.* (2009). A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: Data on 15 Iberian and Latin American populations, *Int. J. Legal Med.* 123: 227–234. [4] L. Excoffier and H. E. L. Lischer (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10: 564–567. [5] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault and L. Busque (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). *J. Forensic Sci.* 43: 1046–1049. [6] A. Caglià *et al.* (2003) A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F(ST) and R(ST) genetic distances. *Hum. Biol.* 75: 313–330. [7] S. A. Tishkoff and S. M. Williams (2002) Genetic analysis of African populations: human evolution and complex disease *Nat. Rev. Genet.* 3: 611–621.

Acknowledgments

FCT

Appendix C. Conference proceeding at the 28th Congress of the International Society for Forensic Genetics (9-13th of September, 2019, Prague, Czech Republic) submitted in the special issue edition of the Forensic Science International: Genetics Supplement Series

Genetic insight into Nigerian population groups using an X-chromosome decaplex system

Gomes C [1, 2, 3], Amorim A [1, 2, 3], Okolie VO [4], Keshinro SO [5], Starke A [6], Vullo C [7], Gusmão L [8], Gomes I *[1,2]

[1] - i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal.

[2] Ipatimup - Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Portugal.

[3] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Portugal.

[4] Molecular Biology Research Laboratory, Lagos University Teaching Hospital (LUTH), Nigeria.

[5] FCIID Annex, Nigeria Police Force, Lagos, Nigeria.

[6] Institute of Medical Microbiology, Immunology and Parasitology, University Hospital of Bonn, Germany.

.[7] DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (EAAF), Córdoba, Argentina.

[8] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil

ABSTRACT

African populations have higher genetic diversity representing an interesting continent for population genetic studies. A sample set from Nigeria, a West African country on the Gulf of Guinea, was genetically characterized using 10 X chromosomal short tandem repeat polymorphisms (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789). 218 unrelated male individuals belong to three of the largest ethnic groups of Nigeria, namely Igbo, Hausa and Yoruba were sampled. Although no evidence for statistically significant LD ($p \leq 0.0011$) between all pairs of markers was detected, it may be due to the low power of the exact test in small sample sizes. When studying the genetic affinity among the three studied ethnic groups, no significant differences were detected ($F_{ST} \leq 0.0069$; $p \geq 0.045$) which supports genetic homogeneity among the Nigerian groups for the studied X-chromosome decaplex markers. Parameters for forensic evaluation were also calculated demonstrating the potential application of these markers in diverse kinship scenarios in Nigeria where X-STR analyses may add value.

Keywords: Short tandem repeat; X-chromosome; Africa; Nigeria

1. Introduction

Nigeria is one of the most populous countries in the world, situated in the Gulf of Guinea in West Africa. It is a highly diverse country with over 250 ethnic groups, of which the Hausa, the Yoruba and the Igbo make up 21%, 20% and 17% of the population, respectively [1]. However, their genetic diversity is still poorly understood as not many population genetic studies have been performed. In recent years the informative power of X chromosome short tandem repeats (X-STRs) has been used in human identification, paternity testing (mainly kinship testing) and in the establishment of databases in different populations around the world. X-STR polymorphisms have also complemented information of phylogeny-based studies that extensively use Y chromosome and mtDNA. The X-chromosome displays a distinctive mode of transmission: in males only one copy is entirely transmitted to female descendants allowing direct haplotype reconstruction, while in females, recombination on the X chromosome shuffles genetic variation in each gametogenesis [2]. The aim of this work was to study a set of highly informative and well-characterized X-STRs in individuals from Nigeria and to construct a population allelic frequency database useful for genetic studies as well as for their application to forensic identification analysis.

*Corresponding author at: IPATIMUP, Tel.: +351 22 5570700; fax: +351 22 5570799. *E-mail:* igomes@ipatimup.pt (I. Gomes)

2. Material and Methods

A total of 218 unrelated males from the three main ethnic groups of Nigeria, namely Igbo (n=94), Hausa (n=87) and Yoruba (n=37), were genotyped for a decaplex X-STR system (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 and DXS6789). PCR conditions and allele assignment were according to Gusmão *et al*. [3]. Products were run in an ABI3500 Genetic Analyzer and genotyped with GeneMapper v5 software (Applied Biosystems). Arlequin v3.5.2 software [4] was used to calculate allelic frequencies, population genetic distances (F_{ST}) and to perform exact test of linkage disequilibrium. Statistic parameters for forensic efficiency evaluation, specifically MEC in trios involving daughters (MEC_T) as well as in father/daughter duos (MEC_D), power of discrimination in females (PD_F) and in males (PD_M) were estimated according to the formulas described by Desmarais *et al*. [5].

3. Results and Discussion

Genetic distance analyses between the three groups was performed using F_{ST}, based on the number of different alleles. This approach follows Caglià *et al.* [6], who compared F_{ST} and R_{ST} genetic distances in Sub-Saharan African populations. Based on their results and on the theoretical argument of the important role that genetic drift had in shaping relationships among African populations (rather than mutation as the principal cause of population differentiation) they concluded that F_{ST} had a better performance in an African context. Comparisons between the three groups showed no statistically significant differences (significance level of 0.017, after Bonferroni's correction for multiple testing; $F_{ST} \leq 0.0069$; $p \geq 0.045$). This supports homogeneity among the groups of Hausa, Yoruba and Igbo for the studied ten X-chromosomal markers. However, despite the fact that these results suggest a common database for Nigeria comprising the three groups, allelic frequencies are presented for each ethnicity here studied. This allows for future database enlargements and comparisons as the number of individuals of each group is relatively small in the present work and more samples are required to confirm the preliminary data here presented. Therefore, allele frequencies and parameters of forensic interest, for the X-STRs studied, were estimated for the three studied ethnic groups of Nigeria (Table 1).

When testing for linkage disequilibrium (LD) no significant association between alleles of any pair of loci was found ($p \geq 0.011$). This lack of LD may be due to a longer evolutionary history of Sub-Saharan populations that allowed extensive recombination (Tishkoff *et al.* [7]) or merely to the low power of detecting LD with small sample sizes.

Powers of discrimination (PD) and mean exclusion chances (MEC) were determined for each locus (Table 1). The highest locus gene diversities were observed for loci DXS6809 (0.860) and DXS6789 (0.830) in all ethnic groups (Table 1). The less discriminating locus was DXS7133 (0.298).

At present, most of the allelic and haplotypic diversity distribution studies in sub-Saharan Africa are demonstrated by autosomal, Y-chromosomal and mtDNA markers, as the number of studies describing X-STRs diversity is very limited. The present results regarding the X-STR genetic characterization of the population of Nigeria support the use of the decaplex system as a powerful tool in human identification in the establishment of complex kinship testing especially in cases where the disputed child is a female.

Acknowledgments

I.G. is funded by the "FCT Scientific Employment Stimulus - Individual Call" (CEECIND/02609/2017).

Conflict of interest

None.

References

[1] T. Falola and M. M. Heaton, *A History of Nigeria*. Cambridge University Press, 2008.
[2] S. F. Schaffner (2004). The X chromosome in population genetics, *Nat. Rev. Genet.* 5: 43–51.
[3] L. Gusmão *et al*. (2009). A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: Data on 15 Iberian and Latin American populations, *Int. J. Legal Med.* 123: 227–234.
[4] L. Excoffier and H. E. L. Lischer. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10: 564–567.
[5] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, and L. Busque (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). *J. Forensic Sci.* 43: 1046–1049.

[6] A. Caglia *et al.* (2003) A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F(ST) and R(ST) genetic distances. *Hum. Biol.* 75: 313–330.
[7] S. A. Tishkoff and S. M. Williams (2002) Genetic analysis of African populations: human evolution and complex disease *Nat. Rev. Genet.* 3: 611–621.

Table 1. Allele frequencies for 10 X-STRs in three ethnic groups of Nigeria and forensic efficiency statistics for each locus: mean exclusion chance in father/daughter duos (MEC_D) and trios involving daughtBusey

DXS8378	IGBO	HAUSA	YORUBA
10	0.245	0.253	0.243
11	0.468	0.391	0.297
12	0.255	0.299	0.405
13	0.032	0.057	0.027
14	–	–	0.027
MEC_D	**0.444**	**0.484**	**0.482**
MEC_T	**0.592**	**0.630**	**0.628**
PD_M	**0.655**	**0.690**	**0.687**
PD_F	**0.818**	**0.844**	**0.843**

DXS9898	IGBO	HAUSA	YORUBA
7	0.021	–	0.027
8.3	0.032	0.034	0.027
9	0.011	0.011	-
10	0.149	0.115	0.135
11	0.191	0.103	0.189
12	0.362	0.414	0.405
13	0.117	0.207	0.135
13.3	0.053	0.080	–
14	0.053	0.034	0.081
14.3	0.011		–
MEC_D	**0.638**	**0.589**	**0.590**
MEC_T	**0.764**	**0.724**	**0.725**
PD_M	**0.789**	**0.753**	**0.756**
PD_F	**0.930**	**0.910**	**0.910**

DXS7133	IGBO	HAUSA	YORUBA
8	0.011	0.023	–
9	0.021	0.057	0.054
10	0.149	0.218	0.189
11	0.681	0.598	0.568
12	0.106	0.069	0.162
13	0.021	0.034	0.027
14	0.011	–	–
MEC_D	0.324	0.395	0.418
MEC_T	0.469	0.543	0.567
PD_M	0.298	0.585	0.612
PD_F	0.719	0.786	0.805

GATA31E08	IGBO	HAUSA	YORUBA
7	0.032	0.023	0.027
8	0.011	0.023	0.027
9	0.170	0.230	0.054
10	0.138	0.138	0.270
11	–	0.069	0.054
12	0.404	0.241	0.270
13	0.181	0.207	0.243
14	0.053	0.069	0.054
15	0.011	–	–
MEC_D	**0.584**	**0.671**	**0.624**
MEC_T	**0.719**	**0.791**	**0.753**
PD_M	**0.752**	**0.817**	**0.785**
PD_F	**0.906**	**0.941**	**0.922**

GATA172D05	IGBO	HAUSA	YORUBA
6	0.255	0.276	0.216
7	0.032	0.069	0.189
8	0.234	0.149	0.054
9	0.309	0.379	0.405
10	0.074	0.080	0.081
11	0.085	0.011	0.054
12	0.011	0.034	–
MEC_D	**0.602**	**0.571**	**0.567**
MEC_T	**0.735**	**0.708**	**0.705**
PD_M	**0.771**	**0.746**	**0.741**
PD_F	**0.911**	**0.897**	**0.896**

DXS7423	IGBO	HAUSA	YORUBA
8	0.011	–	–
12	0.021	0.023	–
13	0.064	0.069	0.135
14	0.553	0.506	0.405
15	0.255	0.299	0.324
16	0.085	0.057	0.108
17	0.011	0.046	0.027
MEC_D	**0.421**	**0.445**	**0.505**
MEC_T	**0.568**	**0.591**	**0.649**
PD_M	**0.617**	**0.644**	**0.700**
PD_F	**0.805**	**0.820**	**0.859**

DXS6809	IGBO	HAUSA	YORUBA
27	0.011	–	–
28	0.011	–	–
29	0.043	–	–
30	0.181	0.115	0.054
31	0.096	0.069	0.270
31.1	–	-	0.054
32	0.128	0.184	0.108
32.1	–	0.011	–
33	0.191	0.276	0.135
34	0.191	0.207	0.243
35	0.096	0.092	0.108
36	0.021	0.034	–
37	0.021	–	0.027
38	0.011	0.011	–
MEC_D	**0.737**	**0.677**	**0.679**
MEC_T	**0.840**	**0.796**	**0.797**
PD_M	**0.856**	**0.819**	**0.820**
PD_F	**0.963**	**0.944**	**0.945**

DXS7132	IGBO	HAUSA	YORUBA
10	–	0.011	–
11	0.0319	0.023	0.027
12	0.0638	0.149	0.108
13	0.2872	0.276	0.216
14	0.4255	0.310	0.270
15	0.1596	0.138	0.243
16	0.0319	0.092	0.108
17	–	–	0.027
MEC_D	**0.514**	**0.612**	**0.640**
MEC_T	**0.658**	**0.744**	**0.766**
PD_M	**0.705**	**0.777**	**0.797**
PD_F	**0.866**	**0.917**	**0.929**

DXS9902	IGBO	HAUSA	YORUBA
8	–	0.023	–
9	0.085	0.069	0.027
10	0.074	0.069	0.108
11	0.362	0.310	0.324
12	0.309	0.391	0.351
13	0.149	0.126	0.162
14	0.021	0.011	0.027
MEC_D	**0.558**	**0.542**	**0.548**
MEC_T	**0.696**	**0.682**	**0.688**
PD_M	**0.738**	**0.725**	**0.732**
PD_F	**0.890**	**0.882**	**0.884**

DXS6789	IGBO	HAUSA	YORUBA
15	0.277	0.264	0.216
16	0.085	0.115	0.108
17	–	0.011	0.027
18	0.032	–	0.027
19	0.074	0.080	0.081
20	0.202	0.195	0.162
21	0.149	0.253	0.270
22	0.160	0.069	0.054
23	0.021	0.011	0.027
24	–	–	0.027
MEC_D	**0.679**	**0.652**	**0.696**
MEC_T	**0.797**	**0.776**	**0.810**
PD_M	**0.821**	**0.804**	**0.830**
PD_F	**0.944**	**0.934**	**0.951**

Appendix D. Published Allelic frequencies of 10 X-STRs in African populations. Ang: Angola; Moz: Mozambique; Alg: Algeria; IC: Ivory Coast; Egy: Egypt; Gha: Ghana; GB: Guinea-Bissau; Mor: Morocco; Tun: Tunisia; Uga: Uganda

DXS8378	Ang	Moz	Alg	IC	Gha	GB	Mor	Tun	Uga
8									0.003
9		0.009	0.009	0.02	0.006	0.003	0.004		
10	0.189	0.295	0.335	0.175	0.247	0.304	0.297	0.345	0.193
11	0.378	0.348	0.345	0.365	0.378	0.358	0.301	0.318	0.407
12	0.324	0.321	0.275	0.375	0.319	0.313	0.354	0.318	0.333
13	0.108	0.027	0.028	0.05	0.036	0.021	0.033	0.016	0.064
14			0.006	0.015	0.014		0.012	0.004	

DXS9898	Ang	Moz	Gha	Mor	Tun	Uga
6.3					0.028	
7	0.014	0.018	0.008	0.004	0.004	0.048
8.3	0.095	0.125	0.076	0.167	0.250	0.064
9	0.014			0.004	0.016	0.010
10	0.122	0.125	0.137	0.033	0.056	0.069
11	0.162	0.143	0.177	0.171	0.171	0.181
12	0.378	0.304	0.401	0.362	0.318	0.433
13	0.149	0.214	0.151	0.187	0.131	0.132
13.3				0.008	0.004	0.005
14	0.054	0.071	0.042	0.061	0.016	0.051
14.3						0.003
15	0.014		0.003	0.004	0.008	0.003
15.3						0.003
16			0.006			

DXS7133	Egy	Gha	Mor	Tun	Uga
6					0.005
7		0.006		0.004	
8	0.002	0.011	0.004		
9	0.460	0.028	0.289	0.337	0.059
10	0.086	0.247	0.215	0.119	0.094
11	0.392	0.569	0.394	0.480	0.682
12	0.051	0.101	0.073	0.040	0.089
13	0.010	0.025	0.020	0.008	0.041
14		0.014	0.004	0.012	0.031

GATA31E08	Egy	Mor	Tun	Uga
7	0.211		0.016	0.046
8	0.023	0.008		0.005
9	0.214	0.244	0.266	0.181
10	0.222	0.024	0.028	0.069
11	0.259	0.138	0.155	0.084
12	0.060	0.207	0.194	0.303
13	0.010	0.244	0.198	0.234
14		0.126	0.139	0.074
15		0.008	0.004	0.005

GATA172 D05	Ang	Moz	Alg	IC	Egy	Mor	Tun	Uga
5			0.006			0.004		
6	0.284	0.25	0.184	0.226	0.234	0.236	0.214	0.239
7	0.054	0.089	0.006	0.080	0.006	0.004	0.024	0.048
8	0.095	0.179	0.193	0.120	0.111	0.199	0.159	0.234
8.2						0.004		
9	0.405	0.33	0.101	0.437	0.080	0.077	0.068	0.290
10	0.135	0.098	0.253	0.105	0.283	0.224	0.218	0.099
11	0.027	0.054	0.184	0.020	0.211	0.159	0.246	0.069
12			0.073	0.010	0.072	0.094	0.064	0.020
13							0.008	
14					0.004			

DXS7423	Ang	Moz	Alg	IC	Egy	Gha	GB	Mor	Tun
8	0.027								
9									
10									
12	0.027	0.009		0.020			0.003		
13	0.122	0.063	0.038	0.121	0.023	0.132	0.042	0.041	0.036
14	0.473	0.518	0.332	0.487	0.359	0.499	0.596	0.337	0.353
15	0.284	0.321	0.421	0.322	0.419	0.280	0.283	0.431	0.381
16	0.068	0.08	0.193	0.050	0.172	0.087	0.066	0.159	0.218
17		0.009	0.016		0.027	0.003	0.009	0.033	0.012
18									
19									

DXS6809	Ang	Moz	Alg	IC	Egy	Mor	Tun	Uga
26						0.004		
27			0.006			0.004		0.005
28	0.041	0.071	0.006	0.015	0.018	0.020	0.016	0.025
29	0.054	0.027	0.044	0.010	0.035	0.024	0.016	0.053
30	0.122	0.063	0.104	0.070	0.062	0.085	0.068	0.122
30.1		0.027						
31	0.149	0.071	0.139	0.200	0.135	0.187	0.079	0.247
31.1		0.009						
32	0.081	0.107	0.13	0.155	0.152	0.155	0.187	0.173
32.1		0.009						
33	0.243	0.161	0.253	0.195	0.275	0.232	0.254	0.160
33.1		0.009						
34	0.189	0.223	0.18	0.220	0.197	0.179	0.234	0.143
34.2			0.003					
35	0.108	0.152	0.092	0.075	0.076	0.069	0.107	0.033
35.1						0.004		
36	0.014	0.036	0.029	0.030	0.041	0.033	0.032	0.025
36.1		0.018						
37		0.018	0.009	0.015	0.008		0.008	0.013
38				0.015	0.002			

DXS7132	Ang	Moz	Alg	IC	Egy	Gha	GB	Mor	Tun	Uga
9	0.014									
10						0.017	0.006	0.008		0.010
11	0.014	0.009	0.032	0.010	0.002	0.028	0.012	0.020	0.036	0.092
12	0.081	0.089	0.111	0.075	0.016	0.101	0.111	0.110	0.079	0.328
13	0.23	0.188	0.301	0.261	0.127	0.266	0.298	0.272	0.274	
13.3								0.004		0.257
14	0.392	0.393	0.38	0.372	0.244	0.350	0.380	0.329	0.361	
14.3										0.242
15	0.23	0.25	0.139	0.246	0.343	0.188	0.154	0.215	0.206	
15.3										0.069
16	0.041	0.054	0.038	0.030	0.209	0.039	0.039	0.037	0.032	
16.3										0.003
17		0.018		0.005	0.058	0.006		0.004	0.008	
18					0.002	0.006			0.004	

DXS9902	Egy	Mor	Tun	Uga
7	0.002			
8	0.010			
9	0.058	0.016	0.036	0.051
10	0.298	0.045	0.040	0.102
11	0.374	0.293	0.282	0.267
11.1		0.008		
12	0.251	0.390	0.397	0.397
12.1		0.029		
13	0.006	0.211	0.226	0.158
14		0.004	0.020	0.023
15		0.004		0.003

DXS6789	Ang	Moz	Alg	IC	Egy	Mor	Tun	Uga
14	0.014	0.009	0.009		0.014	0.012	0.012	
15	0.176	0.179	0.111	0.275	0.099	0.138	0.091	0.294
16	0.176	0.125	0.032	0.085	0.023	0.033	0.020	0.078
17		0.009	0.025		0.002	0.029	0.020	0.059
18	0.014			0.025	0.006	0.008		0.020
19	0.081	0.134	0.041	0.085	0.051	0.057	0.024	0.157
20	0.095	0.152	0.329	0.170	0.419	0.285	0.361	0.177
21	0.23	0.152	0.168	0.185	0.168	0.207	0.210	0.137
22	0.162	0.143	0.184	0.120	0.136	0.134	0.175	0.059
23	0.027	0.098	0.095	0.050	0.072	0.081	0.071	0.020
24	0.027		0.006	0.005	0.010	0.012	0.016	
25						0.004		