



Universidade do Estado do Rio de Janeiro
Centro Biomédico
Instituto de Biologia Roberto Alcântara Gomes

Masinda Pascalina Rafael Nguidi

Herança genética materna dos principais grupos étnicos da Nigéria

Rio de Janeiro

2021

Masinda Pascalina Rafael Nguidi

Herança genética materna dos principais grupos étnicos da Nigéria

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof^a. Dra. Leonor Gusmão

Coorientadora: Dra. Filipa Isabel Figueira Simão

Rio de Janeiro

2021

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/BIBLIOTECA CB-A

N684 Nguidi, Masinda Pascalina Rafael.
Herança genética materna dos principais grupos étnicos da Nigéria / Masinda
Pascalina Rafael Nguidi. - 2021.
113 f.

Orientadora: Prof.^a Dra. Leonor Gusmão
Coorientadora: Dra. Filipa Isabel Figueira Simão

Mestrado (Dissertação) - Universidade do Estado do Rio de Janeiro, Instituto de
Biologia Roberto Alcântara Gomes. Pós-graduação em Biociências.

1. Hereditariedade – Teses. 2. Nigéria – Teses. 3. Genética da população humana
- Teses. 4. Grupo com Ancestrais Nativos do Continente Africano. I. Gusmão,
Leonor. II. Simão, Filipa Isabel Figueira. III. Universidade do Estado do Rio de
Janeiro. Instituto de Biologia Roberto Alcântara Gomes. IV. Título.

CDU 616.411

Bibliotecária: Ana Rachel Fonseca de Oliveira
CRB7/6382

Autorizo apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta
dissertação, desde que citada a fonte.

Assinatura

Data

Masinda Pascalina Rafael Nguidi

Herança genética materna dos principais grupos étnicos da Nigéria

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Aprovada em 01 de setembro de 2021.

Orientadora: Prof.^a Dra. Leonor Gusmão
Instituto de Biologia Roberto Alcântara Gomes – UERJ

Coorientadora: Dra. Filipa Isabel Figueira Simão
Instituto de Biologia Roberto Alcântara Gomes – UERJ

Banca Examinadora: _____
Prof.^a Dra. Cíntia Barros Santos Rebouças
Instituto de Biologia Roberto Alcântara Gomes – UERJ

Prof.^a Dra. Ândrea Kely Campos Ribeiro dos Santos
Universidade Federal do Pará

Prof.^a Dra. Cynthia Chester Cardoso
Universidade Federal do Rio de Janeiro

Rio de Janeiro

2021

DEDICATÓRIA

Às mulheres mais fortes que conheço e às suas linhagens. Ao avô Rafael, por ser grande fonte de inspiração. *Matondo mu lusingu lwaku *nsambo ye malau**

AGRADECIMENTOS

À *Alma mater*, UERJ, por resistir a tantos percalços durante os últimos anos, possibilitando meu crescimento e formação como indivíduo. Grata pelos movimentos revolucionários feitos por esta instituição e por nutrir seus discentes mesmo com os infortúnios ocorridos no último ano com a pandemia.

A meus pais e principais fomentadores pela dedicação, apoio e por serem exemplo de integridade. Às minhas irmãs e melhores amigas, por apenas existirem.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por viabilizar a produção científica através das bolsas de Pós-Graduação.

À Prof^ª. Dr^ª. Leonor Gusmão, pelo aceite à orientação deste trabalho, pelos questionamentos, ensinamentos, cuidado e pelo entusiasmo. Grata por ter a oportunidade de aprender com uma profissional tão dedicada e humana, virtudes cada vez mais raras atualmente.

À Dr^ª. Filipa Simão, pelo aceite à coorientação, pela disponibilidade, conselhos, apoio de sempre e pelo exemplo enquanto profissional acadêmico. Por tornar o trabalho mais agradável e leve.

À intrépida trupe, pela companhia, trocas e paciência, desde a Graduação. E aos outros muitos colegas que tive a oportunidade de conhecer durante esta jornada.

Aos técnicos e alunos do Laboratório de Diagnósticos por DNA, pela companhia, pelos aprendizados e disponibilidade.

Quando eu me for, restarei você

Essler

RESUMO

NGUIDI, Masinda Pascalina Rafael. **Herança genética materna dos principais grupos étnicos da Nigéria**. 2021. 113 f. Dissertação (Mestrado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

A África é o local de origem do homem moderno e comporta hoje a maior diversidade genética, cultural e étnica do mundo. Por isto, tem-se visto cada vez mais estudos, tanto de genética populacional como antropológicos, sobre populações africanas. Parte desta diversidade pode ser observada na Nigéria (também conhecida como o “Gigante da África”), o país mais populoso do continente e sétimo mais populoso no mundo, com cerca de 250 grupos etnolinguísticos distribuídos por seu território. As particularidades da molécula de DNA mitocondrial, como a ausência de recombinação, altas taxas de mutação e transmissão uniparental, tem a tornado cada vez mais útil, essencialmente para estudos de evolução humana, genética de populações e investigações no âmbito forense. Além disto, estudos acerca da genética populacional são cada vez mais atrativos, despertando o interesse de indivíduos que buscam informações sobre a sua ancestralidade, como por exemplo, sobre o provável local de origem dos seus antepassados, antes da Diáspora negra. Este trabalho teve como objetivo caracterizar a diversidade genética de linhagens maternas dos três principais grupos etnolinguísticos da Nigéria, de forma a avaliar se a variabilidade genética encontrada possibilita a diferenciação étnica / etnolinguística de suas populações. Com estes dados, pretendeu-se também aumentar a informação disponível em base de dados forenses sobre linhagens maternas de populações africanas. Assim, amostras de 285 indivíduos do sexo masculino foram analisadas para a região controle do DNA mitocondrial. Foram encontrados altos valores de diversidade haplotípica nos três grupos: Hausa ($HD = 0,9996 \pm 0,0024$), Yoruba ($HD = 0,9992 \pm 0,0013$) e Igbo ($HD = 0,9984 \pm 0,0017$). Os haplótipos encontrados pertencem majoritariamente ao macrohaplogrupo L (98%), característico de populações da África Subsaariana. Foram ainda encontrados quatro haplogrupos característicos do Norte da África e Eurásia, presentes, em sua maioria, no grupo Hausa. A comparação dos três grupos étnicos da Nigéria entre si não demonstrou diferenças estatisticamente significativas ($p \geq 0,5$), como pôde ser observado pelas altas probabilidades de não diferenciação das amostras populacionais, associadas a valores negativos de F_{ST} . As distâncias genéticas encontradas na comparação com outras populações africanas também demonstraram proximidade dos três grupos etnolinguísticos com populações de países geograficamente próximos da Nigéria, ainda que não pertencentes à mesma família linguística. Tais resultados reforçam a persistência de um costume social / institucional que molda as relações entre as populações, bem como, a homogeneização das linhagens maternas das populações desta região do continente africano. Esta homogeneização reflete uma maior migração das mulheres, possivelmente promovida pela patrilocalidade.

Palavras-chave: Nigéria. DNA mitocondrial. África. Ancestralidade materna. Hausa. Yoruba.

Igbo.

ABSTRACT

NGUIDI, Masinda Pascalina Rafael. **Maternal inheritance of the main ethnic groups of Nigeria**. 2021. 113 f. Dissertação (Mestrado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

Africa is the birthplace of modern humans, being the region of the world that today houses the highest genetic, cultural and ethnic diversities. Therefore, an increasing number of studies focusing on African populations have been carried out in the fields of population genetics and anthropology. Part of this diversity can be seen in Nigeria (also known as the “Giant of Africa”), which is the most populous country in the continent and the seventh most populous country in the world, with about 250 ethnolinguistic groups spread across its territory. Specific features of the mitochondrial DNA, such as lack of recombination, high mutation rates and uniparental transmission, have made this molecule increasingly useful in a broad field of applications, such as human evolution, population genetics and forensic studies. In addition, studies on the genetic diversity among human populations are increasingly widespread, arousing the interest of those seeking information about their ancestry, such as the most likely place of origin of their ancestors, before the Black Diaspora. The present work aimed to characterize the genetic diversity of the maternal lineages in three main ethnolinguistic groups in Nigeria. Herein, we also explored whether the genetic variability found allows the differentiation of populations. Besides, we also intended to increase data in forensic databases, regarding the maternal ancestry of African populations. Samples belonging to 285 males were investigated for the mitochondrial DNA control region. High haplotype diversity values were found in the three groups: Hausa ($HD = 0.9996 \pm 0.0024$), Yoruba ($HD = 0.9992 \pm 0.0013$) and Igbo ($HD = 0.9984 \pm 0.0017$). Most haplotypes found belong to macrohaplogroup L (98%), characteristic of Sub-Saharan African populations. Four non-L haplogroups were also found, which are characteristic of North Africa and Eurasia, two of them were present in the Hausa group. The comparison of the three ethnic groups in Nigeria did not show statistically significant differences ($p \geq 0,5$), with high non-differentiation probabilities and negative F_{ST} values between population samples. The genetic distances calculated between Nigerian and other African populations showed a proximity of the three ethnolinguistic groups with populations from countries geographically close to Nigeria, even from different linguistic families. Such results reinforce the persistence of a social/institutional pattern that shaped the relationships between populations, as well as, the homogenization of the maternal lineages of the populations in this region of Africa. This homogeneity reflects a higher female migration, probably motivated by patrilocality practices.

Keywords: Nigeria. Mitochondrial DNA. Africa. Maternal ancestry. Hausa. Yoruba. Igbo.

LISTA DE FIGURAS

Figura 1 – Migração do homo moderno pelo globo explicada pela Teoria <i>Out of Africa</i>	13
Figura 2 – Migrações e grupos etnolinguísticos da África.....	15
Figura 3 – Representação das rotas de expansão Bantu	16
Figura 4 – Continente africano.....	18
Figura 5 – Nigéria pré-colonial.....	20
Figura 6 – Mapa com os maiores grupos etnolinguísticos da Nigéria	23
Figura 7 – Localização da Nigéria e países de fronteira	24
Figura 8 – Teoria endossimbiótica	29
Figura 9 – A mitocôndria e suas estruturas	30
Figura 10 – DNA mitocondrial	31
Figura 11 – Replicação da molécula de DNA mitocondrial	32
Figura 12 – Heredograma do padrão de herança do mtDNA	35
Figura 13 - Comparação de sequências heteroplásmicas.....	36
Figura 14 - Padrões de distribuição do homem moderno permitiu a construção da filogenia do DNA mitocondrial e determinação de haplogrupos.....	37
Figura 15 - Árvore filogenética do DNA mitocondrial e seus haplogrupos	38
Figura 16 - Método de extração química por resina Chelex	43
Figura 17 - Método de extração orgânica por Fenol-Clorofórmio	44
Figura 18 - Alterações de temperatura do termociclador durante a amplificação do DNA	46
Figura 19 - Alterações de temperatura do termociclador durante a reação de sequenciamento	47
Figura 20 - Esquema de purificação por colunas de Sephadex	49
Figura 21 – Representação do número de diferenças médias e network Hausa	55
Figura 22 – Representação do número de diferenças médias e network Yoruba	56
Figura 23 – Representação do número de diferenças médias e network Igbo	57
Figura 24 – Distribuição de amostras do oeste da África usadas para comparação	59
Figura 25 - Representação gráfica por MDS dos valores de distância genética entre os três grupos etnolinguísticos e populações do Centro-oeste da África .	62

LISTA DE ILUSTRAÇÕES

Quadro 1 –	Comparação entre os genomas nuclear e mitocondrial	27
Quadro 2 –	Sequências dos oligonucleotídeos utilizados para PCR	45
Quadro 3 –	Protocolo de amplificação	46
Quadro 4 –	Protocolo para cada reação de sequenciamento	47
Quadro 5 -	Sequências dos oligonucleotídeos utilizados para sequenciamento	48
Quadro 6 -	Populações da região centro-oeste usadas para comparação e seus respectivos valores de diversidade haplotípica	54
Gráfico 1 -	Distribuição dos haplogrupos de mtDNA deste estudo	58
Tabela 1 -	Distâncias genéticas (F_{ST}) pareadas (abaixo da diagonal) e valores de probabilidade de não diferenciação (acima da diagonal) entre as populações do centro-oeste.....	61

LISTA DE ABREVIATURAS E SIGLAS

ATP	Adenosina trifosfato (<i>Adenosine TriPhosphate</i>)
CR	Região controle (<i>Control Region</i>)
CRS	Sequência de Referência de Cambridge (<i>Cambridge Reference Sequence</i>)
DNA	Ácido desoxirribonucleico (<i>DeoxyriboNucleic Acid</i>)
dNTP	Desoxirribonucleotídeos fosfatados (<i>deoxynucleotides</i>)
ddNTP	Didesoxirribonucleotídeos fosfatados (<i>dideoxynucleotides</i>)
EAAF	<i>Equipo Argentino de Antropología Forense</i>
EDNAP	<i>The European DNA Profiling group</i>
EMPOP	<i>EDNAP forensic MtDNA POPulation database</i>
HVI/HV1	Região hipervariável I (<i>HiperVariable Segment I</i>)
HVII/HV2	Região hipervariável II (<i>HiperVariable Segment II</i>)
HVIII/HV3	Região hipervariável III (<i>HiperVariable Segment III</i>)
MDS	Escala multidimensional (<i>Multidimensional scaling</i>)
MRCA	Ancestral comum mais recente (<i>Most Recent Common Ancestor</i>)
mtDNA	DNA mitocondrial (<i>mitochondrial DNA</i>)
pb	par de bases
PCR	Reação em cadeia da polimerase (<i>Polymerase Chain Reaction</i>)
rCRS	Sequência de Referência de Cambridge revista (<i>revised Cambridge Reference Sequence</i>)
RFLP	Polimorfismo de Comprimento de Fragmento de Restrição (<i>Restriction Fragment Length Polymorphism</i>)
RNA	Ácido ribonucleico (<i>RiboNucleic Acid</i>)
rRNA	RNA ribossomal (<i>ribosomal RNA</i>)
SAP	Fosfatase alcalina recombinante de camarão (<i>Shrimp Recombinant Alkaline Phosphatase</i>)
SDS	Dodecil sulfato de sódio (<i>sodium dodecylsulfate</i>)
SNP	Polimorfismo de Nucleotídeo Único (<i>Single Nucleotide Polymorphism</i>)
tRNA	RNA transportador (<i>transfer RNA</i>)

LISTA DE SÍMBOLOS

a.C	Antes de Cristo
d.C	Depois de Cristo
%	Porcentagem
°	Grau
Km	Kilômetro
Km ²	Kilômetro quadrado
µm	Micrômetro
°C	Graus Celsius
pH	Potencial hidrogeniônico
mL	Mililitro
µM	Micromolar
ng	Nanograma
µL	Micro litro
H ₂ O	Molécula da água
<i>HD</i>	Diversidade haplotípica (<i>Haplotype diversity</i>)
<i>F_{ST}</i>	Índice de fixação (<i>Fixation Index</i>)
<i>p</i>	<i>p</i> -valor
±	Mais ou menos
~	Aproximadamente

SUMÁRIO

	INTRODUÇÃO	13
1	OBJETIVOS	41
1.1	Geral	41
1.2	Específicos	41
2	MATERIAL E MÉTODOS	42
2.1	Etapas experimentais	42
2.2	Amostras populacionais	42
2.3	Metodologias de extração	42
2.4	Amplificação	44
2.5	Purificação do produtos amplificados	46
2.6	Sequenciamento de Sanger	47
2.7	Purificação dos produtos sequenciados	48
2.8	Análise das sequências	49
2.8.1	<u>Determinação de haplogrupos</u>	50
2.9	Análises estatísticas	50
2.9.1	<u>Análises de diferenciação populacional</u>	51
3	RESULTADOS	52
3.1	Diversidade haplotípica e de haplogrupos	52
3.2	Análises de diferenciação populacional	58
4	DISCUSSÃO	63
	CONCLUSÕES	69
	REFERÊNCIAS	70
	ANEXO - Lista dos haplótipos encontrados na amostra dos 285 indivíduos totais não relacionados da Nigéria.....	79
	APÊNDICE - The maternal inheritance of the three major ethnic groups of Nigeria (Artigo).....	85

INTRODUÇÃO

TEORIAS DE ORIGEM DO HOMEM MODERNO

A origem evolutiva do homem foi proposta por Charles Darwin no século XIX, a partir de comparações de crânios humanos com os de grandes primatas. Mais tarde, hipóteses foram desenvolvidas, a fim de tentar explicar os achados de Darwin.

Sugere-se, a partir de dados genéticos e arqueológicos, que o gênero *Homo* tenha surgido há aproximadamente 2,5 milhões de anos. O homem anatomicamente moderno pode ter emergido entre 200.000 a 80.000 anos atrás, com a espécie *Homo sapiens* (FORSTER, 2004). Há duas teorias vigentes que tentam explicar tal evento, sendo que ambas convergem para que a origem tenha se dado no continente africano (SHEA, 2013).

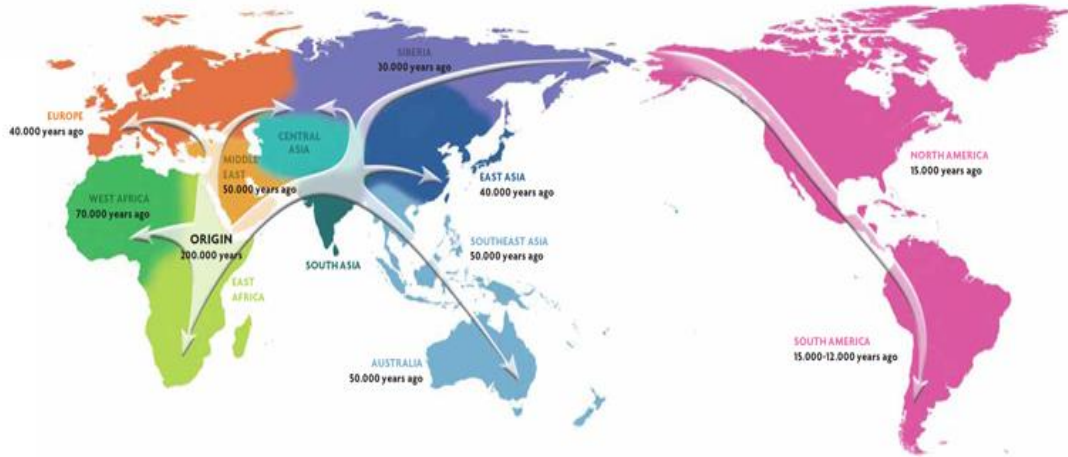
A primeira teoria, dita Teoria Multirregional, sugere que o homem tenha passado por apenas um evento de dispersão. No primeiro momento, o *Homo erectus* teria surgido na África e se dispersado em direção à Eurásia e Oceania. Dali, em locais distintos do globo e simultaneamente, o *Homo sapiens* teria surgido como resultado de forças evolutivas do homo primitivo ou arcaico (WOLPOFF, HAWKS, *et al.*, 2000; SATTA, TAKAHATA, 2002).

A segunda teoria, e atualmente a mais aceita, é denominada *Out of Africa* (**Figura 1**). Segundo esta, acredita-se que tenham ocorrido dois momentos de dispersão. No primeiro, o *Homo erectus* teria surgido na África, saído em direção a Eurásia e ali se estabelecido. As comunidades deste homo teriam evoluído de forma independente nos continentes europeu e asiático. Em paralelo, uma comunidade de *Homo sapiens* arcaico teria evoluído anatomicamente e cognitivamente na África, dando origem ao *Homo sapiens* moderno que, num segundo momento, teria se dispersado novamente em direção à Eurásia, extinguindo as espécies primitivas descendentes de *Homo erectus* que ali estavam (TATTERSALL, 2009). Hoje em dia, sabe-se que o homem moderno possui em seu genoma informações resultantes de processos de miscigenação com outros grupos, durante a sua expansão pela Eurásia, nomeadamente de Neandertais e Denisova, em proporções variáveis dependendo da região geográfica (WALL, YANG, *et al.*, 2013; SANKARARAMAN, PATTERSON, *et al.*, 2012; SANKARARAMAN, MALLICK, *et al.*, 2016).

Mais tarde, a dispersão de homínídeos modernos progrediu por todas as regiões habitáveis do velho continente e, por último, por todo o continente americano. Durante um período de deglaciação, os grupos nômades de homo puderam acessar o norte das Américas com a exposição de um trecho de terra denominado estreito de *Bering* ou Beríngea. Estes

ocuparam e povoaram as Américas, sendo o que conhecemos por ameríndios ou povos nativo-americanos (FORSTER, 2004).

Figura 1 - Migração do homo moderno pelo globo explicada pela Teoria *Out of Africa*



Legenda: O povoamento do globo pelo homem pode ser explicado pela Teoria Out of África, que sugere o surgimento da espécie humana no continente africano há 200 mil anos. Posteriormente, os outros continentes foram colonizados a partir de eventos de dispersão humana por grupos nômades.

Fonte: Comas, 2015.

ÁFRICA

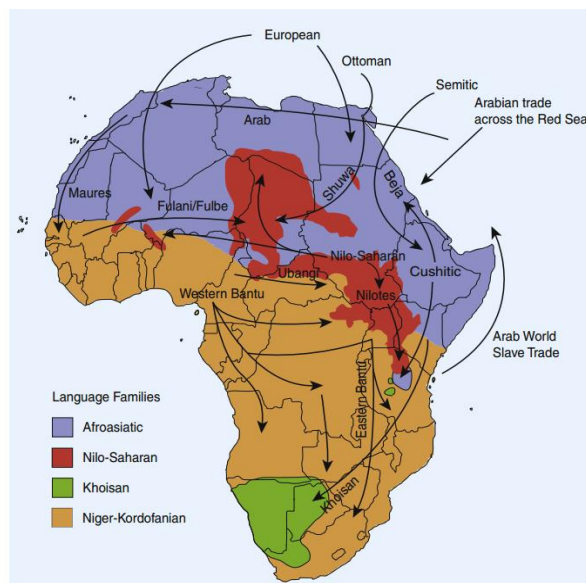
Durante os eventos de dispersão humana, podem ter ocorrido efeitos de deriva ou fundadores, que resultaram em perda de variabilidade genética, pois apenas parte da diversidade presente no local de origem foi transmitida às novas populações. Este fenômeno explica, em parte, a grande diversidade genética presente na África, em comparação àquela do restante dos continentes (COMAS, 2015; HENN, CAVALLI-SFORZA, *et al.*, 2012)

Embora áreas de estudo distintas como a paleoantropologia, a genética e a arqueologia afirmem o surgimento do homem moderno no continente africano, o local exato ainda é incerto. Evidências fósseis, datadas de aproximadamente 200.000 anos atrás, apontam a região leste como local provável de origem do homem moderno. Tal hipótese é apoiada pela árvore filogenética construída a partir de estudos do DNA mitocondrial, devido à grande diversidade encontrada no Leste para o haplogrupo africano mais ancestral - L0 (CEREZO, GUSMÃO, *et al.*, 2016). Por outro lado, outros estudos genéticos indicam o surgimento do homem no sul da África (WATSON *et al.*, 1997; FORSTER, 2004). Sendo assim, têm-se

tornado cada vez mais necessários estudos diversos acerca das populações africanas, de modo que se compreenda melhor os eventos por trás da história da humanidade (RITO, RICHARDS, *et al.*, 2013; REYES-CENTENO, 2016).

O continente africano comporta ambientes diversos como savanas, montanhas, zonas tropicais e o grande deserto do Saara. Fatores climáticos, ambientais e o modo de vida das populações africanas desempenham papel fundamental na seleção e na adaptação dos indivíduos, promovendo um aumento da diversidade fenotípica e genotípica destes grupos. Pelo fato da África ser o local de origem do homem moderno, as populações africanas são as mais antigas do mundo, o que possibilitou o acúmulo de mutações e adaptações, tornando-as mais heterogêneas. Outro fator importante que explica a sua diversidade são os movimentos migratórios que se deram dentro do continente e possibilitaram intercâmbio genético entre os povos (LAMBERT, TISHKOFF, 2009; HENN, CAVALLI-SFORZA, *et al.*, 2012) (**Figura 2**).

Figura 2 - Migrações e grupos etnolinguísticos da África



Legenda: As migrações mais importantes dentro do continente africano estão representadas por setas. As regiões destacadas em diferentes cores representam as quatro grandes famílias linguísticas presentes na África.

Fonte: Campbell e Tishkoff, 2008.

Além disto, o continente também é conhecido por sua diversidade cultural e etnolinguística. São cerca de 2.000 grupos étnicos distribuídos por toda África, compreendidos em quatro grandes famílias linguísticas: **Níger-Cordofiana** ou **Níger-Congo** (renomeada por Bendor-Samuel em 1989), falada em grande parte do território africano,

exceto no norte; **Nilo-Saariana**, que se distribui pelas regiões Central e Leste; **Afro-Asiática**, onde os grupos se localizam a norte e leste; e **Khoisan**, que se concentra ao sul do país, compreendendo as populações caçadoras-coletoras (GREENBERG, 1948; DIMMENDAAL, 2011). Nas três primeiras grandes famílias linguísticas citadas, as populações vivem da agricultura e do pastoreio.

EXPANSÃO BANTU

A distribuição da diversidade genética e fenotípica presente na África pode ser explicada pelos movimentos migratórios intracontinentais. O maior e mais expressivo deles foi a expansão Bantu. A expansão Bantu está associada à dispersão da agricultura, através da África Subsaariana, paralelamente com a disseminação da técnica do ferro/metalurgia, mediada por um grupo de língua proto-Bantu. Estima-se, com base em evidências arqueológicas e linguísticas, que este grupo tenha surgido há aproximadamente 5.000 anos nos entornos do Rios Níger e do lago Chade, onde estão compreendidos Camarões e Nigéria (GABEL, NEWMAN, 1997) (**Figura 3**).

Figura 3 - Representação das rotas de expansão Bantu



Legenda: As setas em vermelho apontam migrações mais antigas, enquanto que as setas em roxo apontam migrações tardias dos povos Bantu. A dispersão deste grupo é a mais significativa do continente pois foi responsável pela disseminação da agricultura e metalurgia, bem como na miscigenação destes povos com populações que já habitavam outras regiões.

Fonte: Google, <https://sites.google.com/site/wwwbantumigration>, agosto de 2021

Durante a Expansão Bantu, este grupo começou a se dispersar, em duas grandes ondas migratórias. A primeira onda se deu para a região Sudoeste, há 3.500 anos, e a segunda para Leste, há 3.000 anos, para a denominada área central onde é hoje o Uganda. Desta área central, ainda se deram outras dispersões para as regiões Sul e para os arredores do Lago Malawi, a leste do Zimbábue. Tais movimentos foram de grande importância para a remodelagem dos primeiros povos que viviam no Sul (Pigmeus e Khoisan, caçadores-coletores), que acabaram por perder parte do território para os Bantu que neste se fixaram. Além disso, possibilitaram interações ou fluxo genético com outros povos, como os grupos de pastores Nilotas e grupos agricultores vizinhos (GABEL, NEWMAN, 1997; PEREIRA, MACAULAY, *et al.*, 2001 ; DE FILIPPO, BOSTOEN, *et al.*, 2012).

Tendo os povos Bantu surgido na área que hoje comporta os países Nigéria e Camarões, espera-se que as populações que ainda residem nestes locais exibam a maior diversidade genética presente na África. Os locais que receberam os povos Bantu comportam, apesar do intercâmbio genético, menor diversidade genética se comparado às populações residentes no local de origem dos Bantu, resultado de efeito fundador, onde há decréscimo da variabilidade genética, como fora explicado anteriormente.

Por consequência, a expansão dos povos Bantu contribuiu fundamentalmente para a diversidade genética presente à oeste, leste e sul da África (PEREIRA, MACAULAY, *et al.*, 2001; CEREZO, GUSMÃO, *et al.*, 2016). A expansão ocorreu em momentos diversos e em direções distintas dentro do continente, tendo grande impacto no desenvolvimento e na disseminação da agricultura e da metalurgia, essencialmente para as regiões Centro-Sul da África (METZ, 1991; Commonwealth, 2021).

NIGÉRIA

O nome do país é oriundo de uma fusão entre o nome do Rio Níger, que adentra o país, e *area*, termo em inglês para 'área', numa tentativa de diferenciar o território do país vizinho Níger (METZ, 1991) (**Figura 4**).

Figura 4 - Continente Africano



Legenda: Destaque ao país Nigéria, à Centro-Oeste, em vermelho.

Fonte: Google Maps, acesso em agosto de 2021.

PERÍODO PRÉ-COLONIAL

A Nigéria possui um passado antigo. Estima-se, a partir de registros históricos, que a área fora habitada há aproximadamente 2.000 anos a.C., por povos que originaram os povos Bantu (METZ, 1991; FALOLA, HEATON, 2008).

A área hoje denominada Nigéria comportava, no ano de 600 a.C., a civilização mais antiga da África Sub-Saariana que se tem registro, a cultura *Nok* ou cultura das estatuetas de *Nok*. Conhecida mundialmente por sua arte em terracota e latão, é prova tangível da cultura tradicional local e evidência arqueológica que aponta o início das civilizações deste país. Este povo residiu na região denominada *Jos Plateau*, que se localiza entre os rios Níger e Benué, entre 300 anos a.C. e 200 anos d.C. (METZ, 1991; FALOLA, HEATON, 2008; Commonwealth, 2021) (**Figura 5**).

No século VI a.C, numa área próxima à civilização *Nok*, surgiu a cultura Daima. Na região a noroeste de Borno, ao sul do Lago Chade, esta população, constituída de pastores neolíticos, começou a se estabelecer em barracas circulares de lama. Eles também produziam

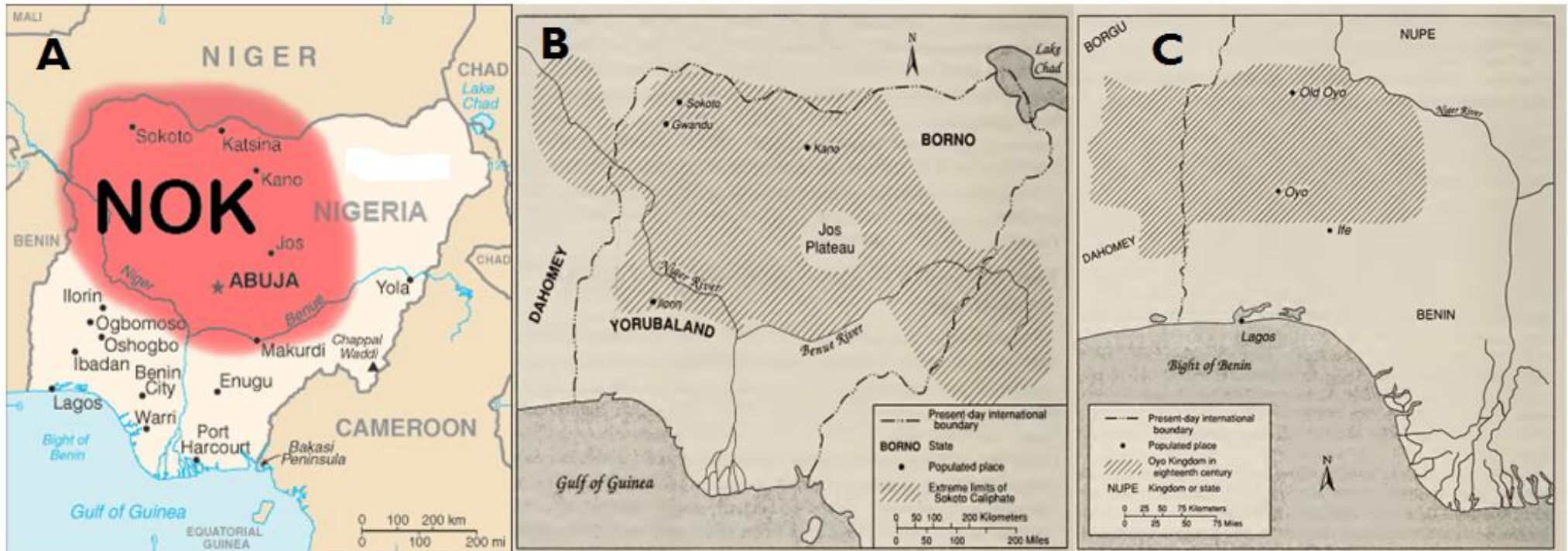
figuras de argila queimada que relatavam atividades cotidianas com reprodução de vacas, ovelhas e cabras, animais selvagens e seres humanos (METZ, 1991).

Evidências arqueológicas, tradições orais e documentos escritos estabelecem a existência de sociedades dinâmicas e sistemas políticos bem estipulados nestas épocas, que foram essenciais para a construção do país. Antes de 1.500, muito do que se tem da divisão política dos estados de Nigéria já existia, dada uma separação promovida pelos próprios grupos étnicos. Estes estados incipientes incluíam os Reinos Yoruba, o Reino Edo de Benin e as cidades de Hausa e Nupe (METZ, 1991).

A partir do ano 1.400, entre o rio Níger e o lago Chade, destacam-se dois reinos ao norte, onde hoje se localizam os povos Hausa: Hausaland e Kanem-Bornu. Adeptos do islamismo, os Hausa viviam de comércio transaariano e trocas entre árabes e povos nômades das florestas (METZ, 1991; FALOLA, HEATON, 2008).

À sul e oeste, o reino Ife originou as primeiras civilizações locais. Acredita-se que este foi fundado pelo Império Yoruba, um grande território cujo *oni* (rei) denominou *Yorubaland* a região oeste do Rio Níger. O surgimento do reino Ife e do reino Benin são convergentes, tendo culturas muito semelhantes. Ambas usavam artes plásticas com esculturas de bustos em bronze, numa reprodução de seus líderes, tidas como patrimônio artístico mundial e admiradas até hoje pelo seu realismo e elegância (METZ, 1991; FALOLA, HEATON, 2008; Commonwealth, 2021).

Figura 5 - Nigéria pré-colonial



Legenda: Provável local onde se estabeleceu o povo da cultura Nok (A); Mapa feito no século XIX da distribuição do território denominado Yorubaland e Região do Borno, onde pode ter se estabelecido a cultura Daima (B); Yorubaland com enfoque nos Reinos Oyo e Ife, próximos ao Rio Níger;

Fontes: Google, <https://i.pinimg.com/originals/c6/36/f1/c636f14b090a4a4cd535bfae4a97d15f.png>, acesso em Agosto de 2021; Metz, 1991.

PERÍODO COLONIAL

A expansão marítima no século XV promoveu a chegada de portugueses à costa africana, à procura de rotas mais curtas para as Índias. Tendo alcançado o Golfo da Guiné, eles faziam trocas entre os povos de Benin e Nigéria, num escambo entre armas, pólvoras e outros bens. Ali também estabeleceram portos costeiros que facilitavam o aumento do tráfico de africanos escravizados para as Américas, principalmente para o Caribe (METZ, 1991; FALOLA, HEATON, 2008).

O século XVIII foi marcado pela chegada de navios britânicos à costa do país, que deixou de estabelecer trocas com Portugal. Resultante do imperialismo europeu, o território nigeriano foi delimitado na Conferência de Berlin (século XIX), com o intuito de impedir a conquista do país por franceses, que tinham sua colônia (Níger) muito próxima. Não só na Nigéria, mas em muitos países africanos, a criação destes Estados artificiais teve grande impacto nas dinâmicas sociais das populações locais, tendo em vista que estas divisões desconsideraram as nações e povos ali já estabelecidos, que resultam em conflitos internos nos países até hoje (ALESINA, EASTERLY, *et al.*, 2011).

Sob posse da Coroa britânica, os povos nigerianos permaneceram sendo escravizados em número massivo e levados a outras colônias na América como fonte de mão-de-obra gratuita. O comércio transatlântico de africanos escravizados foi responsável pela retirada de aproximadamente 3,5 milhões de nigerianos, entre os séculos XVII e XIX, especialmente indivíduos pertencentes aos grupos Yoruba e Igbo, localizados ao sul, perto da costa atlântica, local de fácil negociação e deslocamento de nativos para o novo mundo (METZ, 1991; SALAS, RICHARDS, *et al.*, 2004; FALOLA, HEATON, 2008).

No século XIX, houve forte campanha das missões cristãs britânicas no sul do país, responsáveis por converter nigerianos ao cristianismo, no sentido de exigir o término da escravidão ao parlamento britânico. Com a abolição da escravatura em 1807 pela Coroa Britânica, houve substituição do comércio de escravos pelo comércio de mercadorias (METZ, 1991; FALOLA, HEATON, 2008).

A Nigéria foi um país protetorado da Coroa Britânica até 1960, quando ao fim da Segunda Guerra Mundial, o Reino Unido, impossibilitado de sustentar suas colônias foi pressionado por movimentos internos nigerianos, exigindo a independência do país.

ATUALIDADES

Atualmente, a Nigéria é uma república, sendo governada por um presidente eleito democraticamente por votação direta. O território nigeriano é rico em reservas de petróleo,

sendo esta a maior fonte da economia do país que se encontrava no posto de 26ª maior economia do mundo em 2013, ultrapassando a África do Sul (Commonwealth, 2021).

É considerado um dos países mais diversos do mundo, contendo entre 250 a 400 grupos étnicos reconhecidos que estão divididos em subgrupos de grande importância social e política. Depois da independência do país, a dinâmica estabelecida pela distribuição dos grupos nativos da Nigéria ainda se manteve em voga, influenciando a divisão política dos estados. Os grupos étnicos mais influentes são: Hausa e Fulani, Yoruba, Igbo, Kanuri, Ibibio, Tiv e Ijaw, que se distribuem por todo o território nigeriano (METZ, 1991; SIMPSON, OYÈTÁDÉ, 2008; FALOLA, HEATON, 2008) (**Figura 6**). No presente estudo, foram incluídos os seguintes grupos:

a) Hausa ou Hausa-Fulani

O grupo Hausa é o maior grupo etnolinguístico hoje na Nigéria, totalizando cerca de 29% da população. Localiza-se majoritariamente a nordeste do país, pertencendo à grande família linguística Afro-Asiática. Sua sociedade, por influência do período pré-colonial, segue majoritariamente os costumes islâmicos e à religião muçulmana;

b) Yoruba

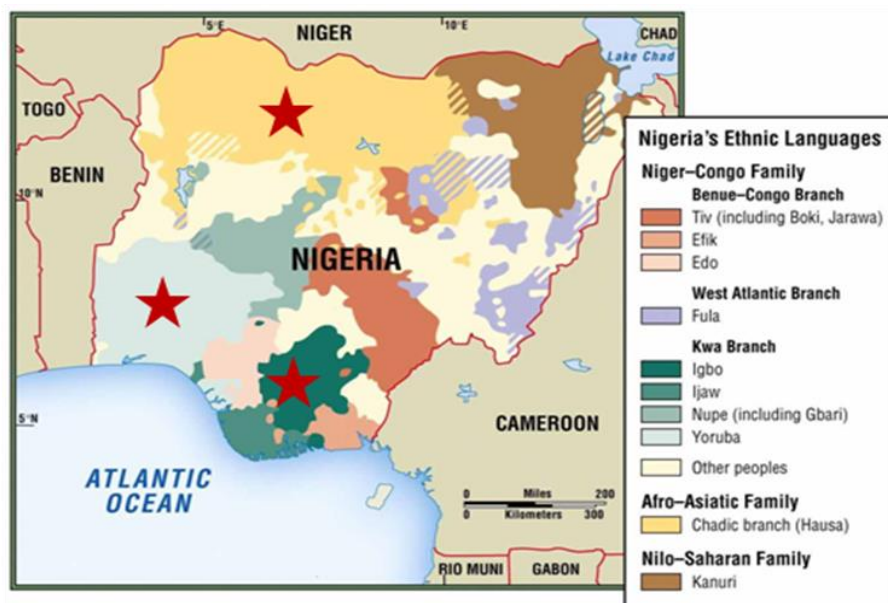
O grupo Yoruba é o segundo maior grupo etnolinguístico, totalizando cerca de 20% da população. Localiza-se majoritariamente a sudoeste do país, pertencendo à grande família linguística Níger-Congo. Sua população segue o cristianismo, em sua maioria, resultante da influência da coroa britânica no período colonial;

c) Igbo

O grupo Igbo é o terceiro maior grupo etnolinguístico de Nigéria, comportando cerca de 17% da população. Localiza-se majoritariamente a sudeste do país, pertencendo à grande família linguística Níger-Congo. Sua população também é majoritariamente adepta do cristianismo.

A religião é um dos principais fatores da identidade cultural das tribos da Nigéria. Atualmente, o país está dividido em muçulmanos (à norte, com aproximadamente 50% da população) e cristãos (ao sul, 40% da população). Ao centro do país, denominado cinturão médio, há seguidores de religiões mistas tradicionais (METZ, 1991; FALOLA, HEATON, 2008; SIMPSON, OYÈTÁDÉ, 2008; Commonwealth, 2021).

Figura 6 - Mapa com os maiores grupos etnolinguísticos da Nigéria



Legenda: Distribuição dos maiores grupos etnolinguísticos presentes na Nigéria especificados por cores segundo o idioma, com destaque aos três maiores grupos com estrela.

Fonte: Modificado de Quora, <https://qph.fs.quoracdn.net/main-qimg-9c73ff310f534255af4595398954de7c>, acesso em agosto de 2021.

A identidade étnica, linguística, religiosa, bem como a história das tribos da Nigéria são traços determinantes do país, visto que atritos entre grupos da população são geralmente de cunho étnico-religioso. Um destes eventos se manifestou durante a *Holy War* entre os anos de 1804 a 1808, liderada por reformadores majoritariamente Fulani, também conhecidos como a jihad islâmica. Este grupo consolidou o Sokoto Califado por toda a região da savana, entre o norte da Nigéria e o sul do Níger, um governo que propunha a fixação e purificação do Islã nos estados Hausa (METZ, 1991; FALOLA, HEATON, 2008).

Outro grande acontecimento importante no país foi a Guerra Civil, também conhecida como Guerra de Biafra (1967-1970), resultante de conflitos étnicos. Logo após a independência do país, a formação de um Parlamento e a tentativa da criação de uma identidade nacional evidenciaram o mosaico étnico e cultural do país. A primeira eleição democrática tinha, como candidatos, representantes de cada um dos três grandes grupos étnicos – Hausa, Yoruba e Igbo. A vitória do candidato do partido Hausa, representante da maioria da população, trouxe descontentamento dos grupos minoritários. Como resultado, em julho de 1960, se iniciou o conflito majoritariamente protagonizado pelos grupos Hausa, até o momento comandando o país, e Igbo, que reivindicavam o resultado das eleições e se auto proclamavam um Estado Independente, Biafra. A Guerra de Biafra é conhecida como o maior

conflito interno do país que culminou na morte de até 3 milhões de pessoas, entre civis e militares nigerianos, mas não resultou na utópica construção da identidade nacional nigeriana (FALOLA, HEATON, 2008).

INFORMAÇÕES GEOGRÁFICAS

A Nigéria ou República Federal da Nigéria está localizada na África Ocidental e faz fronteira com a República do Benin a oeste, Níger ao norte, e Chade e Camarões a leste (**Figura 7**). Ao sul do país, está a Costa no Golfo da Guiné, banhado pelo Oceano Atlântico. O país possui 36 estados e o Território da Capital Federal, onde se encontra Abuja. O território da Nigéria tem uma área total aproximada de 923.768 km² e uma área costeira de 853 km (METZ, 1991).

O país tem clima tropical e semiárido, tendo colinas e planaltos ao centro e a leste está localizado o planalto ou maciço de Adamawa. Ao norte, o clima é árido com vegetação mais semelhante à presente nas savanas e onde também se encontra o limite sul do Deserto do Saara. Na parte sul, há florestas e bosques, sendo o local mais úmido do país, que recebe um período de chuvas vindas do Oceano Atlântico.

O país é dividido ainda em três partes pelos rios Níger, que chega a noroeste, e Benué, que chega a nordeste. No centro do país, próximo à capital Abuja, o rio Benué une-se ao Níger, o qual flui para sul e deságua no Oceano Atlântico (Commonwealth, 2021).

Figura 7 - Localização da Nigéria e países de fronteira



Legenda: Destaque à Capital Federal Abuja com estrela.

Fonte: Google, <https://pt.wikipedia.org/wiki/Nig%C3%A9ria>, acesso em agosto de 2021.

INFORMAÇÕES DEMOGRÁFICAS

A Nigéria é o país mais populoso na África e o 7º mais populoso do mundo, tendo uma população de 186 milhões de habitantes, com expectativa de vida média de 53 anos, segundo o censo de 2016. Mais da metade da população tem menos de 15 anos de idade e, apesar do grande desenvolvimento do país, a população vive em situação de carência (Commonwealth, 2021).

Possui o inglês como língua oficial, sendo usado pelo governo, em escolas e em universidades, no mercado comercial e em meios de comunicação em massa. Mas há grande influência das línguas tradicionais e dialetos, estimados em aproximadamente 400. Os idiomas tradicionais mais importantes são hausa-fulani, yoruba e igbo, falados por cerca de 70% da população do país. As línguas maternas também são usadas na educação primária, a se exemplificar ao norte, onde o árabe é ensinado nas escolas, devido à ampla influência da religião muçulmana (METZ, 1991).

GENÉTICA DE POPULAÇÕES

A genética de populações vem sendo estudada desde as primeiras observações de Mendel, Darwin (século XIX) e, nos anos 1900, por Fisher, Wright e Haldane, que tentaram explicar as variações existentes dentro das populações humanas e não humanas. A generalização da teoria Mendeliana a populações (não tendo por base genealogias conhecidas) foi introduzida de forma independente por Hardy e Weinberg, em 1908, sendo conhecida por equilíbrio (ou princípio) de Hardy-Weinberg. Trata-se de um modelo simples que relaciona as frequências alélicas e as genotípicas, com base nas seguintes premissas: a população é infinita, os indivíduos cruzam-se ao acaso, não existe seleção, migração, deriva e mutação (HARTL; CLARK, 2010).

Com o desenvolvimento da área genômica, foi possível determinar as primeiras sequências de genomas humanos e, assim, entender a variabilidade genética existente entre as populações e sua importância evolutiva e biológica. Assim, foram identificados diversos tipos de polimorfismos presentes nos genomas, caracterizados por substituições, inserções e deleções de uma ou mais bases na sequência de DNA. Entre os polimorfismos mais utilizados em genética de populações, destacam-se os *Single Nucleotide Polymorphisms* (SNPs, em português, Polimorfismos de Nucleotídeo Único) e os *Short Tandem Repeats* (STRs, em português, Repetições Curtas em *Tandem*). Os SNPs são marcadores de evolução mais lenta, dado a sua baixa taxa de mutação entre as gerações. Devido às baixas taxas de mutação, os

SNPs tendem a apresentar menor recorrência que os STRs e, conseqüentemente, uma maior especificidade geográfica. Deste modo, este tipo de marcador é importante para ressaltar as diferenças entre populações, sendo o mais usado para inferir a origem geográfica, ou ancestralidade, de um indivíduo. Por outro lado, os STRs são marcadores que apresentam altas taxas de mutação, sendo por isso altamente polimórficos e discriminativos entre indivíduos de uma mesma população, o que permite a individualização (JEFFREYS, WILSON, *et al.*, 1985; BRUMFIELD, BEERLI, *et al.*, 2003; LAO, VAN DUIJN, *et al.*, 2006).

O desenvolvimento de novas metodologias permitiu o estudo de frequências alélicas e genotípicas para um elevado número de marcadores em diversas populações. Desta forma, tem sido possível determinar as principais características genéticas das populações humanas espalhadas pelo globo e inferir suas diferenças e afinidades, bem como as dinâmicas evolutivas que as moldaram.

Nas últimas décadas, a genética de populações foi revolucionada pela metodologia de *Massive Parallel sequencing* (MPS, em português, Sequenciamento Massivo Paralelo), que permite detectar um elevado número de polimorfismos presentes no genoma e a caracterização de genes determinantes para características humanas. Por outro lado, esta tecnologia tem possibilitado acumular bases de dados extensas que permitem caracterizar melhor as variações das populações humanas por todo o globo, bem como a determinação de genomas completos, tanto de populações atuais como antigas, e inferir a sua história evolutiva com maior precisão (HARTL; CLARK, 2010; OKAZAKI, YAMAZAKI, *et al.*, 2021).

No ramo da genética de populações são utilizados diferentes tipos de marcadores encontrados nos genomas humanos, com especificidades inerentes a sua localização intracelular e modo de transmissão, como indicado no **Quadro 1**.

Os autossomos, contidos no núcleo das células eucariotas, possuem informação de ambos os progenitores, de modo que marcadores localizados nestes cromossomos são amplamente utilizados para identificação humana e também em estudos populacionais. Atualmente, painéis de MPS autossômicos incluem tanto marcadores de microevolução (STR) com alto poder de individualização, como marcadores de macroevolução (SNP), utilizados para diferenciação interpopulacional.

A partir do estudo dos autossomos foi possível datar a origem de várias populações humanas, bem como as migrações durante o *Out of Africa* e consecutivos efeitos fundadores que levaram à perda paulatina de diversidade das populações seguintes (RAMACHANDRAN, DESHPANDE, *et al.*, 2005; DESHPANDE, BATZOGLOU, *et al.*,

2009; LI, DURBIN, 2011). Atualmente, painéis de marcadores autossômicos com elevada diferenciação entre populações têm sido de grande uso em estudos de ancestralidade, garantido, cada vez mais, a compreensão aprofundada das origens de diversas populações. Os dados autossômicos são relevantes não só em genética populacional, como também em outras vertentes da genética, onde é possível, por exemplo, se estudar a etiologia de doenças e suas variantes genéticas e os eventos de seleção que ocorreram ao longo da sua história (CAMPBELL, TISHKOFF, 2008; LAMBERT, TISHKOFF, 2009; BUSTAMANTE, DE LA VEGA, et al., 2011; HENN, CAVALLI-SFORZA, et al., 2012; MENSAH, 2013).

Quadro 1 – Comparação entre os marcadores genéticos usados em genética de populações

Característica	DNA mitocondrial	Cromossomo Y	Cromossomo X	Autossomos
Tamanho em pares de bases (milhão)	0.017	60	150	3.000
Número de <i>loci</i> que contêm informação independente	1	1	Centenas	Milhares
Taxa de mutação (por Mb por geração)	Muito alta (1-300)	Alta (0.033)	Baixa (0.015)	Moderada (0.020)
Emparelhamento cromossômico	Haploide	Haploide	Haploide/Diploide	Diploide
Cópias por célula	Até 1000	0 ou 1	1 ou 2	2
Diversidade	Muito alta (0.4%)	Baixa (0.02%)	Moderado (0.04%)	Alta (0.08%)
Aplicabilidade	Identificação e inferência da origem geográfica da linhagem materna; Análise de amostras antigas ou degradadas.	Identificação e inferência da origem geográfica da linhagem paterna; Identificação de perfis masculinos, essencialmente em crimes sexuais.	Inferência geográfica (ancestralidade); Determinação de características físicas (cor de pele e olhos ...); Identificação individual; Determinação de vínculo genético (por exemplo, paternidade); Identificação de vítimas de desastres em massa.	

Fonte: Adaptado de Schaffner, 2004.

O cromossomo X, por sua vez, é uma boa ferramenta no ramo populacional, pois apresenta padrão de herança diferencial para homens e mulheres, como marcador haploide e

diploide, respectivamente. Pelo fato de marcadores específicos do cromossomo X possuírem padrões de recombinação e mutação dependentes do sexo biológico, é cabível seu uso em estudos populacionais e evolutivos, contando a história do homem moderno, sob diferentes perspectivas. Dados obtidos através do estudo do cromossomo X fomentaram o debate das duas grandes teorias vigentes de origem do homem moderno (SCHAFFNER, 2004). Marcadores localizados no cromossomo X são também vantajosos para estudos de associação gamética e, desta forma, são bons candidatos para a reconstrução de filogenias.

Outra classe de marcadores são os uniparentais, localizados na região não recombinante do cromossomo Y e no DNA mitocondrial, que conferem apenas informação das linhagens paterna e materna, respectivamente. A ausência de recombinação destes marcadores possibilita a inferência da história evolutiva das populações humanas de forma independente, tendo grande aplicação no estudo de populações e suas estratificações, estudos demográficos, assim como em estudos bioantropológicos (BUTLER, 2010). A partir deles, tenta-se explicar as teorias de coalescência e, deste modo, traçar o possível *Most Recent Common Ancestor* (MRCA, em português, Ancestral Comum Mais Recente) das linhagens. O Adão cromossômico Y (há 142 mil anos) e a Eva Mitocondrial (há 150-200 mil anos) teriam surgido na África e seriam os primeiros humanóides que repassaram suas linhagens à prole, dando origem a todas as linhagens paternas e maternas atuais (CANN, STONEKING, *et al.*, 1987; LUCOTTE, GUERIN, *et al.*, 1989; SHEA, 2013).

MITOCÔNDRIA

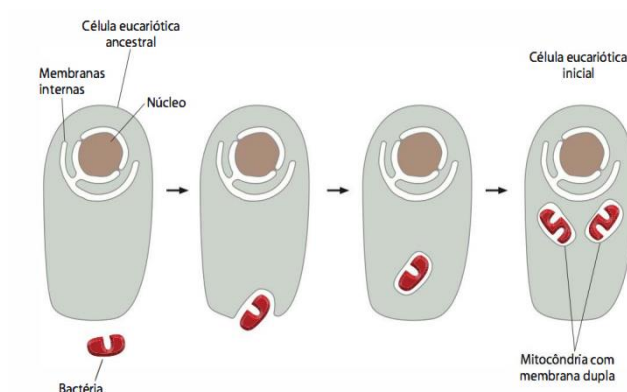
TEORIA DE ORIGEM

A hipótese do surgimento das mitocôndrias, explicada pela teoria endossimbiótica, é cada vez mais aceita, pois essas organelas possuem sistemas genéticos próprios com genoma diferente do nuclear. Além disso, as mitocôndrias têm um tamanho aproximado ao das bactérias, seus próprios ribossomos (diferentes de qualquer outra parte na célula eucariótica) e tRNA específico, assim como as bactérias.

A teoria endossimbiótica, postulada por Lynn Margulis, supõe que as mitocôndrias eram bactérias aeróbias de vida livre que foram endocitadas por uma célula eucarionte ancestral incapaz de fazer uso do oxigênio (anaeróbia) (**Figura 8**). Tal evento teria se dado há aproximadamente 1,5 bilhão de anos, momento em que a atmosfera terrestre se tornou rica em oxigênio. Graças a isto, se estabeleceu uma relação de simbiose entre a bactéria aeróbia e a célula eucarionte ancestral; a primeira recebeu maior proteção ao ser interiorizada por uma

célula maior, de modo a aumentar sua sobrevivência e de sua progênie, enquanto a célula eucarionte anaeróbia foi beneficiada pela maior geração de energia promovida pela bactéria aeróbia e, assim, maior aporte de ATP (em português, adenosina trifosfato) para suas atividades. Outra característica que sustenta a teoria endossimbiótica é que as mitocôndrias possuem duas membranas, sendo a externa com composição semelhante à membrana da célula eucariótica (ALBERTS, JOHNSON, *et al.*, 2017).

Figura 8 - Teoria endossimbiótica



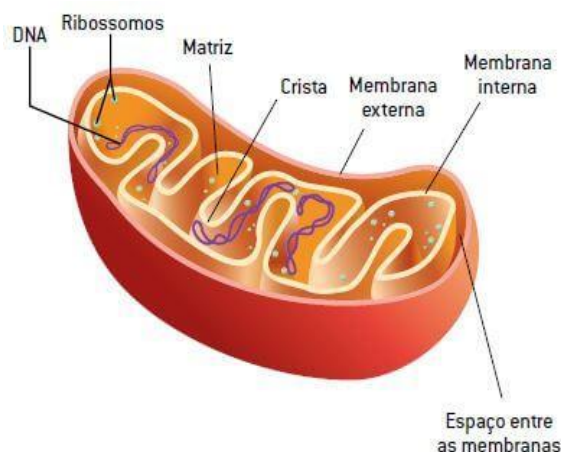
Legenda: Segundo esta teoria, houve endocitose de mitocôndrias de vida livre por uma célula eucarionte anaeróbia há aproximadamente 1,5 bilhão de anos.

Fonte: Albert. *et al.*, 2017.

CARACTERÍSTICAS DA ORGANELA

As mitocôndrias possuem entre 0,5 a 1,0 μm de diâmetro e são compartimentadas, apresentando um espaço interno, denominado matriz, e um espaço intermembranas (**Figura 9**). Tais espaços são separados pela membrana interna, composta por uma bicamada lipídica e proteínas. Esta envolve a matriz, local onde se encontram proteínas, enzimas, RNA e DNA mitocondrial. Na camada interna também se localiza a cadeia respiratória, onde ocorre o processo de fosforilação oxidativa e posterior produção de energia das células sob a forma de ATP. A membrana externa é separada da interna pelo espaço intermembranas e possui proporções muito semelhantes de proteínas e lipídeos em sua composição. Cada um destes compartimentos possui componentes proteicos distintos, devido à diferença de permeabilidade e à presença de translocases, proteínas específicas de membrana (ALBERTS, JOHNSON, *et al.*, 2017).

Figura 9 – A mitocôndria e suas estruturas



Legenda: A organela possui duas membranas, separadas por um espaço intermembranas. O espaço interno, ou matriz, comporta o DNA mitocondrial.

Fonte: Google, acesso em agosto de 2021.

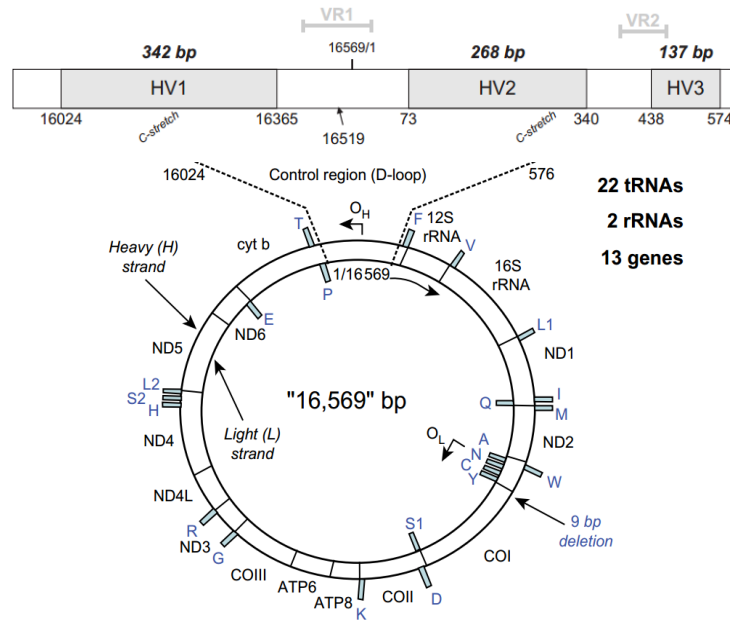
DNA MITOCONDRIAL

O genoma humano, local onde se encontra toda a informação genética de cada indivíduo, é composto por dois tipos de DNA: o nuclear e o mitocondrial. O DNA mitocondrial (mtDNA), também conhecido como mitogenoma, se localiza nas mitocôndrias, compreende em uma molécula circular responsável por codificar componentes da cadeia transportadora de elétrons (ANDERSON, BANKIER, *et al.*, 1981).

O mtDNA possui dupla cadeia, sendo uma dita *light* (L), por possuir maior contingente de citosinas e timinas, e a segunda cadeia, *heavy* (H), que possui mais guaninas e adeninas. A replicação de ambas fitas L e H é unidirecional e tem início em origens específicas, OL (direção de síntese da fita é horária) e OH (direção de síntese da fita é anti-horária), respectivamente (**Figura 10**). Esta molécula também possui duas regiões polimórficas distintas: região codificante e região controle ou não codificante (BUTLER, 2010).

A região codificante é responsável por grande porção do mitogenoma (93%), que contém aproximadamente 16.569 pares de bases, passíveis à variação nos organismos devido à inserções e/ou deleções (BODENTEICH, MITCHELL, *et al.*, 1992; SZIBOR, MICHAEL, *et al.*, 1997).

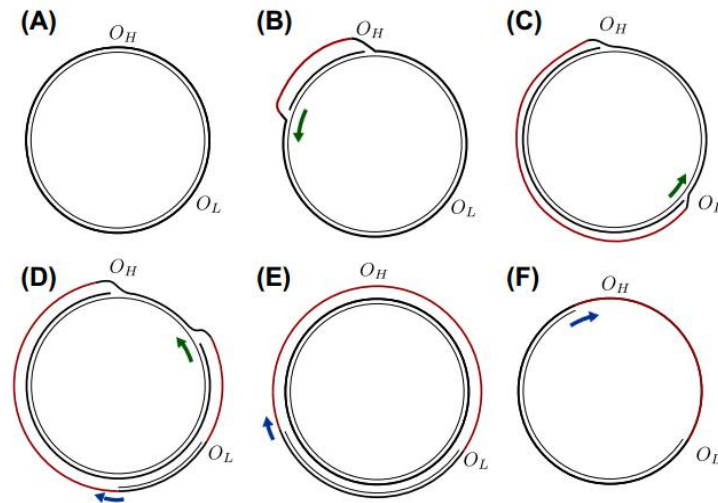
Figura 10 - DNA mitocondrial



Legenda: Cadeia leve ou *light strand* e cadeia pesada ou *heavy strand*. O mtDNA possui duas regiões distintas: controle e codificante. A região controle ou *D-loop* comporta as três regiões hipervariáveis.
 Fonte: Adaptado de Butler, 2010.

A região controle da molécula de DNA mitocondrial se limita das posições 16.024 à 576, contendo trechos ditos hipervariáveis I, II e III, devido à sua alta variabilidade/diversidade (BUTLER, 2010) (**Figura 10**). A região controle recebe este nome por comportar a região reguladora ou *origin region*, onde se inicia a replicação da molécula. O termo alça D (*displacement loop* ou *D-loop*) refere-se à estrutura formada no momento da replicação que pode ser vista por microscopia eletrônica (UPHOLT, DAWID, 1977). A replicação desta molécula em mamíferos se inicia na origem de replicação OH, localizada na fita *heavy* ou reversa (UPHOLT, DAVID, 1977; ANDERSON, BANKIER, *et al.*, 1981; BERNT, BRABAND, *et al.*, 2013; YASUKAWA, KANG, 2018) (**Figura 11**). A região controle não apresenta expressão gênica, fator que favorece seu uso para estudos forenses e populacionais, visto que desta maneira não será revelada nenhuma informação pessoal do indivíduo, como propensão à doenças, por exemplo.

Figura 11 - Replicação da molécula de DNA mitocondrial



Legenda: Replicação da molécula de DNA mitocondrial em humanos (A); formação da alça de deslocamento (B); a replicação da fita H se inicia (seta verde) até alcançar a origem de replicação da fita L (B e C); início da replicação da fita L (seta azul) (D) e fim da replicação da fita H (E); Continuação da replicação da fita L filha (F).

Fonte: Bernt, Barband, *et al.* (2013).

MITOGENOMA

A sequência total do mitogenoma ou genoma mitocondrial humano foi determinada pela primeira vez em 1981 utilizando a metodologia de Sequenciamento de Sanger. A sequência foi descrita nesse mesmo ano por Anderson e colaboradores e nomeada *Cambridge Reference Sequence* (CRS, em português, Sequência de Referência de Cambridge). O mitogenoma é uma ferramenta de comparação que tem grande importância e aplicabilidade em genética de populações, na área da saúde, assim como na resolução de casos no âmbito forense. Sua descrição possibilitou que fossem desenvolvidos diversos estudos décadas após sua divulgação até o presente momento (ANDERSON, BANKIER, *et al.*, 1981).

Dado a sua grande importância e ampla aplicabilidade, a sequência foi reanalisada e corrigida por Andrews e colaboradores em 1999, após análise do mesmo material genético usado no primeiro sequenciamento. Houve acréscimo de 11 nucleotídeos, foi renomeada rCRS ou Sequência de Referência de Cambridge revisada e restabelecida no banco de dados mundial GeneBank (ANDREWS, KUBACKA, *et al.*, 1999).

NOMENCLATURA

Os polimorfismos encontrados, sejam de posição ou de comprimento, comparativamente à rCRS são classificados em substituição, inserção e deleção de um ou mais nucleotídeos.

Substituições ocorrem quando se encontra, numa posição determinada, um nucleotídeo diferente do que se observa na sequência de referência. Neste caso, a posição em questão é anotada seguida da base substituída. Esses polimorfismos podem ser caracterizados por troca de bases nitrogenadas de mesmo tipo (purina-purina ou pirimidina-pirimidina), sendo denominadas transições. Por outro lado, quando ocorrem trocas de bases nitrogenadas de tipos diferentes (purinas-pirimidinas ou pirimidinas-purinas) são denominadas transversões.

As inserções são caracterizadas pela presença de nucleotídeos adicionados na amostra em comparação à rCRS. Devem ser anotadas pela posição nucleotídica que antecede a inserção, seguida de um ponto e o número de inserções.

Deleções, por sua vez, são a ausência de um nucleotídeo em determinada posição da sequência. Devem ser anotadas pela posição nucleotídica onde houve deleção, seguida da abreviação 'del'.

CARACTERÍSTICAS DA MOLÉCULA

Se comparado ao DNA nuclear, existem poucos mecanismos de reparo na mitocôndria, isto porque esta molécula possui uma densidade gênica pequena. Esse fator permite que a molécula de DNA mitocondrial possua altas taxas de mutação. Além disto, a DNA polimerase da mitocôndria não possui mecanismos de revisão à erros, o que aumenta o acúmulo de mutações durante a replicação da molécula e mais alterações na sequência.

As altas taxas de mutação da molécula de DNA mitocondrial ajudam a aumentar a variabilidade genética do genoma possibilitando, em alguns casos, a discriminação de amostras de uma mesma linhagem materna, de modo que se torne viável seu uso como ferramenta de identificação em casos forenses (GABRIEL, HUFFINE, *et al.*, 2001; PARSONS, T. J., COBLE, 2001; PARSONS, 2006).

As mutações no mitogenoma podem ocorrer a nível germinativo ou somático. As mutações germinativas são adquiridas pelo indivíduo através da herança materna do mtDNA, e têm grande relevância especialmente na genética clínica para estudos de etiologias de doenças graves promovidas por polimorfismos do mtDNA (doenças neurológicas e cardiomiopatias, por exemplo) (TAYLOR, TURNBULL, 2005). Por outro lado, as mutações germinativas não associadas à doenças podem ser uma ferramenta útil na resolução de casos

forenses. A exemplo disso, está o misterioso caso da Família Romanov. A presença de uma heteroplasmia na posição 16.169 em comum no mtDNA de Georgij Romanov possibilitou a identificação dos restos mortais de seu irmão, o Czar Nicholas Romanov II, que permaneceu sem paradeiro conhecido após sua morte durante a Revolução Russa (MELTON, 2004; COBLE, 2011).

Por outro lado, existem as mutações das células somáticas, que se dão ao longo da vida do indivíduo através das mitoses. Essas mutações ocorrem com a replicação das células e dependem do metabolismo diferencial das mesmas. As células com maiores taxas de replicação e alta demanda energética (cérebro, músculo e cabelo, por exemplo) necessitam de mais oxigênio e mitocôndrias para atender às suas necessidades respiratórias. Desse modo, essas células são mais expostas aos radicais livres aumentando então suas taxas de mutação e variantes em seu mitogenoma (MELTON, NELSON, 2001; MELTON, 2004).

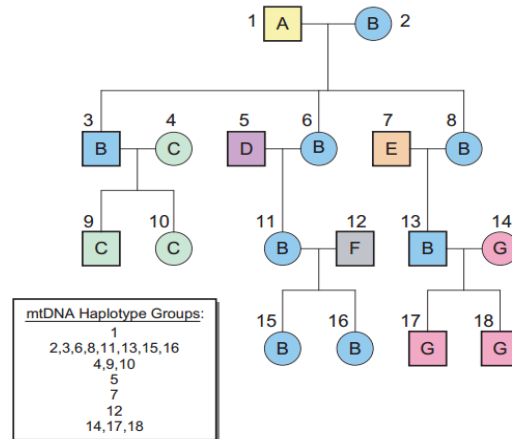
Outra característica importante é que o DNA mitocondrial humano é herdado por mecanismo não mendeliano, ou seja, por herança uniparental, exceto em alguns casos raros descritos na literatura, ainda em constante debate (HAGELBERG, GOLDMAN, *et al.*, 1999; SCHWARTZ, VISSING, 2002; LUO, VALENCIA, *et al.*, 2018; LUTZ-BONENGL, PARSON, 2019). Acredita-se que o genoma mitocondrial seja repassado estritamente da mãe aos filhos de ambos os sexos. Durante a fecundação, há alta produção e gasto de energia pelos espermatozoides que poderá resultar no acúmulo de mutações em função da alta produção de radicais livres. Deste modo, as mitocôndrias dos espermatozoides são eliminadas por ubiquitinação, não contribuindo para a composição do embrião, evitando também que tais mutações sejam herdadas (SUTOVSKY, SONG, 2018). No momento da fecundação, a quantidade de mtDNA masculino que pode ter se isentado do mecanismo de ubiquitinação acaba sendo diluído pela grande quantidade de mtDNA feminino. Ou ainda, o DNA mitocondrial pode ser eliminado de espermatozoides maduros, fator que também explica a ausência de herança paterna (SATO, SATO, 2013).

Deste modo, o genoma mitocondrial é passado integralmente de mãe para filhos de ambos os sexos de forma igual, a exceção de mutações que possam ocorrer durante ou após a sua transmissão (**Figura 12**).

Para além do que foi citado, o genoma das mitocôndrias apresenta várias moléculas de DNA por célula, ao contrário do DNA nuclear. O maior número de cópias e tamanho reduzido do mtDNA aumentam a probabilidade de recuperação dessas sequências completas em casos de material degradado (cenários de crime, acidentes em massa entre outros cenários), se

comparado ao DNA nuclear, com apenas duas cópias por célula (uma materna e outra paterna) e seqüências extensas de material genético (BUTLER, 2010).

Figura 12 - Heredograma do padrão de herança do mtDNA



Legenda: As mulheres são representadas por círculos e homens por quadrados. Os haplótipos únicos estão representados por diferentes cores e letras e demonstram a herança integral da informação do mtDNA das mães aos filhos.

Fonte: Butler, 2010.

HETEROPLASMIA

Acreditava-se que após ser transmitido apenas pela mãe, o DNA mitocondrial possuísse a mesma seqüência em todas as células de um indivíduo, fenômeno conhecido como homoplasmia. Mas, sabe-se atualmente que a presença de mais de um tipo de mtDNA num mesmo indivíduo é um polimorfismo comum que atinge a população global. A ocorrência de mais de uma seqüência num mesmo indivíduo pode ter origem materna (ou seja, já estar presente no óvulo/oócito secundário) ou dever-se a mutações somáticas ocorridas após transmissão materna (durante o desenvolvimento embrionário ou ao longo da vida de um indivíduo) (MELTON, 2004).

Esse fenômeno é chamado de heteroplasmia e pode ser classificado em três níveis:

a) a nível tecidual: onde se encontra mais de um tipo de DNA mitocondrial no mesmo tecido ou em tecidos diferentes;

b) a nível mitocondrial: ocorre quando diferentes mitocôndrias dentro da mesma célula possuem tipos de DNA mitocondrial distintos;

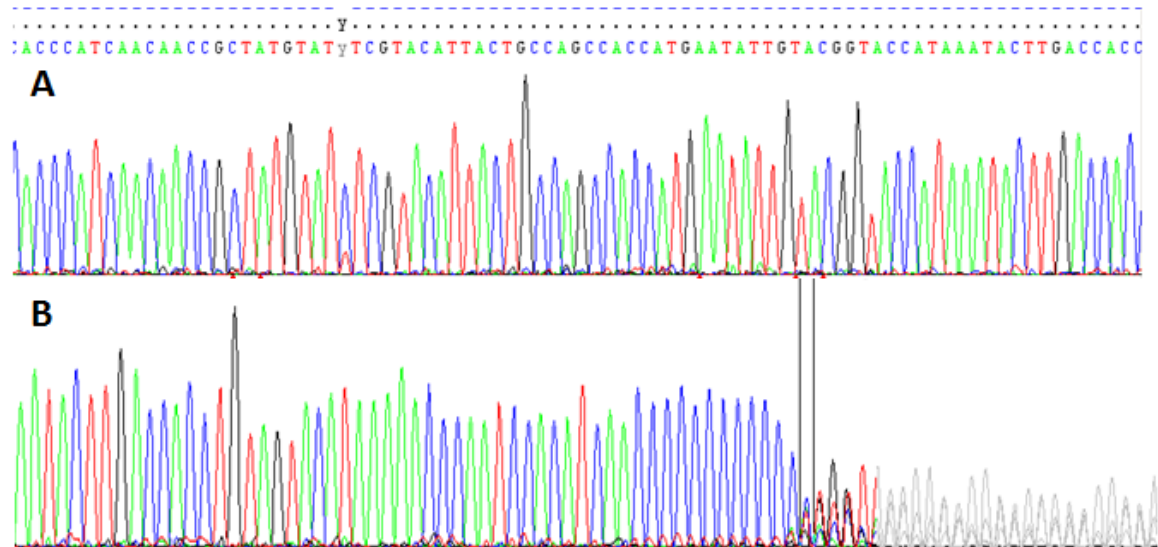
c) a nível ácido nucleico: quando há numa mesma mitocôndria tipos diferentes de DNA mitocondrial.

As heteroplasmias também podem ser classificadas como:

a) de posição: onde há, numa posição da sequência de DNA mitocondrial, a presença de mais de uma base nucleotídica (**Figura 13A**);

b) de comprimento: quando há inserções ou deleções que promovam variação no tamanho da sequência (**Figura 13B**).

Figura 13 - Comparação de sequências heteroplásmicas



Legenda: Sequência com heteroplasmia de posição (A) frente a uma heteroplasmia de comprimento em uma porção homopolimérica de citosinas (B).

Para as sequências com heteroplasmia de comprimento, a determinação da posição nucleotídica é feita de acordo com a complementariedade das fitas leve e pesada, seguindo as diretrizes de Parson e colaboradores (2014).

As heteroplasmias de comprimento ocorrem especialmente nas sequências de citosina (ou *C-stretches*) presentes nas regiões hipervariáveis da região controle, entre as posições 16184-16189 e 303-315, em HVI e HVII, respectivamente. Esse fenômeno pode acontecer por causa da ação promovida pela reação da DNA polimerase na região de Cs consecutivos, que muta a quantidade de bases da região mais facilmente. Por isto, esta região é também chamada de hotspot mutacional. Durante a replicação do DNA mitocondrial pode haver deslizamento da fita ou *strand slippage*; a DNA polimerase pode acidentalmente replicar mais de uma citosina durante a reação, que resulta na alteração do comprimento desta sequência (HAUSWIRTH, CLAYTON, 1985).

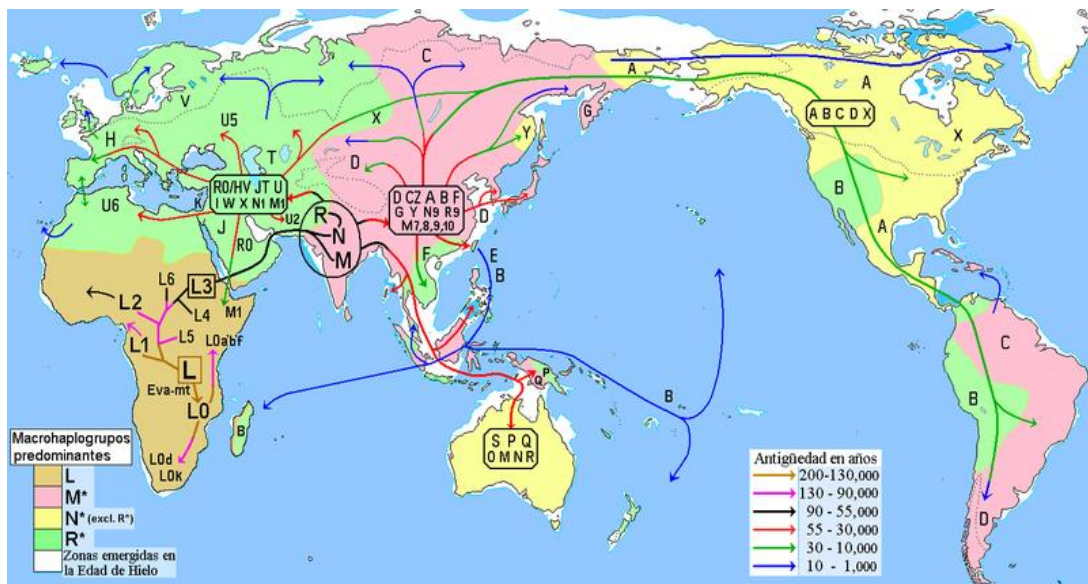
Fenômenos de triplasmia, caracterizados pela presença de duas heteroplasmias num mesmo indivíduo, já foram descritos por Tully e colaboradores (2000), sendo uma condição com baixa frequência se comparada às outras heteroplasmias.

A heteroplasmia é um fator importante no uso de mtDNA em estudos moleculares. Esta característica afeta a interpretação das sequências em forense, dificultando o estabelecimento de uma única sequência de DNA que possa resultar no *match* entre parentes maternos (MELTON, 2004).

FILOGENIA DAS LINHAGENS AFRICANAS DE mtDNA

Em 1987, Wilson e colaboradores desenvolveram estudos que puderam embasar a teoria do MRCA. A partir da Eva mitocondrial foi possível construir a árvore filogenética dos haplogrupos de DNA mitocondrial e das rotas de migração do homem pelo globo (**Figura 14**). É desta antecessora que se reforça a teoria de surgimento do homem moderno há 200.000 anos na África, que teria originado as populações de todos os continentes (BEHAR, VILLEMS, *et al.*, 2008).

Figura 14 - Padrões de distribuição do homem moderno permitiram a construção da filogenia do DNA mitocondrial e determinação de haplogrupos



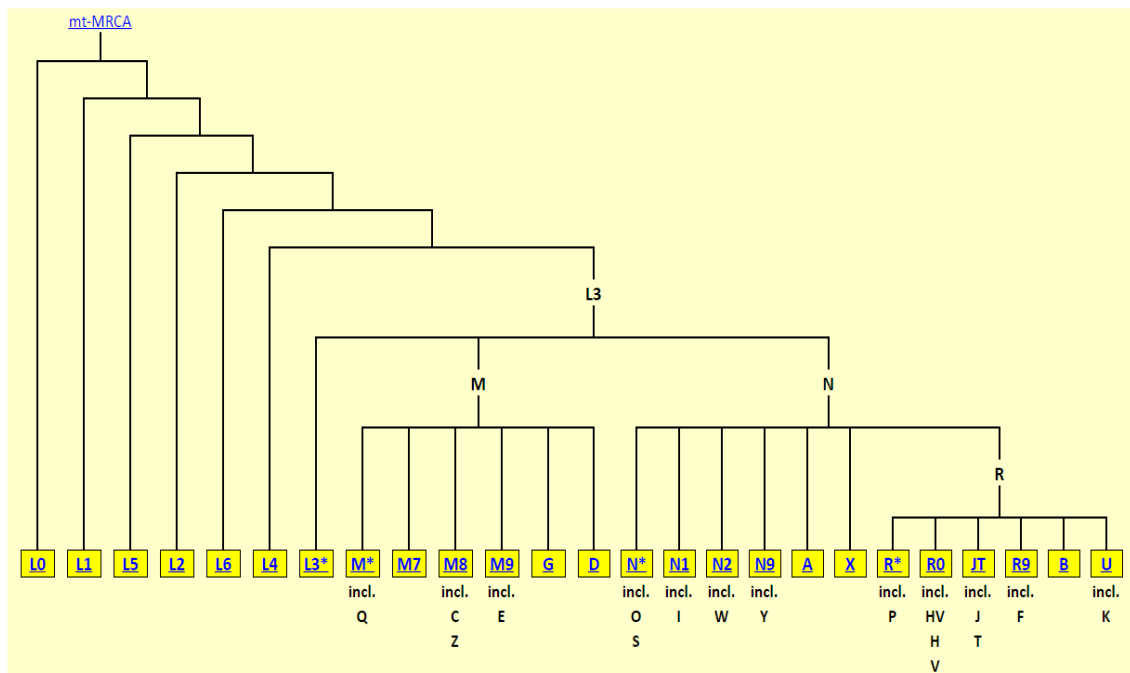
Legenda: Os haplogrupos de mtDNA, representados por letras, são característicos de cada região geográfica do planeta, de modo que, através dele, pode-se inferir a ancestralidade materna de um indivíduo.

Fonte: https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Migraciones_humanas_en_haplogrupos_mitocondriales.PNG/1200px-Migraciones_humanas_en_haplogrupos_mitocondriales.PNG, acesso em agosto de 2021.

A partir da observação dos padrões de migração das populações humanas pode-se definir os haplogrupos mitocondriais. Os haplogrupos são agrupamentos de haplótipos de mtDNA que apresentam polimorfismos específicos, ou seja, representam linhagens características de cada região do globo e próximas umas das outras. A partir de diversos estudos foi construída a árvore filogenética de mtDNA, disponível na plataforma *PhyloTree build 17* (VAN OVEN, KAYSER, 2009) (**Figura 15**). Esta árvore encontra-se em constante atualização, à medida que novas sequências vão sendo descritas.

Com esta ferramenta filogenética, foram descritos e descobertos diversos haplogrupos, entre eles, o macrohaplogrupo L, característico da África Sub-Saariana. Existem sete principais ramificações (sendo seis delas mais relevantes) que se distribuem em frequências diferentes no continente e que caracterizam algumas populações.

Figura 15 - Árvore filogenética do DNA mitocondrial e seus haplogrupos



Legenda: Haplogrupos de mtDNA discriminados na base; mt-MRCA representa a Eva mitocondrial ou antecessora comum às linhagens maternas.

Fonte: *PhyloTree*, <https://www.phyloree.org/tree/index.htm>, acesso em agosto de 2021.

Com esta ferramenta filogenética, foram descritos e descobertos diversos haplogrupos, entre eles, o macrohaplogrupo L, característico da África Sub-Saariana. Existem sete principais ramificações que se distribuem em frequências diferentes no continente e que caracterizam algumas populações.

O haplogrupo L0 é o ramo mais antigo do mtDNA e o mais representativo das regiões Sudeste e Leste da África (32% e 22% de frequência, respectivamente), onde também apresenta os maiores valores de diversidade do continente. Nas regiões Sudoeste (11%) e Centro-oeste (6%), este haplogrupo possui baixas frequências. Sugere-se que tenha surgido há mais de 100.000 anos atrás e é uma linhagem geralmente associada a povos Khoisan (SALAS, RICHARDS, *et al.*, 2002; CERESO, GUSMÃO, *et al.*, 2016). O haplogrupo é caracterizado pela mutação 263A da região controle e das mutações 1048T, 3516A, 5442C, 6185C, 9042T, 9347G, 10589A, 12007A e 12720G da região codificante do mtDNA, segundo a Phylotree (<https://www.phylotree.org/tree/L0.htm>).

O haplogrupo L1 encontra-se amplamente disperso pelo continente africano. Possui altas frequências na região Sudoeste (28%) e é menos presente na região Norte (1%). O haplogrupo está presente também nas regiões Central e Oeste, sendo especialmente característico das populações dos pigmeus da Floresta Equatorial (SALAS, RICHARDS, *et al.*, 2002; CERESO, GUSMÃO, *et al.*, 2016). Esse haplogrupo é caracterizado pelas mutações 3666A, 7055G, 7389C, 13789C, 14178C e 14560A da região codificante do mtDNA, segundo a Phylotree (<https://www.phylotree.org/tree/L1.htm>).

O haplogrupo L2 está altamente difundido por toda África, sendo muito frequente no Sudeste (36%) e na região Centro-oeste (29%), mas é mais diverso nesta última região, onde possui maiores valores de diversidade haplotípica e nucleotídica (SALAS, RICHARDS, *et al.*, 2002; SILVA, ALSHAMALI, *et al.*, 2015; CERESO, GUSMÃO, *et al.*, 2016). Esse haplogrupo é caracterizado pelas mutações 146C, 150T, 152C, 16311T e 16390A da região controle e 2416C, 8206A, 9221G, 10115C e 13590A da região codificante do mtDNA, segundo a Phylotree (<https://www.phylotree.org/tree/L2.htm>).

A linhagem L3 teve origem no Leste da África e está distribuída por todo o continente. Atinge maiores frequências no Sudoeste da África (45%), seguido pelas regiões Centro-oeste (43%) e Leste (34%), mas apresenta maior diversidade de sub haplogrupos na segunda região citada. Ainda que sua provável origem tenha sido na região Leste, este haplogrupo está especialmente difundido entre os povos Bantu da região Sudeste. L3 é também conhecido por ter sido difundido para fora do continente no movimento *Out of Africa*, dando origem a todas as outras linhagens não L (SALAS, RICHARDS, *et al.*, 2002; WATSON, BAUER, *et al.*, 1996; CERESO, GUSMÃO, *et al.*, 2016). Esse haplogrupo é caracterizado pelas mutações 16311T da região controle e 769G e 1018G da região codificante do mtDNA, segundo a Phylotree (<https://www.phylotree.org/tree/L3.htm>).

Os haplogrupos L4 e L5 são mais encontrados no Leste da África do que em suas demais regiões, apresentando frequências de 15% e 6%, respectivamente, nestas regiões. Ainda que pouco difundido, o haplogrupo L5 é um pouco mais diverso do que L4 e também está presente na região Sudeste (CEREZO, GUSMÃO, *et al.*, 2016). O haplogrupo L4 é caracterizado pelas mutações 195C e 16362C da região controle e 5460A da região codificante do mtDNA, segundo a Phylotree (<https://www.phylotree.org/tree/L4.htm>). Por sua vez, o haplogrupo L5 é caracterizado pelas mutações 459.1C, 16148T e 16166G da região controle e 3423C, 7972G, 12432T e 12950G da região codificante do mtDNA, segundo a Phylotree (<https://www.phylotree.org/tree/L5.htm>).

1. OBJETIVOS

1.1 Objetivo geral

Este trabalho almejou caracterizar e comparar as três maiores populações da Nigéria (Hausa, Yoruba e Igbo), através do estudo de linhagens maternas, com análise da sequência da região controle do mtDNA de 285 indivíduos não aparentados.

1.2 Objetivos específicos

a) Analisar a composição dos grupos Hausa, Yoruba e Igbo com o estudo do marcador de linhagem materna (mtDNA), e comparar a estrutura genética entre os três grupos.

b) Comparar a diversidade haplotípica presente nestes três grupos e as distâncias genéticas com populações de países vizinhos à Nigéria.

c) Observar se há estratificação das populações nigerianas em questão, dada a história do país.

d) Contribuir para a elaboração de uma base de dados de linhagens maternas acerca das populações africanas.

2. MATERIAL E MÉTODOS

2.1 Etapas experimentais

O material biológico foi obtido através de Cooperação com o Time Argentino de Antropologia Forense (EAAF - *Equipo Argentino de Antropología Forense*), em Córdoba, na Argentina, o Hospital Universitário de Ensino de Lagos (LUTH - *Lagos University Teaching Hospital*) e pelo Instituto de Polícia FCIID Annex, na Nigéria.

Todos os procedimentos desta pesquisa foram realizados de acordo com a Declaração de Helsinki e do Código de Nuremberg, obedecendo as normas de pesquisas envolvendo seres humanos. O projeto foi submetido ao comitê de ética e aprovado por um Comitê de Ética do país (NHREC: 19/12/2008^a).

A extração de DNA genômico, a obtenção do material amplificado e das sequências de DNA mitocondrial foram feitas no Laboratório de Diagnósticos por DNA (LDD) da Universidade do Estado do Rio de Janeiro e no Laboratório de Genética Forense do EAAF.

2.2 Amostras populacionais

Neste estudo, foi coletado sangue de 285 indivíduos nigerianos voluntários e não relacionados. Todas as amostras foram coletadas sob consentimento informado, assinado pelos participantes. Foram aceitos apenas indivíduos de ascendência materna, de ao menos três gerações (mãe, avó e bisavó), pertencente aos grupos Hausa, Yoruba ou Igbo, segundo autodeclaração.

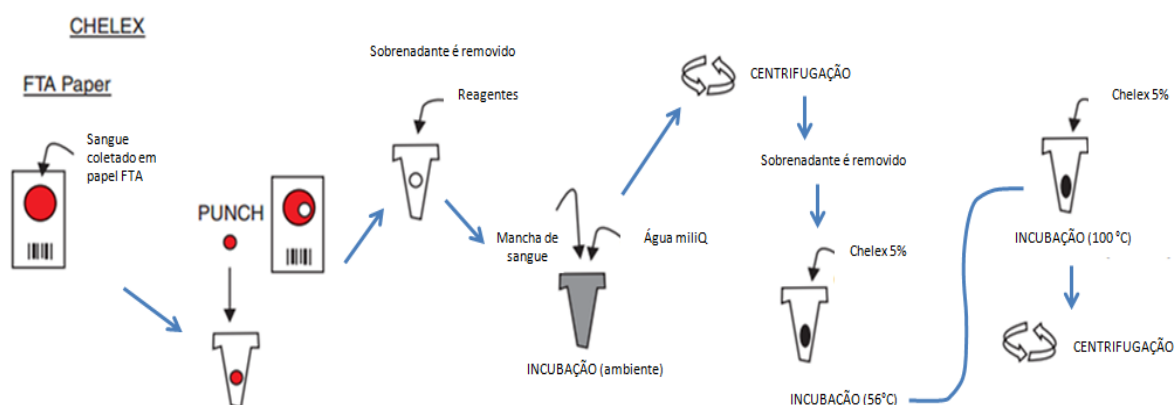
2.3 Metodologias de extração

A extração foi feita a partir de amostras de sangue coletadas em papel FTA, com uso dos métodos químico ou orgânico para cada amostra. Para a maioria das amostras foi utilizado o método químico de extração por Chelex. Para algumas amostras em que não foi possível

obter bons resultados por Chelex, foi realizada uma segunda extração utilizando o método de Fenol-Clorofórmio. Em ambos, uma pequena porção da mancha de sangue armazenado em papel FTA foi utilizado (*punch*). Esta fração do papel contendo a amostra de sangue foi colocada em tubo tipo *ependorf* de 1,5mL, previamente identificado com um código da amostra/indivíduo. Para cada um dos métodos, seguiram-se as etapas descritas abaixo.

No método químico foi usado Chelex 100 a 5%, uma resina quelante que é composta por polímeros de estireno-divinilbenzeno que conseguem estabelecer ligações com íons metálicos, como os presentes nas células (magnésio) (WALSH, METZGER, *et al.*, 1991). Neste processo, há sequestro dos íons celulares pela resina que conseguem atuar na quebra de proteínas e membranas celulares e possibilitam também a inativação de enzimas nucleases presentes na reação (BUTLER, 2010) (**Figura 16**).

Figura 16 - Método de extração química por resina Chelex

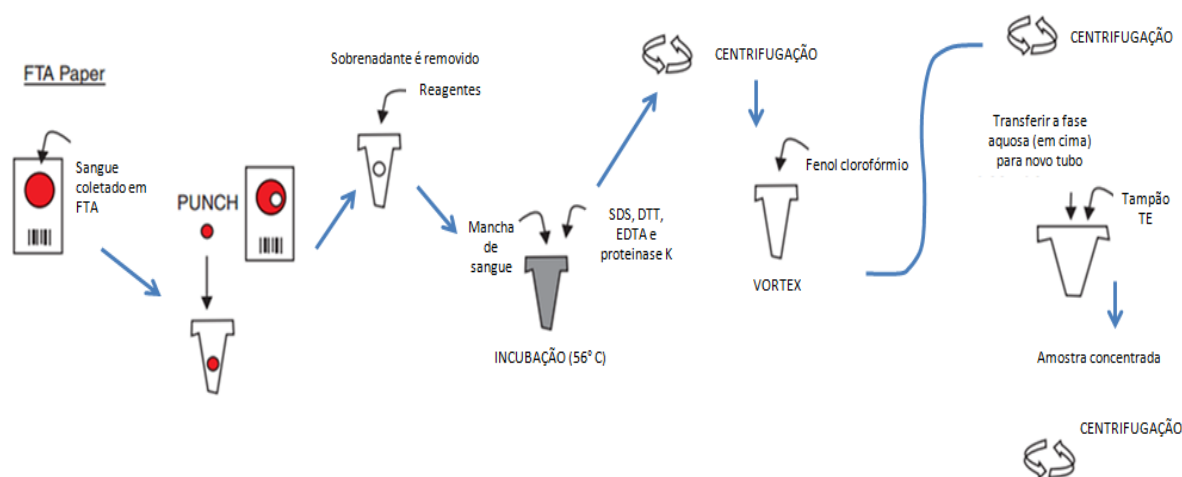


Fonte: Adaptado de Butler, 2010.

A extração orgânica, também referida como Fenol-Clorofórmio, é mais complexa do que a extração por Chelex, pois envolve diversas etapas com a adição de substâncias químicas, o que aumenta o risco de contaminação. Primeiramente, foi adicionado *sodium dodecylsulfate* (SDS, em português, Dodecil Sulfato de Sódio), um tampão e proteinase K, que é responsável por quebrar as membranas celulares e destruir proteínas que protegem o DNA (como as histonas). Nesta solução, foram obtidas duas fases, separadas após centrifugação. Proteínas e remanescentes celulares tendem a se separar na fase orgânica; as moléculas de dupla fita de DNA se concentram na fase aquosa e podem ser usadas nas técnicas de amplificação (BUTLER, 2010) (**Figura 17**).

A extração orgânica, também referida como Fenol-Clorofórmio, é mais complexa do que a extração por Chelex, pois envolve diversas etapas com a adição de substâncias químicas, o que aumenta o risco de contaminação. Primeiramente, foi adicionado *sodium dodecylsulfate* (SDS, em português, Dodecil Sulfato de Sódio), um tampão e proteinase K, que é responsável por quebrar as membranas celulares e destruir proteínas que protegem o DNA (como as histonas). Nesta solução, foram obtidas duas fases, separadas após centrifugação. Proteínas e remanescentes celulares tendem a se separar na fase orgânica; as moléculas de dupla fita de DNA se concentram na fase aquosa e podem ser usadas nas técnicas de amplificação (BUTLER, 2010) (**Figura 17**).

Figura 17 - Método de extração orgânica por Fenol-Clorofórmio



Fonte: Adaptado de Butler, 2010.

2.4 Amplificação

Neste estudo, as amostras foram amplificadas com o uso da técnica de *Polimerase Chain Reaction* (PCR, em português, Reação em Cadeia da Polimerase) (MULLIS, FALOONA, *et al.*, 1986). Foi utilizado o conjunto de oligonucleotídeos (*primers*) L15900-H639, para obtenção da *Control Region* (CR, em português, Região Controle), compreendida entre as posições 16.024 a 576.

Para amostras degradadas em que não se pôde obter produto amplificado da região controle, foram feitas ampliações independentes das regiões HVI (referente as posições

16.024 a 16.569) e HVII/HVIII (compreendidas entre as posições 1 a 576). As regiões HVI e HVII/HVIII foram amplificadas com os conjuntos de oligonucleotídeos L15967-H20 e L16475-H639, respectivamente (**Quadro 2**).

Quadro 2 - Sequências dos oligonucleotídeos utilizados na PCR

<i>Primer</i>	Sequência	Referência
L15900	5'-TAAACTAATACACCAGTCTTGTAACC-3'	Parson <i>et al.</i> , 2007
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'	Parson <i>et al.</i> , 2007
L15967	5'-GTCTTTAACTCCACCATTAGCACC-3'	EAAF
H20	5'-AGCTCCCGTGAGTGGTTAATA-3'	EAAF
L16475	5'-TAGCTAAAGTGAAGTGTATCC-3'	EAAF
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'	Parson <i>et al.</i> , 2007

Nota: EAAF, *comunicação pessoal*;

A região controle do mtDNA foi amplificada com os conjuntos de oligonucleotídeos citados acima, onde cada um dos oligonucleotídeos possuía 2 µM de concentração. A reação de amplificação foi realizada num volume final de 5 µL, com a mistura de reagentes do kit Qiagen Multiplex PCR (Qiagen) (composto por dNTPs, *Taq DNA polimerase* e tampão), mistura de oligonucleotídeos, água livre de RNAses e DNA numa concentração de 1 a 5 ng/µL (**Quadro 3**). O mesmo protocolo foi seguido para amostras em que foram feitas duas amplificações independentes das regiões hipervariáveis.

Para a amplificação, foi feita uma mistura contendo o volume final dos reagentes Qiagen, oligonucleotídeos e água, posteriormente distribuída pelos tubos identificados previamente. Nestes tubos individualizados foram acrescentados os 0.5 µL de DNA extraído. A mistura dos reagentes possuía volume excedente para um controle negativo e um positivo. Ao controle negativo, foi adicionada água ao invés de DNA e, ao controle positivo, foi adicionado DNA que já se havia obtido produto amplificado. Os controles passaram pelo mesmo procedimento de amplificação, ajudando a identificar possíveis contaminantes na reação e a integridade dos reagentes utilizados.

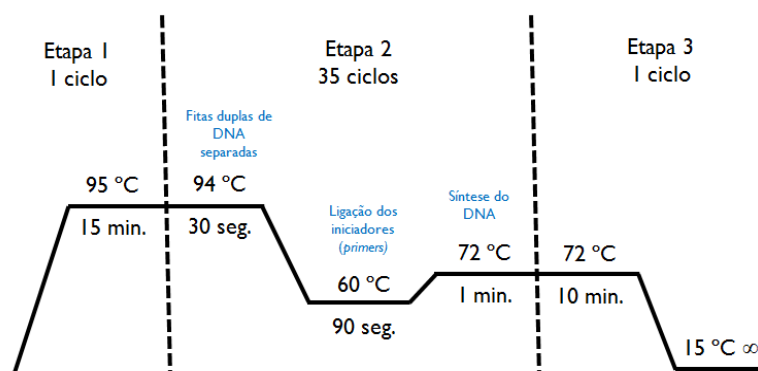
Quadro 3 – Protocolo de amplificação

Reagente	Volume PCR
Qiagen Master Mix 2x	2,5 µL
Primer Mix 2 µM	0,5 µL
H ₂ O	1,5 µL
MIX	4,5 µL
DNA	0,5 µL
Volume Final	5,0 µL

Nota: para amostras com concentração de DNA abaixo de 2ng/µL usou-se 1 µL de DNA, e 1 µL de H₂O.

A amplificação da região controle se deu em um termociclador (GeneAmp® PCRSystem 9700 Thermal Cycler) programado previamente com as rampas de temperatura necessárias para a reação (**Figura 18**).

Figura 18 - Alterações de temperatura do termociclador durante a amplificação do DNA



Após a reação de amplificação, foi realizado um teste qualitativo das amostras através de eletroforese em gel de agarose a 1,2%. Para tal, foi misturado 1,0 µL de corante *safer* (Kasvi) para cada 2,0 µL de amostra, que eram aplicados em cada poço do gel. Ao final da eletroforese, a visualização de uma banda no gel permitiu confirmar a amplificação da sequência de DNA mitocondrial e o tamanho aproximado do fragmento.

2.5 Purificação dos produtos amplificados

As amostras amplificadas com sucesso passaram por um processo de purificação com ExoSAP (Thermo Fisher Scientific), uma solução hidrolítica que retira os remanescentes da reação de amplificação. Este reagente é composto por duas enzimas distintas. A primeira enzima Exonuclease I consome os oligonucleotídeos residuais e a segunda enzima, *Shrimp*

recombinant alkaline phosphatase (SAP, em português, Fosfatase Alcalina de Camarão Recombinante), destrói as bases nucleotídicas remanescentes.

Para esta reação, foi adicionado 1,2 µL do reagente ExoSAP (Thermo Fisher Scientific) a 3,0 µL de produto amplificado, seguido de incubação, com ativação por 15 minutos a 37°C e inativação de 15 minutos a 85°C.

2.6 Sequenciamento de SANGER

A reação de sequenciamento de Sanger possibilitou a determinação da posição das bases nucleotídicas na sequência de mtDNA.

Para a reação de sequenciamento, com volume final de 5 µL, foi feita uma mistura com o kit *BigDye Terminator v3.1* (Thermo Fisher Scientific), que contém os quatro ddNTPs, dNTPs e *TaqDNA polimerase*; tampão (Thermo Fisher Scientific), responsável por manter o pH da reação constante; oligonucleotídeos à 2 µM e água livre de *RNAses* (**Quadro 4**).

Quadro 4 - Protocolo para cada reação de Sequenciamento

Reagente	Volume
Kit BigDye 5X	0,5 µL
Buffer 5X	0,75 µL
Primer 2 µM	0,5 µL
H ₂ O	2,5 µL
MIX	4,5 µL
DNA	0,5 µL

Cada amostra passou por pelo menos duas reações de sequenciamento separadas, com diferentes oligonucleotídeos, que possibilitaram a cobertura das regiões hipervariáveis da região controle do DNA mitocondrial. Dependendo da demanda da amostra (estado das sequências, heteroplasmias de comprimento e/ou de posição), foram feitos sequenciamentos com outros oligonucleotídeos (nos sentidos senso e/ou anti-senso) que possibilitaram a identificação da posição ou região dúbia (**Quadro 5**).

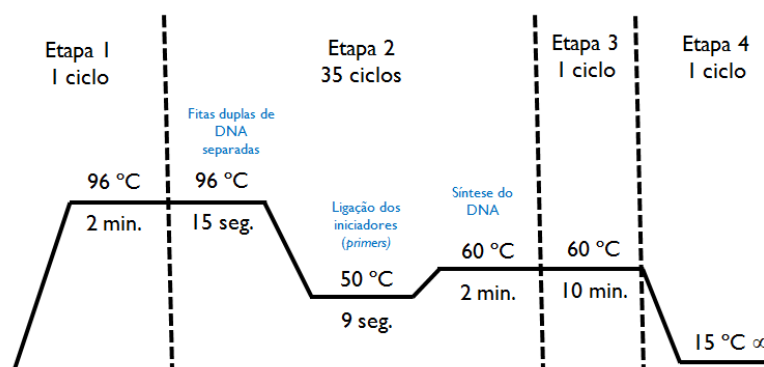
Condições do Sequenciamento: a reação foi feita no termociclador GeneAmp® PCRSystem 9700 Thermal Cycler (**Figura 19**).

Quadro 5 - Sequências dos oligonucleotídeos utilizados para sequenciamento

Primer	Sequência	Referência
L15900	5'-TAAACTAATACACCAGTCTTGTAACC-3'	Parson <i>et al.</i> , 2007
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'	Parson <i>et al.</i> , 2007
L15967	5'-GTCTTTAACTCCACCATTAGCACC-3'	EAAF
H20	5-AGCTCCCGTGAGTGGTTAATA-3'	EAAF
L16475	5'-TAGCTAAAGTGAAGTGTATCC-3'	EAAF
L15997	5'-CACCATTAGCACCCAAAGCT-3'	Alonso <i>et al.</i> , 2003
L16268	5'-CACTAGGATACCAACAACC-3'	Gabriel <i>et al.</i> , 2001
L16555	5'-CCCACACGTTCCCCTTAAAT-3'	Simão <i>et al.</i> , 2018
L00314	5'-CCGCTTCTGGCCACAGCACT-3'	Parson <i>et al.</i> , 2007
H16322	5'-TGGCTTTATGTACTATGTAC-3'	Gabriel <i>et al.</i> , 2001
H00016	5'-CCCGTGAGTGGTTAATAGGGT-3'	Parson <i>et al.</i> , 2007
H00159	5'-AAATAATAGGATGAGGCAGGAATC-3'	Gabriel <i>et al.</i> , 2001
H00388	5'-TGGTTAGGCTGGTGTAGGG-3'	Hwan <i>et al.</i> , 2008
H00599	5'-TTGAGGAGGTAAGCTACATA-3'	Parson <i>et al.</i> , 2007

Nota: EAAF, comunicação pessoal;

Figura 19 - Alterações de temperatura do termociclador durante a reação de sequenciamento



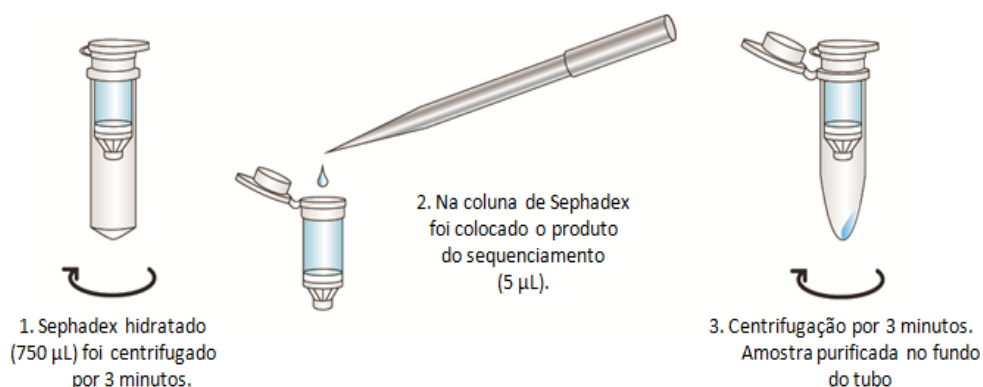
2.7 Purificação dos produtos sequenciados

Após o sequenciamento, foi feita nova purificação para a retirada dos remanescentes da reação. Para tal, foi feita purificação por coluna de Sephadex (Ilustra Sephadex DNA Grade; GE Healthcare) (**Figura 20**). Esta coluna é construída por polímero de sephadex

hidratado com água miliQ, que constrói uma barreira física à passagem dos componentes residuais.

Por fim, cada reação de sequenciamento foi distribuída separadamente em placas de 96 poços com formamida e colocadas sob eletroforese em capilares individualmente no sequenciador automático *ABI 3500* (Thermo Fisher Scientific).

Figura 20 - Esquema de purificação por colunas de Sephadex



Fonte: Adaptado de Google, http://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQxC13Xxb5fhcz1B5KyxU2Id8h869oakCRlg8OOBsLsNMffVUk-5hTGv2dNSR878k_Rjqx1T3Vgw1E&usqp=CAU, acesso em agosto de 2021.

2.8 Análise das sequências

Foram utilizados dois *softwares* para análise das sequências. Primeiramente, as amostras foram observadas individualmente através do software *Sequencing Analysis v5.4*, que permite ver intensidade dos fluorocromos e, conseqüentemente, os picos do eletroferograma. Além disto, analisa as sequências apresentando dando uma cor que simboliza a qualidade da sequência e a amplitude dos picos.

Depois, as sequências obtidas com oligonucleotídeos diferentes para a mesma amostra foram reunidas no *software SeqScape v2.7*, que torna possível o alinhamento da região controle completa da amostra e comparação com a Sequência rCRS.

A região controle da amostra em questão foi alinhada à sequência de referência de Cambridge revista e desta comparação foram obtidas as diferenças entre ambas sequências. O conjunto das diferenças encontradas na amostra em estudo em comparação com a rCRS é

denominado haplótipo e anotado de acordo com as regras de nomenclatura padronizadas pela União Internacional de Química Pura e Aplicada (IUPAC) e às normas aplicadas para a determinação de polimorfismos pelo Comitê da Sociedade Internacional de Genética Forense (ISFG) (PARSON, W., GUSMÃO, *et al.*, 2014).

2.8.1 Determinação de haplogrupos

Após análise das sequências da região controle, foram obtidos os haplótipos, ou seja, o conjunto de polimorfismos encontrados nas amostras, comparativamente à rCRS.

Para cada haplótipo, pôde-se determinar seu haplogrupo e provável local de origem com a plataforma EMPOP, de acordo com a árvore filogenética *PhyloTree build 17* (VAN OVEN, KAYSER, 2009). Ainda na plataforma EMPOP, foi realizado o controle de qualidade das amostras usando as ferramentas “*EMPCheck*” e “*Network*” (ZIMMERMANN, RÖCK, *et al.*, 2014).

A frequência de haplogrupos foi calculada por contagem direta.

2.9 **Análises estatísticas**

A diversidade haplotípica (*haplotype diversity* ou *HD*) foi calculada com o *software* Arlequin 3.5.2.2. (EXCOFFIER, LAVAL, *et al.*, 2005). Este parâmetro é dado pela probabilidade de dois indivíduos ou haplótipos escolhidos ao acaso numa população serem diferentes, ou seja, mede a heterogeneidade das linhagens maternas encontradas. A diversidade genética então varia entre zero e um, onde um demonstra a diversidade máxima encontrada numa dada população.

Para cada um dos três grupos populacionais da Nigéria foi feito um estudo filogeográfico através do *software* Network v10.1.0.0 (<https://www.fluxus-engineering.com>) usando o método “median-joinig” (BANDELT, FORSTER, *et al.*, 1999). As redes ou árvores obtidas neste estudo mostram as relações entre as linhagens da população e a diversidade intrapopulacional das mesmas.

Com o intuito de observar as proximidades genéticas entre as populações do estudo e outras populações da região, foi feito o cálculo da distancia genética (F_{ST}) usando o software Arlequin ver. 3.5.2.2 (EXCOFFIER, LAVAL, *et al.*, 2005). Este parâmetro determina as proximidades genéticas entre as populações do estudo. A obtenção destes resultados possibilitou a construção de um gráfico de dispersão *Multidimensional Scaling* (MDS, em português, Escala Multidimensional) no software STATISTICA (data analysis software system), ver.8.0 (www.statsoft.com). O MDS é uma representação visual das relações de afinidade genética observada entre as populações.

2.9.1 Análises de diferenciação populacional

De forma a verificar se existiam diferenças significativas entre os três grupos estudados, foi feito um teste exato de não diferenciação (p valor), com base nos valores de F_{ST} , utilizando o software Arlequin ver. 3.5.1.2. Os valores de F_{ST} variam entre zero e um, onde um é o valor máximo de distância entre duas populações.

Para possibilitar a padronização das comparações de variabilidade molecular e distâncias genéticas entre as populações, todas as sequências foram reduzidas à extensão das posições 16.030 a 576 da região controle do mtDNA, de modo a se ajustarem ao comprimento das sequências de populações de referência obtidas em Cerezo *et al.* (2016). Com o mesmo fim, também foram excluídos outros polimorfismos do tipo inserção e deleção nas posições 16093.xC, 309.xC, 315.xC, 524.xC, 523-524del, 573.xC., ausentes nas sequências da base de dados *1000 Genome Project*, que acabam por variar no comprimento das mesmas. Para estas sequências, foi utilizada *Next-Generation Sequencing* (NGS, em português, Sequenciamento de Nova Geração), que faz leituras do genoma em quantidade massiva e tende a obter resolução e alinhamento diferentes do Sequenciamento de Sanger, fatores que interferem na determinação de heteroplasmias de comprimento. Além disso, como foram utilizadas amostras de outros estudos, feitos em laboratórios diferentes, o uso de protocolos no sequenciamento automatizado ABI também pode interferir na detecção destes polimorfismos (PARSON, GUSMÃO, *et al.*, 2014; JUST, IRWIN, *et al.*, 2015).

Os valores de p obtidos foram confrontados para um nível de significância obtido após a correção de Bonferroni (ARMSTRONG, 2014; CHEN, FENG, *et al.*, 2017), onde o valor de 0,05 foi dividido pelo número de comparações feitas entre as populações.

3. RESULTADOS

A lista dos haplótipos e haplogrupos correspondentes encontrados nas 285 amostras de indivíduos dos grupos da Nigéria é apresentada no Apêndice 1.

Cabe citar que o presente trabalho acerca das linhagens maternas da Nigéria gerou um manuscrito (em fase de finalização), o qual será submetido a uma das revistas de circulação internacional: *Human Genetics* ou *European Journal of Human Genetics* (Anexo 1).

3.1 Diversidade haplotípica e de haplogrupo

Os resultados das análises de diversidade haplotípica (*HD*) definem a probabilidade de dois indivíduos escolhidos ao acaso numa população serem diferentes. Os três grupos etnolinguísticos exibiram alta diversidade intrapopulacional. Os valores de diversidade haplotípica (**Quadro 6**) foram maiores do que aqueles relatados anteriormente em outras populações da região Centro-oeste da África por Cerezo *et al.* (0,984) (2016), Čížková *et al.* (2017) (entre 0,7252 e 0,9985) e Fendt *et al.* (0,98) (2012).

Ao observar a distribuição do número de diferenças par a par entre os haplótipos dos três grupos estudados (**Figuras 21-23**), pôde-se perceber que, em mais de 85% das comparações, os haplótipos foram diferentes entre si em mais de 10 posições polimórficas. A distribuição de haplótipos em cada população encontra-se representada em formato de *Network*, nas Figuras 21-23. Uma alta diversidade foi detectada não apenas entre, mas também dentro dos haplogrupos, como pode ser observado nas linhagens L2 e L3 dos três grupos étnicos. Especificamente no grupo Hausa (**Figura 21**), pôde-se observar uma maior uniformidade no espalhamento dos haplótipos e maior diversidade de haplótipos e haplogrupos, representado pelo compartilhamento de um único haplótipo entre duas amostras/indivíduos. Na representação do grupo Yoruba (**Figura 22**), por sua vez, a árvore demonstrou grande compartilhamento de haplótipos nos haplogrupos L1 e L3 e menor diferenças médias entre os haplótipos dentro de L2, o que levou a maior aglomeração dos dados. Por último, o grupo étnico Igbo (**Figura 23**) foi o que obteve maior quantidade de compartilhamento de haplótipos na comparação entre os outros grupos, especialmente dentro

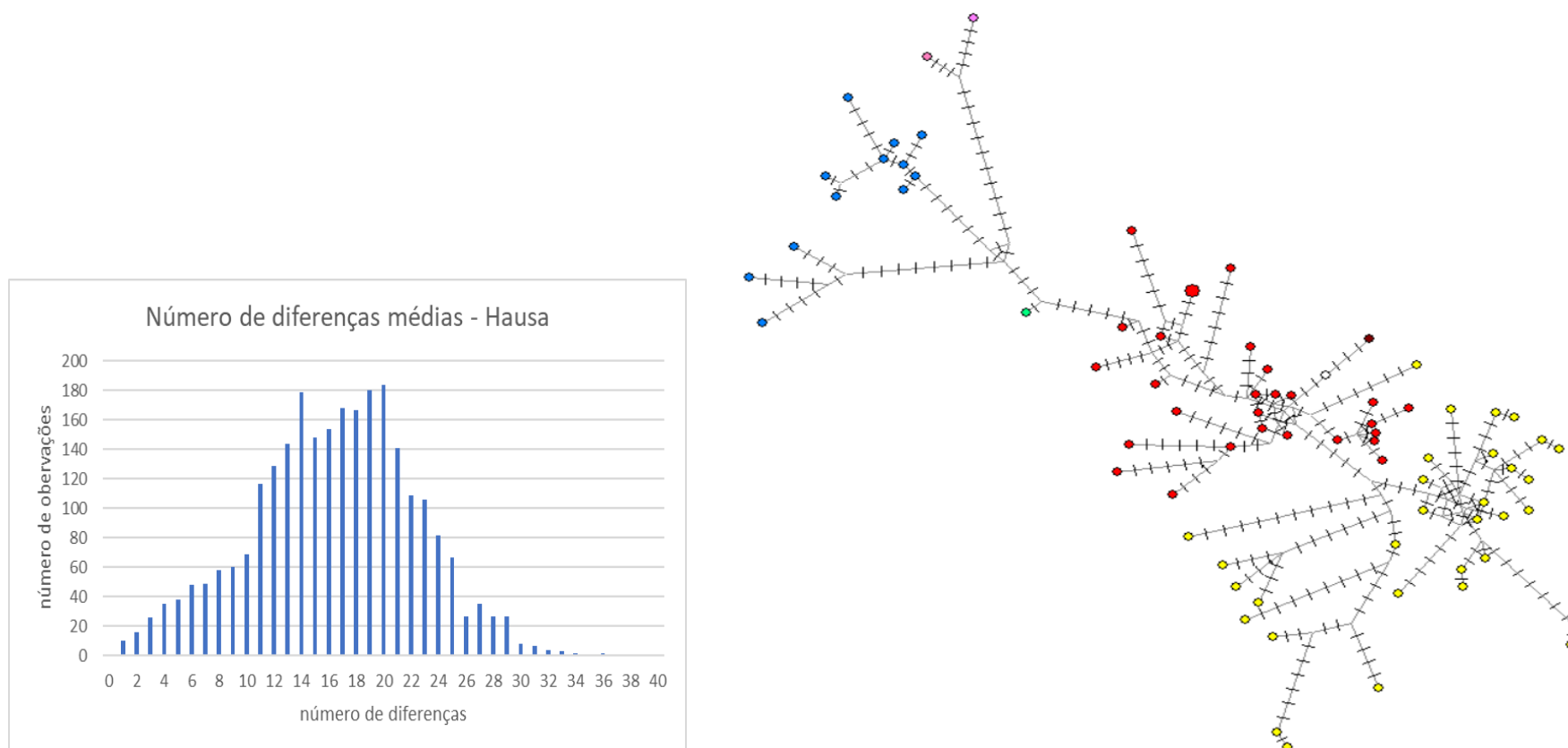
dos haplogrupos L1 e L3. Também foi observado que as menores diferenças médias entre haplótipos desse grupo foi encontrada em L2 e L3, como pode ser visto com a maior concentração de dados na árvore.

Quadro 6 - Populações da região Centro-oeste usadas para comparação e seus respectivos valores de diversidade haplotípica

País	Grupo étnico	Referência	N	Haplótipos únicos	Diversidade haplotípica (HD)
Gana	-	Göbel <i>et al.</i> , 2020	22	22	1,0000 +/- 0,0137
Costa do Marfim	-	Göbel <i>et al.</i> , 2020	27	27	1,0000 +/- 0,0101
Nigéria	-	Göbel <i>et al.</i> , 2020	28	28	1,0000 +/- 0,0095
Camarões	-	Göbel <i>et al.</i> , 2020	11	11	1,0000 +/- 0,0388
Togo	-	Göbel <i>et al.</i> , 2020	34	34	1,0000 +/- 0,0071
Nigéria	Yoruba	Este estudo	112	102	0,9982 +/- 0,0015
Nigéria	Hausa	Este estudo	73	68	0,9977 +/- 0,0029
Serra Leoa	Mende	1000 GP	68	62	0,9974 +/- 0,0030
Nigéria	Yoruba	1000 GP	108	97	0,9971 +/- 0,0021
Burquina Faso	Mossi	Čížková <i>et al.</i> , 2017	36	34	0,9968 +/- 0,0075
Nigéria	Esan	1000 GP	99	86	0,9965 +/- 0,0022
Senegal	Halpularen	Čížková <i>et al.</i> , 2017	47	43	0,9963 +/- 0,0053
Nigéria	Igbo	Este estudo	100	84	0,9960 +/- 0,0022
Gâmbia	-	1000 GP	113	93	0,9959 +/- 0,0019
Chade	Kanembu	Cerezo <i>et al.</i> , 2016	50	43	0,9943 +/- 0,0052
Burquina Faso	Gourmantche	Čížková <i>et al.</i> , 2017	36	33	0,9937 +/- 0,0090
Burquina Faso	Gurunsi	Čížková <i>et al.</i> , 2017	33	29	0,9924 +/- 0,0094
Senegal	Serrere	Čížková <i>et al.</i> , 2017	48	40	0,9920 +/- 0,0061
Gana	Akan	Fendt <i>et al.</i> , 2012	192	126	0,9915 +/- 0,0020
Chade	Arabs Abala	Čížková <i>et al.</i> , 2017	51	42	0,9914 +/- 0,0060
Guiné	Fulani Fouta Djallon	Čížková <i>et al.</i> , 2017	51	42	0,9914 +/- 0,0060
Senegal	Bedik	Čížková <i>et al.</i> , 2017	41	33	0,9878 +/- 0,0086
Chade	Maba	Čížková <i>et al.</i> , 2017	42	31	0,9849 +/- 0,0084
Chade	Dangaleat	Čížková <i>et al.</i> , 2017	50	35	0,9820 +/- 0,0079
Camarões	Tcheboua Fulani	Cerezo <i>et al.</i> , 2016	40	30	0,9795 +/- 0,0117
Chade	Daza	Čížková <i>et al.</i> , 2017	40	27	0,9769 +/- 0,0109
Níger	Tuareg	Cerezo <i>et al.</i> , 2016	27	19	0,9601 +/- 0,0250
Burquina Faso	Fulani Ziniare	Čížková <i>et al.</i> , 2017	55	27	0,9455 +/- 0,0155
Mali	Tuareg	Cerezo <i>et al.</i> , 2016	14	11	0,9341 +/- 0,0611
Burquina Faso	Tuareg	Cerezo <i>et al.</i> , 2016	20	8	0,8421 +/- 0,0595
Senegal	Fulani Ferlo	Čížková <i>et al.</i> , 2017	56	24	0,8403 +/- 0,0476

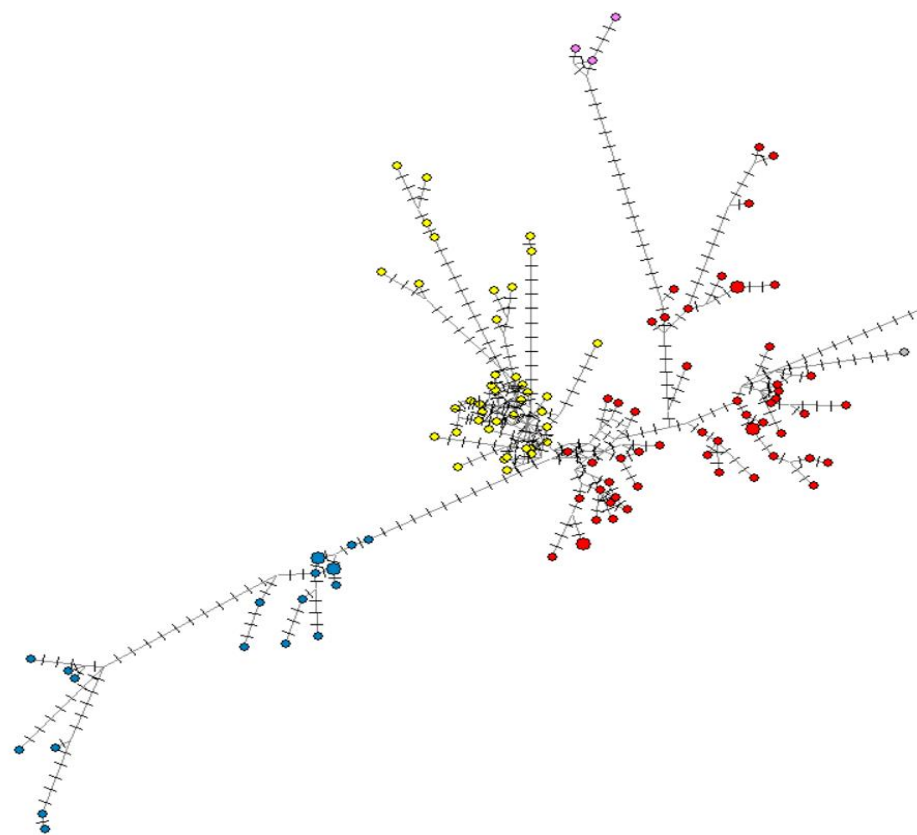
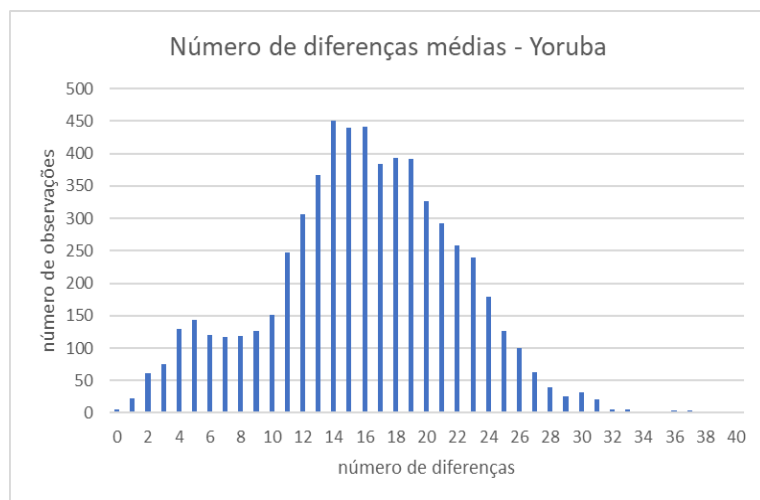
Nota: As amostras do 1000 Genome Project estão identificadas com a sigla 1000 GP.

Figura 21 - Representação do número de diferenças médias par a par (à esquerda) e *median joining network* (à direita), dos haplótipos da região controle do mtDNA encontrados no grupo Hausa



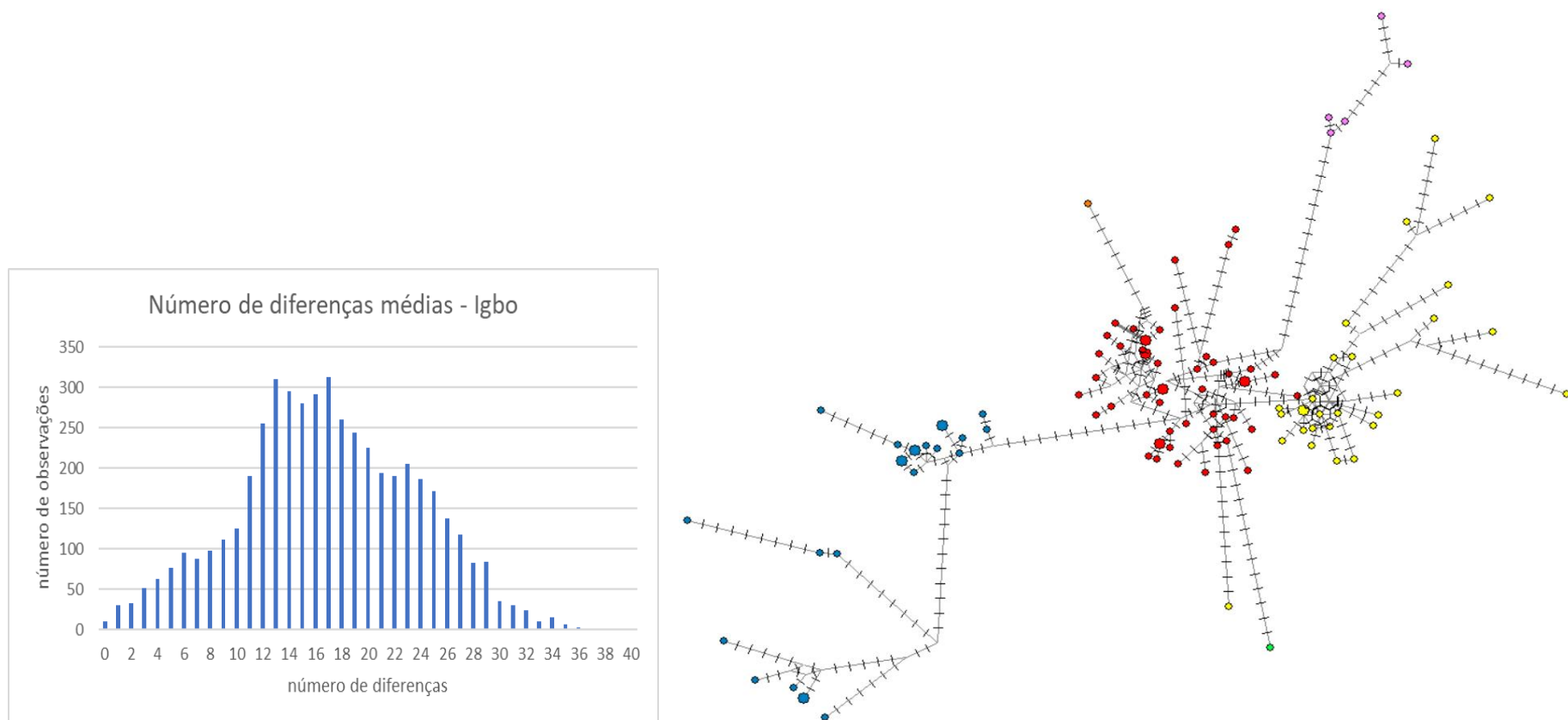
Nota: os haplogrupos são representados pelas cores - marrom (H), rosa (L0); azul (L1); amarelo (L2), vermelho (L3), verde (L4) e branco (R). Os círculos representam os haplótipos e suas respectivas frequências de acordo com tamanho. Cada traço / linha representa os passos mutacionais ou polimorfismos entre os haplótipos.

Figura 22 - Representação do número de diferenças médias par a par (à esquerda) e *median joining network* (à direita), dos haplótipos da região controle do mtDNA encontrados no grupo Yoruba



Nota: os haplogrupos são representados pelas cores - rosa (L0); azul (L1); amarelo (L2), vermelho (L3), verde (L4) e cinza (M). Os círculos representam os haplótipos e suas respectivas frequências de acordo com tamanho. Cada traço / linha representa os passos mutacionais ou polimorfismos entre os haplótipos.

Figura 23 - Representação do número de diferenças médias par a par (à esquerda) e median joining network (à direita), dos haplótipos da região controle do mtDNA encontrados no grupo Igbo

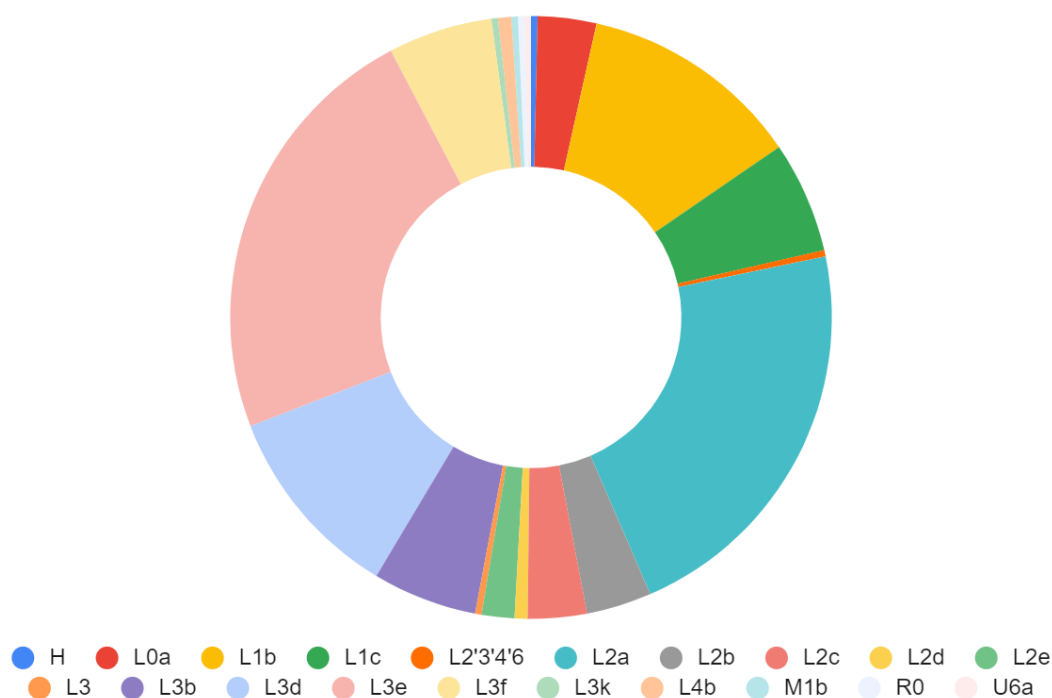


Nota: os haplogrupos são representados pelas cores - rosa (L0); azul (L1); amarelo (L2), vermelho (L3), verde (L4) e laranja (U). Os círculos representam os haplótipos e suas respectivas frequências de acordo com tamanho. Cada traço / linha representa os passos mutacionais ou polimorfismos entre os haplótipos.

Uma análise detalhada da composição dos haplogrupos foi realizada para melhor compreensão da estrutura genética das linhagens maternas nas populações da África Ocidental. A maioria das linhagens maternas encontradas na Nigéria pertence ao macro haplogrupo L (98,6%), característico das populações da África Subsaariana. O haplogrupo L3 (45,6%) foi a linhagem mais frequentemente encontrada nos três grupos étnicos, seguido pelos haplogrupos L2 (31,2%), L1 (17,9%), L0 (3,2%) e L4 (0,7%). Todas as linhagens africanas foram classificadas nos ramos L0a, L1b-c, L2a-e, L3b, L3d-f, L3k e L4b, como detalhado no Gráfico 1. Os subclados L3e, L2a, L1b e L3d representam mais de 60% das linhagens maternas da Nigéria. A relação dos haplogrupos resultante da junção das amostras dos três grupos etnolinguísticos também apresentou amostras não pertencentes aos haplogrupos característicos dessa região do continente (**Figura 24; Apêndice 1**).

Gráfico 1 - Distribuição dos haplogrupos de mtDNA deste estudo

Distribuição geral dos haplogrupos



3.2 Análises de diferenciação populacional

De forma a verificar se existiam diferenças significativas entre os grupos estudados, foram determinadas as distâncias genéticas (F_{ST}) par a par entre as amostras dos três grupos

étnolinguísticos e calculada a probabilidade de não diferenciação (p valor) correspondente. Outras populações da região Centro-oeste da África (**Figura 24**) foram também incluídas nesta análise, para efeitos de comparação.

Figura 24 - Distribuição de amostras do oeste da África usadas para comparação neste estudo reunidas por país



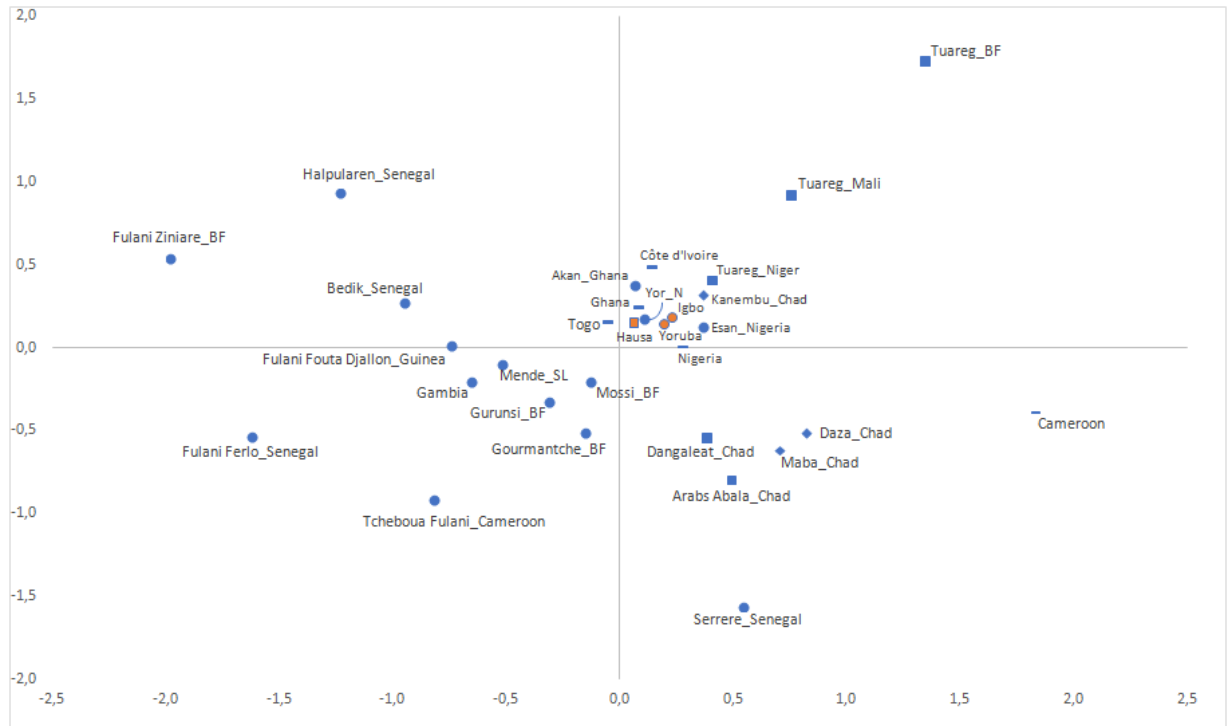
Fonte: Mapa adaptado de Google, <https://www.pngitem.com/search.html?k=africa+continent+>, agosto de 2021.

Nota: Burquina Faso é representado pelo código BF com amostras de Cerezo ($n = 20$) e Čížková ($n = 160$); Camarões de Cerezo ($n = 40$) e Göbel ($n = 11$); Chade de Cerezo ($n = 50$) e Čížková ($n = 183$); Costa do Marfim de Göbel ($n = 27$); Gana de Fendt ($n = 192$) e Göbel ($n = 22$); Guiné de Čížková ($n = 51$); Mali de Cerezo ($n = 14$); Níger de Cerezo ($n = 27$); Nigéria de Göbel ($n = 28$) e 1000 Genome Project ($n = 207$); Senegâmbia representando os países Senegal de Čížková ($n = 192$) e Gâmbia do Projeto Genoma 1000 ($n = 113$); Serra Leoa do 1000 Genome Project ($n = 68$); Togo de Göbel ($n = 34$);

A comparação dos três grupos estudados por meio de distâncias genéticas F_{ST} não mostrou diferenças estatisticamente significativas entre eles. Os valores de F_{ST} foram todos negativos e as probabilidades de não diferenciação estavam acima do nível de significância (Igbo vs. Yoruba, $F_{ST} = -0,00397$, $p = 0,92209 \pm 0,0024$; Yoruba vs. Hausa, $F_{ST} = -0,00409$, $p = 0,86546 \pm 0,0037$; Igbo vs. Hausa, $F_{ST} = -0,00089$, $p = 0,51074 \pm 0,0054$). (**Tabela 1**).

A fim de compreender a dinâmica entre os três grupos etnolinguísticos da Nigéria e outras populações, foram reunidos dados já publicados para outros países da região Centro-oeste da África (**Figura 25** e **Quadro 6**). As distâncias genéticas (F_{ST}) foram calculadas entre as populações (**Tabela 1**) e representadas em um gráfico MDS (**Figura 25**).

Figura 25 - Representação gráfica por MDS dos valores de distância genética (F_{ST}) entre os três grupos etnolinguísticos e populações da região centro-oeste da África, referenciadas no Quadro 6



Legenda: Família linguística

- Afro-Asiática
- Níger-Congo
- ◆ Nilo-Saariana
- Não informado

Nota: Amostras de Burquina Faso e Serra Leoa estão representadas pelos códigos BF e SL, respectivamente.

4 DISCUSSÃO

Os resultados do presente trabalho permitiram revelar parte da grande diversidade presente no continente africano, a partir do estudo dos grupos etnolinguísticos mais representativos da Nigéria. As amostras dos três grupos apresentaram altos valores de diversidade dentro da região, assim como em comparação com populações de países vizinhos.

Os grupos Hausa, Yoruba e Igbo não se mostraram significativamente diferentes quanto às suas linhagens maternas, ainda que se tratem de grupos étnicos diferentes. Ademais, apesar do povo Hausa pertencer a uma família linguística diferente dos grupos Igbo e Yoruba, e ter uma história de conflitos e exclusão intergrupar durante a Guerra Civil, foi detectada uma homogeneidade das linhagens maternas das três populações. Esses resultados estão de acordo com estudos demográficos, referentes ao casamento interétnico entre grupos como prática normalizada na Nigéria (OLANIYAN, 2009), especialmente nas cidades de Kano (Yoruba-Hausa), Ibadan e Abeokuta (Hausa-Yoruba). Contudo, cabe destacar que em um passado recente, a maioria dos casamentos nigerianos se deu entre integrantes de um mesmo grupo (METZ, 1991). Essas regiões apresentam evidências de familiaridade com diferentes grupos, com adoção de mecanismos de manejo de tensões que possibilitam/facilitam a convivência entre diferentes tribos.

As amostras nigerianas apresentaram valores de F_{ST} baixos com outras amostras de populações nigerianas relatadas anteriormente (Esan e Yoruba do *1000 Genome Project* e Nigéria (GÖBEL, BODNER, *et al.*, 2020), e com populações geograficamente próximas, nomeadamente do Gana (Göbel e Fendt), Costa do Marfim, Togo e Chade (Göbel e Cerezo). Em contraste, dentro da região Centro-oeste, as populações mais distantes geograficamente (Senegal, Guiné, Gana, Camarões) apresentaram-se mais distantes e dispersas dos três grupos étnicos deste estudo, de acordo com o gráfico MDS (**Figura 25**).

A alta homogeneidade entre populações geograficamente próximas, independentemente de sua afiliação etnolinguística, pode ser explicada pelo alto fluxo gênico mediado por mulheres. Por exemplo, nas populações de Burquina Faso, uma subestrutura no componente genético das linhagens de origem paterna paralelo a uma quase homogeneidade no componente das linhagens maternas já foi descrita (SEIELSTAD, MINCH, *et al.*, 1998; BARBIERI, WHITTEN, *et al.*, 2012). A ocorrência de casamentos interétnicos é reconhecido na história antiga da região (antes do século XVI), uma vez que as trocas comerciais entre comunidades vizinhas eram comuns (FALOLA, HEATON, 2008). O fluxo genético entre

diferentes grupos étnicos está também presente nos resultados deste estudo, tal como pode ser observado no gráfico MDS, onde há um agrupamento central que inclui populações geneticamente próximas, mas pertencentes a diferentes grupos etnolinguísticos. Além disso, no século 16, o estabelecimento e exploração dos britânicos na África centro-oeste resultaram na divisão do território atualmente conhecido como Nigéria. A divisão política das nações/países na África durante a Conferência de Berlim (no século 19) não respeitou as fronteiras políticas já estabelecidas impostas pelos nativos e suas afiliações etnolinguísticas, o que levou à fragmentação dos grupos (ALESINA, EASTERLY, *et al.*, 2011).

Deste modo, nem sempre é possível prever a composição genética materna das populações com base em sua afiliação etnolinguística ou distribuição geográfica na região Centro-oeste da África. A falta de subestrutura da origem materna na Nigéria e em países vizinhos, é o oposto do que pode ser encontrado para marcadores de linhagem paterna presentes nos grupos étnicos Hausa, Yoruba e Igbo. Os resultados obtidos pela análise do componente masculino da Nigéria (dados não mostrados, comunicação pessoal) destacam diferenças entre eles.

Embora não haja dados étnicos para as populações de Togo e Gana, é mais provável que elas pertençam à família linguística Níger-Congo, dada a sua predominância na região. A proximidade dessas populações com o grupo Hausa pode ser explicada pela disseminação do dialeto hausa como língua materna nessas regiões da África Ocidental (SIMPSON, OYÈTÁDÉ, 2008). Entre os séculos XIV e XVI, as etnias, apesar de um tanto delimitadas, eram econômica e socialmente interdependentes. A comunicação entre os grupos era tão restrita que as trocas sociais e culturais eram mantidas por meio de negócios ou mesmo de casamentos entre si. Um desses grupos itinerantes foi o Hausa. Este grupo era muito conhecido no comércio Transaariano por seu perfil mercantil, o que possibilitou a eventual disseminação do dialeto para além de seus estados, com fins comerciais. Os Hausa também se estabeleceram fora da Hausaland, para facilitar os negócios, e esse movimento resultou em parentesco e linguagem compartilhados com outras comunidades (FALOLA, HEATON, 2008).

A amostra nigeriana do grupo Yoruba do *1000 Genome Project* também apresentou maior proximidade com os Hausa do que com a população Yoruba deste estudo. Isso pode ser explicado pelo local de coleta, já que amostras do *1000 Genome Project* foram coletadas em Ibadan, no estado Yoruba. Embora seja uma região ocupada principalmente por Yorubas, existe uma considerável população Hausa estabelecida na área, o que explica tal proximidade.

A proximidade entre a amostra Yoruba e Igbo deste estudo e a população Esan do *1000 Genome Project* era esperada, visto que esses grupos pertencem à mesma família linguística, Níger-Congo. Além disso, o povo Esan está presente em grande proporção no estado de Edo, localizado entre o sudoeste e o sudeste, onde os grupos Yoruba e Igbo vivem.

Numa perspectiva mais ampla, as amostras do Chade foram as únicas que se agruparam de forma independente à etnia, com exceção dos Kanembu. A proximidade com este grupo pode ser explicada pela integração da Nigéria com a África Ocidental pela Bacia do Chade. A área da Bacia do Chade é conhecida por ser uma região altamente diversificada em termos etnolinguísticos, com populações Afro-Asiáticas, Nilo-Saarianas e Níger-Congolesas vivendo ali (ČERNÝ, SALAS, *et al.*, 2007). Assim, essa região pode ter sido facilitadora do fluxo genético de forma a permitir a troca e homogeneização das linhagens maternas dos grupos nigerianos e Kanembu.

A partir do gráfico do MDS, também é possível observar a separação das amostras pertencentes à família linguística Níger-Congo, independentemente do país, como visto nas amostras do Senegal, Gâmbia e Burquina Faso.

Filogenias

Uma análise detalhada da composição dos haplogrupos foi realizada para melhor compreensão da estrutura genética das linhagens maternas nas populações da África Ocidental.

A maioria das linhagens maternas encontradas na Nigéria pertence ao macro haplogrupo L (98,6%), característico das populações da África Subsaariana. O haplogrupo L3 foi a linhagem mais frequentemente encontrada nos três grupos étnicos, seguido pelos haplogrupos L2, L1, L0 e L4. Todas as linhagens africanas foram classificadas nos ramos L0a, L1b-c, L2a-e, L3b, L3d-f, L3k e L4b, que possuem frequências mais altas em populações da região Centro-oeste da África, em sua maioria. A exceção é o sub haplogrupo L2a, que é mais frequente na região Sudeste do continente (SALAS, RICHARDS, *et al.*, 2002; CEREZO, GUSMÃO, *et al.*, 2016).

O haplogrupo de mtDNA mais antigo, L0, data de 150-180.000 anos atrás, com origem no sul da África. Este macro haplogrupo é mais frequente nas regiões Sul e Leste, em ambos os grupos de língua KhoeSan / KhoiSan e Bantu (TORRONI, ACHILLI, *et al.*, 2006; ROSA, BREHEM, 2011; RITO, RICHARDS, *et al.*, 2013). A linhagem L0a2, conhecida como marcador Bantu, não foi encontrada nas amostras estudadas. Mesmo que essas três populações representem grupos não-Bantu, foi mostrado anteriormente que essa sublinhagem

representa 68% de todos os sub-haplogrupos L0 mtDNA no centro-oeste (CEREZO, GUSMÃO, *et al.*, 2016). Contrariamente, todos os haplótipos L0 encontrados neste estudo pertencem à linhagem L0a1, a mais frequente no sudeste da África

O macro haplogrupo L1 tem entre 140-150.000 anos (HORAI, HAYASAKA, *et al.*, 1995; TORRONI, ACHILLI, *et al.*, 2006; BEHAR, VILLEMS, *et al.*, 2008) e é mais frequente no sudoeste da África. Ambas as sublinhagens L1 representadas neste estudo, L1b e L1c, têm uma origem provável na África Central. As amostras Yoruba, Igbo e Hausa apresentaram altas frequências dos sub-clados L1b (11,9%), frequentes nas populações de nômades Fulani (ČERNÝ, SALAS, *et al.*, 2007), mostrando a influência deste grupo de língua Níger-Congo. Os Fulani são tradicionalmente pastoralistas nômades e estão amplamente distribuídos por todo o continente, desde a África Ocidental e Central, do Sahel ao Sudão (FAN, KELLY, *et al.*, 2019). Em contrapartida, a linhagem L1c, presente em 17 amostras nigerianas, demonstra um contato com populações sul-africanas, uma vez que a sua origem mais provável é em Angola e países vizinhos (BATINI, COIA, *et al.*, 2007; ČERNÝ, SALAS, *et al.*, 2007).

O macro haplogrupo L2 tem os valores mais altos de diversidades haplotípica e nucleotídica na região Centro-oeste (SALAS, RICHARDS, *et al.*, 2002; SILVA, ALSHAMALI, *et al.*, 2015; CEREZO, GUSMÃO, *et al.*, 2016). Estima-se que a linhagem L2 se originou entre 90-105.000 anos atrás (GONDER, MORTENSEN, *et al.*, 2007; BEHAR, VILLEMS, *et al.*, 2008). O clado L2a, presente com uma frequência de 21,8% na Nigéria, é o mais frequente no continente africano, totalizando um quarto de todos os macro haplogrupos L2. Os sub-clados L2 menos representados, nomeadamente L2b-d, provavelmente têm origem nas regiões Centro-oeste e Oeste da África, uma vez que se encontram amplamente restritos a essas regiões, com poucas ocorrências no sudeste (SALAS, RICHARDS, *et al.*, 2002). Um dos sub-clados L2 menos representados, L2e, tem origem provável no oeste da África, sendo frequente nesta região e disperso pela região do Sahel (SILVA, ALSHAMALI, *et al.*, 2015).

Estima-se que o macro haplogrupo L3 surgiu na região da África Central ou Leste a ~ 65-80.000 anos atrás (SALAS, RICHARDS, *et al.*, 2002; FORSTER, 2004; TORRONI, ACHILLI, *et al.*, 2006; BEHAR, VILLEMS, *et al.*, 2008), mas é mais diversificado na região Centro-oeste (CEREZO, GUSMÃO, *et al.*, 2016). As sublinhagens L3e (0,7%), L3e2 (14%) e L3e1 (4,9%), encontradas na Nigéria, são comuns entre as populações de língua Bantu e estão associadas à dispersão Bantu do centro-oeste ao sudeste da África (SALAS, RICHARDS, *et al.*, 2002). Na região Centro-oeste, L3e1 foi anteriormente encontrado em baixas frequências, embora na Nigéria tenha sido a segunda sublinhagem mais representada, dentro do ramo L3e.

Ainda dentro de L3, o haplogrupo L3d (10,5% na Nigéria) também é frequente na região centro-oeste (CEREZO, GUSMÃO, *et al.*, 2016). Esta linhagem, em conjunto com a L3b, é comum em populações de nômades Fulani (ČERNÝ, SALAS, *et al.*, 2007). Esses haplogrupos provavelmente se espalharam na Nigéria pela fusão do grupo Hausa-Fulani, do norte da Nigéria (DUCROTOY, MAJEKODUNMI, *et al.*, 2018). O haplogrupo L3f surgiu no leste (ČERNÝ, SALAS, *et al.*, 2007; ERN, FERNANDES, *et al.*, 2009). O ramo L3f1, com uma dispersão inicial na região Oeste, foi encontrado em todas as nossas amostras pertencentes a este haplogrupo. Finalmente, L3k, o subclado L3 menos representado, é raro e está presente principalmente na África do Norte (SOARES, ALSHAMALI, *et al.*, 2012).

O haplogrupo L4 foi observado em apenas duas amostras deste estudo. Este clado totaliza 15% de todas as linhagens de DNA mitocondrial presentes no leste da África, sendo mais frequente nesta região do que em outras áreas do continente (CEREZO, GUSMÃO, *et al.*, 2016).

As linhagens restantes, encontradas nas amostras nigerianas, pertencem a haplogrupos não L (M1, R0 e U6) e sua presença pode ser explicada por uma influência norte/mediterrânea, que foi facilitada pelo Corredor da Bacia do Chade (SALAS, RICHARDS, *et al.*, 2002; OLIVIERI, ACHILLI, *et al.*, 2006; ČERNÝ, SALAS, *et al.*, 2007; CEREZO, ČERNÝ, *et al.*, 2011). Entre os três grupos nigerianos, a maior proporção dos haplogrupos não L foi encontrada nos Hausa, que pode ser resultante de um maior contato com a região Norte da África. O haplogrupo H, que é altamente diverso na atual Europa Ocidental (BROTHERTON, HAAK, *et al.*, 2013), também foi encontrado em uma amostra do grupo Hausa. De acordo com Ottoni *et al.* (2010), este haplogrupo tem altas frequências em populações Tuareg, sendo a fonte mais provável do haplogrupo na amostra nigeriana.

CONSIDERAÇÕES FINAIS

Apesar de ser o local que comporta a maior diversidade humana em diversos aspectos, a África ainda é pouco representada nas bases de dados genéticos. A Nigéria, o país mais populoso do continente, retrata grande parte da diversidade Africana, apresentando ao menos 250 grupos etnolinguísticos em seu território.

Com o intuito de estabelecer uma base de dados sobre as populações nigerianas de modo a contemplar suas particularidades, foram estudados os três maiores grupos etnolinguísticos do país: Hausa, Igbo e Yoruba.

Com os resultados obtidos do estudo da região controle do DNA mitocondrial foi possível destacar uma homogeneidade entre as três populações nigerianas, independentemente da sua localização geográfica ou grupo etnolinguístico. Apesar do histórico de conflitos do país, majoritariamente de cunho étnico e religioso, as linhagens maternas de suas populações apresentaram-se homogêneas, evidenciando a presença de fluxo gênico entre os grupos étnicos analisados.

Pôde-se também observar que os grupos da Nigéria obtiveram baixos valores de distância genética com populações próximas geograficamente, mesmo que com diferentes afiliações linguísticas. Tais resultados demonstram que há alto fluxo gênico do componente materno entre as populações do Centro-oeste africano, a partir do qual pode-se inferir que a patrilocalidade contribuiu significativamente na história e migração diferencial das mulheres desta região. Ainda, a presença na Nigéria de linhagens não características da região subsaariana apontam para uma influência de populações do Norte da África. No entanto, dados sobre as linhagens paternas presentes na Nigéria serão essenciais para corroborar estes achados.

Acreditamos que a disponibilidade dos dados obtidos no presente trabalho irá contribuir para a elaboração de bases de dados mais detalhadas e precisas, no que diz respeito à descrição da ancestralidade materna africana.

CONCLUSÕES

- a) A alta heterogeneidade dos três maiores grupos étnicos da Nigéria pode ser demonstrada pelos altos valores de diversidade haplotípicas encontrados, condizentes com o observado na região Centro-Oeste;
- b) Os três grupos analisados não demonstraram diferenças estatisticamente significativas entre eles, como visto pelas baixas distâncias genéticas observadas e pode ser explicado pela homogeneidade das linhagens maternas;
- c) As linhagens maternas das três populações analisadas também apresentaram homogeneidade com países vizinhos, apesar do contraste de distintos grupos étnicos e famílias linguísticas presentes na região;
- d) Conclui-se, portanto, que a homogeneidade observada das linhagens maternas entre os três grupos da Nigéria e países do Centro-oeste são grande indicativo da migração diferencial das mulheres da região, possivelmente promovida pela patrilocalidade.

REFERÊNCIAS

- ALBERTS, B., JOHNSON, A., LEWIS, J., *et al.* **Biologia molecular da célula**. [S.l.: s.n.], 2017.
- ALESINA, A., EASTERLY, W., MATUSZESKI, J. "Artificial states", **Journal of the European Economic Association**, v. 9, n. 2, p. 246–277, 2011. DOI: 10.1111/j.1542-4774.2010.01009.x.
- ANDERSON, S., BANKIER, A. T., BARRELL, B. G., *et al.* "Sequence and organization of the human mitochondrial genome", **Nature**, p. 457–465, 1981. DOI: 10.1038/290457a0. .
- ANDREWS, R. M., KUBACKA, I., CHINNERY, P. F., *et al.* "Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA [5]", **Nature Genetics**, v. 23, p. 147, 1999. DOI: 10.1038/13779.
- ARMSTRONG, R. A. "When to use the Bonferroni correction", **Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)**, v. 34, n. 5, p. 502–508, 2014. DOI: 10.1111/opo.12131.
- BANDELT, H. J., FORSTER, P., RÖHL, A. "Median-joining networks for inferring intraspecific phylogenies", **Molecular Biology and Evolution**, v. 16, n. 1, p. 37–48, 1999. DOI: 10.1093/oxfordjournals.molbev.a026036.
- BARBIERI, C., WHITTEN, M., BEYER, K., *et al.* "Contrasting maternal and paternal histories in the linguistic context of Burkina Faso", **Molecular Biology and Evolution**, v. 29, n. 4, p. 1213–1223, 2012. DOI: 10.1093/molbev/msr291.
- BATINI, C., COIA, V., BATTAGLIA, C., *et al.* "Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa", **Molecular Phylogenetics and Evolution**, v. 43, n. 2, p. 635–644, 2007. DOI: 10.1016/j.ympev.2006.09.014.
- BEHAR, D. M., VILLEMS, R., SOODYALL, H., *et al.* "The Dawn of Human Matrilineal Diversity", **American Journal of Human Genetics**, v. 82, n. 5, p. 1130–1140, 2008. DOI: 10.1016/j.ajhg.2008.04.002. .
- BERNT, M., BRABAND, A., SCHIERWATER, B., *et al.* "Genetic aspects of mitochondrial genome evolution", **Molecular Phylogenetics and Evolution**, v. 69, n. 2, p. 328–338, 2013. DOI: 10.1016/j.ympev.2012.10.020. .
- BODENTEICH, A., MITCHELL, L. G., POLYMERPOULOS, M. H., *et al.* "Dinucleotide repeat in the human mitochondrial D-loop", **Human Molecular Genetics**, v. 1, n. 2, p. 140, 1992. DOI: 10.1093/hmg/1.2.140-a. .
- BROTHERTON, P., HAAK, W., TEMPLETON, J., *et al.* "Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans", **Nature Communications**, v. 4, 2013. DOI: 10.1038/ncomms2656. .

BRUMFIELD, R. T., BEERLI, P., NICKERSON, D. A., *et al.* "The utility of single nucleotide polymorphisms in inferences of population history", **Trends in Ecology and Evolution**, v. 18, n. 5, p. 249–256, 2003. DOI: 10.1016/S0169-5347(03)00018-1. .

BUSTAMANTE, C. D., DE LA VEGA, F. M., BURCHARD, E. G. **Genomics for the world**. **Nature**. [S.l: s.n.], 2011

BUTLER, J. M. **Fundamentals of Forensic DNA Typing**. [S.l: s.n.], 2010.

CAMPBELL, M. C., TISHKOFF, S. A. "African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping", **Annual Review of Genomics and Human Genetics**, v. 9, p. 403–433, 2008. DOI: 10.1146/annurev.genom.9.081307.164258. .

CANN, R. L., STONEKING, M., WILSON, A. C. "Mitochondrial DNA and human evolution", **Nature**, v. 325, p. 31–36, 1987. DOI: 10.1038/325031a0. .

CEREZO, M., ČERNÝ, V., CARRACEDO, Á., *et al.* "New insights into the lake chad basin population structure revealed by high-throughput genotyping of mitochondrial dna coding snps", **PLoS ONE**, v. 6, n. 4, 2011. DOI: 10.1371/journal.pone.0018682. .

CEREZO, M., GUSMÃO, L., ČERNÝ, V., *et al.* "Comprehensive Analysis of Pan-African Mitochondrial DNA Variation Provides New Insights into Continental Variation and Demography", **Journal of Genetics and Genomics**, v. 43, n. 3, p. 133–143, 2016. DOI: 10.1016/j.jgg.2015.09.005. .

ČERNÝ, V., SALAS, A., HÁJEK, M., *et al.* "A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: A history revealed by the mitochondrial DNA genome", **Annals of Human Genetics**, v. 71, n. 4, p. 433–452, 2007. DOI: 10.1111/j.1469-1809.2006.00339.x. .

CHEN, S. Y., FENG, Z., YI, X. "A general introduction to adjustment for multiple comparisons", **Journal of Thoracic Disease**, v. 9, n. 6, p. 1725–1729, 2017. DOI: 10.21037/jtd.2017.05.34. .

ČÍŽKOVÁ, M., MUNCLINGER, P., DIALLO, M. Y., *et al.* "Genetic structure of the western and eastern African Sahel/Savannah belt and the role of nomadic pastoralists as inferred from the variation of D-loop mitochondrial DNA sequences", **Human Biology**, v. 89, n. 4, p. 281–302, 2017. DOI: 10.13110/humanbiology.89.4.02. .

COBLE, M. D. "The identification of the Romanovs: Can we (finally) put the controversies to rest?", **Investigative Genetics**, v. 2, n. 20, 2011. DOI: 10.1186/2041-2223-2-20. .

COMAS, D. "The genetics of human migration: Tracing migrations through the genome", **Metode**, v. 5, p. 43–49, 2015. DOI: 10.7203/metode.81.3088. .

COMMONWEALTH, <https://thecommonwealth.org/our-member-countries/nigeria>, acesso em agosto de 2021.

DE FILIPPO, C., BOSTOEN, K., STONEKING, M., *et al.* **Bringing together linguistic and genetic evidence to test the Bantu expansion. Proceedings of the Royal Society B: Biological Sciences.** [S.l: s.n.], 2012

DESHPANDE, O., BATZOGLOU, S., FELDMAN, M. W., *et al.* "A serial founder effect model for human settlement out of Africa", **Proceedings of the Royal Society B: Biological Sciences**, v. 276, n. 1655, p. 291–300, 2009. DOI: 10.1098/rspb.2008.0750. .

DIMMENDAAL, G. J. **Historical linguistics and the comparative study of African languages.** [S.l: s.n.], 2011.

DUCROTOY, M. J., MAJEKODUNMI, A. O., SHAW, A. P. M., *et al.* "Patterns of passage into protected areas: Drivers and outcomes of Fulani immigration, settlement and integration into the Kachia Grazing Reserve, northwest Nigeria", **Pastoralism**, v. 8, n. 1, 2018. DOI: 10.1186/s13570-017-0105-1. .

ERN, V., FERNANDES, V., COSTA, M. D., *et al.* "Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup", **BMC Evolutionary Biology**, 2009. DOI: 10.1186/1471-2148-9-63. .

EXCOFFIER, L., LAVAL, G., SCHNEIDER, S. "Arlequin (version 3.0): An integrated software package for population genetics data analysis", **Evolutionary Bioinformatics**, v. 1, p. 47–50, 2005. DOI: 10.1177/117693430500100003. .

FALOLA, T., HEATON, M. M. **A history of Nigeria.** [S.l: s.n.], 2008.

FAN, S., KELLY, D. E., BELTRAME, M. H., *et al.* "African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations", **Genome Biology**, v. 20, n. 1, 2019. DOI: 10.1186/s13059-019-1679-2. .

FENDT, L., RÖCK, A., ZIMMERMANN, B., *et al.* "MtDNA diversity of Ghana: A forensic and phylogeographic view", **Forensic Science International: Genetics**, v. 6, n. 2, p. 244–249, 2012. DOI: 10.1016/j.fsigen.2011.05.011. .

FORSTER, P. "Ice Ages and the mitochondrial DNA chronology of human dispersals: A review", **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 359, n. 1442, p. 255–264, 2004. DOI: 10.1098/rstb.2003.1394. .

GABEL, C., NEWMAN, J. L. **The Peopling of Africa: A Geographic Interpretation.** [S.l: s.n.], 1997.

GABRIEL, M. N., HUFFINE, E. F., RYAN, J. H., *et al.* "Improved MtDNA Sequence Analysis of Forensic Remains Using a “Mini-Primer Set” Amplification Strategy", **Journal of Forensic Sciences**, v. 46, n. 2, p. 247–253, 2001. DOI: 10.1520/jfs14957j. .

GÖBEL, T. M. K., BODNER, M., ROBINO, C., *et al.* **Mitochondrial DNA variation in Sub-Saharan Africa: Forensic data from a mixed West African sample, Côte d’Ivoire (Ivory Coast), and Rwanda. Forensic Science International: Genetics.** [S.l: s.n.], 2020

GONDER, M. K., MORTENSEN, H. M., REED, F. A., *et al.* "Whole-mtDNA genome sequence analysis of ancient african lineages", **Molecular Biology and Evolution**, v. 24, n. 3, p. 757–768, 2007. DOI: 10.1093/molbev/msl209. .

GOOGLE, <https://sites.google.com/site/wwwbantumigration>, acesso em agosto de 2021.

GOOGLE, <https://i.pinimg.com/originals/c6/36/f1/c636f14b090a4a4cd535bfae4a97d15f.png>, acesso em agosto de 2021.

GOOGLE, <https://pt.wikipedia.org/wiki/Nig%C3%A9ria>, acesso em agosto de 2021.

GOOGLE, <https://www.coladaweb.com/wp-content/uploads/2014/12/20180328-mitocondria.jpg>, acesso em agosto de 2021.

GOOGLE, https://upload.wikimedia.org/wikipedia/commons/thumb/d/da/Migraciones_humanas_en_haplogrupos_mitocondriales.PNG/1200px-Migraciones_humanas_en_haplogrupos_mitocondriales.PNG, acesso em agosto de 2021.

GOOGLE, <http://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQxC13Xxb5fhcz1B5KyxU2Id8h869oakCRlg8OOBsLsNMffVUk>, acesso em agosto de 2021.

GOOGLE, <https://www.pngitem.com/search.html?k=africa+continent+>, acesso em agosto de 2021.

GOOGLE MAPS, <https://www.google.com/maps/place/%C3%81frica/@1.7400339,-16.2444486,3z/data=!3m1!4b1!4m5!3m4!1s0x10a06c0a948cf5d5:0x108270c99e90f0b3!8m2!3d-8.783195!4d34.508523>, acesso em agosto de 2021.

GREENBERG, J. H. "The Classification of African Languages", **American Anthropologist**, v. 50, n. 1, p. 24–30, 1948. DOI: 10.1525/aa.1948.50.1.02a00050. .

HAGELBERG, E., GOLDMAN, N., LIÓ, P., *et al.* "Evidence for mitochondrial DNA recombination in a human population of island Melanesia", **Proceedings of the Royal Society B: Biological Sciences**, v. 266, n. 1418, p. 485–492, 1999. DOI: 10.1098/rspb.1999.0663. .

HARTL, D.L.; CLARK, A. G. **Princípios de Genética de Populações**. [S.l: s.n.], 2010.

HAUSWIRTH, W. W., CLAYTON, D. A. "Length heterogeneity of a conserved displacement-loop sequence in human mitochondrial DNA", **Nucleic Acids Research**, v. 13, n. 22, p. 8093–8104, 1985. DOI: 10.1093/nar/13.22.8093. .

HENN, B. M., CAVALLI-SFORZA, L. L., FELDMAN, M. W. "The great human expansion", **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 44, p. 17758–17764, 2012. DOI: 10.1073/pnas.1212380109. .

HORAI, S., HAYASAKA, K., KONDO, R., *et al.* "Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs", **Proceedings of the National Academy of Sciences of the United States of America**, v. 92, n. 2, p. 532–536, 1995. DOI: 10.1073/pnas.92.2.532. .

JEFFREYS, A. J., WILSON, V., THEIN, S. L. "Individual-specific “fingerprints” of human DNA", **Nature**, v. 316, n. 6023, p. 76–79, 1985. DOI: 10.1038/316076a0. .

JUST, R. S., IRWIN, J. A., PARSON, W. "Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing", **Forensic Science International: Genetics**, v. 18, p. 131–139, 2015. DOI: 10.1016/j.fsigen.2015.05.003. Disponível em: <http://dx.doi.org/10.1016/j.fsigen.2015.05.003>.

LAMBERT, C. A., TISHKOFF, S. A. "Genetic structure in African populations: Implications for human demographic history", **Cold Spring Harbor Symposia on Quantitative Biology**, v. 74, p. 395–402, 2009. DOI: 10.1101/sqb.2009.74.053. .

LAO, O., VAN DUIJN, K., KERSBERGEN, P., *et al.* "Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry", **American Journal of Human Genetics**, v. 78, n. 4, p. 680–690, 2006. DOI: 10.1086/501531. .

LI, H., DURBIN, R. "Inference of human population history from individual whole-genome sequences", **Nature**, v. 475, n. 7357, p. 493–496, 2011. DOI: 10.1038/nature10231. .

LUCOTTE, G., GUERIN, P., HALLE, L., *et al.* "Y chromosome DNA polymorphisms in two African populations", **American Journal of Human Genetics**, v. 45, n. 1, p. 16–20, 1989. .

LUO, S., VALENCIA, C. A., ZHANG, J., *et al.* "Biparental inheritance of mitochondrial DNA in humans", **Proceedings of the National Academy of Sciences of the United States of America**, v. 115, n. 51, p. 13039–13044, 2018. DOI: 10.1073/pnas.1810946115. .

LUTZ-BONENGL, S., PARSON, W. **No further evidence for paternal leakage of mitochondrial DNA in humans yet. Proceedings of the National Academy of Sciences of the United States of America**. [S.l: s.n.]. , 2019

MELTON, T. "Mitochondrial DNA Heteroplasmy.", **Forensic science review**, v. 982, p. 577–594, 2004. .

MELTON, T., NELSON, K. "Forensic mitochondrial DNA analysis: Two years of

commercial casework experience in the United States", **Croatian Medical Journal**, v. 42, n. 3, p. 298–303, 2001. .

MENSAH, G. A. "Descriptive epidemiology of cardiovascular risk factors and diabetes in sub-saharan Africa", **Progress in Cardiovascular Diseases**, v. 56, n. 3, p. 240–250, 2013. DOI: 10.1016/j.pcad.2013.10.014. .

METZ, H. C. **Nigeria: A Country Study**. [S.l: s.n.], 1991. Disponível em: <https://www.loc.gov/item/92009026/%0Ahttp://countrystudies.us/Nigeria>.

MULLIS, K., FALOONA, F., SCHARF, S., *et al.* "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.", **Biotechnology (Reading, Mass.)**, v. 51, p. 263–273, 1986. .

OKAZAKI, A., YAMAZAKI, S., INOUE, I., *et al.* **Population genetics: past, present, and future. Human Genetics**. [S.l: s.n.], 2021

OLANIYAN, A. O. "From accommodation to discrimination and exclusion : The changing pattern of inter-group relations in Nigeria", **African Journal of Political Science and International Relation**, v. 3, n. 11, p. 521–525, 2009. .

OLIVIERI, A., ACHILLI, A., PALA, M., *et al.* "The mtDNA legacy of the levantine early Upper Palaeolithic in Africa", **Science**, v. 314, n. 5806, p. 1767–1770, 2006. DOI: 10.1126/science.1135566. .

OTTONI, C., PRIMATIVO, G., KASHANI, B. H., *et al.* "Mitochondrial haplogroup H1 in north Africa: An early Holocene arrival from Iberia", **PLoS ONE**, v. 5, n. 10, 2010. DOI: 10.1371/journal.pone.0013378. .

PARSON, W., GUSMÃO, L., HARES, D. R., *et al.* "DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing", **Forensic Science International: Genetics**, v. 13, p. 134–142, 2014. DOI: 10.1016/j.fsigen.2014.07.010. .

PARSON, Walther, BANDELT, H. J. "Extended guidelines for mtDNA typing of population data in forensic science", **Forensic Science International: Genetics**, v. 1, n. 1, p. 13–19, 2007. DOI: 10.1016/j.fsigen.2006.11.003. .

PARSONS, T. J., COBLE, M. D. "Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome", **Croatian Medical Journal**, 2001. .

PARSONS, Thomas J. "Mitochondrial DNA Genome Sequencing and SNP Assay Development for Increased Power of Discrimination", **U.S. Department of Justice**, v. s/v, n. s/n, p. 1–119, 2006. .

PEREIRA, L., MACAULAY, V., TORRONI, A., *et al.* "Prehistoric and historic traces in the mtDNA of Mozambique: Insights into the Bantu expansions and the slave trade", **Annals of Human Genetics**, v. 65, p. 439–58, 2001. DOI: 10.1046/j.1469-1809.2001.6550439.x. .

PHYLOTREE, <https://www.phylotree.org/tree/index.htm>, acesso em agosto de 2021.

PHYLOTREE, <https://www.phylotree.org/tree/L0.htm>, acesso em setembro de 2021.

PHYLOTREE, <https://www.phylotree.org/tree/L1.htm>, acesso em setembro de 2021.

PHYLOTREE, <https://www.phylotree.org/tree/L2.htm>, acesso em setembro de 2021.

PHYLOTREE, <https://www.phylotree.org/tree/L3.htm>, acesso em setembro de 2021.

PHYLOTREE, <https://www.phylotree.org/tree/L4.htm>, acesso em setembro de 2021.

PHYLOTREE, <https://www.phylotree.org/tree/L5.htm>, acesso em setembro de 2021.

QUORA, <https://qph.fs.quoracdn.net/main-qimg-9c73ff310f534255af4595398954de7c>, acesso em agosto de 2021.

RAMACHANDRAN, S., DESHPANDE, O., ROSEMAN, C. C., *et al.* "Support from the relationship of genetic and geographic in human populations for a serial founder effect originating in Africa", **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 44, p. 15942–15947, 2005. DOI: 10.1073/pnas.0507611102. .

REYES-CENTENO, H. "Out of Africa and into Asia: Fossil and genetic evidence on modern human origins and dispersals", **Quaternary International**, v. 416, p. 249–262, 2016. DOI: 10.1016/j.quaint.2015.11.063. .

RITO, T., RICHARDS, M. B., FERNANDES, V., *et al.* "The first modern human dispersals across Africa", **PLoS ONE**, v. 8, n. 11, 2013. DOI: 10.1371/journal.pone.0080031. .

ROSA, A., BREHEM, A. **African human mtDNA phylogeography at-a-glance**. **Journal of Anthropological Sciences**, v. 89, p. 25-58, 2011.

SALAS, A., RICHARDS, M., DE LA FE, T., *et al.* "The making of the African mtDNA landscape", **American Journal of Human Genetics**, p. 1082–1111, 2002. DOI: 10.1086/344348.

SALAS, A., RICHARDS, M., LAREU, M. V., *et al.* "The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade", **American Journal of Human Genetics**, v. 74, n. 3, p. 454–465, 2004. DOI: 10.1086/382194. .

SANKARARAMAN, S., MALLICK, S., PATTERSON, N., *et al.* "The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans", **Current Biology**, v. 26, n.

9, p. 1241–1247, 2016. DOI: 10.1016/j.cub.2016.03.037. .

SANKARARAMAN, S., PATTERSON, N., LI, H., *et al.* "The Date of Interbreeding between Neandertals and Modern Humans", **PLoS Genetics**, v. 8, n. 10, 2012. DOI: 10.1371/journal.pgen.1002947. .

SATO, M., SATO, K. "Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA", **Biochimica et Biophysica Acta - Molecular Cell Research**, v. 1833, n. 8, p. 1979–1984, 2013. DOI: 10.1016/j.bbamcr.2013.03.010. .

SATTA, Y., TAKAHATA, N. **Out of Africa with regional interbreeding? Modern human origins.** **BioEssays**. [S.l: s.n.]. , 2002

SCHAFFNER, S. F. "The X chromosome in population genetics", **Nature Reviews Genetics**, v. 5, n. 1, p. 43–51, 2004. DOI: 10.1038/nrg1247. .

SCHWARTZ, M., VISSING, J. "Paternal Inheritance of Mitochondrial DNA", **New England Journal of Medicine**, p. 576–580, 2002. DOI: 10.1056/nejmoa020350. .

SEIELSTAD, M. T., MINCH, E., CAVALLI-SFORZA, L. L. "Genetic evidence for a higher female migration rate in humans", **Nature Genetics**, v. 20, p. 278–280, 1998. DOI: 10.1038/3088. .

SHEA, J. J. " Lone Survivors: How We Came to Be the Only Humans on Earth . By Chris Stringer. New York: Henry Holt and Company (Times Books). ISBN: 978-0-8050-8891-5. 2012. ", **The Quarterly Review of Biology**, 2013. DOI: 10.1086/669325. .

SILVA, M., ALSHAMALI, F., SILVA, P., *et al.* "60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2", **Scientific Reports**, 2015. DOI: 10.1038/srep12526. .

SIMÃO, F., FERREIRA, A. P., DE CARVALHO, E. F., *et al.* "Defining mtDNA origins and population stratification in Rio de Janeiro", **Forensic Science International: Genetics**, v. 34, p. 97–104, 2018. DOI: 10.1016/j.fsigen.2018.02.003. .

SIMPSON, A., OYÈTÁDÉ, B. A., "Nigeria: Ethno-linguistic competition in the Giant of Africa". **Language and national identity in Africa**, [S.l: s.n.], 2008. p. 172–198.

SOARES, P., ALSHAMALI, F., PEREIRA, J. B., *et al.* **The expansion of mtDNA haplogroup L3 within and out of Africa.** **Molecular Biology and Evolution**. [S.l: s.n.]. , 2012

SUTOVSKY, P., SONG, W. H. "Post-fertilisation sperm mitophagy: The tale of Mitochondrial Eve and Steve", **Reproduction, Fertility and Development**, p. 56–63, 2018. DOI: 10.1071/RD17364. .

SZIBOR, R., MICHAEL, M., SPITSYN, V. A., *et al.* "Mitochondrial D-loop 3' (CA)(n) repeat polymorphism: Optimization of analysis and population data", **Electrophoresis**, v. 18, n. 15, p. 2857–2860, 1997. DOI: 10.1002/elps.1150181523. .

TATTERSALL, I. **Human origins: Out of Africa.** **Proceedings of the National Academy of Sciences of the United States of America**. [S.l: s.n.]. , 2009

TAYLOR, R. W., TURNBULL, D. M. **Mitochondrial DNA mutations in human disease.** *Nature Reviews Genetics*. [S.l.: s.n.], 2005

TORRONI, A., ACHILLI, A., MACAULAY, V., *et al.* "Harvesting the fruit of the human mtDNA tree", *Trends in Genetics*, v. 22, n. 6, p. 339–345, 2006. DOI: 10.1016/j.tig.2006.04.001. .

TULLY, L. A., PARSONS, T. J., STEIGHNER, R. J., *et al.* "A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region", *American Journal of Human Genetics*, v. 67, n. 2, p. 432–443, 2000. DOI: 10.1086/302996. .

UPHOLT, W. B., DAWID, I. B. "Mapping of mitochondrial DNA of individual sheep and goats: Rapid evolution in the D loop region", *Cell*, v. 11, n. 3, p. 571–583, 1977. DOI: 10.1016/0092-8674(77)90075-7. .

VAN OVEN, M., KAYSER, M. "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation.", *Human mutation*, v. 30, n. 2, p. 386–394, 2009. DOI: 10.1002/humu.20921. .

WALL, J. D., YANG, M. A., JAY, F., *et al.* "Higher levels of Neanderthal ancestry in east Asians than in Europeans", *Genetics*, v. 194, p. 199–209, 2013. DOI: 10.1534/genetics.112.148213. .

WALSH, P. S., METZGER, D. A., HIGUCHI, R. "Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material", *BioTechniques*, v. 10, p. 506–513, 1991. DOI: 10.2144/000114018. .

WATSON, E., BAUER, K., AMAN, R., *et al.* "mtDNA sequence diversity in Africa", *American Journal of Human Genetics*, v. 59, p. 437–444, 1996. .

WATSON, E., FORSTER, P., RICHARDS, M., *et al.* "Mitochondrial footprints of human expansions in africa", *American Journal of Human Genetics*, v. 61, n. 3, p. 691–704, 1997. DOI: 10.1086/515503. .

WOLPOFF, M. H., HAWKS, J., CASPARI, R. "Multiregional, not multiple origins", *American Journal of Physical Anthropology*, v. 112, n. 1, p. 129–136, 2000. DOI: 10.1002/(SICI)1096-8644(200005)112:1<129::AID-AJPA11>3.0.CO;2-K. .

YASUKAWA, T., KANG, D. "An overview of mammalian mitochondrial DNA replication mechanisms", *Journal of Biochemistry*, v. 164, n. 3, p. 183–193, 2018. DOI: 10.1093/jb/mvy058. .

ZIMMERMANN, B., RÖCK, A. W., DÜR, A., *et al.* "Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software", *Croatian Medical Journal*, v. 55, n. 2, p. 115–120, 2014. DOI: 10.3325/cmj.2014.55.115. .

ANEXO - Lista dos haplótipos encontrados na amostra dos 285 indivíduos totais não relacionados da Nigéria

A001	L1b1a10	16126C	16136C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	151T	152C	182T	185T	195C	247A	263G	309.1C	315.1C	357G	523del	524del							
A004	L1b1a18	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	185T	189G	195C	247A	263G	315.1C	357G	523del	524del									
A005	L2a1	16223T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	263G	309.1C	315.1C																
A007	L3e1a	16093C	16185T	16223T	16527T	73G	150T	152C	189G	200G	263G	309.1C	315.1C																		
A008	L1c3a1b	16093T	16129A	16183C	16189C	16193.1C	16215G	16223T	16278T	16294T	16311C	16355T	16360T	16519C	73G	151T	152C	182T	189G	186A	189C	204C	247A	257G	263G	309.1C	315.1C	316A	523del	524del	
A010	L3e3	16093C	16223T	16262T	16265T	16519C	73G	150T	195C	263G	315.1C	316A	523del	524del																	
A011	L1b1a+189	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	189G	195C	247A	263G	309.1C	315.1C	357G	523del	524del							
A012	M1b2	16129A	16182C	16183C	16189C	16223T	16249C	16311C	16399G	16519C	73G	195C	263G	315.1C	489C																
A018	L3f1b1a	16129A	16209C	16223T	16291T	16292T	16295T	16311C	16519C	73G	189G	200G	207A	263G	309.1C	315.1C															
A020	L2a1	16189C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	153G	195C	263G	309.1C	315.1C															
A022a	L2a1	16172C	16189C	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C																
A022b	L3e2a1	16223T	16320T	16519C	73G	150T	195C	198T	263G	309.1C	315.1C																				
A023	L3d4	16124C	16223T	73G	152C	189G	195C	263G	315.1C	523del	524del																				
A024	L3b2	16086C	16124C	16189C	16223T	16278T	16311C	16362C	16527T	73G	263G	315.1C	523del	524del																	
A027	L2a1	16223T	16271C	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C																	
A029	L2b1a	16114A	16129A	16213A	16223T	16278T	16355T	16362C	16390A	73G	150T	152C	182T	195C	198T	204C	263G	315.1C	418T	523del	524del										
A030	L1b1a+189	16099T	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	152C	182T	185T	189G	195C	247A	263G	315.1C	357G	521del	522del	523del	524del						
A039	L3b2	16124C	16223T	16278T	16362C	16527T	73G	263G	315.1C	523del	524del																				
A043	L3b	16093C	16124C	16223T	16278T	16362C	16519C	73G	263G	308.1C	315.1C	523del	524del																		
A045	L3e2	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																						
A048	L3e3	16223T	16265T	16519C	73G	150T	195C	263G	315.1C	523del	524del																				
A061	L2a1	16223T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	189G	195C	263G	309.1C	315.1C	523del	524del													
A062	L2a1	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C	523del	524del															
A063	L3e2b	16172C	16189C	16223T	16234T	16320T	16519C	73G	150T	195C	263G	315.1C																			
A067	L3e1	16185T	16189C	16223T	16264T	16327T	73G	150T	152C	189G	200G	207A	263G	315.1C																	
A068	L3e2	16192T	16223T	16320T	16519C	73G	150T	195C	263G	309.1C	315.1C																				
A074	L3e2b	16172C	16189C	16223T	16284G	16320T	16519C	73G	150T	195C	263G	315.1C																			
A076	L3d1d	16124C	16223T	16256T	16368C	16519C	73G	152C	263G	315.1C	523del	524del																			
A078	L3d	16124C	16223T	16278T	16362C	16519C	73G	263G	315.1C	523del	524del																				
A084	L3e2	16223T	16320T	16519C	73G	150T	195C	263G	309.1C	315.1C																					
A085	L3e2b	16172C	16189C	16223T	16257T	16320T	16519C	73G	150T	195C	262T	263G	282C	315.1C																	
A086	L3b2a	16124C	16150T	16183C	16189C	16193.1C	16223T	16278T	16362C	16527T	73G	263G	315.1C	523del	524del																
A090	L1c3b	16129A	16169G	16187T	16189C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	385G	523del	524del						
A092	L3e1	16185T	16189C	16193del	16223T	16264T	16327T	73G	150T	152C	189G	200G	207A	263G	309.1C	315.1C															
A093	L2a1	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	198T	263G	309.1C	315.1C																
A094a	L1c1	16129A	16148T	16175G	16187T	16189C	16192T	16223T	16255A	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A	523del	524del	
A094b	L3d1c	16124C	16166G	16223T	73G	152C	263G	309.1C	315.1C	523del	524del																				
A094c	L3d5	16124C	16223T	16362C	73G	152C	263G	315.1C	523del	524del																					
A096	L2a1	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	523del	524del																
A097	L1b1a+189	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	189G	195C	247A	263G	315.1C	357G	523del	524del								
A101	L0a1a+200	16129A	16148T	16168T	16172C	16187T	16188A	16189C	16223T	16230G	16311C	16320T	93G	185A	189G	200G	236C	247A	263G	309.1C	309.2C	315.1C	523del	524del							
A102	L2a1	16189C	16192T	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C															
A103	L2e	16111A	16145A	16184T	16223T	16239T	16278T	16355T	16390A	16399G	16400T	16519C	73G	146C	150T	152C	182T	195Y	263G	315.1C	479G										
A104	L1b1a18	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	152C	185T	189G	195C	263G	309.1C	315.1C	357G	523del	524del										
A105	L2a1	16111A	16145A	16184T	16223T	16239T	16274A	16278T	16292T	16355T	16390A	16399G	16400T	16519C	73G	146C	150T	182T	263G	315.1C	479G	568T									
A111	L3f1b1a	16129A	16209C	16223T	16292T	16295T	16311C	16519C	73G	189G	200G	263G	309.1C	315.1C																	
A112	L1b2a	16093C	16111T	16126C	16187T	16189C	16223T	16239T	16270T	16278T	16293G	16311C	16519C	73G	146C	182T	185T	189G	247A	263G	315.1C	357G	523del	524del							
A113	L2e	16111A	16145A	16184T	16223T	16239T	16278T	16292T	16355T	16362C	16390A	16399G	16400T	16519C	73G	146C	150T	152C	182T	263G	315.1C										
A114	L2c2	16223T	16264T	16278T	16311C	16390A	73G	93G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523del	524del												
A115	L3f1b+16292+150	16209C	16223T	16292T	16311C	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																		
A117	L3f1b+16292	16209C	16223T	16292T	16295T	16311C	16519C	73G	150T	152C	189G	200G	263G	309.1C	315.1C	316A															
A118	L2c2	16217C	16223T	16264T	16278T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	204C	263G	309.1C	315.1C	325T	523del	524del										
A120	L1b1a+189	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T																

APÊNDICE - The maternal inheritance of the three major ethnic groups of Nigeria

The maternal inheritance of the three major ethnic groups of Nigeria

Masinda Nguidi^{1*}; Laura Catelli³; Carlos Vullo³; Beatriz Martínez²; Micaela Longaray³; Victoria O. Okolie⁴; Samuel O. Keshinro⁵; Leonor Gusmão¹; Filipa Simão¹

¹DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil. ²Molecular Genetics Lab, Institute for Immunological Research, University of Cartagena, Cartagena, Colombia.

³DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (EAAF), Córdoba, Argentina.

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

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

*masi.nguidi@gmail.com

Abstract

Africa is the homeland of the most ancient human populations and, for that reason, harbors the most genetically diverse populations. Nonetheless, African groups are still scarcely represented in worldwide forensic and populational genetic databases. Inside Africa, Nigeria, located in the Gulf of Guinea region, is one of the most diverse countries in the world. The present study aimed to help fill the existing gap in the genetic description of Nigerian populations. For this purpose, 285 samples were collected from unrelated individuals belonging to the three major Nigerian groups: Hausa (n=73), Yoruba (n=112) and Igbo (n=100). The entire mtDNA control region (16024-639) was analyzed and data was submitted to the EMPOP database. Most of the maternal lineages found in Nigeria belong to the macro haplogroup L (98.6%), characteristic of Sub-Saharan African populations. The F_{ST} values calculated among the three studied groups did not show statistically significant differences between them. Also, the Nigerian samples showed low F_{ST} values with other Nigerian samples and with neighboring populations from Central West Africa. The high homogeneity among geographically close populations, regardless of their ethnolinguistic affiliation, can be explained by a high gene flow mediated by females. These results reinforce the difficulty of predicting geographic or linguistic affiliations based on mtDNA data. The data presented here can also be useful in future investigations to elucidate the admixture history of South American populations, where Nigeria is known to have played an important role.

Introduction

Africa is the homeland of the most ancient human populations and, therefore, harbors the most genetically diverse populations. Nonetheless, this continent is still underrepresented in worldwide forensic and populational genetic databases (Göbel *et al.*, 2020). The genetic characterization of African population groups is crucial to understand the history of the continent. Furthermore, detailed information on the structure of these groups is also important to clarify the history of the Americas, since West-Central Africa played an importance role in the transatlantic slave trade (ex: Calabar region, a big commercial center of enslaved people) and, subsequently, in the genetic composition of the colonies, especially from the New World and Caribs (Rotimi *et al.*, 2016). A thorough description of the genetic composition of African populations is also important in the forensic field. To give statistical weight to a genotypic match, population databases need to be established and widely representative of different populations. This is particularly relevant for criminal cases in conflicting areas, very common in pluralistic countries such as Nigeria.

Nigeria is a country located in western Africa, in the Gulf of Guinea region. It is thought to be the Bantu natives' homeland (around 4,000 years ago) and, therefore, it has been the focus of several linguistic studies (Newman, 1995; Pereira *et al.*, 2001). Bantu people were responsible for the biggest migration movement inside the continent, which led to the spread of agriculture and metallurgy across sub-Saharan Africa. Ultimately, this migration resulted in genetic admixture with indigenous hunter-gather groups inhabiting southern and central Africa - the Khoisan/Khoesan and Pygmies (Newman, 1995; Pereira *et al.*, 2001; Berniell-Lee *et al.*, 2009; de Filippo *et al.*, 2012). Later, Nigeria region was also the place where the Nok people emerged, the most ancient civilization of Sub-Saharan Africa (400 BCE until 200 CE) (Falola and Heaton, 2008; Metz, 1991).

Nigeria is currently the most populated African country, comprising at least 250 distinct ethnolinguistic groups, with subgroups and dialects of political and social importance. The predominant groups are the Hausa-Fulani (29%), the Yoruba (20%), and the Igbo (17%) (Simpson and Oyètádé, 2007). The existence of these groups and their kingdoms is dated before 1500. Due to their divergent beliefs and varied cultures, different dynamics and political organizations were implemented, which led to the geographic separation of the groups. These divisions would result in the current States' limits. The Hausa people live mainly in the north, in the savannas of the Hausaland kingdom (Figure 1). Their location allowed contact with people of Middle East and North Africa, who introduced Islam to their culture as a result of

exchanges with the Trans-Saharan trade (Metz, 1991; Falola and Heaton, 2008). Yoruba and Igbo people emerged in South Nigeria (Figure 1), in a region called Yorubaland, and had great contact with Christianity with the arrival of the British on the country coast and posterior exploitation of the natives and the territory.

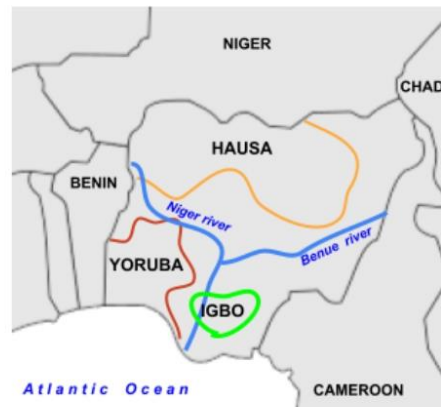


Figure 1 - Location of Nigeria in the West-Central region, and neighboring countries. The current geographic distribution of its major ethnolinguistic groups is highlighted in different colors: Yoruba in red, Igbo in green and Hausa in orange. The Niger and Benue rivers are represented in blue.

Nowadays, Nigeria population self-identifies as Christians, non-Muslims or as adherents of other indigenous religions. In the Middle belt – the area between north and south – the people have mixed faith. Given the wide diversities of ethnic, linguistic, and religious beliefs, identity is an important factor for the Nigerian population. This is reflected in the history of Nigeria, where several conflicts emerged due to religious and ethnic divergences. This was the case of the Nigerian Civil War (or Biafran War, 1967-1970), which started shortly after the Independence (1960), as a result of inter-ethnic and inter-religious turmoil episodes, essentially involving Igbo and Hausa groups (Metz, 1991; Falola and Heaton, 2008).

Because mtDNA has a non-Mendelian inheritance and is transmitted through generations as haplotypes, it is a useful tool in population studies, allowing the identification of maternal lineages. Moreover, it can provide additional information such as population demographic history and migration patterns around the globe and allow macro and microevolution studies (Parsons and Coble, 2001; Zhao *et al.*, 2003; Underhill, 2007).

Attempts have been made to disclose genetic variation in different population groups in Africa, and to correlate geography and ethnicity with patterns of genetic diversity (Wood *et al.*, 2005; Tishkoff *et al.*, 2009; Choudhury *et al.*, 2020). However, more studies on the African continent

are still needed to uplift the already known diversity of its populations and emphasize its particularities. Although Nigeria is one of the most diverse countries in the world, few studies have been so far dedicated to disclose the genetic diversity of its current populations. The existing data on maternal ancestry of Nigeria includes a study of mtDNA in samples with no ethnolinguistic affiliation or geographic specification (Göbel *et al.*, 2020). Furthermore, although information is available for some ethnolinguistic groups, most studies only included partial information of hypervariable regions (Watson *et al.*, 1996; Wood *et al.*, 2005; Cerny *et al.*, 2007; Veeramah *et al.*, 2010), sometimes complemented with coding mtSNPs (Cerezo *et al.*, 2011). Mitogenomes are available only for Yoruba and Esan populations (1000 Genome Project), not comprising all major groups of Nigeria.

Aiming a deeper and more comprehensive study of the diversity present in Nigeria, mtDNA control region haplotypes were described for Yoruba, Hausa and Igbo ethnic groups. Taking into consideration the high diversity previously reported in several aspects of Nigerian populations, we also aim to evaluate possible substructure or genetic homogeneity among these groups. Furthermore, a comparative analysis with the surrounding populations allowed to investigate whether geographic or linguistic barriers exist. This study contributed to increase the existing forensic database of mtDNA that is still not representative of most African populations. This data can also be used as reference in future population studies aiming a more comprehensive understanding of the American population gene pool, since both historical and genetic reports point to West Africa as one of the greatest contributors, during the transatlantic slave trade.

Material and Methods

Population sample: A total of 285 blood stains were collected in FTA cards from unrelated males of the three major Nigerian groups: Hausa (73 samples), Yoruba (112 samples) and Igbo (100 samples). The ethnolinguistic affiliation of the selected individuals was traced back to three generations, with parents and grandparents all belonging to the same group. All samples were obtained after informed consent and the study was approved by the Health Research Ethics Committee from the Lagos University Teaching Hospital (NHREC: 19/12/2008a). DNA was extracted using Phenol-chloroform (Corney *et al.*, 1994; Vandenberg *et al.*, 1997) or Chelex methods (Walsh *et al.*, 1991).

mtDNA Typing: The entire control region (16024-639) was amplified with the Qiagen Multiplex PCR Kit using the primers L15900/H639, as described by Simão *et al.* (2018) (details in Supplementary Table 1). When no amplification was obtained due to sample degradation, a

new amplification was performed for two separate fragments, using primer pairs 15967F/20R and 16497/639R (Supplementary Table 1) (Gabriel *et al.*, 2001). Amplicons were purified using ExoSAP enzymes (Applied Biosystems). Sequences were obtained using the BigDye v3.1 cycle Sequencing kit (Applied Biosystems), following the manufacturer's guidelines, and using the primers described in Supplementary Table 1. To eliminate sequencing remains, the products were purified through Sephadex columns (Illustra Sephadex DNA Grade; GE Healthcare). Capillary Electrophoresis on ABI 3500 (Applied Biosystems) was used in fragment separation and detection.

Haplogroup assignment: Haplotypes were classified with SeqScape v2.7 software (Thermo Fisher Scientific) after comparison to the Revised Cambridge Reference Sequence (rCRS) (Andrews *et al.*, 1999). Quality control was performed using the "EMPcheck" and "Network" tools available on the EMPOP database (Zimmermann *et al.*, 2014). Haplogroups were assigned on EMPOP, according to the Phylotree build 17, February 2016 (van Oven *et al.*, 2009). Data was submitted to EMPOP for final quality control (Parson and Dür, 2007) and will be available for research.

Statistical Analysis

Haplogroup frequencies were calculated by direct counting. Sequences conversion and alignment was performed on Haplosearch (Fregel and Delgado, 2011).

The statistical parameters of haplotype diversity (HD), F_{ST} genetic distance, mean number of pairwise differences and non-differentiation probabilities were calculated using Arlequin ver. 3.5.1.2 software (Excoffier *et al.*, 2005). Analyses were performed after discarding indels in positions inside 16030-16193, 16194-309, 310-315, 316-522, 525-573 and 574-576.

Pairwise genetic distances were visualized in three-dimensional space using the multidimensional scaling (MDS) analysis included in STATISTICA data analysis software system, ver. 8.0 (www.statsoft.com).

Phylogenetic networks were designed with median-joining methods (Bandelt *et al.*, 1999) on Network v10.1.0.0 software (<https://www.fluxus-engineering.com>).

Results and Discussion

mtDNA diversity among groups

A list of the haplotypes, and corresponding haplogroups, found in the 285 samples from Nigeria is presented in Supplementary Table 2.

The three ethnolinguistic groups displayed high intrapopulation diversity. The values of haplotype diversity (Table 1) were higher than those previously reported in other West Central African populations by Cerezo *et al.* (0.984), Čížková *et al.* (between 0.7252 and 0.9985) and Fendt *et al.* (0.98). Looking at the distribution of the number of pairwise differences between haplotypes in the three studied groups (Supplementary Figures 1-3), it is possible to see that, in more than 85% of the pairwise comparisons, the haplotypes are different from each other in more than 10 polymorphic positions. The haplotype distribution in each population is depicted in the networks represented in Supplementary Figures 1-3. A high diversity is detected not only among but also within haplogroups.

A comparison of the three studied groups by mean of pairwise F_{ST} genetic distances showed no statistically significant differences between them. The F_{ST} values were all negative and non-differentiation probabilities were above the significance level (Igbo *vs.* Yoruba, $F_{ST} = -0.00397$, $p = 0.92209 \pm 0.0024$; Yoruba *vs.* Hausa, $F_{ST} = -0.00409$, $p = 0.86546 \pm 0.0037$; Igbo *vs.* Hausa, $F_{ST} = -0.00089$, $p = 0.51074 \pm 0.0054$).

Despite Hausa people has a different linguistic background from the Igbo and Yoruba groups, and an history of conflicts and inter-group exclusion during the Civil War, an homogeneity was detected in the maternal lineages of the three populations. These results are in accordance with demographic studies referring to the inter-ethnic marriage between groups as a normalized practice in Nigeria (Azeez, 2009), specially in Kano (Yoruba-Hausa), Ibadan and Abeokuta (Hausa-Yoruba) cities, although the majority of Nigerian marriages were within groups in recent past (Metz, 1991). These regions have evidence of familiarity with different groups with adoption of mechanisms of tension management to enable the coexistence between different tribes.

Table 1 - West African populations used for comparison and corresponding haplotype diversity indices.

Note: Samples from 1000 Genome Project are identified with the code 1000 GP.

Country	Ethnic group	Reference	n	unique haplotypes	haplotype diversity (HD)
Ghana	-	Göbel <i>et al.</i> , 2020	22	22	1.0000 +/- 0.0137
Côte d'Ivoire	-	Göbel <i>et al.</i> , 2020	27	27	1.0000 +/- 0.0101
Nigeria	-	Göbel <i>et al.</i> , 2020	28	28	1.0000 +/- 0.0095
Cameroon	-	Göbel <i>et al.</i> , 2020	11	11	1.0000 +/- 0.0388
Togo	-	Göbel <i>et al.</i> , 2020	34	34	1.0000 +/- 0.0071
Nigeria	Yoruba	This study	112	102	0.9982 +/- 0.0015
Nigeria	Hausa	This study	73	68	0.9977 +/- 0.0029
Sierra Leone	Mende	1000 GP	68	62	0.9974 +/- 0.0030
Nigeria	Yoruba	1000 GP	108	97	0.9971 +/- 0.0021
Burkina Faso	Mossi	Čížková <i>et al.</i> , 2018	36	34	0.9968 +/- 0.0075
Nigeria	Esan	1000 GP	99	86	0.9965 +/- 0.0022
Senegal	Halpularen	Čížková <i>et al.</i> , 2018	47	43	0.9963 +/- 0.0053
Nigeria	Igbo	This study	100	84	0.9960 +/- 0.0022
Gambia	-	1000 GP	113	93	0.9959 +/- 0.0019
Chad	Kanembu	Cerezo <i>et al.</i> , 2016	50	43	0.9943 +/- 0.0052
Burkina Faso	Goumantche	Čížková <i>et al.</i> , 2018	36	33	0.9937 +/- 0.0090
Burkina Faso	Gurunsi	Čížková <i>et al.</i> , 2018	33	29	0.9924 +/- 0.0094
Senegal	Serrere	Čížková <i>et al.</i> , 2018	48	40	0.9920 +/- 0.0061
Ghana	Akan	Fendt <i>et al.</i> , 2012	192	126	0.9915 +/- 0.0020
Chad	Arabs Abala	Čížková <i>et al.</i> , 2018	51	42	0.9914 +/- 0.0060
Guinea	Fulani Fouta Djallon	Čížková <i>et al.</i> , 2018	51	42	0.9914 +/- 0.0060
Senegal	Bedik	Čížková <i>et al.</i> , 2018	41	33	0.9878 +/- 0.0086
Chad	Maba	Čížková <i>et al.</i> , 2018	42	31	0.9849 +/- 0.0084
Chad	Dangaleat	Čížková <i>et al.</i> , 2018	50	35	0.9820 +/- 0.0079
Cameroon	Tcheboua Fulani	Cerezo <i>et al.</i> , 2016	40	30	0.9795 +/- 0.0117
Chad	Daza	Čížková <i>et al.</i> , 2018	40	27	0.9769 +/- 0.0109
Niger	Tuareg	Cerezo <i>et al.</i> , 2016	27	19	0.9601 +/- 0.0250
Burkina Faso	Fulani Ziniare	Čížková <i>et al.</i> , 2018	55	27	0.9455 +/- 0.0155
Mali	Tuareg	Cerezo <i>et al.</i> , 2016	14	11	0.9341 +/- 0.0611
Burkina Faso	Tuareg	Cerezo <i>et al.</i> , 2016	20	8	0.8421 +/- 0.0595
Senegal	Fulani Ferlo	Čížková <i>et al.</i> , 2018	56	24	0.8403 +/- 0.0476

Nigeria in Central West Africa

To better understand the dynamics of the studied populations with the surrounding countries (Supplementary Figure 4), genetic distances (F_{ST}) were calculated among Nigeria and other populations from Central West Africa (Supplementary Table 3) and depicted in a MDS plot (Figure 2).

The Nigerian samples showed lower F_{ST} values with other Nigerian population samples previously reported (Esan and Yoruba from 1000 Genome Project and Nigeria from Göbel *et al.*, 2020), and with the geographically close populations from Ghana (Göbel and Fendt), Côte d'Ivoire, Togo and Chad (Göbel and Cerezo). In contrast, the more geographically distant populations from the West region - Senegal, Guinea, Ghana, Cameroon - were more distant and dispersed from the three ethnic groups of this study (Figure 2).

The high homogeneity among geographically close populations, regardless of their ethnolinguistic affiliation, can be explained by the high female mediated gene flow. For instance, in Burkina Faso populations, a substructure in the paternal background and a near homogeneity in the maternal background was described (Barbieri *et al.*, 2012; Seielstad *et al.*, 1998). The interethnic marriages are known to have occurred in ancient history of the region (before 1500), since commercial trades between neighboring communities were common (Falola and Heaton, 2008). In fact, looking at the MDS plot, the central cluster includes genetically close populations belonging to different ethnolinguistic backgrounds. Also, in the 16th century, the British establishment and exploitation of Central West Africa led to the division of territory known as Nigeria. The political division of nations/countries in Africa during the Berlin Conference (in the 19th Century) did not respect the already established political boundaries imposed by the natives and their ethnolinguistic affiliations, which led to the fragmentation of groups (Alesina *et al.*, 2011). Thus, it is not always possible to predict the maternal genetic composition of populations based on their ethnolinguistic affiliation or geographic distribution in Central West Africa. The lack of substructure in the maternal background in Nigeria and surrounding countries, is opposite to what can be found for the paternal heritage. The results obtained by analyzing the male component of Nigeria (data not shown, personal communication) highlights differences between Hausa, Yoruba and Igbo.

Although no ethnicity was detailed for the Togo and Ghana populations, they most probably have a Niger-Congo background. The proximity of these populations with Hausa, can be explained by the dissemination of the Hausa dialect as a mother tongue in these regions of West Africa (Simons and Oyètádé, 2008). Between the 14th and 16th centuries, the ethnic groups, despite being somewhat delimited, were economically and socially interdependent. The

communication between them was so close that social and cultural exchanges were maintained through business or even intermarriages. One of these itinerant groups was the Hausa. They were very well known in the Trans-Saharan trade for their mercantile profile/acumen that the dialect eventually spread beyond their states for trade purposes. The Hausa also settled outside Hausaland to facilitate the business and that movement resulted in shared kinship and language with other communities (Falola and Heaton, 2008).

The Nigerian sample of Yoruba from the 1000 Genome Project also presented more proximity with the Hausa than with the Yoruba population from this study. This can be due to the collection place, since samples from 1000 Genome Project were collected from Ibadan, a Yoruba-state. Although it is a region occupied mainly by the Yoruba, there is a considerable population of Hausa established in the area, which explains such proximity.

The proximity among the Yoruba and Igbo sample of this study and the Esan population from the 1000 Genome Project was expected because these groups belong to the same linguistic family, Niger-Congo. Also, the Esan people are present in great proportion in the Edo State, located between southwest and southeast, where Igbo and Yoruba people live.

From a wider perspective, the samples from Chad were the only ones that grouped independently to the ethnic group, with exception of the Kanembu. The proximity with this group can be explained by the integration of Nigeria with West Africa by the Chad basin. The Chad Basin area is known to be a highly diversified region in ethnolinguistic terms, with Afro-Asiatic, Nilo-Saharan and Niger-Congo populations living there (Cerny *et al.*, 2007). Thus, this region may have been a facilitator of the genetic flow and allowed the exchange and homogenization of the maternal lineages from Kanembu and Nigerian groups.

From the MDS graphic, it is also possible to discern the separation of samples belonging to the Niger-Congo linguistic family, regardless of the country, as seen in samples from Senegal, Gambia and Burkina Faso.

Figure 2 - MDS plot of F_{ST} genetic distances between the three ethnolinguistic groups of this study and West Central Africa populations from other publications.

Note: Samples from Burkina Faso and Sierra Leone are named with the codes BF and SL, respectively.



Phylogenies

A detailed analysis of haplogroups composition was performed for better understanding the genetic structure of the maternal lineages in the West African populations.

Most of the maternal lineages found in Nigeria belong to the macro haplogroup L (98.6%), characteristic of Sub-Saharan African populations. The L3 (45.6%) haplogroup was the most frequent lineage among the three ethnic groups, followed by L2 (31.2%), L1 (17.9%), L0 (3.2%) and L4 (0.7%) haplogroups. All African lineages were classified inside branches L0a, L1b-c, L2a-e, L3b, L3d-f, L3k and L4b. See detailed information on lineage frequency in Figure 3. The sub-clades L3e, L2a, L1b and L3d account for more than 60% of the maternal ancestry in Nigeria. Most of these branches show the highest frequencies in populations from the Central-West African region. The exception is L2a which is more frequent in the South-East region of Africa (Salas *et al.*, 2002; Cerezo *et al.*, 2016).

Haplogroup distribution

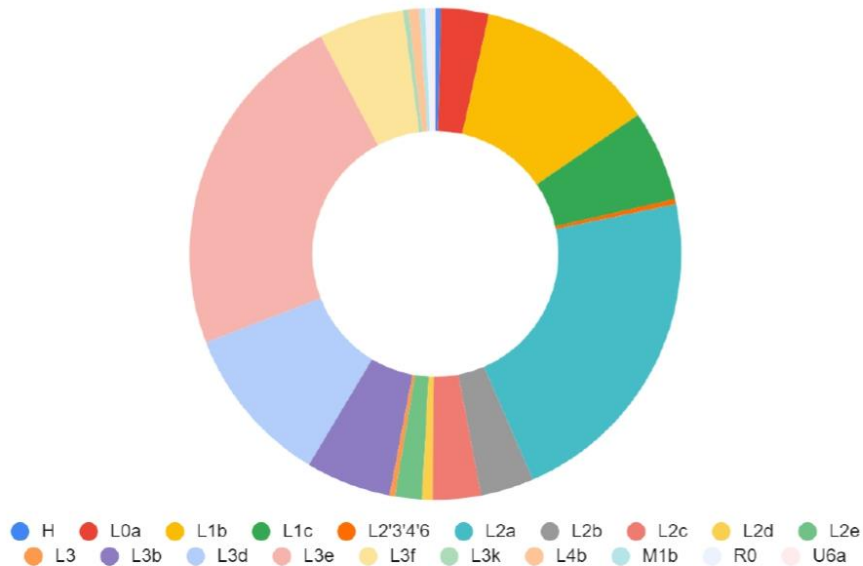


Figure 3 - Mitochondrial DNA haplogroup distribution.

The most ancient mtDNA haplogroup, L0, dates from 150-180.000 years ago with a southern African origin. This macro haplogroup is more frequent in the south and east regions, in both KhoeSan/KhoiSan and Bantu-speaking groups (Torroni *et al.*, 2006; Rosa *et al.*, 2011; Rito *et al.*, 2013). The lineage L0a2, known as a Bantu marker, was not found on the studied samples. Even though these three populations are non-Bantu groups, it was previously shown that this sub-lineage represent 68% of all L0 mtDNA sub-haplogroups in West-Central. On the contrary, all L0 haplotypes found in this study belong to L0a1 lineage, the most frequent in Southeast Africa (Cerezo *et al.*, 2016).

The L1 macro haplogroup is between 140-150.000 years old (Horai *et al.*, 1995; Torroni *et al.*, 2006; Behar *et al.*, 2008) and more frequent in Southwest Africa. Both L1 sub lineages represented in this study, L1b and L1c, have a probable Central Africa origin. The Yoruba, Igbo and Hausa samples presented high frequencies of the sub-clades L1b (11.9%), frequent in Fulani nomads' populations (Cerny *et al.*, 2007), showing the influence from this Niger-Congo-speaking group that are traditionally nomadic pastoralists and broadly distributed across the continent from Senegal, West and Central Africa, the Sahel to Sudan (Fan *et al.*, 2019). In contrast, the L1c lineage, present in 17 Nigerian samples, shows a contact with South African

populations, since its most likely origin is in Angola and neighboring countries (Batini *et al.*, 2007; Cerny *et al.*, 2007).

The macro haplogroup L2 has the highest values of haplotype and nucleotide diversities in the West-Central region (Salas *et al.*, 2002; Silva *et al.*, 2015; Cerezo *et al.*, 2016). The L2 lineage could have been emerged around 90-105.000 years ago (Gonder *et al.*, 2007; Behar *et al.*, 2008). The clade L2a, present at a frequency of 21.8% in Nigeria, is the most frequent in the continent Africa, totalizing a quarter of all these L2 macro haplogroups. The less represented L2 sub-clades, L2b-d, have likely origin in West and western Central Africa, since they are largely restricted to these regions with few occurrences in southeast (Salas *et al.*, 2002). One of the least represented L2 sub-clades, L2e, has a probable origin in West Africa being frequent in this region and dispersed across the Sahel region (Silva *et al.*, 2015).

According to studies the macro haplogroup L3 emerged in the Central Africa/Sudan region ~ 65-80.000 years ago (Salas *et al.*, 2002; Forster, 2004; Torroni *et al.*, 2006; Behar *et al.*, 2008) but is more diverse in the West-Central region (Cerezo *et al.*, 2016). The L3e (0,7%) sub lineages, L3e2 (14%) and L3e1 (4,9%), found in Nigeria, are common among Bantu-speaking populations and are associated with Bantu dispersal from Central West to southeastern Africa (Salas *et al.*, 2002). In the Central West region, L3e1 was previously found at low frequencies, although in Nigeria it was the second most represented sub-lineage found in the L3e branch. Still inside L3, the L3d haplogroup (10.5% in Nigeria) is also frequent in the West-Central region (Cerezo *et al.*, 2016). This lineage, together with L3b, is common in Fulani nomad's populations (Cerny *et al.*, 2007). These haplogroups were probably spread in Nigeria by the fusion of the Hausa-Fulani group, from Northern Nigeria (Ducrottoy *et al.*, 2018). The haplogroup L3f emerged in the East. The branch L3f1, with an early dispersion into the West region, was found in all our samples belonging to this haplogroup. Finally, L3k, the less represented L3 sub-clade, is rare and mainly present in North Africa (Soares *et al.*, 2012).

The L4 haplogroup was observed in only two samples from this study. This clade totalizes 15% of all mitochondrial DNA lineages present in East Africa, being more frequent in this region than in other areas of the continent (Cerezo *et al.*, 2016).

The remaining lineages from non-L haplogroups U6, R0 and M1 found in the Nigerian samples can be explained by a Northern/Mediterranean influence which was facilitated by the Chad Basin Corridor (Salas *et al.*, 2002; Olivieri *et al.*, 2006; Cerny *et al.*, 2007; Cerezo *et al.*, 2011). Among the three Nigerian groups, the major proportion of the non-L haplogroups was found in the Hausa, resulting from a greater contact with the North African region.

The haplogroup H, which is highly diverse in present-day West Europe (Brotherlon *et al.*, 2013), was also found in one Hausa sample. According to Ottoni *et al.* (2010), this haplogroup has high frequencies in Tuareg populations, being the most probable source of the haplogroup in the Nigerian sample.

Conclusion

The present study aimed to help fill the existing gap on the genetic description of Nigeria. Studying the total control region of mtDNA, we evaluated how the interactions between the distinct ethnic groups helped shaping the current genetic background of Nigeria. With this study we were able to demonstrate the great haplotype diversity present in the three Nigerian groups. The homogeneity of the maternal lineages on the west central populations, especially the neighboring populations of Nigeria, showed a high genetic flow mediated by the females, to which patrilocality may have contributed. Some lineages were found in populations from different regions along the continent. Also, as Nigeria was part of the slavery exploitation route and had great impact in the current African pool of admixed south American populations, this study will also contribute to further investigations that intend to elucidate the shaping of admixed populations.

Acknowledgements: LG was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (ref. 306342/2019-7) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ (CNE-2018), Brazil.

References

- 1) Alesina, Alberto *et al.* "Artificial States." *Journal of the European Economic Association* 9: 246-77 (2011)
- 2) Andrews, R. M. *et al.* "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA." *Nature genetics* vol. 23,2 (1999): 147. doi:10.1038/13779
- 3) Azeez O. Olaniyan, 2009. "From accommodation to discrimination and exclusion the changing pattern of inter-group relations in Nigeria." *African Journal of Political Science and International Relations* Vol. 3 (11), pp. 521-525
- 4) Bandelt, H J *et al.* "Median-joining networks for inferring intraspecific phylogenies." *Molecular biology and evolution* vol. 16,1 (1999): 37-48. doi:10.1093/oxfordjournals.molbev.a026036

- 5) Barbieri, Chiara *et al.* “Contrasting maternal and paternal histories in the linguistic context of Burkina Faso.” *Molecular biology and evolution* vol. 29,4 (2012): 1213-23. doi:10.1093/molbev/msr291
- 6) Batini, Chiara *et al.* “Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa.” *Molecular phylogenetics and evolution* vol. 43,2 (2007): 635-44. doi:10.1016/j.ympev.2006.09.014
- 7) Behar, Doron M *et al.* “The dawn of human matrilineal diversity.” *American journal of human genetics* vol. 82,5 (2008): 1130-40. doi:10.1016/j.ajhg.2008.04.002
- 8) Bermiell-Lee, Gemma *et al.* “Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages.” *Molecular biology and evolution* vol. 26,7 (2009): 1581-9. doi:10.1093/molbev/msp069
- 9) Brotherton, Paul *et al.* “Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans.” *Nature communications* vol. 4 (2013): 1764. doi:10.1038/ncomms2656
- 10) Butler, John *Forensic DNA Typing*. Academic Press, 2004
- 11) Cerezo, María *et al.* “New insights into the Lake Chad Basin population structure revealed by high-throughput genotyping of mitochondrial DNA coding SNPs.” *PloS one* vol. 6,4 e18682. 20 Apr. 2011, doi:10.1371/journal.pone.0018682
- 12) Cerezo, María *et al.* “New insights into the Lake Chad Basin population structure revealed by high-throughput genotyping of mitochondrial DNA coding SNPs.” *PloS one* vol. 6,4 e18682. 20 Apr. 2011, doi:10.1371/journal.pone.0018682
- 13) Cerný, V. *et al.* “A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome.” *Annals of human genetics* vol. 71,Pt 4 (2007): 433-52. doi:10.1111/j.1469-1809.2006.00339.x
- 14) Cerný, V. *et al.* “A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome.” *Annals of human genetics* vol. 71,Pt 4 (2007): 433-52. doi:10.1111/j.1469-1809.2006.00339.x
- 15) Choudhury, A. *et al.* High-depth African genomes inform human migration and health. *Nature* 586, 741–748 (2020). <https://doi.org/10.1038/s41586-020-2859-7>
- 16) Comey, C. T., *et al.* DNA extraction strategies for amplified fragment length polymorphism analysis. *Journal of Forensic Sciences*, 39 (1994): 1254–1269

- 17) de Filippo, Cesare *et al.* "Bringing together linguistic and genetic evidence to test the Bantu expansion." *Proceedings. Biological sciences* vol. 279,1741 (2012): 3256-63. doi:10.1098/rspb.2012.0318
- 18) Ducrottoy, Marie J. *et al.* "Patterns of passage into protected areas: Drivers and outcomes of Fulani immigration, settlement and integration into the Kachia Grazing Reserve, northwest Nigeria." *Pastoralism : research, policy and practice* vol. 8,1 (2018): 1. doi:10.1186/s13570-017-0105-1
- 19) Excoffier, Laurent *et al.* "Arlequin (version 3.0): an integrated software package for population genetics data analysis." *Evolutionary bioinformatics online* vol. 1 47-50. 23 Feb. 2007
- 20) Falola, T.; Heaton, M. M. *A history of Nigeria*. Cambridge University Press, 2008.
- 21) Fan, Shaohua *et al.* "African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations." *Genome biology* vol. 20,1 82. 26 Apr. 2019, doi:10.1186/s13059-019-1679-2
- 22) Fregel, Rosa; Sergio Delgado. "HaploSearch: a tool for haplotype-sequence two-way transformation." *Mitochondrion* vol. 11,2 (2011): 366-7. doi:10.1016/j.mito.2010.11.001
- 23) Gabriel, M. N. *et al.*. Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. *J Forensic Sci.* 2001 Mar;46(2):247-53. PMID: 11305426.
- 24) Göbel, Tanja *et al.* "Mitochondrial DNA variation in Sub-Saharan Africa: Forensic data from a mixed West African sample, Côte d'Ivoire (Ivory Coast), and Rwanda." *Forensic science international. Genetics* vol. 44 (2020): 102202. doi:10.1016/j.fsigen.2019.102202
- 25) Gonder, Mary Katherine *et al.* "Whole-mtDNA genome sequence analysis of ancient African lineages." *Molecular biology and evolution* vol. 24,3 (2007): 757-68. doi:10.1093/molbev/msl209
- 26) Horai, S *et al.* "Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs." *Proceedings of the National Academy of Sciences of the United States of America* vol. 92,2 (1995): 532-6. doi:10.1073/pnas.92.2.532
- 27) Metz, H. C. *Nigeria: a country study*. Washington D. C. Library of Congress, 1991
- 28) Newman J. L. *The Peopling of Africa: A Geographic Interpretation*. Yale University Press: New Haven, London; 1995.

- 29) Olivieri, Anna *et al.* "The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa." *Science* (New York, N.Y.) vol. 314,5806 (2006): 1767-70. doi:10.1126/science.1135566
- 30) Ottoni, Claudio *et al.* "Mitochondrial haplogroup H1 in north Africa: an early holocene arrival from Iberia." *PloS one* vol. 5,10 e13378. 21 Oct. 2010, doi:10.1371/journal.pone.0013378
- 31) Parsons, T. J.; Coble, M. D. "Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome." *Croatian medical journal* vol. 42,3 (2001): 304-9.
- 32) Pereira, L. *et al.* "Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade." *Annals of human genetics* vol. 65,Pt 5 (2001): 439-58. doi:10.1017/S0003480001008855
- 33) Rito, Teresa *et al.* "The first modern human dispersals across Africa." *PloS one* vol. 8,11 e80031. 13 Nov. 2013, doi:10.1371/journal.pone.0080031
- 34) Rosa A, Brehem A. African human mtDNA phylogeography at-a-glance. *J Anthropol Sci.* 2011;89:25-58. doi: 10.4436/jass.89006. Epub 2011 Mar 15. PMID: 21368343.
- 35) Rotimi, Charles N. *et al.* "The African diaspora: history, adaptation and health." *Current opinion in genetics & development* vol. 41 (2016): 77-84. doi:10.1016/j.gde.2016.08.005
- 36) Salas, Antonio *et al.* "The making of the African mtDNA landscape." *American journal of human genetics* vol. 71,5 (2002): 1082-111. doi:10.1086/344348
- 37) Salas, Antonio *et al.* "The making of the African mtDNA landscape." *American journal of human genetics* vol. 71,5 (2002): 1082-111. doi:10.1086/344348
- 38) Seielstad, M. T.; Minch, E.; Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in humans. *Nat Genet.* 1998 Nov;20(3):278-80. doi: 10.1038/3088. PMID: 9806547.
- 39) Silva, M. *et al.* 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. *Sci Rep* 5, 12526 (2015). <https://doi.org/10.1038/srep12526>
- 40) Simão, F. *et al.* Defining mtDNA origins and population stratification in Rio de Janeiro. *Forensic Science international. Genetics.* May;34 (2018):97-104. DOI: 10.1016/j.fsigen.2018.02.003.

- 41) Simpson, A.; Oyètádé, A. Nigeria ethno-linguistic competition in the Giant of Africa, in *Language and National Identity in Africa*. 172-98, Oxford: Oxford University Press, 2007.
- 42) Soares, Pedro *et al.* "The Expansion of mtDNA Haplogroup L3 within and out of Africa." *Molecular biology and evolution* vol. 29,3 (2012): 915-27. doi:10.1093/molbev/msr245
- 43) Tishkoff, Sarah A. *et al.* "The genetic structure and history of Africans and African Americans." *Science* (New York, N.Y.) vol. 324,5930 (2009): 1035-44. doi:10.1126/science.1172257
- 44) Torroni, Antonio *et al.* "Harvesting the fruit of the human mtDNA tree." *Trends in genetics : TIG* vol. 22,6 (2006): 339-45. doi:10.1016/j.tig.2006.04.001
- 45) Underhill, Peter A; Toomas Kivisild. "Use of y chromosome and mitochondrial DNA population structure in tracing human migrations." *Annual review of genetics* vol. 41 (2007): 539-64. doi:10.1146/annurev.genet.41.110306.130407
- 46) van Oven, Mannis; Manfred Kayser. "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation." *Human mutation* vol. 30,2 (2009): E386-94. doi:10.1002/humu.20921
- 47) Vandenberg, N. *et al.* An evaluation of selected DNA extraction strategies for short tandem repeat typing. *Electrophoresis*, 18 (1997): 1624–1626
- 48) Veeramah, K. R. *et al.* Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evol Biol* 10, 92 (2010). <https://doi.org/10.1186/1471-2148-10-92>
- 49) Walsh, P. S. *et al.* Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, 10 (1991): 506–513
- 50) Watson, E. *et al.* "mtDNA sequence diversity in Africa." *American journal of human genetics* vol. 59,2 (1996): 437-44.
- 51) Wood, Elizabeth T. *et al.* "Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes." *European journal of human genetics : EJHG* vol. 13,7 (2005): 867-76. doi:10.1038/sj.ejhg.5201408
- 52) Zhao, Hongyu *et al.* "Haplotype analysis in population genetics and association studies." *Pharmacogenomics* vol. 4,2 (2003): 171-8. doi:10.1517/phgs.4.2.171.22636
- 53) Zimmermann, Bettina *et al.* "Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software." *Croatian medical journal* vol. 55,2 (2014): 115-20. doi:10.3325/cmj.2014.55.115

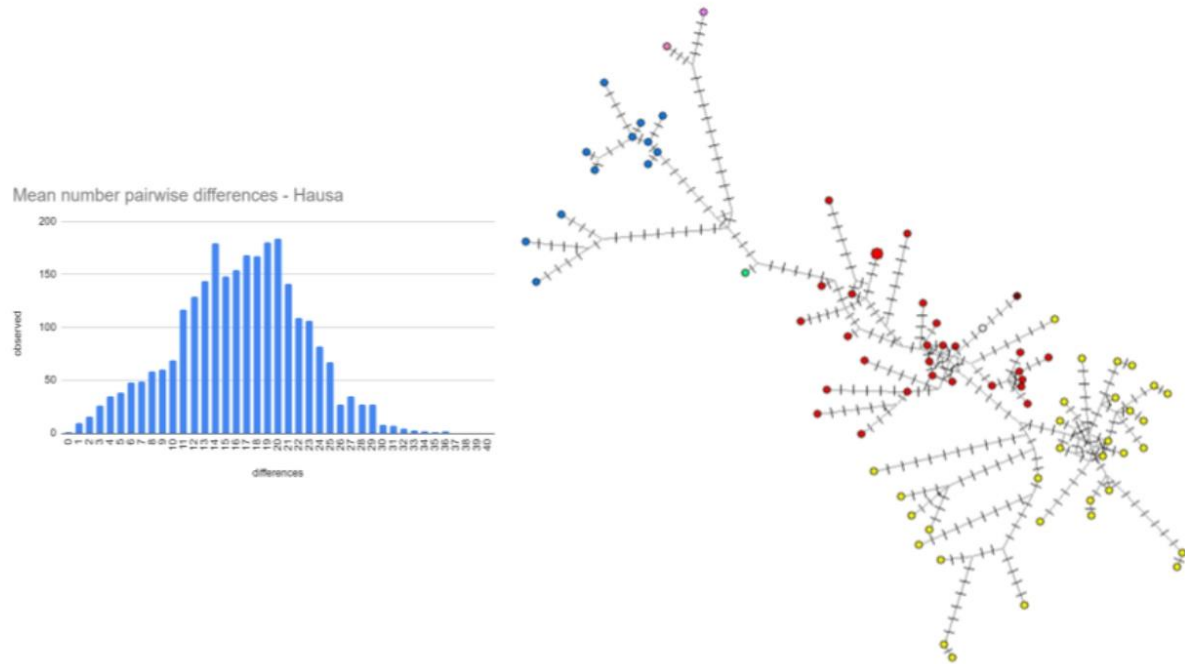
SUPPLEMENTARY MATERIAL

Supplementary Table 1 - Primer sequences used in PCR and Sanger sequencing.

Primer	Sequence
L15900	5'-TAAACTAATACACCAGTCTTGTAACC-3'
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'
L15967	5'-GTCTTAACTCCACCATTAGCACC-3'
H20	5'-AGCTCCCGTGAGTGGTTAATA-3'
L16475	5'-TAGCTAAAGTGAAGTGTATCC-3'
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'
L15900	5'-TAAACTAATACACCAGTCTTGTAACC-3'
L15997	5'-CACCATTAGCACCCAAAGCT-3'
L16268	5'-CACTAGGATACCAACAAACC-3'
L16555	5'-CCCACACGTTCCCCTTAAAT-3'
L00314	5'-CCGCTTCTGGCCACAGCACT-3'
H16322	5'-TGGCTTTATGTAAGTATGTAC-3'
H00016	5'-CCCGTGAGTGGTTAATAGGGT-3'
H00159	5'-AAATAATAGGATGAGGCAGGAATC-3'
H00388	5'-TGGTTAGGCTGGTGTAGGG-3'
H00599	5'-TTGAGGAGGTAAGCTACATA-3'
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'

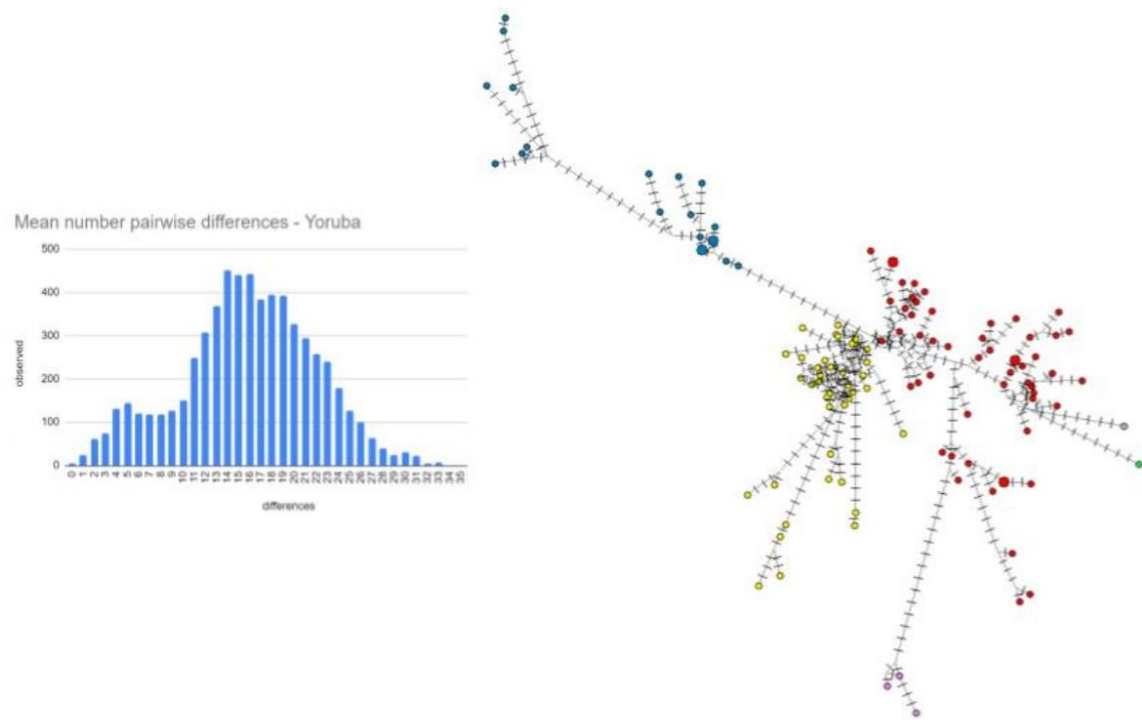
Supplementary Table 2 - List of haplotypes found in a sample of 285 unrelated individuals from Nigeria.

A001	L18a10	1612C	1619C	1618T	1618C	1622T	1624T	1627T	1628T	1611C	1619C	79G	111T	112C	112T	112T	112C	112T	112C	112T	112C	247A	269G	309.1C	315.1C	357G	3236f	3246f
A004	L18a10	1612C	1618T	1619C	1622T	1624T	1627T	1628T	1611C	1619C	79G	111T	112C	112T	112T	112C	112T	112C	112T	112C	247A	269G	309.1C	315.1C	357G	3236f	3246f	
A005	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	309.1C	315.1C													
A007	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	309.1C	315.1C													
A008	L1c3a8	1626T	1619A	1618C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C	1619C
A110	L2a1	1622T	1622T	1622T	1627T	1639G	79G	150T	199C	269G	315.1C	316f	3236f	3246f														
A011	L18a1189	1612C	1618T	1619C	1622T	1624T	1627T	1628T	1611C	1619C	79G	112C	112T	112T	112T	112C	112T	112C	112T	112C	247A	269G	309.1C	315.1C	357G	3236f	3246f	
A012	M11d	1619A	1618C	1619C	1619C	1622T	1624C	1611C	1619C	1619C	79G	199C	269G	315.1C	499C													
A018	L3T10a	1612A	1629C	1623T	1629T	1629T	1629T	1611C	1619C	79G	149G	200G	207A	269G	309.1C	315.1C												
A020	L2a1	1619C	1623T	1627T	1629T	1639G	1639G	79G	143A	148C	152C	159G	199C	269G	309.1C	315.1C												
A022a	L2a1	1612C	1619C	1622T	1627T	1629G	1639A	1639C	79G	148C	152C	199C	269G	315.1C														
A022b	L2a1	1622T	1639G	79G	150T	199C	199T	269G	309.1C	315.1C																		
A023	L2a1	1622T	79G	152C	199G	199C	269G	315.1C	3236f	3246f																		
A024	L2a1	1619C	1623T	1627T	1629T	1639G	1639G	1639G	79G	148C	152C	199C	269G	315.1C	3236f	3246f												
A027	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	148C	152C	199C	269G	309.1C	315.1C														
A029	L2a1	1619A	1612A	1612A	1623T	1627T	1611T	1611T	1611T	1611T	79G	110T	152C	112T	199C	199T	207C	269G	315.1C	411T	5236f	5246f						
A031	L18a1189	1609T	1619C	1618T	1619C	1622T	1624T	1627T	1628T	1611C	1619C	79G	112C	112T	112T	112T	112C	112T	112C	112T	112C	247A	269G	309.1C	315.1C	357G	3236f	3246f
A033	L2a1	1612C	1623T	1627T	1629C	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A034	L2a1	1619C	1612C	1623T	1627T	1629G	1639G	79G	150T	199C	269G	315.1C																
A035	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A036	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A037	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A038	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A039	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A040	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A041	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A042	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A043	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A044	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A045	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A046	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A047	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A048	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A049	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A050	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A051	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A052	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A053	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A054	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A055	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A056	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A057	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A058	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A059	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A060	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A061	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A062	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A063	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A064	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A065	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A066	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A067	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A068	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A069	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A070	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A071	L2a1	1622T	1627T	1629T	1639G	1639G	1639G	79G	143A	148C	152C	199C	269G	315.1C	3236f	3246f												
A072	L2a1	1622T	1627T	1629T	1639G	1639G	16																					



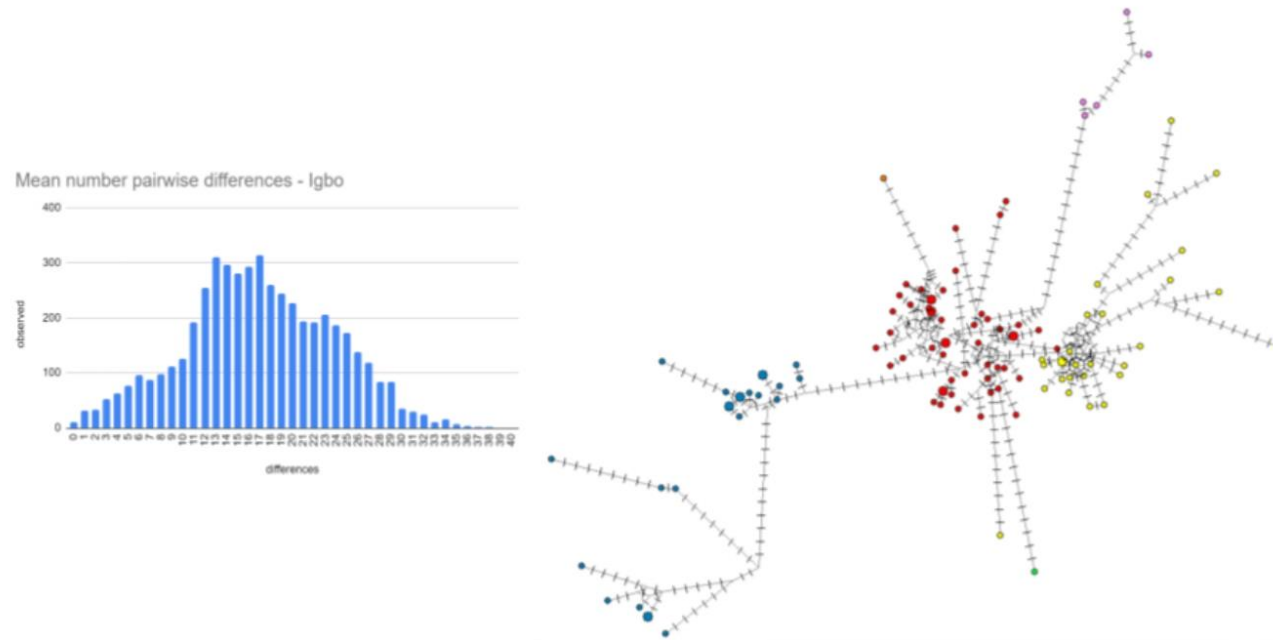
Supplementary Figure 1 - Mean number pairwise differences and median joining network representation of mitochondrial DNA control region haplotypes from Hausa.

Note: the haplogroups are represented by colors - brown (H), pink (L0); blue (L1); yellow (L2), red (L3), green (L4), grey (M), white (R) and orange (U). The circles represent the haplotypes their consecutive frequencies according to their size. Each dash/line represents the mutational steps/polymorphisms between the haplotypes.



Supplementary Figure 2 - Mean number pairwise differences and median joining network representation of mitochondrial DNA control region haplotypes from Yoruba.

Note: the haplogroups are represented by colors - brown (H), pink (L0); blue (L1); yellow (L2), red (L3), green (L4), grey (M), white (R) and orange (U). The circles represent the haplotypes and their consecutive frequencies according to their size. Each dash/line represents the mutational steps/polymorphisms between the haplotypes.



Supplementary Figure 3 - Mean number pairwise differences and median joining network representation of mitochondrial DNA control region haplotypes from Igbo.

Note: the haplogroups are represented by colors - brown (H), pink (L0); blue (L1); yellow (L2), red (L3), green (L4), grey (M), white (R) and orange (U). The circles represent the haplotypes and their consecutive frequencies according to their size. Each dash/line represents the mutational steps/polymorphisms between the haplotypes.



Supplementary Figure 4 - Distribution of the West African samples used for comparison in this study grouped by country. Adapted from Google.

Note: Burkina Faso is represented by the code BF with samples from Cerezo (n=20) and Čížková (n=160); Cameroon from Cerezo (n=40) and Göbel (n=11); Chad from Cerezo (n=50) and Čížková (n=183); Côte d'Ivoire from Göbel (n=27); Ghana from Fendt (n=192) and Göbel (n=22); Guinea from Čížková (n=51); Mali from Cerezo (n=14); Niger from Cerezo (27); Nigeria from Göbel (n=28) and 1000 Genome Project (n=207); Senegambia representing the countries Senegal from Čížková (n=192) and Gambia from 1000 Genome Project (n=113); Sierra Leone from 1000 Genome Project (n=68); Togo from Göbel (n=34);

Supplementary Table 3 - Pairwise F_{ST} genetic distances (below diagonal) among west central populations, and non-differentiation p -values (above diagonal) (s.e. ≤ 0.0001 , after 10100 permutations).

Note: Codes stands for: TU = Tuareg; BF = Burkina Faso; TF = Tcheboua Fulani; KAN = Kanembu; SL = Sierra Leone; DAN = Dangaleat; ARA = Arabs Abala; GOU = Gourmantche; FF = Fulani Ferlo; FFD = Fulani Fouta Djallon; FZ = Fulani Ziniare; HAL = Halpularen.

	Yoruba	Igbo	Hausa	TU_BF	TF_Cameroon	KAN_Chad	TU_Mali	TU_Niger	Akan_Ghana	Esoo_Nigeria	Yoruba_Niger	Gambia	Mende_SL	DAN_Chad	Maba_Chad	ARA_Chad	Daza_Chad	Bedik_Senegal	Gurusi_BF	GOU_BF	Mossi_BF	Senegal_Senegal	FF_Senegal	FFD_Guinea	FZ_BF	HAL_Senegal	Cameroon	Cote d'Ivoire	Ghana	Nigeria	Togo		
Yoruba	*	0.92158	0.87764	0.00000	0.02237	0.03772	0.10385	0.09601	0.47669	0.90922	0.00000	0.00000	0.00020	0.00010	0.00000	0.00000	0.00000	0.00000	0.00475	0.02158	0.02267	0.00000	0.00000	0.00000	0.00000	0.00000	0.00505	0.07801	0.57014	0.89338	0.14593		
Igbo	-0.00410	*	0.48362	0.00010	0.00000	0.01782	0.04485	0.05633	0.02079	0.88248	0.53205	0.00000	0.00000	0.00030	0.00020	0.00000	0.00000	0.00000	0.00089	0.00050	0.01059	0.00000	0.00010	0.00000	0.00000	0.00000	0.01911	0.15098	0.44303	0.98663	0.07821		
Hausa	-0.00438	-0.00061	*	0.00010	0.00020	0.06069	0.03792	0.10415	0.21998	0.20701	0.99944	0.00000	0.00287	0.00010	0.00000	0.00000	0.00020	0.00050	0.02129	0.01495	0.04029	0.00000	0.00178	0.00000	0.00000	0.00426	0.16513	0.68577	0.45590	0.43788			
TU_BF	0.00962	0.09150	0.00815	*	0.00000	0.00079	0.03544	0.00842	0.00010	0.00000	0.00000	0.00000	0.00000	0.00010	0.00010	0.00000	0.00149	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
TF_Cameroon	0.05572	0.04772	0.04879	0.17285	*	0.00010	0.00089	0.00010	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00010	0.00000	0.00020	0.00000	0.00037	0.00010	0.00000	0.00109	0.00139	0.01812	0.00745	0.00940			
KAN_Chad	0.01378	0.01530	0.01090	0.05751	0.06299	*	0.04346	0.57341	0.05138	0.01337	0.03349	0.00000	0.00000	0.01040	0.00168	0.00010	0.01830	0.00000	0.021337	0.02297	0.02356	0.00000	0.00030	0.00000	0.00000	0.00426	0.22928	0.66508	0.17699	0.31532			
TU_Mali	0.03227	0.02780	0.03418	0.07225	0.10345	0.03385	*	0.03336	0.02277	0.05207	0.02881	0.00059	0.00079	0.00277	0.00917	0.00277	0.01436	0.00069	0.00089	0.00178	0.01109	0.00020	0.00050	0.00248	0.00059	0.00069	0.01010	0.04376	0.05306	0.11048	0.02633		
TU_Niger	0.01114	0.01637	0.01227	0.06376	0.08465	-0.00084	0.04784	*	0.28096	0.02782	0.06742	0.00000	0.00059	0.01802	0.01574	0.00059	0.01703	0.00010	0.05306	0.01307	0.05009	0.00000	0.00000	0.00139	0.00000	0.00000	0.00554	0.33799	0.73438	0.09415	0.48431		
Akan_Ghana	0.01633	0.01782	0.00241	0.06194	0.05855	0.01008	0.04175	0.01340	*	0.00158	0.04216	0.00000	0.00010	0.00000	0.00000	0.00000	0.00010	0.00000	0.01337	0.01068	0.01089	0.00000	0.00000	0.00000	0.00000	0.00000	0.00129	0.41243	0.61304	0.08118	0.22859		
Esoo_Nigeria	-0.00048	-0.00556	0.00305	0.09404	0.05699	0.01770	0.02958	0.02205	0.01536	*	0.13294	0.00000	0.00000	0.00030	0.00010	0.00000	0.00010	0.00000	0.00030	0.00040	0.00594	0.00000	0.00000	0.00000	0.00000	0.00000	0.04485	0.11286	0.23404	0.91664	0.02020		
Yoruba_Nigeria	-0.00385	-0.00086	-0.00716	0.05754	0.04917	0.01267	0.01616	0.01343	0.00626	0.00094	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00941	0.0525	0.03643	0.00000	0.00169	0.00000	0.00000	0.00043	0.10830	0.54688	0.72488	0.34987			
Gambia	0.02534	0.04010	0.02628	0.15226	0.05813	0.05512	0.09104	0.06024	0.04041	0.04958	0.02962	*	0.19137	0.00000	0.00000	0.00000	0.00000	0.00010	0.00812	0.06138	0.02851	0.01228	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000		
Mende_SL	0.03605	0.03726	0.02365	0.13754	0.06050	0.04163	0.08283	0.05003	0.02907	0.04383	0.03016	0.00323	*	0.00020	0.00020	0.00000	0.00020	0.10464	0.16108	0.07937	0.06356	0.00000	0.00000	0.00000	0.00000	0.00000	0.11197	0.00000	0.00010	0.01247	0.04604	0.02568	0.05099
DAN_Chad	0.08328	0.09436	0.03809	0.11114	0.08324	0.02493	0.07240	0.09058	0.04370	0.03575	0.03614	0.04651	0.03844	*	0.13099	0.00000	0.14741	0.13236	0.00000	0.02426	0.02772	0.02727	0.00000	0.00000	0.00000	0.00000	0.00000	0.00186	0.00158	0.02426	0.04237	0.01188	
Maba_Chad	0.05592	0.05237	0.05641	0.12968	0.10111	0.04374	0.08932	0.04005	0.05297	0.05020	0.05783	0.07206	0.04713	0.00589	*	0.02841	0.07722	0.00000	0.00950	0.02030	0.01921	0.00000	0.00000	0.00000	0.00000	0.00000	0.01699	0.01089	0.01168	0.00990	0.01337		
ARA_Chad	0.04925	0.04649	0.05087	0.13795	0.07794	0.04914	0.07949	0.05777	0.06713	0.04934	0.03013	0.06731	0.05995	0.00773	0.02341	*	0.05693	0.00000	0.00040	0.02138	0.00190	0.00050	0.00000	0.00000	0.00000	0.00000	0.00000	0.01614	0.00000	0.00287	0.03455	0.00020	
Daza_Chad	0.05189	0.05786	0.05205	0.10476	0.12285	0.03132	0.07128	0.09959	0.05571	0.05876	0.05314	0.06999	0.05466	0.01025	0.01817	0.01921	*	0.00020	0.02782	0.05118	0.06415	0.00000	0.00000	0.00020	0.00000	0.00000	0.00916	0.00109	0.00752	0.00723	0.00782		
Bedik_Senegal	0.00547	0.00607	0.04746	0.17313	0.07509	0.06489	0.09492	0.08054	0.05815	0.06476	0.05212	0.02051	0.00906	0.06733	0.07415	0.08625	0.09045	*	0.00907	0.00109	0.00980	0.00000	0.00000	1.00118	0.00129	0.00178	0.00000	0.00396	0.00832	0.00099	0.01020		
Gurusi_BF	0.03080	0.00990	0.02241	0.14658	0.09147	0.02902	0.10888	0.02425	0.02372	0.05906	0.02707	0.01336	0.00779	0.02559	0.04118	0.05754	0.05102	0.04583	*	0.78616	0.38115	0.00000	0.00000	0.04505	0.00000	0.00000	0.00010	0.00812	0.16688	0.00762	0.28571		
GOU_BF	0.08391	0.00233	0.02299	0.13979	0.08473	0.03756	0.09527	0.03997	0.03941	0.04704	0.03876	0.01649	0.01299	0.02325	0.03367	0.04708	0.02336	0.04904	-0.00946	*	0.21216	0.00000	0.00000	0.00723	0.00000	0.00000	0.00089	0.00455	0.05019	0.00851	0.18335		
Mossi_BF	0.01809	0.02207	0.01841	0.13147	0.06962	0.02286	0.06495	0.02463	0.02211	0.02882	0.01625	0.02139	0.01230	0.01447	0.03055	0.02511	0.01990	0.03582	0.00060	0.00705	*	0.00000	0.00000	0.13682	0.00010	0.00129	0.01944	0.13147	0.16830	0.16137			
Senegal_Senegal	0.07351	0.07078	0.07020	0.17960	0.10772	0.05976	0.09456	0.11299	0.10234	0.07027	0.07710	0.08969	0.09814	0.06035	0.07872	0.04033	0.07827	0.11914	0.10783	0.08599	0.07829	*	0.00000	0.00000	0.00000	0.00000	0.00989	0.00000	0.00000	0.00020	0.00000		
FF_Senegal	0.07515	0.07524	0.08057	0.22715	0.07532	0.10346	0.13106	0.12435	0.08492	0.08471	0.07005	0.05349	0.07184	0.08899	0.13122	0.10806	0.13748	0.07927	0.09756	0.10514	0.07971	0.11600	*	0.00842	0.00099	0.00000	0.00020	0.00000	0.00030	0.00069	0.00020		
FFD_Guinea	0.09990	0.04007	0.03099	0.15724	0.04510	0.04643	0.08431	0.05689	0.03955	0.04780	0.03016	0.00780	0.00862	0.04887	0.06863	0.06803	0.07325	0.01217	0.02364	0.03680	0.01007	0.09740	0.03523	*	0.01445	0.00257	0.00050	0.00455	0.03554	0.01079	0.03643		
FZ_BF	0.11469	0.10290	0.10815	0.25144	0.10405	0.11888	0.13591	0.14566	0.10277	0.10845	0.09979	0.09424	0.08669	0.13480	0.15111	0.14845	0.17197	0.05876	0.13348	0.15005	0.08874	0.17604	0.06702	0.03362	*	0.00089	0.00010	0.00010	0.00010	0.00010	0.00000		
HAL_Senegal	0.08995	0.08252	0.08010	0.19435	0.09006	0.09718	0.10943	0.11242	0.09267	0.08740	0.08854	0.06684	0.06092	0.09545	0.10888	0.10871	0.12653	0.04575	0.09241	0.09515	0.06833	0.12941	0.08388	0.04555	0.06530	*	0.00010	0.00000	0.00040	0.00030	0.00000		
Cameroon	0.07084	0.05037	0.08237	0.21732	0.11664	0.08059	0.08441	0.10382	0.10612	0.03775	0.07347	0.15935	0.14356	0.07142	0.08124	0.06203	0.11532	0.14336	0.18414	0.14584	0.10850	0.08826	0.17632	0.13925	0.19002	0.13958	*	0.00663	0.00604	0.15147	0.00208		
Cote d'Ivoire	0.01263	0.00820	0.00810	0.09616	0.06116	0.00556	0.04066	0.00389	-0.00015	0.01089	0.01136	0.05135	0.02956	0.04922	0.04088	0.07291	0.06773	0.04566	0.04378	0.04737	0.05539	0.11495	0.11912	0.04608	0.11079	0.06549	0.08283	*	0.49133	0.10930	0.27948		
Ghana	-0.00955	-0.00065	-0.00667	0.08039	0.04114	-0.00657	0.04631	-0.01104	-0.00500	0.00600	-0.00337	0.02671	0.02271	0.03076	0.04913	0.05256	0.05608	0.04579	0.01399	0.													