

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL



**CARACTERIZAÇÃO GENÉTICA DA POPULAÇÃO
BRASILEIRA IMIGRANTE EM LISBOA ATRAVÉS DO
ESTUDO DE DNA MITOCONDRIAL**

Miguel António Faria Marcelino

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Professor Doutor António Amorim

Índice

Índice de Tabelas	III
Índice de Figuras	IV
Agradecimentos	V
Resumo	VI
Abstract	VII
Lista de Abreviaturas	VIII
1. Introdução	1
1.1. Mitocôndria	1
1.2. Origem da mitocôndria	1
1.3. DNA Mitocondrial	2
1.4. Genoma mitocondrial	2
1.5. Características do DNAm	4
1.6. Heteroplasmia	5
1.6. Aplicações do estudo do DNAm	7
1.6.1. Aplicações forenses	7
1.6.2. Genética Populacional	8
1.6.3. Genética médica	9
1.7. Base de Dados	9
1.8. Contexto migratório que deu origem à população brasileira atual	10
2. Objetivos	11
3. Materiais e métodos	12
3.1. Seleção das Amostras	13
3.2. Extração de DNA	13
3.3. Amplificação do DNA	13
3.4. Purificação dos Produtos Amplificados	15
3.5. Sequenciação	15
3.6. Purificação dos Produtos Sequenciados	17
3.7. Detecção dos Produtos sequenciados	17
3.8. Análise dos produtos Sequenciados	18
4. Resultados e Discussão	19
4.1 Estudo dos Haplótipos	19
4.3. Heteroplasmias	24
4.4. Análise filogenética	25
5. Conclusão	28
6. Referências	29
	2

Índice de Tabelas

Tabela 1.1: Comparação das características entre o DNA nuclear e DNA mitocondrial, adaptado de [5]	2
Tabela 3.1: Sequência dos primers usados na amplificação da região controlo total do DNAMt	14
Tabela 3.2: Volumes usados na reação de amplificação da região controlo total do DNA	14
Tabela 3.3: Condições de temperatura para a amplificação da região controlo do DNAMt	14
tabela 3.4: Condições de temperatura para a purificação dos produtos amplificados	15
Tabela 3.5: Reagentes e volumes utilizados na sequenciação da região controlo total do DNAMt	16
Tabela 3.6: Condições de temperatura para a sequenciação da região controlo do DNAMt	17
Tabela 3.7: Volumes utilizados no processo de purificação dos produtos sequenciados	17
Tabela 3.8: Código IUPAC utilizado na nomenclatura das bases nucleotídicas	18
Tabela 4.1: Distribuições dos polimorfismos observados na região controlo total do DNAMt de imigrantes brasileiros residentes em Lisboa	20
Tabela 4.2: Heteroplasmias presentes na região controlo total do DNAMt da população brasileira imigrante em Lisboa	25
Tabela 4.3: Populações de estudos anteriores onde se analisou a região controlo do DNAMt, utilizadas para fins comparativos.	25
Tabela 4.4: Valores de Fst e p obtidos através do programa arlequin para a totalidade da região controlo do DNAMt.	26

Índice de Figuras

Figura 1.1: Representação do DNAm (adaptado de [22])	4
Figura 1.2: Representação de uma heteroplasmia de comprimento	6
Figura 1.3: Representação de um heteroplasmia de posição	6
Figura 1.4: Representação das migrações do Homem moderno e respetiva distribuição geográfica dos haplogrupos (adaptado de [66])	9
Figura 3.1: Fluxograma do procedimento experimental realizado	12
Figura 3.2: Esquema representativo da sequenciação com ddNTPs marcados com fluorescência, adaptado de [83].	16

Agradecimentos

Gostaria de em primeiro lugar agradecer profundamente à minha orientadora Dra. Heloísa Afonso Costa por todo o apoio prestado e por toda a ajuda incondicional ao longo do processo de realização deste trabalho.

Ao meu orientador, Dr. António Amorim, pela oportunidade que me proporcionou em realizar este trabalho e por toda a ajuda na realização do mesmo.

A toda a restante equipa, que compõe o Serviço de Genética e Biologia Forenses do INMLCF, por toda a paciência que tiveram para comigo, bem como em toda a ajuda prestada no esclarecimento de diversas técnicas.

Aos meus amigos que sempre mostraram a capacidade de escutar as minhas frustrações e sempre me conseguiram motivar de modo a continuar a trabalhar.

A toda a minha família, em especial aos meus pais e avós, por estarem sempre ao meu lado em todas as decisões por mim tomadas e terem tornado a realização deste trabalho um processo muito mais fácil.

Resumo

A migração de indivíduos é um dos principais contribuidores para a introdução de variabilidade genética dentro das populações. Nas últimas duas décadas em Portugal, mais particularmente Lisboa, recebeu um grande número destes migrantes. A comunidade de imigrantes brasileiros é uma das principais comunidades estrangeiras atualmente a residir em Portugal, constituindo, em 2020, cerca de 184 000 indivíduos.

As características únicas do DNAm_t, ao ser transmitido exclusivamente por via materna e não sofrendo recombinação, fazem deste um marcador muito útil, tendo várias aplicações como o estudo da evolução das populações, a utilização em perícias forenses e a genética médica.

Neste estudo foi realizada a sequenciação de DNAm_t de 75 imigrantes brasileiros a residir em Lisboa, de modo a determinar o impacto que o *pool* genético desta população poderá ter no gene *pool* da população lisboeta. Para isso foi amplificada a região controlo do DNAm_t com dois pares de primers L15971 / H016 e L16555 / H639. Os produtos amplificados foram sequenciados com BigDye®Terminator v.3.1 Cycle Sequence (AB) e detetados no SeqStudio™ Genetic Analyzer (AB). Os resultados obtidos foram analisados nos softwares Sequencing Analysis v7. e SeqScape v4. (AB), onde as sequências obtidas foram comparadas com a rCRS obtendo haplótipos que, recorrendo à Phylotree, build 17, podem ser classificados em haplogrupos. Sendo que a maioria dos haplótipos obtidos neste estudo não apresentaram correspondência na base de dados EMPOP. Tratando-se de uma população extremamente heterogénea, com cerca de metade dos haplótipos obtidos pertencendo ao macrohaplogrupo L, característico de populações africanas, um quarto dos haplótipos correspondem a haplótipos sul americanos, sendo o quarto restante pertencente a populações euroasiáticas.

Palavras-chave: DNAm_t; Região controlo total; Brasil

Abstract

Migration is one of the main factors for genetic variability within populations. Currently, the Portuguese population, and particularly the population from Lisbon, welcomes a considerable number of immigrants. Brazilian immigrants are the main foreign community in Portugal, with about 184 000 individuals in 2020.

Mitochondrial DNA (mtDNA), due to its unique characteristics such as being exclusively maternal inheritance and suffering no recombination, which results in its slow evolution, is a useful genetic marker in various areas such as forensic the study the evolution of populations, forensic genetics and medical genetics.

In this study mtDNA sequencing analysis of 75 Brazilian immigrants who currently live in Lisbon were carried out in order to assess the impact of this population on the Portuguese gene pool. The mtDNA control region were amplified using two pairs of primers - L15971 / H016 and L16555 / H639. The amplified products were then sequenced using BigDye®Terminator v.3.1 Cycle Sequence (AB) and detected in the SeqStudio™ Genetic Analyzer (AB). The results were analyzed with the Sequencing Analysis v7. and SeqScape v4. (AB) softwares, where the obtained sequences were compared with the rCRS in order to obtain haplotypes that, with Phylotree, build 17, can be converted in haplogroups. From our results, it is possible to confirm that this Brazilian population living in Lisbon presents a high number of unique haplotypes, most of them had no coincidence on EMPOP Forensic database. It's an extremely heterogeneous population with half of the studied haplotypes belonging to macrohaplogroup L, characteristic mainly of Sub-Saharan region of Africa, and a quarter of the studied haplotypes belonging to South American population and the other quarter belonging to Euro Asiatic population.

Key Words: mtDNA; Control Region; Brazil

Lista de Abreviaturas

- A: Adenina
- AB: Applied Biosystems
- AMOVA: Análise da variância molecular
- C: Citosina
- CRS: Sequência de Cambridge
- ddATP: Didesoxiadenina trifosfatada
- ddCTP: Didesoxicitosina trifosfatada
- ddGTP: Didesoxiguanina trifosfatada
- ddNTPs: Didesoxinucleótidos trifosfatados
- ddTTP: Didesoxitimina trifosfatada
- DEL: Deleção
- D-Loop: *Displacement Loop*
- DNA: Ácido desoxirribonucleico
- DNAmt: DNA mitocondrial
- DNAn: DNA nuclear
- dNTP: Desoxinucleótidos trifosfatados
- EDNAP: *European DNA Profiling Group*
- EMPOP: *EDNAP mitochondrial DNA population database*
- FBI: *Federal Bureau of Investigation*
- Fst: *Fixation rate*
- G: Guanina
- H: Cadeia pesada
- HVI: Região Hipervariável 1
- HVII: Região Hipervariável 2
- HVIII: Região Hipervariável 3
- IUPAC: *International Union of Pure and Applied Chemistry*
- L: Cadeia Leve
- M: Adenina ou Citosina
- MgCl₂: Cloreto de Magnésio
- ml: Mililitro
- °C: Graus Celsius
- O_H: Origem de replicação da cadeia pesada
- p: Nível de significância

pb: Pares de bases

PCR: *Polymerase Chain Reaction*

poli-C: Poli-citosinas

R: Guanina ou Adenina

rCRS: Sequência de referência de Cambridge revista

RFLP: *Restriction Fragment Length Polymorphism*

ROS: Espécies Reativas de Oxigênio

rpm: Rotações por minuto

SGBF: Serviço de Genética e Biologia Forenses

SNPs: *Single Nucleotide Polymorphisms*

STR: Short Tandem Repeat

SWGDM: Scientific Working Group for DNA Analysis Methods

T: Timina

Y: Citosina e Timina

μL: Microlitro

1. Introdução

1.1. Mitocôndria

As Mitocôndrias são organelos celulares semiautónomos, pois possuem um genoma próprio independente do DNA nuclear, o DNAm (DNA mitocondrial) que existe no citoplasma de todas as células eucarióticas (com exceção dos glóbulos vermelhos e dos protozoários)[1]. São constituídas por um sistema membranar duplo, com uma membrana interna e uma membrana externa, as quais se envolvem numa matriz mitocondrial, local onde se encontram o DNAm, e são essenciais num variado número de processos celulares que contribuem para a manutenção da homeostasia celular como a fosforilação oxidativa, processo responsável pela produção da energia celular essencial em diversos processos metabólicos, sendo também essenciais na morte celular programada (apoptose) e no metabolismo de ácidos gordos, lípidos, colesterol e esteróis[2].

1.2. Origem da mitocôndria

A teoria aceita relativamente à origem das mitocôndrias reitera que estas são o resultado de um processo evolutivo de uma bactéria ancestral, capaz de metabolizar oxigénio, que estabeleceu uma relação de simbiose com uma célula eucariótica primitiva, que até então vivia em condições anaeróbias e terá a partir deste momento adquirido a capacidade de sobreviver em meios aeróbios, situação que potenciou o seu crescimento e a sua expansão, em ambientes com crescentes concentrações de oxigénio, que até então era um elemento tóxico para as mesmas[1].

Com o passar do tempo a bactéria capaz de metabolizar oxigénio ter-se-á tornado parte integral da célula eucariótica primitiva o que resultou na formação da mitocôndria. Os genes presentes na bactéria ancestral terão na sua maioria sido transferidos para o núcleo das células eucarióticas tendo um pequeno número de genes permanecido no organelo, o qual se viria a tornar a mitocôndria[3,4]. O facto de possuir o seu próprio genoma conserva a independência genética deste organelo relativamente à restante célula.

Várias evidências suportam esta teoria. Dentro destas destacam-se algumas como a existência de seres eucariotas unicelulares que servem como hospedeiros de bactérias, estabelecendo uma relação endossimbiótica, e o facto de as mitocôndrias terem a capacidade de replicarem o seu DNA, possuindo também um tamanho semelhante às bactérias atuais.

1.3. DNA Mitocondrial

O genoma humano é constituído pelo DNA mitocondrial (DNAMt) e pelo DNA nuclear (DNAn)

O DNAn está organizado em 23 cromossomas e encontra-se localizado no núcleo das células, ao passo que o DNAMt é constituído por uma molécula circular e está localizado na mitocôndria. A tabela 1.1 resume as diferenças entre o DNAMt e o DNAn.

Tabela 1.1: Comparação das características entre o DNA nuclear e o DNA mitocondrial adaptado de [5]

Características	DNA nuclear	DNA mitocondrial
Localização	Núcleo	Mitocôndria
Número de cópias por célula	2	Poderá ser superior a 1000
Funcionamento	Autónomo	Semiautónomo
Herança	Materna e Paterna	Materna
Recombinação	Sim	Não
Mecanismo de reparação	Sim	Não
Taxa de mutação	Baixa	Elevada (5 a 10x superior ao DNA nuclear)
Poder de individualização	Único para cada indivíduo com exceção gémeos monozigóticos	Não individualiza indivíduos da mesma linhagem materna
Sequencia de referência	Obtida em 2001 (Projeto genoma humano)	Obtida em 1981 (Anderson e colaboradores)

1.4. Genoma mitocondrial

O DNAMt foi identificado e isolado pela primeira vez em 1963 por Margit Nass and Sylvan Nass, com recurso à microscopia eletrónica [6]. Mas, apenas em 1981, a *Cambridge Reference Sequence* (CRS), a primeira sequência completa de DNAMt, foi publicada [7]. Em 1999, foi feita nova sequenciação da molécula de DNAMt usada para criar a CRS, tendo revelado a existência de dez erros de substituição e dois resíduos de citosina (C) nas posições 3 106 e 3 107, quando apenas deveria existir um e, assim, depois desta revisão, a CRS passou a ser denominada de *Revised Cambridge Reference Sequence* (rCRS)[8].

As diferenças entre nucleótidos, entre os perfis de DNAm quando comparados com a rCRS, são designados de haplótipos [9], sendo estes identificados pela base nucleotídica mutada e pelo número da posição alterada. Estes podem ser agrupados em haplogrupos, quando partilham certas mutações características de um haplogrupo que são detetadas com base em técnicas como RFLPs do DNAm total [10] ou com base em polimorfismos na região de controlo do DNAm [11–13]. Foi considerado que as pessoas que partilham um haplogrupo têm um antecessor comum [14]. De acordo com a nomenclatura, os haplogrupos são organizados por ordem alfabética tendo em conta o momento da sua descoberta [15], ou seja, os haplogrupos A, B, C e D foram os primeiros a serem identificados e assim sequencialmente.

O DNAm é uma molécula circular de cadeia dupla, com um comprimento de 16 560 pares de base (pb) e tem um peso de 10^7 Daltons [16], como se encontra representada na figura 1.1.

As duas cadeias que compõem o DNAm são a cadeia pesada (H) e uma cadeia complementar a esta, a cadeia leve (L); estas diferem entre si na composição, uma vez que a cadeia leve (L) é rica em pirimidinas (timinas e citosinas), de menor peso molecular, enquanto a cadeia pesada (H) é rica em purinas (guaninas e adeninas), de maior peso molecular.

O genoma mitocondrial pode ser dividido em duas regiões distintas, uma região codificante e uma região não codificante, também conhecida como D loop ou região de controlo.

A região codificante é aquela que possui maior dimensão, cerca de 15 435 pb [17] e possui 37 genes que codificam 13 proteínas, que estão envolvidas na oxidação fosforilativa[18], 22 RNAt e 2 RNAr[7]. A maioria das proteínas mitocondriais são codificadas por genes nucleares, sendo sintetizadas em ribossomas e transportados até à mitocondrial[19].

A D loop possui um tamanho de 1 122 pb, o que corresponde a cerca de 7% do DNAm total [17], e está contida entre as posições 16 024 e 576 do DNAm também codifica 8 RNAt e 1 polipéptido e está ainda envolvida na regulação da transcrição e replicação da molécula [7,17,20–22]. Esta região apesar de ser não codificante contém promotores para a cadeia leve (L) e para a cadeia pesada (H), elementos de regulação transcricional, locais de ligação de fatores de transcrição mitocondriais e a origem de replicação da cadeia pesada (O_H). Sendo também a região onde estão presentes os segmentos mais polimórficos do DNAm os segmentos Hipervariáveis, HVI, HVII e HVIII.

Reactive oxygen species (ROS), produzidas na cadeia transportadora de elétrons e que atacam o DNAm, que é bastante vulnerável a estes ataques, uma vez que não possui histonas que o possam proteger [1,19]. Outro fator que explica a elevada taxa de mutação do DNAm é a baixa fidelidade apresentada pela DNAm polimerase que, no processo de replicação, é menos eficiente no processo de detecção e correção de erros [4,27]. As mutações que ocorrem podem ser classificadas como substituições, deleções e inserções. Por seu lado, as substituições podem ser classificadas como transversões, quando uma purina é substituída por uma pirimidina ou vice-versa, já as transições ocorrem caso surja uma substituição entre purinas ou pirimidinas.

Ao contrário do DNA nuclear, o DNAm é herdado exclusivamente por via materna [28], ou seja, a sequência de DNAm de toda a linhagem materna é idêntica, uma vez que os oócitos possuem uma dimensão superior aos espermatozoides e conseqüentemente um maior número de mitocôndrias e também devido ao facto de a maioria das mitocôndrias presentes nos espermatozoides desaparecem durante a fase inicial da embriogénese, durante um processo denominado de ubiquitinação [24,29,30]. Esta característica torna-se especialmente útil em estudos forenses, de linhagens e na evolução de populações. No entanto, recentemente têm-se levantado algumas questões sobre o DNAm ser herdado exclusivamente pela via materna, especialmente por se ter apurado que num doente de 28 anos que apresentava intolerância extrema ao exercício, o seu DNAm foi sequenciado e constatou-se que era idêntico ao encontrado no seu pai e tio paterno [31], concluindo-se, deste modo, que o DNAm do seu músculo apresentava origem paternal, o que pôs em causa o que era assumido até então. Contudo, após o estudo de mais pessoas com doenças mitocondriais revelou-se que não havia evidência de existir transmissão paterna de DNAm [24,32,33].

Outra característica distintiva do DNAm é a sua falta de recombinação [34], o que faz com que as mutações ocorridas ao longo da vida sejam a única fonte de variabilidade do genoma mitocondrial e que seja possível fazer a associação do DNAm com um ancestral feminino comum [24,28–30]. Porém, recentemente esta evidência também foi posta em causa, designadamente com alguns relatos de recombinação mitocondrial [35,36]. Após mais estudos foi concluído que a molécula de DNAm tem uma recombinase funcional e pode, por isso, sofrer recombinação, mas as probabilidades de isso acontecer são ínfimas, algo que pode ser explicado pela ausência de mitocondriais de origem paterna, pois iria resultar, mesmo que ocorresse recombinação, uma molécula de DNAm não muito diferente da molécula original [24,37,38].

Estas características, em conjunto com a grande abundância de DNAm, já que cada mitocôndria possui entre a 4 a 5 destas moléculas e cada célula poderá possuir centenas de mitocôndrias, fazem deste um marcador genético muito útil.

1.6. Heteroplasmia

Apesar de a maioria da população apresentar apenas uma população de mitocôndrias, ou seja, é homoplásmico, existem alguns indivíduos que possuem diferentes populações de mitocôndrias que diferem na sua composição nucleotídica numa única mitocôndria, célula ou tecido, fenómeno que é designado de heteroplasmia [39,40]. É inclusive possível que um indivíduo possua uma heteroplasmia num tecido enquanto é homoplásmico noutro tecido [41]. Por outro lado, também é possível que um indivíduo apresente um tipo de heteroplasmia num tecido e outro tipo noutro tecido [41], sendo que as heteroplasmas são mais frequentemente

encontradas nas regiões HVI e HVII [42]. Alguns estudos reiteram que a probabilidade de ocorrerem heteroplasmias é superior nos tecidos com elevadas taxas metabólicas e quando o DNAm passa por *bottlenecks*, como por exemplo no processo de formação de cabelos (sendo este um local onde será esperado um número elevado de heteroplasmias) [14]. Também foi teorizado que com o aumento da idade a frequência de heteroplasmias presentes em cada indivíduo fosse aumentando[43]; contudo, mais tarde, esta teoria foi rejeitada com o estudo de 4 mulheres ao longo de 10 anos, ficando demonstrado que as heteroplasmias são transmitidas entre gerações e mantêm-se entre níveis semelhantes ao longo do tempo de vida de cada indivíduo[44].

Existem dois tipos de heteroplasmias, as heteroplasmias de comprimento (Figura 1.2) e as heteroplasmias de posição (Figura 1.3). As heteroplasmias de posição são mutações pontuais numa das sequências de DNAm, o que faz com que na mesma posição existam diferentes nucleótidos quando comparadas duas populações de mitocôndrias, o que resulta na sobreposição de dois picos nos eletroferograma obtidos após a sequenciação[45]. As heteroplasmias de comprimento são observadas normalmente em zonas de poli-C quando estes excedem as 8 citosinas (C) consecutivas [46], zonas homopoliméricas de citosinas (C), sendo estes locais hotspots mutacionais e ocorrem com especial frequência na zona HVI entre as posições 16184 e 16193, sendo que quando existe uma heteroplasmia de comprimento se observa a transição de uma timina para uma citosina (C) na posição 16189, enquanto que na zona HVII as heteroplasmias têm mais probabilidade de serem encontradas entre as posições 302 e 310, sendo na posição 310 a transição de uma timina por uma citosina (C) na posição 309, sinal de a existência de uma heteroplasmia de comprimento; por fim, as heteroplasmias de comprimento poderão ocorrer na região HVIII, apesar de apresentarem uma menor frequência, entre posições 568 e 573, sendo que a sua função biológica ainda é desconhecida [40,41].

Há, contudo, vários aspetos sobre as heteroplasmias que ainda permanecem desconhecidos e necessitam, por isso, de serem mais estudados, como os seus mecanismos e a probabilidade de ocorrerem consoante o sexo, a idade e o grupo populacional dos indivíduos [41]. Por essa razão, diferenças de heteroplasmias entre dois indivíduos não são evidência suficiente para excluir a hipótese de estes não pertencerem à mesma linhagem materna [47].

Figura 1.2: Representação de uma heteroplasmia de comprimento

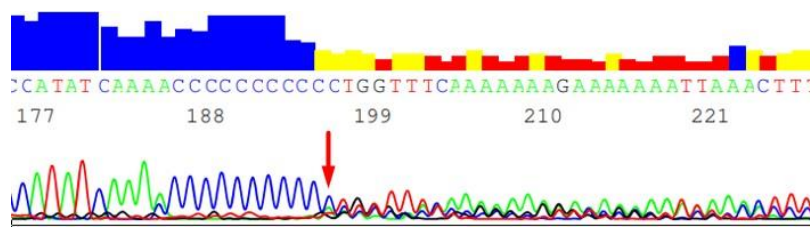
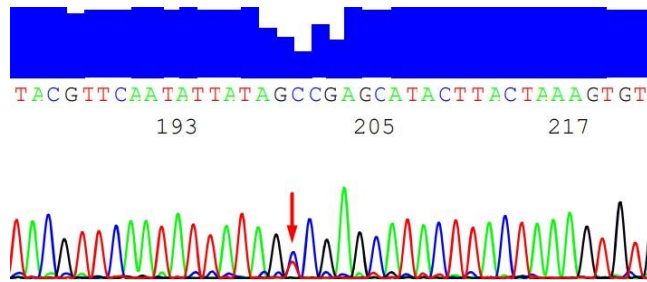


Figura 1.3: Representação de uma heteroplasmia de posição



1.6. Aplicações do estudo do DNAm

O DNAm é uma ferramenta essencial em diversas áreas, desde a antropologia molecular, a arqueologia, a genética populacional, a forense e a médica. Os estudos iniciais ao DNAm começaram a ser efetuados usando *Restriction Fragment Length Polymorphisms* (RFLPs) [11,48,49]. Mais tarde, com a descoberta do PCR foi possível iniciar a análise de produtos de PCR com RFLP e também com o recurso a técnicas de sequenciação efetuar a sequenciação dos mesmos.

O estudo do DNAm começou a ser executado apenas pela sequenciação da região HVI [50,51]. Tendo mais tarde o estudo e a sequenciação do DNAm sido estendida às regiões HVI e HVII[49] e posteriormente foi estendido à região controlo total [52], ou seja, ao conjunto dos segmentos HVI, HVII e HVIII, o que se tem tornado uma prática comum [53], uma vez que apenas o estudo das regiões da região HVI individualmente ou o estudo das regiões HVI e HVII podem não apresentar um grau de discriminação individual suficiente [42], mostrando-se mais precisa a atribuição de haplogrupos com a extensão do estudo à região controlo total do DNAm.

1.6.1. Aplicações forenses

Atualmente o método analítico mais utilizado que permite o estudo do DNA para a obtenção de perfis genéticos é o estudo de STR (*short tandem repeats*) que tem por base polimorfismos presentes no DNA nuclear.

Porém, quando o DNA nuclear não pode ser utilizado, pois poderá não existir em quantidade suficiente, por estar muito degradado (como em amostras de cabelo) ou por ser muito antigo, tornando impossível a extração e, conseqüentemente, não permitir traçar um perfil genético, pode recorrer-se ao DNAm [42,54]. Uma vez que este existe em grande abundância em todas as células e também o facto de apenas serem necessárias, apenas, 10 cópias de DNAm para produzir um perfil de DNAm [55].

Por este motivo, o DNAm torna-se num marcador genético bastante útil na identificação de linhagens maternas. Em casos de identificação de indivíduos é necessária uma amostra de comparação de um possível parente da mesma linhagem materna, ou seja, entre irmãos e descendentes maternos, de modo a que seja possível fazer a comparação de

haplogrupos com o indivíduo desconhecido, visto que estes iram apresentar o mesmo haplogrupo [56,57].

Apesar de útil, o estudo do DNAm_t apresenta algumas desvantagens dentro das quais se destacam o seu baixo poder de discriminação e individualização entre indivíduos que pertençam a mesma linha materna, já que estes irão partilhar o mesmo haplótipo, além de o facto de diferentes tecidos poderem apresentar diferentes heteroplasmias e a existência de heteroplasmias diferentes dentro da mesma linhagem materna que podem confundir as análises [58].

Um dos casos práticos onde se recorreu à análise do DNAm_t foi na análise de uma cova comum encontrada na Polónia, em 2012, após o período comunista polaco e no encobrimento de vítimas que era frequente neste regime, que devastou o País, tendo como resultado a morte de mais de 50 000 pessoas. Através da sequenciação direta das regiões HVI e HVII foi possível identificar cerca de 50 vítimas encontradas nesta cova comum [59].

1.6.2. Genética Populacional

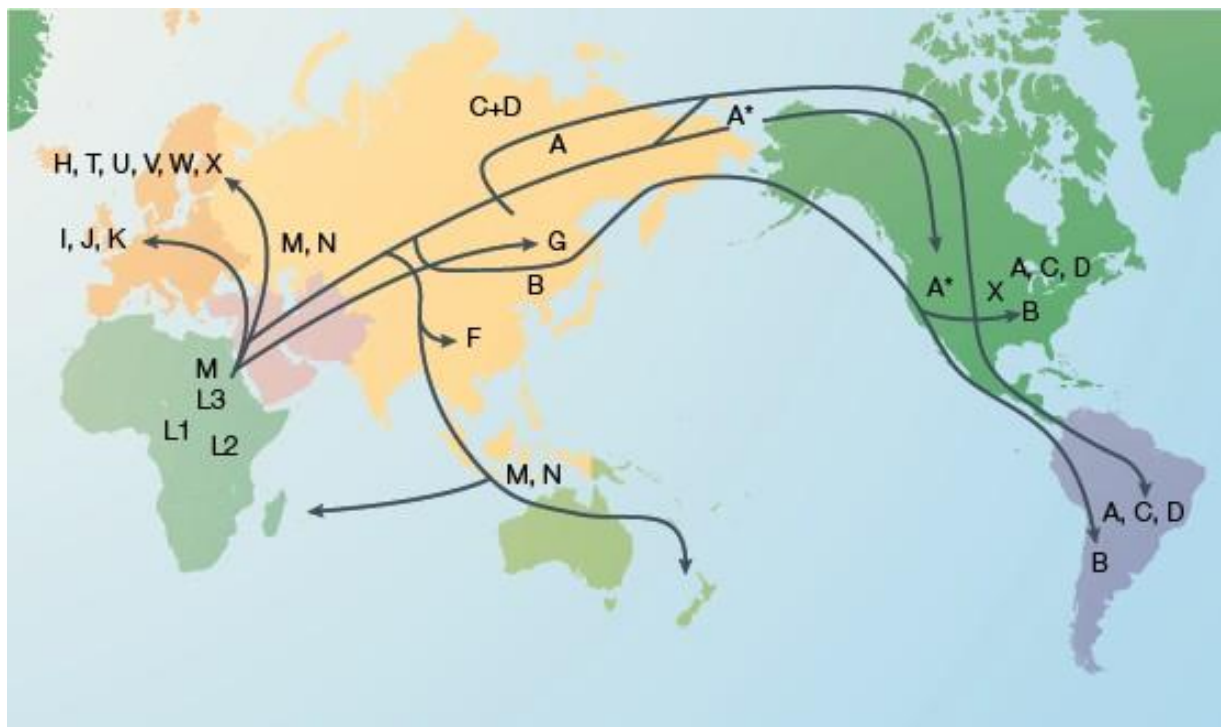
O DNAm_t é um “registo molecular”[60] das migrações do ser humano moderno, devido às características específicas, tais como apresentar uma maior taxa de mutação quando comparado com o DNA nuclear e ser estritamente transmitido por via materna, sendo por isso um marcador haploide [42], e pelo facto de não sofrer recombinação, o que faz com que os polimorfismos ocorridos sejam específicos de uma linhagem materna. De modo a estudar estas migrações recorre-se à comparação de haplogrupos entre as diferentes populações, uma vez que cada população terá o seu conjunto de polimorfismos específicos e consequentemente um haplogrupo distinto [24,60].

A hipótese *out of Africa* é atualmente a teoria mais aceite sobre a origem do homem moderno, pois esta indica que ele teve origem em África e foi a partir deste continente que se dispersou pelo mundo [61] [11]. Através da análise do DNAm_t de 147 pessoas não relacionadas de várias partes do mundo foi possível corroborar esta hipótese e chegar à existência de uma árvore filogenética bem como a uma ancestral feminina de origem em África, que ficou conhecida como Eva mitocondrial, sendo que a partir dela foi possível, através de estudos filogenéticos, identificar haplogrupos específicos de populações e assim identificar os processos migratórios que ocorreram [11].

Todas as mutações do DNAm_t podem ser representadas por uma árvore filogenética [62], sendo a *PhyloTree* [63] a ferramenta mais frequentemente usada neste processo [63]. Esta contém todos os polimorfismos da região de controlo e da região codificante do DNAm_t, algo muito útil pois existem alguns haplogrupos que não podem ser unicamente atribuídos quando nos baseamos somente na análise da região de controlo [63]. Esta apresenta na sua base todos os haplogrupos característicos de populações africanas, como o macrohaplogrupo L que se divide nos subhaplogrupos L0 L1, L2, L3, L4, L5 e L6 sendo estes os precursores dos restantes haplogrupos que se formaram [64]. O haplogrupo L0 é o mais antigo, sendo reiterado que se formou há cerca de 140 a 160 mil anos[64]. Por outro lado, o haplogrupo L3 ter-se-á dividido mais tarde nos haplogrupos M e N (figura 4), que são os haplogrupos que foram formados pela migração para fora de África, e são característicos da população euroasiática [65,66]. Por sua vez, a população da Europa partilha 9 haplogrupos (H, I, J, K, T, U, V, W, X) que derivam do

haplogrupo N (figura 1.4), enquanto a população sul americana apresenta como haplogrupos característicos os haplogrupos A, B, C, D, que tiveram origem na Ásia [65].

Figura 1.4: Representação das migrações do Homem moderno e respetiva distribuição geográfica dos haplogrupos (adaptado de [66])



2.6.3. Genética médica

O estudo do DNAm_t também tem o seu impacto na área clínica, uma vez que várias patologias associadas a diversas áreas, como a neurológica, a cardíaca e a respiratória estão associadas a mutações naquele[67]. Além disso, diversos estudos estão a ser efetuados tendo por base a possibilidade de certos haplótipos terem uma maior predisposição para o desenvolvimento de doenças como Parkinson ou Alzheimer[67].

Os dados da literatura acerca dos haplogrupos de risco (*risk haplogroups*) são controversos e dependem da patologia que está a ser estudada. Há estudos que mostram que há associação entre haplogrupos específicos e cancro de mama [68], glioma [69] e osteoartrite [70].

1.7. Base de Dados

Com o aumento dos estudos de DNAMt foi necessário criar bases de dados, de modo a armazenar os dados resultantes destes estudos e também uma forma de controlo de qualidade de todos os estudos realizados, uma vez que muitos dos estudos efetuados até a criação destas bases de dados apresentavam erros [71]. Para serem aceites nestas bases de dados os resultados dos estudos têm de possuir índices que garantam a sua qualidade, situação que levará à mitigação da ocorrência de erros inerentes à análise. Estas bases de dados também permitem determinar a frequência de haplótipos em cada população e avaliar a sua raridade [42][], algo necessário no processo de identificação de indivíduos [72].

Uma destas bases de dados internacionais mais conceituada e mais utilizada na recolha de haplótipos de DNAMt é a EMPOP (*EDNAP mitochondrial DNA population*) que recolhe dados da região controlo total, como também da genoma mitocondrial total [71]. Atualmente, a EMPOP armazena um total de 38361 haplótipos [71]. Esta base de dados tem como principal objetivo a recolha de sequências garantindo a sua qualidade, sendo que com a recolha de dados de sequências de diversas partes do mundo é possível realizar análises estatísticas e calcular parâmetros populacionais importantes para estudos de genética populacional e no âmbito forense.

Também a base de dados SWGDAM (*Scientific Working Group for DNA Analysis Methods*), gerida pelo FBI possui dados de sequências de DNAMt de diversos grupos e etnias tornando-se por isso muito útil.

Outra base de dados importante no armazenamento de dados recorrentes do estudo do DNAMt é a Mitomap [73]. Esta contém os polimorfismos característicos de acordo com a origem geográfica das populações e faz a sua ligação com doenças degenerativas que estão associadas ao DNAMt, afetando principalmente o sistema nervoso central, o coração, os músculos e os sistemas endócrino e urinário [42,73,74]. Esta base de dados torna-se assim essencial tanto para investigadores como para médicos e geneticistas [75]. A Mitomap contém as sequências de DNAMt de pessoas saudáveis e doentes e permite recolher informações sobre o loci e qual o nucleótido que se alterou [42], sendo assim possível ver qual a mutação responsável pela doença. Deste modo, também será possível fazer o cálculo da frequência destas mutações na população presente na base de dados [42].

1.8. Contexto migratório que deu origem à população brasileira atual

Atualmente a comunidade brasileira residente em Portugal é, a comunidade imigrante, que apresenta um maior número de indivíduos, 184 000 [76].

A população brasileira é uma das mais heterógenas a nível mundial, muito devido ao facto de este país ter sido uma colónia portuguesa, o que permitiu a miscigenação do povo

nativo brasileiro com o povo português, mas também com escravos vindos de África. Após o momento da descoberta do Brasil pelo povo português, no ano de 1500, cerca de 500 000 colonos portugueses ali se fixaram [77–80].

A população africana começou por ser introduzida no Brasil como escravos para trabalharem nas minas de diamantes e nas plantações de café, estimando-se que cerca de 4 milhões de africanos foram movidos para o Brasil [78,79].

A partir do momento da independência do Brasil, este País recebeu uma grande afluência de migrantes de todo o mundo, de onde se destacam primeiramente os imigrantes portugueses, italianos e espanhóis [78,79]. Todos estes dados vêm comprovar a grande heterogeneidade do povo brasileiro.

Atualmente a comunidade brasileira residente em Portugal é a que apresenta um maior número de indivíduos

2. Objetivos

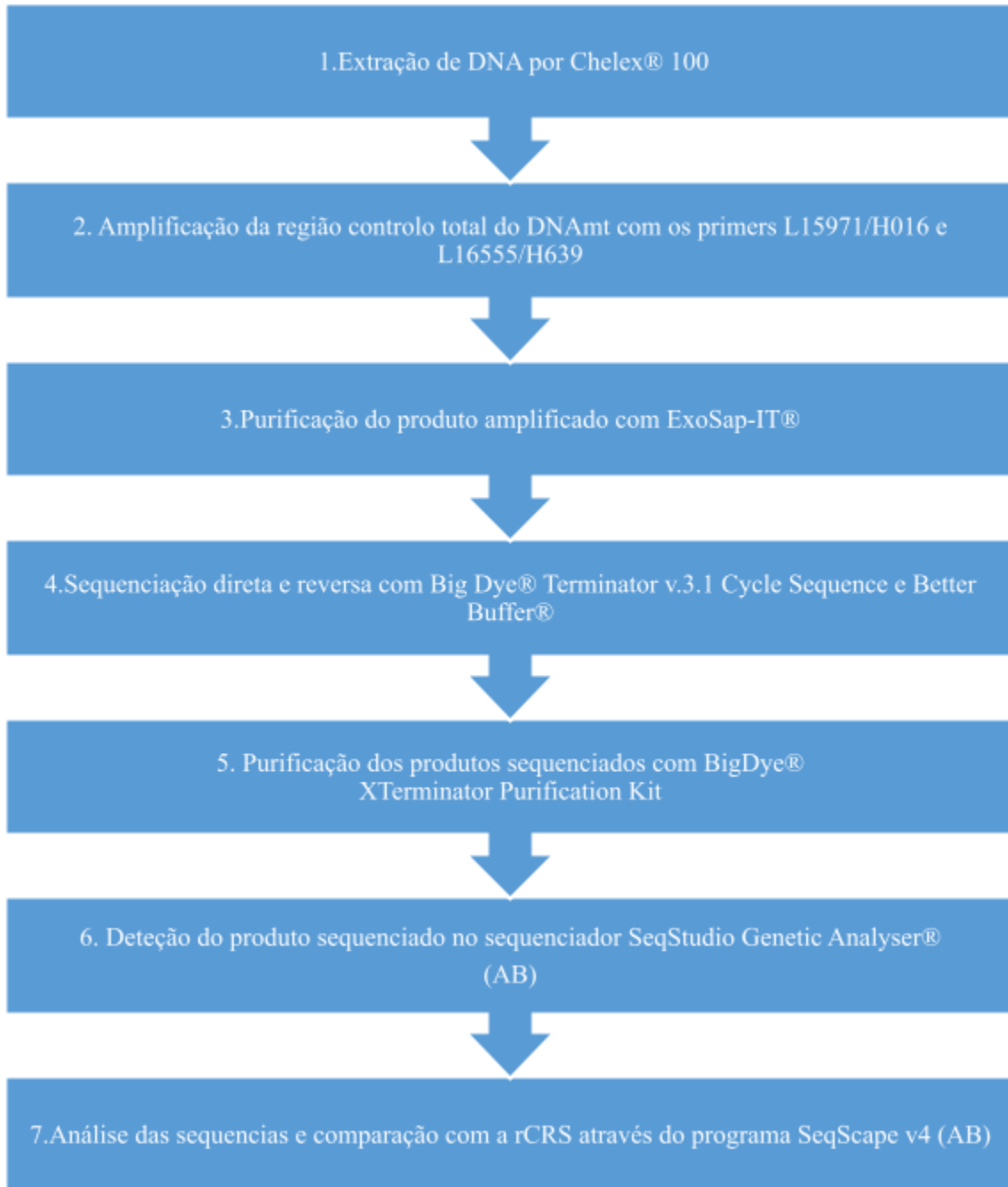
O principal objetivo deste trabalho é o apuramento da heterogeneidade genética presente na população de origem brasileira residente em Lisboa, através da caracterização, recorrendo à análise da região controlo total do DNAm, de modo que seja possível perceber o impacto que esta população irá ter na genética da população residente atualmente em Lisboa. Assim, tendo isso em conta, foram estabelecidos os seguintes objetivos:

- Determinar os haplótipos correspondentes a cada indivíduo e o seu respetivo haplogrupo;
- Identificar com base nos haplogrupos a origem geográfica de cada indivíduo;
- Compreender o impacto que a introdução da população de imigrantes brasileiros a residir em Lisboa terá no pool genética da população atualmente a residir em Lisboa;
- Publicar os haplogrupos identificados na base de dados EMPOP, a principal base de dados de DNAm a nível forense usada mundialmente.

3. Materiais e métodos

As várias etapas do procedimento experimental realizado estão representadas na figura 3.1.

Figura 3.1: Fluxograma do procedimento experimental realizado



3.1. Seleção das Amostras

Neste estudo foram analisadas 75 amostras de sangue de imigrantes brasileiros, não relacionados entre si, que residem atualmente na região de Lisboa. As amostras estudadas foram provenientes de investigações de parentesco biológico realizadas anteriormente, que decorreram no Serviço de Genética e Biologia Forenses (SGBF) da Delegação Sul do Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF – DS) até 2019, tendo sido selecionadas aleatoriamente.

3.2. Extração de DNA

A extração de DNA a partir de manchas de sangue foi realizada utilizando o método de extração Chelex® 100 (Bio-Rad), técnica desenvolvida em 1991 por P. Sean Walsh[81]. Esta técnica tem como base a função de uma resina, o Chelex que se liga a iões metálicos como o magnésio e quando submetido a altas temperaturas origina a lise de células dando-se assim a libertação do DNA, que fica protegido da ação das nucleases, inativando-as com a remoção dos iões de magnésio.

Este procedimento tem como primeiro passo para cada amostra o recorte de um círculo a partir de uma mancha de sangue, com o auxílio de um *paper punch single hole*. O fragmento recortado foi colocado num tubo *eppendorf* de 1,5 mililitro (mL) e foi adicionado a este 1 mL de água mili-Q. Os tubos foram de seguida encubados à temperatura ambiente, durante cerca de 20 minutos, sofrendo posteriormente uma agitação ligeira no vórtex, durante cerca de 5 a 10 segundos. Após a incubação as amostras foram centrifugadas a 14 000 rotações por minuto (rpm), durante 3 minutos, a 20°C. E recorrendo a uma micropipeta retirou-se o sobrenadante a cada tubo, aproximadamente 980 µL, e adicionou-se 180 µL de Chelex® 100 a 5%. De seguida, colocaram--se os tubos num banho de água aquecido, a 56°C, durante 15 minutos.

Posteriormente, os tubos foram sujeitos a nova agitação no vórtex, a alta velocidade, durante cerca de 15 segundos, e colocados em incubação a 100°C, durante 8 minutos. Por fim, os tubos foram novamente agitados no vórtex, a alta velocidade, e centrifugados a 14 000 rpm, durante 5 minutos, a 20°C. As amostras foram armazenadas a -20°C até serem usadas.

3.3. Amplificação do DNA

Para a amplificação da região controlo total do DNAMt, constituída pelas regiões HVI, HVII e HVIII, foram utilizados dois pares de *primers* acoplados o L15971/H016 e L16555/H639 (tabela 3.1), os *primers* tem esta denominação consoante a cadeia ao qual se ligam sendo os *primers* L integrados na cadeia leve e os *primers* H na cadeia pesada. Os *primers* L15971/H016 permitiram realizar a amplificação da zona entre as posições 16024 e 16569, já os *primers* L16555/H639 permitiram a realização da amplificação entre as posições 1 e 576 do DNAMt.

Tabela 3.1: Sequência dos *primers* usados na amplificação da região controle total do DNAm_t

L15997	5' CAC CAT TAG CAC CCA AAG CT 3'
H016	5' TGA TAG ACC TGT GAT CCA TCG TGA 3'
L16555	5' CCC ACA CGT TCC CCT TAA AT 3'
H639	5' GGG TGA TGT GAG CCC GTC TA 3'

Para a amplificação foi preparada uma solução de volume final 10 μ L contendo Multiplex PCR Master Mix ®(Qiagen), uma solução que contém *HotStarTaq Plus* DNA Polymerase, PCR Buffer com MgCl₂ e dNTPs - desoxinucleótidos trifosfatados, água ultrapura e o conjunto de *primers* L15971/H016 e L16555/H639 a uma concentração de 2 μ M. Os volumes usados na reação estão descritos na figura 3.2

Tabela 3.2: Volumes usados na reação de amplificação da região controle total do DNAm_t

Reagente	Volume (μL)
Multiplex PCR Master Mix	5
Água mili-Q	3
Primers a 2 μ M	1
DNA	1
Total	10

Para cada reação foi preparado um controle negativo para cada região a amplificar, num conjunto de 2 controles negativos aos quais foi adicionado 1 μ L de água em detrimento de DNA. A reação de amplificação foi efetuada num termociclador GeneAmp PCR system 9700 (AB) de acordo com as temperaturas representadas na tabela 3.3, durante 35 ciclos.

Tabela 3.3: Condições de temperatura para a amplificação da região controle do DNAm_t

1 ciclo	35 ciclos				1 ciclo	
94 °C	94 °C	60 °C	60 °C	72 °C	72 °C	4 °C

5 min	30 seg	90 seg	90 seg	60 seg	10 min	∞
--------------	--------	--------	--------	--------	--------	----------

3.4. Purificação dos Produtos Amplificados

De modo a retirar possíveis resíduos de elementos indesejados resultantes da amplificação foi necessário realizar uma purificação dos produtos amplificados. Esta foi realizada com o reagente ExoSap-IT (*Affymetrix*) que com as suas duas enzimas hidrolíticas removeu *primers* residuais e dNTPs não utilizados durante o processo de amplificação.

Esta técnica consiste na adição de 2 μL de ExoSAP-TI® por cada 5 μL de produto amplificado. Sendo as amostras posteriormente colocadas num termociclador GeneAmp PCR system 9700 (AB), seguindo o programa indicado na tabela 3.4.

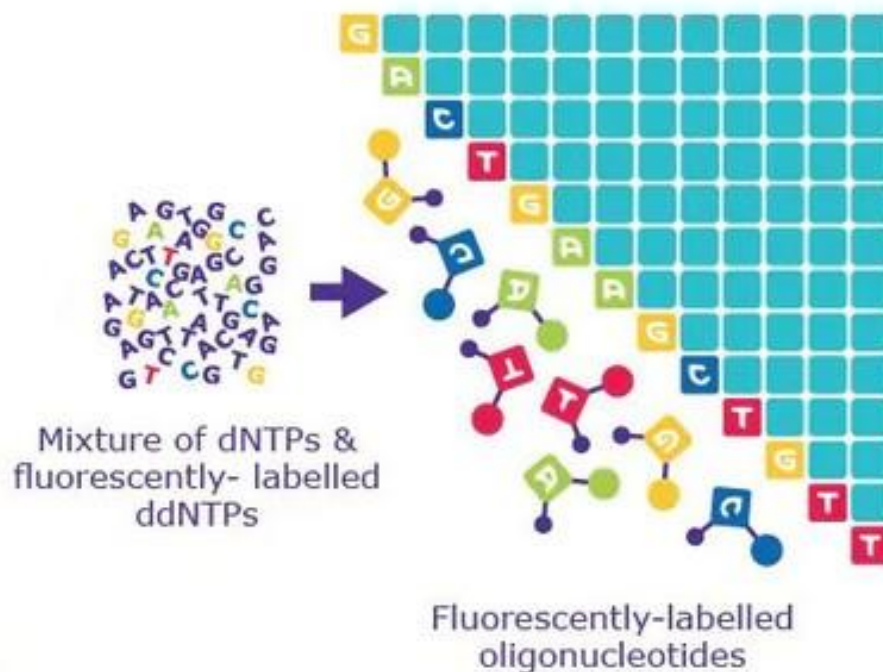
Tabela 3.4: Condições de temperatura para a purificação dos produtos amplificados

1 ciclo	
37 °C	80 °C
15 min	15 min

3.5. Sequenciação

A sequenciação permite determinar quais as bases nucleotídicas de cada posição, neste caso dentro da região controlo do DNAm. Sendo a técnica de sequenciação usada neste estudo a sequenciação de Sanger [82]. Esta baseia-se na incorporação de didesoxinucleotídeos trifosfatados (ddNTPs) na extremidade 3', do DNA presente nas amostras, o que resulta na terminação da síntese da cadeia na posição onde ocorre a ligação destes. Os ddNTPs apresentam na sua composição fluorocromos, com diferentes cores, que permitem, após a análise das sequências, a sua deteção e identificação. Sendo que o ddATP se encontra marcado a verde, o ddCTP a azul, o ddGTP marcado a amarelo, mas visualizado a preto, e o ddTTP marcado a vermelho (figura 3.2).

Figura 3.2: Esquema representativo da sequenciação com ddNTPs marcados com fluorescência, adaptado de [83].



Para a realização desta etapa foi necessário preparar quatro misturas, uma por cada *primer* (L15971, L16555, H016 e H639). Sendo cada uma destas constituída por Better Buffer® (Microzone Ltd.) que é um tampão, BigDye® Terminator v.3.1 Cycle Sequence (AB), uma solução que contém DNA polimerase, MgCl₂, dNTPs e ddNTPs marcados com Dye Terminator, e do respetivo *primer* a 2,5 mM, num volume total de 4 µL que foram distribuídas pelos poços de uma placa de 96 poços, sendo, por fim, adicionado 1 µL de DNA das amostras previamente amplificadas e purificadas, estando os volumes necessárias para esta etapa descritos na tabela 3.5. O processo de sequenciação da região foi realizado num termociclador GeneAmp PCR system 9700 (AB) programado para as temperaturas e os tempos descritos na tabela 3.6.

Tabela 3.5: Reagentes e volumes utilizados na sequenciação da região controlo total do DNAMt

Reagentes	Volume (µL)
Better Buffer	2
BigDye® Terminator v.3.1 Cycle Sequence (AB)	1
Primers a 2,5µM	1
DNA	1
Total	5

Tabela 3.6: Condições de temperatura para a sequenciação da região controlo do DNAm

1 ciclo	35 ciclos			1 ciclo	
96 °C	96 °C	50 °C	60 °C	60 °C	4 °C
2 min	15 seg	9 seg	10 min	10 min	∞

3.6. Purificação dos Produtos Sequenciados

A etapa de purificação foi realizada com BigDye® XTerminator Purification Kit (AB), que contém dois reagentes, o XTerminator Solution e a SAM TM Solution. Este passo permite a remoção do excesso de dNTPs e ddNTPs não incorporados durante a reação de sequenciação, excesso de *primers* e sais livres. Foi, então, preparada uma mistura de acordo com os volumes representados na tabela 3.7, sendo distribuídos 50 µL desta mistura por cada poço utilizado da placa de 96 poços. Por fim, a placa é colocada num agitador durante aproximadamente 20 minutos e é centrifugada a 2800 rpm, durante 2 minutos.

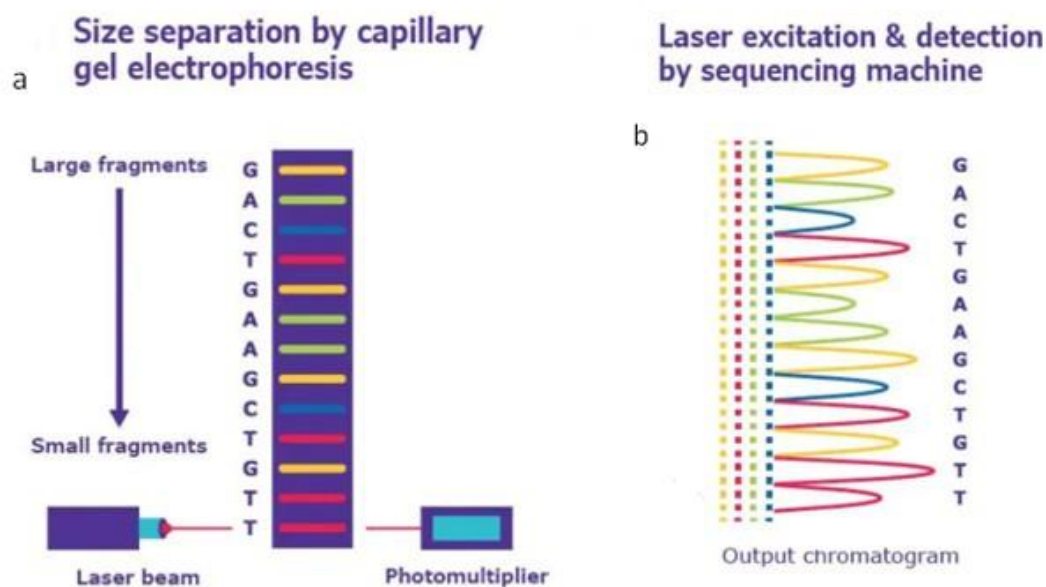
Tabela 3.7: Volumes utilizados no processo de purificação dos produtos sequenciados

Reagente	Volume (µL)
San Solution	16,4
Xterminator Solution	3,7
Água	30
Total	50,1

3.7. Detecção dos Produtos sequenciados

Após purificação dos produtos sequenciados, as amostras foram colocadas num sequenciador automático, SeqStudio (AB) de 4 capilares, de modo que a eletroforese capilar pudesse decorrer. Durante este processo, os fragmentos migraram do polo negativo (cátodo) até ao polo positivo (ânodo), de acordo com os seus pesos moleculares sendo que os fragmentos de menor peso molecular irão migrar a uma velocidade maior quando comparados com os fragmentos que apresentam um maior peso molecular. Quando estes se encontram no processo de migração é lhes incidido um feixe de laser que excita os eletrões presentes nos fluorocromos dos ddNTPs, o que origina a emissão de fluorescência num comprimento de onda conhecido e específico para cada ddNTP, este processo está descrito na figura 3.2. Os resultados deste processo são obtidos num eletroferograma, onde estão diferenciadas as bases nucleotídicas presentes em cada posição.

Figura 3.2: Esquema representativo do processo de deteção dos produtos sequenciados, adaptado de [83]



3.8. Análise dos produtos Sequenciados

A análise dos resultados da sequenciação da região controlo do DNAm foi efetuada com auxílio a dois programas, o Sequencing Analysis v.5.2, que avalia a qualidade das sequências obtidas, e o SeqScape v.4 (AB), que permitiu o alinhamento e a posterior comparação das sequências obtidas com a rCRS, permitindo a identificação dos polimorfismos presentes nas amostras, sendo que o conjunto destes se designam de haplótipo. As sequências apresentadas neste programa seguem as guidelines da European DNA Profiling Group (EDNAP) [84] e as bases nucleotídicas estão codificadas de acordo com o código descrito na tabela 3.8.

Tabela 3.8: Código IUPAC utilizado na nomenclatura das bases nucleotídicas

Código IUPAC	Base Correspondente
A	Adenina
C	Citosina
G	Guanina
T	Timina
B	Citosina, Guanina ou Timina

D	Adenina, Guanina ou Timina
H	Adenina, Citosina ou Timina
R	Adenina ou Timina
Y	Citosina ou Timina
K	Guanina ou Timina
M	Adenina ou Citosina
N	Qualquer base
V	Adenina, Citosina ou Guanina
S	Guanina ou Citosina
W	Adenina ou Timina

É possível ocorrerem 3 fenômenos diferentes que levam a alteração da sequência de DNAMt, transições, quando ocorre uma substituição entre purinas (A e G) ou entre pirimidinas (C e T), transversões quando existe uma substituição de uma purina por uma pirimidina e vice-versa, ou inserções e por fim deleções. As substituições são identificadas e registradas, anotando-se a posição que foi alterada seguida da base nucleotídica agora presente na sequência. Enquanto as inserções são identificadas pela posição nucleotídica imediatamente anterior à alteração, seguida de um ponto e do número um (.1) e do nucleótido inserido, para a primeira inserção, de um ponto e do número dois (.2) e do nucleótido inserido, na segunda inserção, e assim sucessivamente. As deleções são identificadas pela posição do nucleótido deletado relativamente à rCRS, seguida do sufixo “d”, “DEL” ou “-”.

Após a determinação dos haplótipos de cada amostra recorrendo à *PhyloTree* [63] foi possível corresponder estes aos seus respetivos haplogrupos.

4. Resultados e Discussão

4.1 Estudo dos Haplótipos

A partir das 75 amostras analisadas, foram obtidos 69 haplótipos únicos e 6 haplótipos duplicados. O que demonstra o fraco poder de discriminação individual do DNAMt, uma vez que existem indivíduos que não são relacionados, não partilhando, assim, a mesma linha materna e apresentam o mesmo haplótipo. A sua utilização como marcador depende da frequência de cada haplótipo na população, sendo que a identificação de um haplótipo pouco frequente pode contribuir com mais informação para a investigação em curso.

Todos os haplótipos obtidos foram submetidos para pesquisa na base de dados EMPOP, de modo que fosse possível obter uma frequência destes entre as 38361 sequências presentes nesta base de dados. Tendo sido possível averiguar que, dos haplótipos obtidos, 32 não apresentavam qualquer correspondência, tendo por isso uma frequência de 0.

Com base na comparação com a rCRS, foram possíveis identificar 1050 polimorfismos dentro de 139 posições. Dos polimorfismos encontrados, as substituições foram a forma que apresentou uma maior prevalência, tendo sido encontradas em 885 ocasiões, das quais 810 correspondem a transições, 465 entre pirimidinas, 361 entre purinas e 75 transversões. O que contrasta com o menor número de deleções (71) e inserções (107) encontradas.

Tabela 4.1: Distribuições dos polimorfismos observados na região controlo total do DNAmT de imigrantes brasileiros residentes em Lisboa

Alterações	HVI	HVII	HVIII	Fora HV	TOTAL
Número total de polimorfismos	345	453	118	150	1050
Número de posições polimórficas	59	54	11	15	139
Transições	318	316	42	134	810
Pirimidinas	282	119	16	48	465
C→T	177	32	2	4	215
T→C	105	87	14	44	250
Purinas	36	197	26	102	361
A→G	25	182	10	69	286
G→A	11	15	16	17	59
Transversões	26	16	17	16	75
C → G	4	0	0	0	4
A → T	1	2	0	0	3
A → C	21	4	1	0	26
G → T	0	3	0	0	3
C→A	0	4	0	0	4
G→A	0	0	16	16	32
T→G	0	1	0	0	1
G→C	0	2	0	0	2
Deleções	0	25	46	0	71
A	0	25	23	0	48
C	0	0	23	0	23
Inserções	0	95	12	0	107
C	0	95	6	0	101
A	0	0	6	0	6
Heteroplasmia de posição	1	1	1	0	3
M (A e C)	1	0	0	0	1

Y (A e G)	0	1	0	0	1
R (T e C)	0	0	1	0	1

Foi também possível identificar que a região HVI, que se encontra entre as posições 16024 e 16365, da região controlo do DNAMt, apresenta 345 polimorfismos, sendo dentro desta região a posição 16223 que apresenta um maior número de substituições, tendo sofrido alterações em 51 das 75 sequências estudadas, passando de uma Citosina (C) para uma Timina (T). Também será de destacar a elevada frequência de transições encontradas (24) de uma T por uma C na posição 16278, a transição de uma C por T, em 13 das sequências analisadas e a substituição de uma C para uma T na posição 16294 em 16 vezes e ainda a substituição em 11 sequências de uma C por uma T nas posições 16325 e 16362 e a substituição de uma C para uma T na posição 16294 16 vezes. É possível denotar a ausência de inserções e de deleções nesta região do DNAMt.

Dentro da região HVI ainda se realça o intervalo entre as posições 16182 e 16192 por apresentar uma zona de poli C, que tem como consequência a existência de 70 polimorfismos nesta região. Sendo a sua maioria localizadas na posição 16189, onde em 32 sequências das sequências analisadas ocorreu a alteração de uma Timina (T) por uma Citosina (C), enquanto as restantes alterações que apresentam uma maior frequência são de transições de uma C para uma T, sendo a posição 16187 aquela que apresenta uma maior frequência destas alterações, 7.

Também é nesta região de poli C que existe a maioria das transversões encontradas em HVI. As transversões que apresentam uma maior frequência encontram-se nesta zona, onde ocorre a substituição em 15 das sequências analisadas na posição 16182 de uma Adenina (A) por uma Citosina (C) e na posição 16183 onde também existe a substituição de uma Adenina (A) por uma Citosina (C) na posição 16182 em 6 sequências.

Na região HVII, que se encontra compreendida entre as posições 73 e 340, destaca-se por ser a região hipervariável que apresenta a maior frequência de polimorfismos, 453. Sendo aqueles que apresentam uma maior frequência transições de uma Adenina (A) para uma Guanina (G), como é o caso na posição 73, onde em 66 das 75 sequências das estudadas e na posição 263 onde existem 72 substituições de uma A para uma G. Também será de realçar a presença de uma substituição na posição 152 de uma T para uma C em 32 das sequências analisadas, a transição, na posição 195, de uma T para uma C em 25 das sequências e a substituição na posição 146 de uma timina (T) para uma Citosina (C) em 15 sequências e na posição 150 onde se verificou a existência de uma transição de uma Timina (T) para uma Citosina (C) em 18 sequências.

Esta região hipervariável também se realça por ser aquela que apresenta uma maior frequência de inserções, dentro das amostras analisadas, estando estas localizadas entre as posições 310 e 315; esta zona, caracterizada por ser uma região poli C, possui a repetição de 5 Citosinas (C). Encontram-se, assim, 95 inserções na posição 315, passando a nomenclatura desta posição a ser 315.1C e 26 inserções na posição 309, tendo em 2 amostras se verificado a inserção de 4 citosinas (C) nesta posição ficando, por isso, de acordo com a nomenclatura denominadas como 309.1C, 309.2C, 309.3C e 309.4C.

Já na região HVIII que se encontra estendida entre as posições 438 e 574 encontram-se 118 polimorfismos, sendo assim a região que apresenta uma menor frequência de polimorfismos, transições (42), deleções (46) e inserções (26). As alterações mais vezes encontradas nesta região são a transição de uma Guanina (G) para uma Adenina (A) em 9

sequências na posição 499, a inserção na posição 524 de 6 Adeninas (A) e de 6 Citosinas (C) e as deleções nas posições 523 e 524, que sempre apareceram acopladas num número de 23 vezes totalizando as 46 deleções encontradas nesta região.

Também foram encontrados polimorfismos fora do intervalo pertencente à região controle do DNAmT que está compreendido entre as posições 1 a 72, 341 a 437 e 16365 a 16569, mas mesmo sendo esse o caso, este intervalo possui 150 polimorfismos, destacando-se aquele que apresenta uma maior frequência e que se encontra localizado na posição 16519, onde houve uma transição de uma Timina (T) para uma Citosina (C) em 42 das sequências estudadas. Também de notar as inserções presentes nesta zona tendo em 2 sequências existido a inserção de 4 citosinas (C) na posição 534 e em 2 sequências a inserção de 2 citosinas (C) nesta mesma posição.

Nenhuma das sequências estudadas está em total concordância com a rCRS. O haplótipo que apresenta se apresenta mais próximo da sequência de referência é o GENPOP 1002, possuindo 5 alterações. Por outro lado, o haplótipo que apresenta um maior número de polimorfismos é o GENPOP1266 e o GENPOP 1258, sendo que ambos apresentam 25 alterações.

Foi também possível a partir da análise de 75 amostras inferir e atribuir a cada uma destas um haplogrupo, tendo sido identificados 57 haplogrupos distintos. Esta atribuição tem por base a Phylotree, que atribui os diferentes polimorfismos encontrados no DNAmT aos seus respectivos haplogrupos e subhaplogrupos, ou seja, para que um haplótipo seja incorporado num determinado haplogrupo tem de possuir certos polimorfismos característicos deste haplogrupo, sendo por isso possível que dois indivíduos mesmo não possuindo um haplótipo igual pertencem ao mesmo haplogrupo desde que partilhem estes polimorfismos identificativos de um haplogrupo [63]. Cada haplogrupo é identificado com letras sendo a posterior divisão em subhaplogrupos identificada por números.

Os haplogrupos presentes na população (Figura 8) apresentam frequências diferentes, tendo a sua maioria sido atribuída ao macrohaplogrupo L 31 indivíduos (54,3%), macrohaplogrupo característico das populações africanas. Dentre deste macrohaplogrupo, o haplogrupo que apresentou uma maior frequência foi o haplogrupo L3, com 13 sequências (22,8%), seguindo-se o haplogrupo L2 (17,5%) e por fim o haplogrupo L1 (14%).

O haplogrupo L1 surgiu e começou a suas ramificações nos haplogrupos L1a, L1b, L1c, L1d, L1e, L1k, há cerca de 140-150 mil anos. A linhagem L1b está maioritariamente presente em populações do Ocidente africano, sendo também encontrada em cerca de 27% das populações afro-americanas que é resultante do tráfico de escravos[64]. Também a linhagem L1c está presente em populações da África central, apresentando, no entanto, uma elevada frequência em populações afroamericanas [64]. A linhagem L1c teve origem na África Central e ter-se-á distribuído para a costa atlântica estando atualmente presente nas populações de Angola e do Delta do Congo, 4 haplótipos obtidos enquadraram-se dentro deste haplogrupo [64]. O haplogrupo L2a é o mais encontrado na população africana, estando presente em todo o continente, e na população afro americana, sendo também o haplogrupo que apresenta uma maior frequência dentro da população estudada, com 12 amostras, pertencentes a este haplogrupo [64]. As restantes ramificações do haplogrupo L2, os haplogrupos L2b, L2c, 2 dos haplótipos estudados pertencem a este haplogrupo, e L2d circunscritos a população residente no Oeste e Centro Africano.

De entre as amostras estudadas o haplogrupo que apresenta uma maior representação é o macrohaplogrupo L, cerca de 54%, o que sendo este macrohaplogrupo característico das populações africanas mostra a grande proximidade da população brasileira em estudo com a população africana. Estes resultados são coincidentes com estudos anteriormente realizados na população residente no Brasil onde foi possível verificar a grande predominância dos haplogrupos L0, L1, L2, L3 e L4 nesta população[85–87].

Seguindo-se os haplogrupos característicos das populações nativas do continente americano, A, B, C e D[88]. Estes são encontrados em aproximadamente 21% da população estudada demonstrando alguma presença de ancestrais americanos nativos na população em estudo. Com efeito, apurou-se que estão presentes em 12 amostras estudadas.

Por fim, os restantes haplogrupos encontrados na população em estudo, os haplogrupos H, J, R, T, U e X, os quais estão presentes em cerca de 25% das amostras e são característicos das populações euroasiáticas. Esta frequência mais reduzida de haplogrupos euroasiáticos presentes nesta população demonstra a baixa influencia das linhagens femininas das populações europeias e asiáticas na população em estudo.

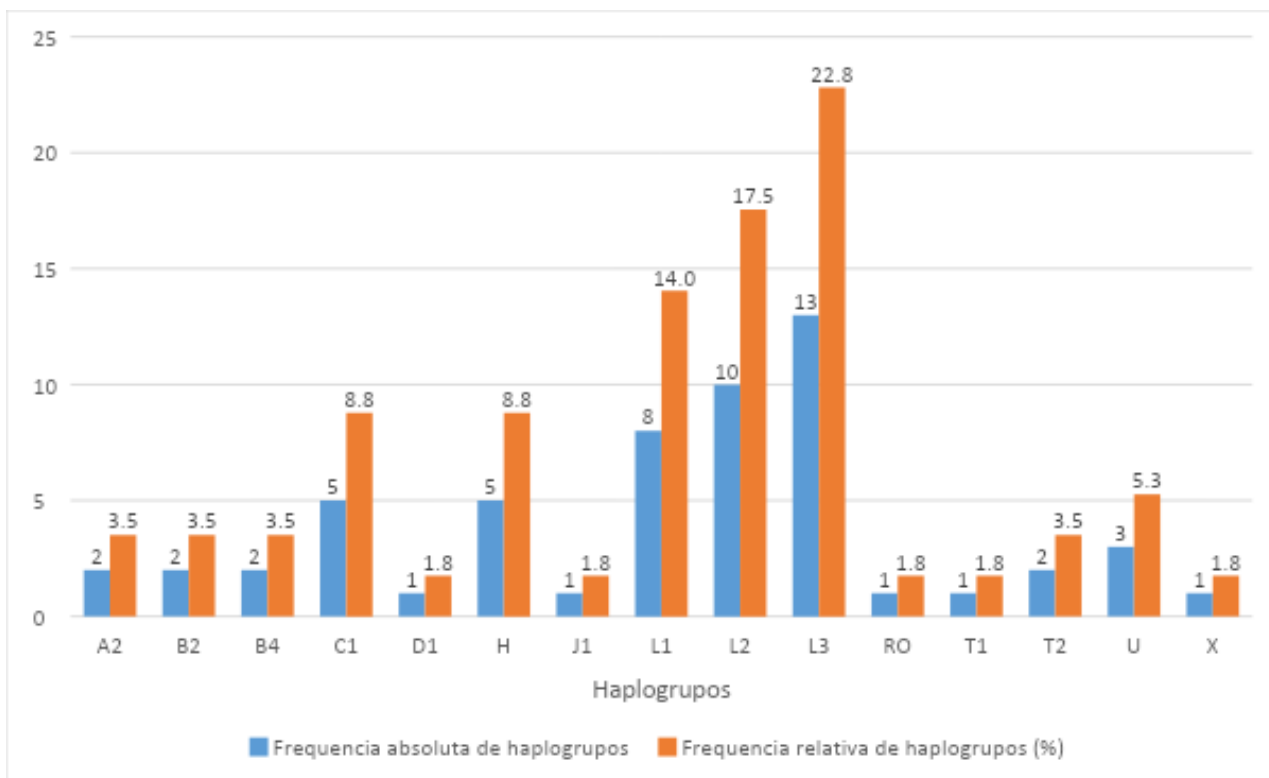
Os resultados obtidos, descritos na figura 4.1, confirmam a grande heterogeneidade apresentada por esta população resultante de entre outros fatores da colonização levada a cabo pelos portugueses a partir da descoberta do Brasil, no século XVI, que se manteve colónia até aos inícios do século XIX, data da sua independência. Foi encorajada a miscigenação entre os cerca de 500 000 colonos portugueses, que se estabeleceram no Brasil, com os povos nativos do território, o que com os dados obtidos, mais em concreto a pouca influência de haplótipos europeus presentes na população em estudo, é possível aferir que esta miscigenação se deu maioritariamente entre homens portugueses e mulheres indígenas. Como, posteriormente, o território começou a ser invadido por outros povos europeus nomeadamente por ingleses, por franceses e por holandeses e que a partir do século XX o Brasil tem vindo a receber migrantes de diversos países entre os quais se destacam a Espanha, a Itália, a Alemanha e o Japão. O que pode justificar os resultados obtidos, apesar de apresentarem uma frequência baixa, existe uma grande heterogeneidade de haplótipos caracteristicamente euroasiáticos na população em estudo.

Esta colonização teve como resultado a diminuição drástica do número de indivíduos indígenas, devido aos conflitos entre colonos e estes e também devido a introdução de novas doenças nestas populações por parte dos colonos cujos povos nativos desta região não possuíam defesas imunitárias para se defenderem destas enfermidades, o que é coerente com os resultados obtidos na pouca frequência de haplótipos nativo-americanos presentes na população em estudo.

A partir do século XVI, este período de colonização também ficou marcado pela introdução de escravos vindos de África para funcionarem como mão de obra nas plantações de açúcar e nas minas de ouro e diamantes. Tendo sido transportados para o Brasil cerca de 4 milhões de escravos africanos, este grande fluxo de pessoas levou a que os haplótipos característicos das populações africanas apresentassem uma grande frequência atualmente como é comprovado por este estudo.

Também foi possível verificar o impacto que a introdução desta população irá ter no pool genético da população atualmente residente em Lisboa que apresenta como haplogrupos maioritários haplogrupos característicos de populações europeias como os H, R e U[89], o que contrasta com a maioria de haplogrupos identificados na população em estudo.

Figura 4.1: Representação dos haplogrupos presentes na população brasileira imigrante em Lisboa e as suas respectivas frequências



4.3. Heteroplasmias

Com a análise das 75 sequências obtidas foi possível verificar a existência de 22 amostras heteroplásmicas, cerca de 29% das sequências obtidas, sendo que 18 destas apresentam heteroplasmias de comprimento, 16 encontravam-se na região HVI e 2 na região HVII que foram confirmadas pela sequenciação em duplicado da no sentido *forward* e *reverse*, estando também presente uma heteroplasmias em ambas as regiões HVI e HVII. As restantes heteroplasmias encontradas são heteroplasmias de posição 16183M, 185Y e 499R. A tabela 4.2 resume as heteroplasmias encontradas nas regiões hipervariáveis das diferentes amostras estudadas.

As 19 heteroplasmias de comprimento identificadas na região HVI tiveram todas como origem a transição de uma Timina (T) para uma citosina (C) na posição 16189, ficando nesta região uma sequência de 10 Citosinas(C) contínuas, o que resulta na inserção de diversas citosinas (C), causando a sobreposição de picos nos diversos eletroferogramas dificultando a análise dos mesmos.

Em contrapartida, as 3 heteroplasmias de comprimento, presentes na região HVII, tiveram como origem a introdução de citosinas (C) entre as posições 302 e 310.

Por fim, as 3 heteroplasmias de posição encontradas, uma na posição 16183, onde em algumas sequências existe uma transversão de uma Adenina (A) por uma Citosina (C), ao passo

que noutras esta não ocorreu, sendo esta posição designada como 16183M. Outra heteroplasmia de posição foi identificada na posição 185, onde algumas sequências apresentam uma transição para uma Guanina (G), enquanto outras uma transição para uma Citosina (C), sendo a sua designação 185Y. A última heteroplasmia de posição foi encontrada na posição 499, onde em algumas sequências existe uma transição de uma Guanina (G) para uma Adenina (A), enquanto em outras a Guanina (G) se mantém na sequência, sendo, então, esta posição designada de 499R.

Tabela 4.2: Heteroplasmias presentes na região controlo total do DNAMt da população brasileira imigrante em Lisboa

Haplótipo	Região Hiper variável	Frequência	Percentagem
Homoplásmicos	-	54	72,00%
	-	-	-
	HVI	16	21,33%
	HVII	2	2,67%
Heteroplásmicos	HVII	0	0,00%
	HVI+HVII	1	1,33%
	Posição	3	4,00%
	Total:	22	29,33%

4.4. Análise filogenética

Através dos haplótipos obtidos foi possível fazer a comparação entre os dados obtidos em estudos feitos anteriormente, de maneira a determinar a proximidade entre a população brasileira imigrante em Lisboa e a restantes (tabela 4.3). De modo a realizar essa comparação foi necessário recorrer à análise da variância molecular (AMOVA), de forma a que fosse possível averiguar a variabilidade genética entre as populações.

Tabela 4.3: Populações de estudos anteriores onde se analisou a região controlo do DNAMt, utilizadas para fins comparativos

Populações	Tamanho da amostra	Referência
Cabo Verde	103	[90]
Angola	129	[91]
Gana	193	[92]
Somália	190	[19]

AngolaLX	173	[93]
MoçambiqueLX	83	[94]
BrasilLX	75	Presente Estudo

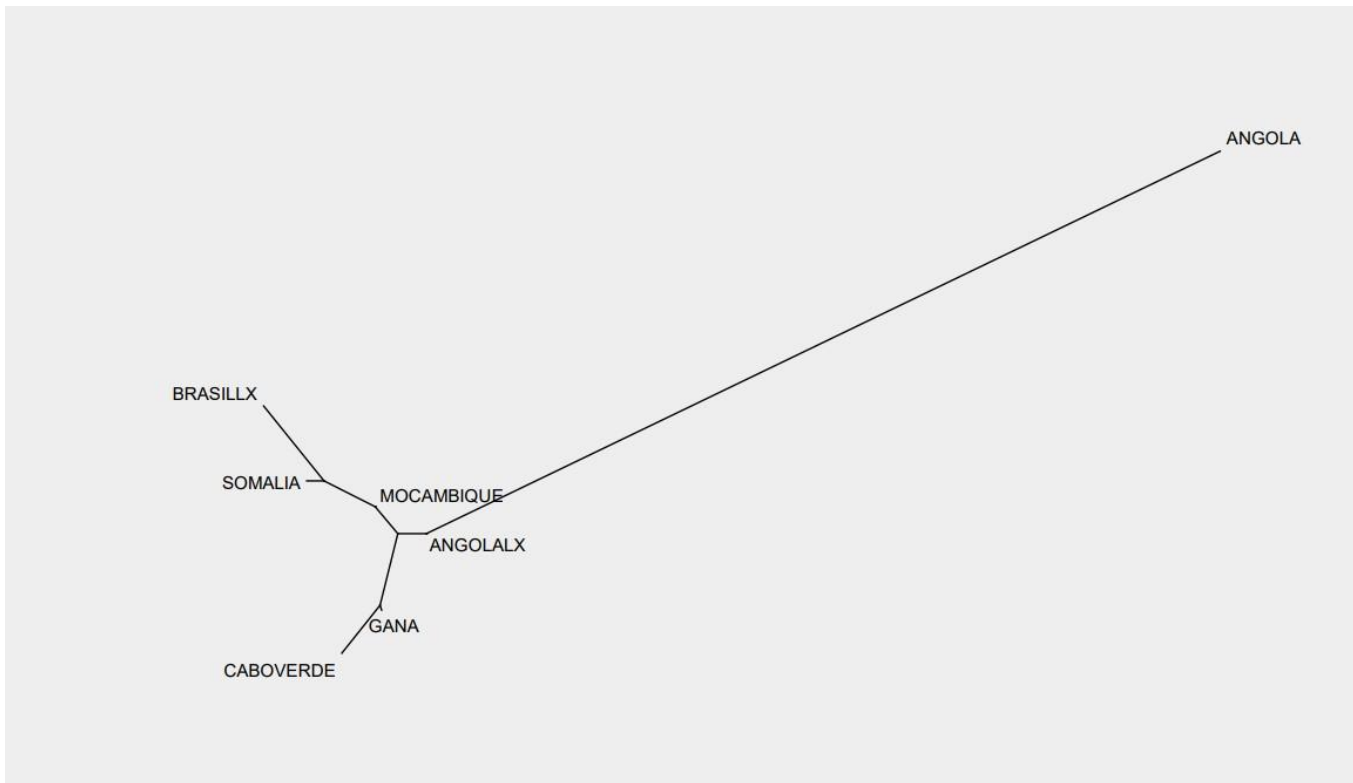
O programa Arlequin foi utilizado para calcular a AMOVA e determinar as distâncias genéticas entre as populações, obtendo, assim, valores do índice de fixação (Fst) (tabela 4.4) e os respetivos *pvalue* entre cada uma das populações que indicam o nível de significância dos dados calculados. Através dos dados de Fst obtidos e recorrendo ao programa *Neighbor Joining* pertencente ao Software Phylip v.3.695 foi possível obter as distâncias moleculares entre as diferentes populações e posterior representação destas numa árvore filogenética com recurso ao programa *TreeView* (figura 9).

Tabela 4.4: Valores de FST e p obtidos através do programa Arlequin para a totalidade da região controlo do DNAm

Populações	Cabo Verde	Angola	Gana	Somália	AngolaLX	MoçambiqueLX	BrasilLX
Cabo Verde	-	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Angola	0.30102	-	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Gana	0.01912	0.27936	-	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Somália	0.06419	0.28884	0.05610	-		0.0000±0.0000	0.0000±0.0000
AngolaLX	0.04388	0.24685	0.02606	0.03572	-		0.0000±0.0000
MoçambiqueLX	0.06223	0.25879	0.04070	0.02327	0.01273	-	0.0000±0.0000
BrasilLX	0.06915	0.34497	0.05887	0.03098	0.04857	0.03946	-

A partir das matrizes de distâncias obtidas foi possível fazer a representação filogenética (figura 4.1) entre as populações, tornando-se possível denotar as diferenças apresentadas entre todas as populações e a população brasileira imigrante residente em Lisboa, através de uma árvore filogenética como está representada na figura 4.1.

Figura 4.2: Representação filogenética de distâncias genéticas entre populações de estudos anteriores e a população imigrante brasileira residentes em lisboa à região controlo do DNAm



Através do estudo filogenético foi possível realizar a comparação da população brasileira imigrante em Lisboa com as restantes populações previamente estudadas, uma população de imigrantes moçambicanos, angolanos e cabo-verdianos em Lisboa e também populações do Gana da Somália e, por fim, de uma população indígena angolana. Foram selecionadas populações de origem africanas devido à grande prevalência de haplótipos característicos da população africana presentes na população em estudo.

Os resultados obtidos a partir da análise da matriz de distâncias interpopulacionais permitiram concluir que a população brasileira atualmente a residir em Lisboa encontra-se geneticamente mais distante da população de Angola, o que seria de esperar tratando-se esta de uma população indígena, denominada Kohen-San, que apresenta como haplogrupos mais frequentemente presentes na população L0d e L0k que se restringem maioritariamente a esta população, estando até, inclusivamente, muito distante dos imigrantes angolanos a residir atualmente em Lisboa.

Também é possível averiguar a proximidade genética entre a população brasileira imigrante em Lisboa com a população residente na Somália ($F_{st}=0.03098$; $pvalue=0.0000\pm 0.0000$). Neste estudo também foi possível verificar a grande frequência dos

haplogrupos L2 e L3, aqueles que na população brasileira imigrante em Lisboa também se verificam.

As comparações com as outras populações também permitem observar a grande distância existente entre a população residente em Angola e todas as restantes, sendo que é a população brasileira residente em Lisboa que apresenta uma maior distância desta ($F_{st}=0.34497$; $pvalue=0.0000\pm 0.0000$), resultados esperados uma vez que se trata de uma população indígena que apresenta como haplótipos dominantes L0.

Relativamente à população do Gana e de Cabo Verde será de notar a sua proximidade filogenética ($F_{st}=0.01912$; $pvalue=0.0000\pm 0.0000$), facto que pode ser justificado pela grande proximidade geográfica entre estas duas regiões, partilhando haplótipos L1, L2 e L3. Sendo também duas das populações mais distantes da população brasileira imigrantes em Lisboa provavelmente devido a alguma frequência de haplótipos nativo-americanos e euroasiáticos que esta possui.

Comparativamente as populações imigrantes residentes em Lisboa, a que mais se aproxima da população em estudo é a de Moçambique ($F_{st}=0.03946$; $pvalue=0.0000\pm 0.0000$), sendo a seguinte a população Angolana ($F_{st}=0.04857$; $pvalue=0.0000\pm 0.0000$) o que permite verificar a grande influência que o grande êxodo de escravos africanos, de antigas colónias portuguesas como Moçambique e Angola e de outros países africanos circundantes, para o Brasil, enquanto este país ainda era uma colónia portuguesa, tiveram na pool genética dos brasileiros.

5. Conclusão

Foi estudada a região controlo total do DNAmT de 75 indivíduos brasileiros não relacionados imigrantes em Lisboa.

Foram amplificados 1122pb, correspondente à região controlo total do DNAmT, através de dois pares de primers L15997/H016 e L16555/H639. A posterior sequenciação dos produtos amplificados foi realizada através da sequenciação de Sanger, método que provou ser adequado para a realização deste estudo.

O estudo da região controlo total do DNAmT provou ser eficaz e permitiu um maior grau de discriminação entre amostras relativamente a outros estudos onde a região controlo do DNAmT não foi estudada na sua totalidade.

Os resultados obtidos foram submetidos e posteriormente integrados na base de dados EMPOP, comprovando a qualidade dos mesmos, e estão disponíveis para consulta com o número de acesso EMP.

Existe uma grande variabilidade genética nesta população o que pode ser comprovado pelos 63 haplótipos únicos encontrados nas 75 sequências analisadas.

Foram encontrados em 139 posições alteradas um número total de 1050 polimorfismos.

O maior número de polimorfismos, quando comparado com a rCRS, encontrado numa sequência foram 25, ao passo que o menor número registado foi 5.

A taxa de heteroplasmia total determinada neste estudo foi de 29,33%. Tendo sido encontradas 3 heteroplasmias de posição que são classificadas consoante as normas da IUPAC como 499R, 16183M e 185Y. Foram, ainda, observadas 19 heteroplasmias de comprimento, sendo que a sua maioria dos casos a ocorre na região HVI.

Cerca de 54% dos haplótipos determinados neste estudo são característicos de populações africanas. Sendo dentro destes os haplótipos L1, L2 e L3 os que apresentam uma maior frequência e dentro destes o L3 apresentou, também, relevante frequência, estando presente em 22,80% das sequências analisadas.

Também foram identificados cerca de 25% de haplótipos característicos de populações euroasiáticas, demonstrando a pouca influência destas populações no pool genético desta população.

Em suma, o estudo filogenético permitiu observar as relações filogenéticas entre populações selecionadas de estudos realizados e a população brasileira imigrante em Lisboa. Concluiu-se, também, que a população da Somália é geneticamente mais próxima da dita população brasileira, sendo a população indígena Khoe-San de Angola a que apresenta uma distância genética maior para a população em estudo.

O presente estudo confirma, ainda, que a introdução desta população nos residentes em Lisboa, irá introduzir bastante variabilidade genética a estes, levando a um aumento da frequência do haplogrupo L na população lisboeta. Será, assim, importante compreender, em estudos futuros, o real impacto da introdução desta variabilidade genética na população de Lisboa.

6. Referências

- [1] J. Cummins, Mitochondrial DNA in mammalian reproduction, *Rev Reprod.* 3 (1998) 172–182. <https://doi.org/10.1530/ror.0.0030172>.
- [2] M. Sato, K. Sato, Biochimica et Biophysica Acta Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA, *BBA - Molecular Cell Research.* 1833 (2013) 1979–1984. <https://doi.org/10.1016/j.bbamcr.2013.03.010>.
- [3] Molecular Cell Biology Fifth Edition by Lodish, Harvey; Berk, Arnold; Matsudaira, Paul; Kaiser, Chri published by W. H. Freeman Hardcover: Amazon.com: Books, (n.d.). <https://www.amazon.com/Molecular-Biology-Matsudaira-published-Hardcover/dp/B008BWZC5C> (accessed August 27, 2022).
- [4] J. Boyle, Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M., Biochemistry and Molecular Biology Education. 33 (2005) 74–75. <https://doi.org/10.1002/BMB.2005.494033010419>.
- [5] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, *Advanced Topics in Forensic DNA Typing: Methodology.* (2012). <https://doi.org/10.1016/C2011-0-04189-3>.
- [6] M.M. NASS, S. NASS, Intramitochondrial Fibers With Dna Characteristics. I. Fixation and, *J Cell Biol.* 19 (1963) 593–611. <https://doi.org/10.1083/jcb.19.3.593>.
- [7] S. Anderson, A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, *Nature.* 290 (1981) 457–465. <https://doi.org/10.1038/290457a0>.
- [8] D.M. Turnbull, R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA - *Nature Genetics, Nat Genet.* 23 (1999) 147. <http://www.nature.com/doi/10.1038/13779%5Cnpapers3://publication/doi/10.1038/13779>.

- [9] S.L. Mitchell, R. Goodloe, K. Brown-Gentry, S.A. Pendergrass, D.G. Murdock, D.C. Crawford, Characterization of mitochondrial haplogroups in a large population-based sample from the United States, *Hum Genet.* 133 (2014) 861–868. <https://doi.org/10.1007/S00439-014-1421-9>.
- [10] M. Denaro, H. Blanc, M.J. Johnson, K.H. Chen, E. Wilmsen, L.L. Cavalli-Sforza, D.C. Wallace, Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA, *Proc Natl Acad Sci U S A.* 78 (1981) 5768–5772. <https://doi.org/10.1073/PNAS.78.9.5768>.
- [11] R.L. Cann, M. Stoneking, A.C. Wilson, Mitochondrial DNA and human evolution, *Nature.* 325 (1987) 31–36. <https://doi.org/10.1038/325031a0>.
- [12] L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A.C. Wilson, African populations and the evolution of human mitochondrial DNA, *Science.* 253 (1991) 1503–1507. <https://doi.org/10.1126/SCIENCE.1840702>.
- [13] Y.S. Chen, A. Olckers, T.G. Schurr, A.M. Kogelnik, K. Huoponen, D.C. Wallace, mtDNA Variation in the South African Kung and Khwe—and Their Genetic Relationships to Other African Populations, *Am J Hum Genet.* 66 (2000) 1362. <https://doi.org/10.1086/302848>.
- [14] B. Budowle, M.W. Allard, M.R. Wilson, R. Chakraborty, Forensics and Mitochondrial DNA: Applications, Debates, and Foundations, *Annu Rev Genomics Hum Genet.* 4 (2003) 119–141. <https://doi.org/10.1146/annurev.genom.4.070802.110352>.
- [15] A. Torroni, T.G. Schurr, M.F. Cabell, M.D. Brown, J. v. Neel, M. Larsen, D.G. Smith, C.M. Vullo, D.C. Wallace, Asian affinities and continental radiation of the four founding Native American mtDNAs., *Am J Hum Genet.* 53 (1993) 563. [/pmc/articles/PMC1682412/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/1682412/) (accessed November 7, 2021).
- [16] J.W. Taanman, The mitochondrial genome: Structure, transcription, translation and replication, *Biochim Biophys Acta Bioenerg.* 1410 (1999) 103–123. [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3).
- [17] M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and the origin of modern humans, *Nature* 2000 408:6813. 408 (2000) 708–713. <https://doi.org/10.1038/35047064>.
- [18] J.L. Boore, Animal mitochondrial genomes, *Nucleic Acids Res.* 27 (1999) 1767–1780. <https://doi.org/10.1093/NAR/27.8.1767>.
- [19] M. Mikkelsen, L. Fendt, A.W. Röck, B. Zimmermann, E. Rockenbauer, A.J. Hansen, W. Parson, N. Morling, Forensic and phylogeographic characterisation of mtDNA lineages from Somalia, *Int J Legal Med.* 126 (2012) 573–579. <https://doi.org/10.1007/S00414-012-0694-6>.
- [20] J.W. Taanman, The mitochondrial genome: structure, transcription, translation and replication, *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* 1410 (1999) 103–123. [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3).
- [21] J.L. Boore, Animal mitochondrial genomes, 27 (1999) 1767–1780.
- [22] S. Lutz, H.J. Weisser, J. Heizmann, S. Pollak, Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany, *Int J Legal Med.* 111 (1998) 67–77. <https://doi.org/10.1007/s004140050117>.
- [23] M. Stoneking, Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots, (2000) 1029–1032.
- [24] B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution, *Annu Rev Genomics Hum Genet.* 6 (2005) 165–183. <https://doi.org/10.1146/annurev.genom.6.080604.162249>.
- [25] R. Piercy, K.M. Sullivan, N. Benson, P. Gill, The application of mitochondrial DNA typing to the study of white Caucasian genetic identification, *Int J Legal Med.* 106 (1993) 85–90. <https://doi.org/10.1007/BF01225046>.

- [26] N. Acids, H.J. Gross, Human Mitochondrial DNA and the Evolution of Homo sapiens, 2006. <https://doi.org/10.1007/3-540-31789-9>.
- [27] E. Scheurer, M. Ith, C. Boesch, Application of MRS in forensic pathology, *Molecular Forensics*. (2007) 221–239. <http://www.wiley-vch.de/publish/dt/books/bySubjectLS00/bySubSubjectLS60/0-470-02495-X/?sID=> (accessed August 27, 2022).
- [28] R.E. Giles, H. Blanc, H.M. Cann, D.C. Wallace, Maternal inheritance of human mitochondrial DNA, 77 (1980) 6715–6719.
- [29] H. Shitara, H. Kaneda, A. Sato, K. Inoue, A. Ogura, H. Yonekawa, J. Hayashi, Selective and Continuous Elimination of Mitochondria Microinjected Into Mouse Eggs From Spermatids , but Not From Liver Cells , Occurs Throughout Embryogenesis, (2000).
- [30] P. Sutovsky, R.D. Moreno, T. Dominko, C. Simerly, Ubiquitinated Sperm Mitochondria , Selective Proteolysis , and the Regulation of Mitochondrial Inheritance in Mammalian Embryos 1, 590 (2000) 582–590.
- [31] M. Arianne, S. Chwartz, J. Ohn, V. Issing, Paternal Inheritance of Mitochondrial DNA, <Http://Dx.Doi.Org/10.1056/NEJMoa020350.347> (2009) 576–580. <https://doi.org/10.1056/NEJMoa020350>.
- [32] M. Filosto, M. Mancuso, S. Shanske, M. Hirano, A.L. Andreu, S. Dimauro, Lack of Paternal Inheritance of Muscle Mitochondrial DNA in Sporadic Mitochondrial Myopathies, (2003) 524–526.
- [33] R.W. Taylor, M.T. McDonnell, E.L. Blakely, P.F. Chinnery, G.A. Taylor, N. Howell, M. Zeviani, E. Briem, F. Carrara, Genotypes from Patients Indicate No Paternal Mitochondrial DNA Contribution, (2003) 521–524.
- [34] J.W.O. Ballard, M.C. Whitlock, The incomplete natural history of mitochondria, *Mol Ecol*. 13 (2004) 729–744. <https://doi.org/10.1046/J.1365-294X.2003.02063.X>.
- [35] E. Hagelberg, N. Goldman, P. Lió, S. Whelan, W. Schiefenhöel, J.B. Clegg, D.K. Bowden, S. Whelan, W. Schiefenho, Evidence for mitochondrial DNA recombination in a human population of island Melanesia Evidence for mitochondrial DNA recombination in a human population of island Melanesia, (1999). <https://doi.org/10.1098/rspb.1999.0663>.
- [36] A. Eyre-walker, N.H. Smith, J.M. Smith, How clonal are human mitochondria ?, (1999).
- [37] C. Freyer, B.J. Battersby, J.B. Stewart, E. Hagstro, No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline, 42 (2014) 1111–1116. <https://doi.org/10.1093/nar/gkt969>.
- [38] B. Thyagarajan, R.A. Padua, C. Campbell, Recombination Activity *, *J Biol Chem*. 271 (1996) 27536–27543. <https://doi.org/10.1074/jbc.271.44.27536>.
- [39] A. Brandst, W. Parson, Anita Brandstätter · Walther Parson Mitochondrial DNA heteroplasmy or artefacts — a matter of the amplification strategy ?, (2003) 180–184. <https://doi.org/10.1007/s00414-002-0350-7>.
- [40] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, E. Pokorak, M. Prinz, A. Salas, P.M. Schneider, T.J. Parsons, DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci Int Genet*. 13 (2014) 134–142. <https://doi.org/10.1016/j.fsigen.2014.07.010>.
- [41] J. Naue, S. Hörer, T. Sängler, C. Strobl, P. Hatzgrubwieser, W. Parson, S. Lutz-bonengel, Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA, *MITOCH*. (2014). <https://doi.org/10.1016/j.mito.2014.12.002>.
- [42] A. Amorim, T. Fernandes, N. Taveira, Mitochondrial DNA in human identification: A review, *PeerJ*. 7 (2019). <https://doi.org/10.7717/peerj.7314>.

- [43] C.D. Calloway, R.L. Reynolds, G.L. Herrin, W.W. Anderson, The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age, *Am J Hum Genet.* 66 (2000) 1384–1397. <https://doi.org/10.1086/302844>.
- [44] N. Sondheimer, C.E. Glatz, J.E. Tirone, M.A. Deardorff, A.M. Krieger, H. Hakonarson, Neutral mitochondrial heteroplasmy and the influence of aging, *Hum Mol Genet.* 20 (2011) 1653–1659. <https://doi.org/10.1093/HMG/DDR043>.
- [45] K.A. Roberts, C. Calloway, Characterization of Mitochondrial DNA Sequence Heteroplasmy in Blood Tissue and Hair as a Function of Hair Morphology * , , à, 56 (2011) 46–60. <https://doi.org/10.1111/j.1556-4029.2010.01540.x>.
- [46] W. Parson, T.J. Parsons, R. Scheithauer, M.M. Holland, Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence analysis to a forensic case, *International Journal of Legal Medicine* 1998 111:3. 111 (1998) 124–132. <https://doi.org/10.1007/S004140050132>.
- [47] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, E. Pokorak, M. Prinz, A. Salas, P.M. Schneider, T.J. Parsons, DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci Int Genet.* 13 (2014) 134–142. <https://doi.org/10.1016/j.fsigen.2014.07.010>.
- [48] M. Johnson, D.C. Wallace, S.D. Ferris, M.C. Rattazzi, Radiation of Human Mitochondria DNA Types Analyzed by Restriction Endonuclease Cleavage Patterns, (1983) 255–271.
- [49] H.A. Costa, V. Lopes, F. Balsa, A.M. Bento, S. Pantoja, M.J. Anjos, M. Carvalho, M.C. Vide, D.N. Vieira, F. Corte-Real, Mitochondrial DNA sequence analysis of native Bolivians population, *Forensic Sci Int Genet Suppl Ser.* 1 (2008) 259–261. <https://doi.org/10.1016/J.FSIGSS.2007.10.090>.
- [50] A. Brandstätter, T.J. Parsons, W. Parson, Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups, *Int J Legal Med.* 117 (2003) 291–298. <https://doi.org/10.1007/S00414-003-0395-2>.
- [51] L. Castrì, S. Tofanelli, P. Garagnani, C. Bini, X. Fosella, S. Pelotti, G. Paoli, D. Pettener, D. Luiselli, mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa, *Am J Phys Anthropol.* 140 (2009) 302–311. <https://doi.org/10.1002/AJPA.21070>.
- [52] F.S. International, F.P. Genetics, F.P. Genetics, F.P. Genetics, F.P. Genetics, F. Strs, F. Strs, T. Qc, *Forensic Science International : Genetics*, 4 (2010) 145–147. <https://doi.org/10.1016/j.fsigen.2010.02.001>.
- [53] S. Lutz, H. Wittig, H. Weisser, J. Heizmann, A. Junge, W. Parson, J. Edelmann, K. Anslinger, S. Jung, C. Augustin, Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions ?, 113 (2000) 97–101.
- [54] B.C.L. Davis, *Mitochondrial DNA : State of Tennessee v . Paul Ware*, (1998) 6–7.
- [55] M. Allen, A.-S. Engström, S. Meyers, O. Handt, T. Saldeen, A. von Haeseler, S. Pääbo, U. Gyllensten, Mitochondrial DNA Sequencing of Shed Hairs and Saliva on Robbery Caps: Sensitivity and Matching Probabilities, *J Forensic Sci.* 43 (1998) 16169J. <https://doi.org/10.1520/JFS16169J>.
- [56] P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, I. Evett, E. Hagelberg, K. Sullivan, Identification of the remains of the romanov family by DNA analysis, *Nat Genet.* 6 (1994) 130–135. <https://doi.org/10.1038/NG0294-130>.
- [57] E. Jhaes, R. Decorte, A. Peneau, J.H. Petrie, P.A. Boiry, A. Gilissen, J.P. Moisan, H. van den Berghe, O. Pascal, J.J. Cassiman, Mitochondrial DNA analysis on remains of a putative son of Louis XVI, King of France and Marie-Antoinette, *European Journal of Human Genetics.* 6 (1998) 383–395. <https://doi.org/10.1038/SJ.EJHG.5200227>.

- [58] K. Sekiguchi, K. Kasai, B.C. Levin, Inter- and intragenerational transmission of a human mitochondrial DNA heteroplasmy among 13 maternally-related individuals and differences between and within tissues in two family members, *Mitochondrion*. 2 (2003) 401–414. [https://doi.org/10.1016/S1567-7249\(03\)00028-X](https://doi.org/10.1016/S1567-7249(03)00028-X).
- [59] M. Bykowska-witowska, M. Sc, D. Ph, T. Dembi, D. Ph, Genetic Identification of Communist Crimes ' Victims (1944 – 1956) Based on the Analysis of One of Many Mass Graves Discovered on the Powazki Military Cemetery in Warsaw , Poland, 61 (2016) 1450–1455. <https://doi.org/10.1111/1556-4029.13205>.
- [60] A. Torroni, A. Achilli, V. Macaulay, M. Richards, H.J. Bandelt, Harvesting the fruit of the human mtDNA tree, *Trends in Genetics*. 22 (2006) 339–345. <https://doi.org/10.1016/j.tig.2006.04.001>.
- [61] Z. A, U. Gyllensten, M. Ingman, H. Kaessmann, S. Pa, Mitochondrial genome variation and the origin of modern humans, 2173 (2000).
- [62] N. Howell, I. Kubacka, D.A. Mackey, How rapidly does the human mitochondrial genome evolve?, *Am J Hum Genet*. 59 (1996) 501. [/pmc/articles/PMC1914922/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/1914922/) (accessed November 7, 2021).
- [63] M. van Oven, M. Kayser, Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation, (2008) 386–394. <https://doi.org/10.1002/humu.20921>.
- [64] A. Salas, M. Richards, B. Sobrino, P. Sa, V. Macaulay, The Making of the African mtDNA Landscape, (2002) 1082–1111.
- [65] D. Mishmar, E. Ruiz-pesini, P. Golik, V. Macaulay, A.G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, M.D. Brown, R.I. Sukernik, A. Olckers, D.C. Wallace, Natural selection shaped regional mtDNA variation in humans, 100 (2003) 171–176.
- [66] L. Quintana-murci, O. Semino, H. Bandelt, G. Passarino, K. Mcelreavey, A.S. Santachiara-benerecetti, Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa, 23 (1999) 437–441.
- [67] R.W. Taylor, D.M. Turnbull, MITOCHONDRIAL DNA MUTATIONS IN HUMAN DISEASE, *Nat Rev Genet*. 6 (2005) 389. <https://doi.org/10.1038/NRG1606>.
- [68] S. Tommasi, P. Favia, S. Weigl, A. Bianco, B. Pilato, L. Russo, A. Paradiso, V. Petruzzella, Mitochondrial DNA variants and risk of familial breast cancer: An exploratory study, *Int J Oncol*. 44 (2014) 1691–1698. <https://doi.org/10.3892/IJO.2014.2324/HTML>.
- [69] M. del M. González, C. Santos, C. Alarcón, A. Ramos, M. Cos, G. Catalano, J.J. Acebes, M.P. Aluja, Mitochondrial DNA haplogroups J and T increase the risk of glioma, *Mitochondrion*. 58 (2021) 95–101. <https://doi.org/10.1016/J.MITO.2021.02.013>.
- [70] P. Ramos-Louro, R. Daniel, A. Pérez Vertti, A.L. Reyes, G. Angélica Martínez-Nava, R. Espinosa, C. Pineda, F. Francisco González Galarza, R. Argüello Astorga, L.S. Aguilar Muñoz, F. Hernández Terán, N. Marbella, P. Torres, A. Durán Sotuela, M. Fernández-Moreno, V.B. Barreiro, F.J. Blanco, I. Rego-Pérez, mtDNA haplogroup A enhances the effect of obesity on the risk of knee OA in a Mexican population, *Scientific Reports* | . 12 (123AD) 5173. <https://doi.org/10.1038/s41598-022-09265-y>.
- [71] W. Parson, A. Du, EMPOP — A forensic mtDNA database, 1 (2007) 88–92. <https://doi.org/10.1016/j.fsigen.2007.01.018>.
- [72] P.M. Schneider, Scientific standards for studies in forensic genetics, *Forensic Sci Int*. 165 (2007) 238–243. <https://doi.org/10.1016/j.forsciint.2006.06.067>.
- [73] M.C. Brandon, M.T. Lott, K.C. Nguyen, S. Spolim, S.B. Navathe, P. Baldi, D.C. Wallace, MITOMAP : a human mitochondrial genome database — 2004 update, 33 (2005) 2004–2006. <https://doi.org/10.1093/nar/gki079>.

- [74] M.T. Lott, O. Derbeneva, H.M. Xie, D.C. Wallace, NIH Public Access, 2014. <https://doi.org/10.1002/0471250953.bi0123s44.mtDNA>.
- [75] E. Ruiz-pesini, M.T. Lott, V. Procaccio, J.C. Poole, M.C. Brandon, D. Mishmar, C. Yi, J. Kreuziger, P. Baldi, D.C. Wallace, An enhanced MITOMAP with a global mtDNA mutational phylogeny, 35 (2007) 823–828. <https://doi.org/10.1093/nar/gkl927>.
- [76] SEF/GEPF, Relatório de Imigração, Fronteiras e Asilo 2021, 2022.
- [77] F.S.G. Kehdy, M.H. Gouveia, M. Machado, W.C.S. Magalhães, A.R. Horimoto, Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations, 112 (2015) 8696–8701. <https://doi.org/10.1073/pnas.1504447112>.
- [78] S.D.J. Pena, F.R. Santos, E. Tarazona-santos, Genetic admixture in Brazil, (2020) 928–938. <https://doi.org/10.1002/ajmg.c.31853>.
- [79] IBGE, Brasil 500 anos de povoamento, Rio de Janeiro: Instituto Brasileiro de Geografia Estatística. (2000).
- [80] H. Hoetink, Race Mixture in the History of Latin America, *Hispanic American Historical Review*. 48 (1968) 669–670. <https://doi.org/10.1215/00182168-48.4.669>.
- [81] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material., *Biotechniques*. 10 (1991) 506–513. <https://doi.org/10.2144/000114018>.
- [82] F. Sanger, S. Nicklen, DNA sequencing with chain-terminating, 74 (1977) 5463–5467.
- [83] Sequencing, (n.d.). <https://www.sigmaaldrich.com/PT/en/applications/genomics/sequencing> (accessed August 26, 2022).
- [84] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, E. Pokorak, M. Prinz, A. Salas, P.M. Schneider, T.J. Parsons, DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci Int Genet*. 13 (2014) 134–142. <https://doi.org/10.1016/J.FSIGEN.2014.07.010>.
- [85] F. Simão, A.P. Ferreira, E.F. de Carvalho, W. Parson, L. Gusmão, Defining mtDNA origins and population stratification in Rio de Janeiro, *Forensic Sci Int Genet*. 34 (2018) 97–104. <https://doi.org/10.1016/J.FSIGEN.2018.02.003>.
- [86] A. Paula Schaan, L. Costa, D. Santos, A. Modesto, M. Amador, C. Lopes, S. Helena Rabenhorst, R. Montenegro, B.D. A Souza, T. Lopes, F. Keiko Yoshioka, G. Pinto, V. Silbiger, Â. Ribeiro-dos-Santos, mtDNA structure: the women who formed the Brazilian Northeast, (n.d.). <https://doi.org/10.1186/s12862-017-1027-7>.
- [87] S. Bernardo, R. Hermida, M. Desidério, D.A. Silva, E.F. de Carvalho, MtdNA ancestry of Rio de Janeiro population, Brazil, *Mol Biol Rep*. 41 (2014) 1945–1950. <https://doi.org/10.1007/S11033-014-3041-9>.
- [88] C.J. Mulligan, K. Hunley, S. Cole, J.C. Long, Population genetics, history, and health patterns in native americans, *Annu Rev Genomics Hum Genet*. 5 (2004) 295–315. <https://doi.org/10.1146/ANNUREV.GENOM.5.061903.175920>.
- [89] S.L. Marques, A. Goios, A.M. Rocha, M.J. Prata, A. Amorim, L. Gusmão, C. Alves, L. Alvarez, Portuguese mitochondrial DNA genetic diversity—An update and a phylogenetic revision, *Forensic Sci Int Genet*. 15 (2015) 27–32. <https://doi.org/10.1016/J.FSIGEN.2014.10.004>.
- [90] P. Morais, A. Amorim, C. Vieira Da Silva, T. Ribeiro, J. Costa Santos, H. Afonso Costa, Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis, *J Genet*. 94 (2015) 509–512. <https://doi.org/10.1007/s12041-015-0552-7>.
- [91] L. Fendt, G. Huber, A.W. Röck, B. Zimmermann, M. Bodner, R. Delport, K. Schmidt, W. Parson, Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages, *Forensic Sci Int Genet*. 6 (2012) 662–663. <https://doi.org/10.1016/J.FSIGEN.2012.02.010>.

- [92] L. Fendt, A. Röck, B. Zimmermann, M. Bodner, T. Thye, F. Tschentscher, E. Owusu-Dabo, T.M.K. Göbel, P.M. Schneider, W. Parson, MtDNA diversity of Ghana: a forensic and phylogeographic view, *Forensic Sci Int Genet.* 6 (2012) 244. <https://doi.org/10.1016/J.FSIGEN.2011.05.011>.
- [93] F. Simão, H.A. Costa, C.V. da Silva, T. Ribeiro, M.J. Porto, J.C. Santos, A. Amorim, Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA, *Forensic Sci Int Genet.* 15 (2015) 33–38. <https://doi.org/10.1016/J.FSIGEN.2014.09.013>.
- [94] M. Proença de Campos, H. Afonso Costa, C. Vieira da Silva, V. Bogas, T. Ribeiro, M.J. Porto, A. Amorim, The immigrant population from Mozambique in Lisbon: Updated mitochondrial DNA portrait, *Forensic Sci Int Genet Suppl Ser.* 6 (2017) e298–e300. <https://doi.org/10.1016/J.FSIGSS.2017.09.134>.