

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL



**Estudo do genoma mitocondrial com metodologia NGS:
Contributo para a caracterização da população do Sul de
Portugal**

Maria Margarida Cristino Domingos

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Professor Doutor António Amorim

2024

AGRADECIMENTOS

No final deste projeto, olhar para trás e lembrar este caminho traz-me um profundo sentimento de realização e gratidão para com todos que contribuíram, oferecendo orientação, motivação e apoio para a sua concretização.

À minha querida orientadora, Mestre Heloísa Costa, deixo o meu mais sincero agradecimento pela partilha comigo o seu vasto conhecimento sobre ADN mitocondrial. Agradeço por toda a positividade e tranquilidade transmitida ao longo do ano e por sempre acreditar que conseguimos fazer mais e melhor.

Ao Professor Doutor António Amorim, o meu orientador, agradeço pelas incríveis aulas de genética forense, que aumentaram a minha paixão por esta área e pela oportunidade de trabalhar próximo de pessoas tão inspiradoras.

Ao Rui, o meu muito obrigada por todo o acompanhamento e orientação, tanto na parte prática como na escrita.

À direção de mestrado, especialmente à Professora Doutora Andreia Figueiredo e ao Professor Doutor Francisco Dionísio, um obrigada pela serenidade e atenção ao longo destes dois anos.

Ao Serviço de Genética e Biologia Forense, obrigada pelo acolhimento e ajuda técnica.

À Mariana, uma das minhas melhores amigas e parceira de laboratório, obrigada pelos risos, pelos abraços, pelas conversas constantes e pelo apoio e amizade incondicionais.

Aos meus amigos de mestrado, Bea, Sara, Barbára, Vasco e João, agradeço por partilharem esta jornada comigo. Obrigada por nunca deixarem o nosso “barco” naufragar e por todas as aventuras que tornaram este ano mais leve e especial.

Aos meus amigos Maria João Cunha e João Bodião, o meu especial agradecimento pela paciência, compreensão e por todas as valiosas dicas que me deram.

À minha família, um agradecimento especial pelo amor incondicional e pela oportunidade que me proporcionaram de seguir este caminho. À minha mãe por ser o meu porto seguro, o meu mais profundo agradecimento. Alcançar este objetivo só foi possível graças aos sacrifícios e esforços que fizeram para me oferecer as condições necessárias para estudar, crescer e ser feliz.

E, por fim, a todos que contribuíram e me ajudaram de alguma forma ao longo do meu percurso, o meu mais sincero reconhecimento. Esta dissertação é, também, uma conquista vossa.

RESUMO

O ADN mitocondrial (ADNmt), uma molécula de ADN circular encontrada nas mitocôndrias, é herdado exclusivamente por via materna sem sofrer recombinação genética significativa. Estas características tornam-no um marcador genético fundamental na investigação da origem populacional e dispersão humana, sendo também amplamente utilizado em casos forenses onde o ADN nuclear é inadequado ou inexistente.

Esta dissertação teve como principal objetivo a análise do ADNmt para caracterizar a diversidade genética e a ancestralidade da população do Sul de Portugal. O estudo envolveu a análise de 155 amostras através da sequenciação da nova geração (NGS), que permite obter o genoma mitocondrial completo com maior eficiência. Paralelamente, para a validação e implementação da técnica de NGS, no Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF), esta foi comparada com a metodologia de sequenciação de Sanger (SS). As amostras foram previamente estudadas por SS, tendo-se verificado uma concordância, na região controlo do ADNmt, entre os dois métodos de sequenciação. A NGS demonstrou várias vantagens, incluindo maior precisão e cobertura do genoma, simplificando o processo de sequenciação. Relativamente ao estudo filogenético, a população do Sul de Portugal revelou uma maior divergência genética, face às restantes regiões, refletindo possíveis influências históricas e padrões migratórios distintos.

Os resultados obtidos contribuem significativamente para o conhecimento da diversidade genética da população na região do Sul de Portugal e promovem avanços na área da genética forense. O uso do ADNmt, aliado às tecnologias de sequenciação avançada, mostra-se essencial tanto na caracterização de populações, como em outras áreas, reforçando a sua aplicabilidade em contextos forenses e científicos.

Palavras-chave: ADN mitocondrial, Sequenciação de Nova Geração, Genética Forense, População de Lisboa, Caracterização Filogenética.

ABSTRACT

Mitochondrial DNA (mtDNA), a circular DNA molecule found in mitochondria, is inherited exclusively by the mother without undergoing significant genetic recombination. These characteristics make it a fundamental genetic marker in the investigation of population origin and human dispersion, and it is also widely used in forensic cases where nuclear DNA is inadequate or non-existent.

The main aim of this dissertation was to analyse mtDNA to characterise the genetic diversity and ancestry of the population of southern Portugal. The study involved analysing 155 samples using next generation sequencing (NGS), which makes it possible to obtain the complete mitochondrial genome more efficiently. At the same time, in order to validate and implement the NGS technique at the National Institute of Legal Medicine and Forensic Sciences (INMLCF), it was compared with the Sanger sequencing (SS) methodology. The samples had previously been studied by SS and there was concordance between the two sequencing methods in the control region of the mtDNA. NGS demonstrated several advantages, including greater precision and coverage of the genome, simplifying the sequencing process. Regarding the phylogenetic analysis, the population from southern Portugal exhibited greater genetic divergence compared to other regions, reflecting possible historical influences and distinct migratory patterns.

The results obtained significantly contribute to the knowledge of the genetic diversity of the population in the southern region of Portugal and promote advances in the area of forensic genetics. The use of mtDNA, combined with advanced sequencing technologies, is proving to be essential both in population characterisation and in other areas, reinforcing its applicability in forensic and scientific contexts.

Key-words: Mitochondrial DNA, Next-Generation Sequencing, Forensic Genetic, Lisbon Population, Phylogenetic Characterization

ÍNDICE

AGRADECIMENTOS	I
RESUMO.....	II
ABSTRACT	III
LISTA DE FIGURAS	VI
LISTA DE QUADROS.....	VII
LISTA DE TABELAS.....	VII
LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS	VIII
1. INTRODUÇÃO	- 1 -
1.1. Mitocôndria	- 1 -
1.2. Genoma Mitocondrial.....	- 1 -
1.2.1. Características de ADN mitocondrial	- 1 -
1.2.2. Heteroplasmias.....	- 2 -
1.2.3. Aplicações do ADN mitocondrial	- 2 -
1.3. Base de dados	- 5 -
1.4. Sequenciação de ADN mitocondrial.....	- 5 -
1.4.1. Sequenciação de Sanger.....	- 5 -
1.4.2. Sequenciação de Nova Geração.....	- 6 -
1.5. Validação dos métodos de sequenciação de primeira e segunda geração.....	- 6 -
2. OBJETIVOS.....	- 7 -
3. METODOLOGIA.....	- 7 -
3.1. Seleção das Amostras	- 7 -
3.2. Sequenciação de Sanger	- 7 -
3.2.1. Amplificação:.....	- 7 -
3.2.2. Purificação dos produtos amplificados (<i>ExoSap</i>):	- 8 -
3.2.3. Sequenciação:	- 8 -
3.2.4. Purificação BigDye Xterminator:	- 8 -
3.3. Sequenciação de Nova Geração (NGS).....	- 8 -
3.3.1. Extração e Quantificação de ADN das amostras selecionadas	- 8 -
3.3.2. Construção e Quantificação das Bibliotecas.....	- 9 -
3.3.3. Preparação do <i>Template</i> (carregamento do <i>chip</i>).....	- 10 -
3.3.4. Sequenciação de Semicondutores de Iões (<i>chip</i> carregado)	- 10 -
3.3.5. Análise de dados (Bioinformática)	- 10 -

3.4.	Análise Filogenética	- 11 -
4.	RESULTADOS E DISCUSSÃO	- 11 -
4.1.	Validação da Sequenciação de Sanger – Especificidade	- 11 -
4.2.	Análise das amostras de ADN mitocondrial – Sequenciação NGS	- 12 -
4.3.	Região controlo de ADN mitocondrial: Tecnologia NGS <i>versus</i> Sequenciação de Sanger ...	15 -
4.4.	Estudo Populacional	- 17 -
4.4.1.	Estudo dos Haplogrupos	- 17 -
4.4.2.	Diversidade Genética	- 22 -
4.5.	Análise filogenética	- 22 -
4.6.	Aplicações Clínicas e Genética Médica da Sequenciação do ADN mitocondrial	- 26 -
4.6.1.	Genoma mitocondrial – função e aplicação	- 26 -
4.6.2.	Mutações no ADN mitocondrial associadas a doenças.....	- 26 -
4.6.3.	Relação entre Haplogrupos Mitocondriais e Doenças	- 27 -
5.	CONCLUSÃO	- 28 -
6.	REFERÊNCIAS	- 30 -
7.	ANEXOS	- 41 -

LISTA DE FIGURAS

- Figura 1.1** - Mapa da distribuição geográfica dos haplótipos do DNA mitocondrial humano (letras no mapa) deduzindo a origem dos humanos anatomicamente modernos e as possíveis rotas das migrações humanas “fora de África”. A localização a vermelho marca a origem da “Mitochondrial Eve” (Criada no BioRender)..... - 4 -
- Figura 3.1** - Representação do *equipamento HID Ion ChefTM* carregado para realização da pool de bibliotecas, retirada do manual “*Precision ID mtDNA Whole Genome Panel with HID Ion S5TM*”. Utilizando o *kit Precision ID DL8*: 1- Embalagem de *Precision ID DL8 Solutions*; 2 – Embalagem de *Precision ID DL8 Reagents*; 3 – Embalagem de pontas *L8 Ion AmpliSeqTM* ; 4 – Selo de Alumínio Moldado para a *PCR*; 5 – Placa *Precision ID IonCode Barcode adapters* com 96 poços; 6 – Embalagem de pontas vazias; 7 – Cartucho de Enriquecimento. - 9 -
- Figura 4.1** - Esquemas do relatório pré-alinhamento criado pelo *software Torrent SuiteTM do Ion 530TM Chip 1*. (A) Densidade de ISPs. (B) Resumo do estado das ISPs, sequências utilizáveis. (C) Histograma do comprimento de leituras..... - 12 -
- Figura 4.2** - Relatório do Alinhamento, relativo ao *Ion 530TM Chip 1* (alinhado a *PrecisionID_mtDNA-rCRS*). (A) Gráfico do alinhamento das leituras. (B) Precisão do alinhamento das leituras..... - 13 -
- Figura 4.3** - Esquemas do relatório pré-alinhamento criado pelo *software Torrent SuiteTM do Ion 530TM Chip 2*. (A) Densidade de ISPs. (B) Resumo do estado das ISPs, sequências utilizáveis. (C) Histograma do comprimento de leituras. - 13 -
- Figura 4.4** - Relatório do Alinhamento, relativo ao *Ion 530TM Chip 2* (alinhado a *PrecisionID_mtDNA-rCRS*). (A) Gráfico do alinhamento das leituras. (B) Precisão do alinhamento das leituras..... - 14 -
- Figura 4.5** - Esquemas do relatório pré-alinhamento criado pelo *software Torrent SuiteTM do Ion 530TM Chip 3*. (A) Densidade de ISPs. (B) Resumo do estado das ISPs, sequências utilizáveis. (C) Histograma do comprimento de leituras. - 14 -
- Figura 4.6** - Relatório do Alinhamento, relativo ao *Ion 530TM Chip 3* (alinhado a *PrecisionID_mtDNA-rCRS*). (A) Gráfico do alinhamento das leituras. (B) Precisão do alinhamento das leituras..... - 15 -
- Figura 4.7** - Gráfico com os diferentes haplogrupos e respectivas frequências identificados nas 155 sequências de ADNmt de indivíduos da população do sul de Portugal - 18 -
- Figura 4.8** - Representação filogenética das distâncias genéticas entre a população do Sul, as populações africanas imigrantes em Lisboa e a população de Portugal, selecionadas da literatura. - 24 -
- Figura 4.9** - Representação filogenética das distâncias genéticas entre as populações portuguesas (população das regiões Norte, Centro e Sul) estudadas com a metodologia NGS..... - 25 -

LISTA DE QUADROS

- Quadro 3.1-** Código IUPAC utilizado na nomenclatura das bases nucleotídicas de ADNmt. - 11 -
- Quadro 4.1** - Comparação do Haplogrupo nas amostras (GENPOP8594; GENPOP8595; GENPOP8596; GENPOP8597; GENPOP8599; GENPOP8600; GENPOP8601; GENPOP8603) de ADN mitocondrial com os dois métodos de sequenciação: Sanger VS NGS. (A azul estão as mutações localizadas na região controlo e a preto estão as mutações localizadas na região codificante) - 16 -
- Quadro 4.2** - Comparação dos haplótipos das amostras GENPOP8594 e GENPOP8656. (A negrito está marcado a variante característica do haplogrupo H1 e a sublinhado estão os polimorfismos partilhados pelas amostras)..... - 18 -
- Quadro 4.3** - Comparação dos haplótipos das amostras GENPOP8694 e GENPOP8565. (A negrito estão marcadas as variantes características do haplogrupo U4 e a vermelho estão os polimorfismos partilhados pelas amostras). - 18 -
- Quadro 4.4-** População do presente estudo e populações selecionadas da literatura (populações africanas imigrantes e população de Portugal), tendo por base a região controlo total do ADNmt. - 23 -
- Quadro 4.5** - População do presente estudo e populações portuguesas das regiões do Centro e Norte, com amostras que foram sequenciadas com NGS para análise filogenética, tendo por base a região controlo do ADNmt..... - 24 -

LISTA DE TABELAS

- Tabela 4.1** - Valores de F_{st} e o p relativos às populações do estudo filogenético conseguidos através do programa Arlequin para a totalidade da região controlo do ADNmt: Sul de Portugal; Imigrantes de Cabo Verde; Imigrantes de Guiné-Bissau; Imigrantes de Angola; Imigrantes de Moçambique; Portugal - 23 -
- Tabela 4.2** - Valores de F_{st} e os respetivos valores de p , recolhidos através do software Arlequin para a totalidade da região controlo do ADNmt, relativos aos haplótipos das regiões portuguesas (Norte, Centro e Sul de Portugal), obtidos com a sequenciação de nova geração. (* significa que o valor de p é inferior a 0,01) - 25 -

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

A	Adenina
ADN	Ácido Desoxirribonucléico
ADNg	ADN genómico
ADNmt	ADN mitocondrial
AMOVA	Análise de variância molecular
BAM	<i>Binary Alignment Map</i>
C	Citosina
C_t	<i>Cycle Threshold</i>
CE	Eletroforese capilar (Inglês, <i>capillary electrophoresis</i>)
CRS	<i>Cambridge Reference Sequence</i>
ddATP	Didesoxiadenina trifosfatada
ddCTP	Didesoxicitosina trifosfatada
ddGTP	Didesoxiguanina trifosfatada
ddNTPs	Didesoxinucleótidos trifosfatados
ddTTP	Didesoxitimina trifosfatada
del	Deleção
D-loop	<i>Displacement Loop</i>
dNTPs	Desoxinucleótidos Trifosfatos
EDNAP	<i>European DNA Profiling Group</i>
EMPOP	<i>EDNAP mitochondrial DNA population database</i>
ePCR	<i>Emulsion PCR</i> (PCR de emulsão)
G	Guanina
HVI	Região hipervariável 1
HVII	Região hipervariável 2
HVIII	Região hipervariável 3

INMLCF	Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.
ISP	<i>Ion Spheres Particules</i>
IUPAC	<i>International Union of Pure & Applied Chemistry</i>
LHON	Neuropatia ótica hereditária de Leber (em Inglês, <i>Leber Hereditary Optic Neuropathy</i>)
LHP	Heteroplasma de comprimento (em Inglês, <i>Length heteroplasmy</i>)
MERRF	Epilepsia Mioclônica com fibras rotas vermelhas (em Inglês, <i>Myoclonic Epilepsy and Ragged Red Fibers</i>)
MIDD	<i>Maternally Inherited Diabetes and Deafness</i>
mL	Mililitro
NGS	Sequenciação de Nova Geração (em Inglês, <i>Next- Generation Sequencing</i>) ou Sequenciação Massiva Paralelo (MPS)
pb	Par de bases (em Inglês, <i>Base pair</i>)
PCR	<i>Polymerase chain reaction</i> (Reação em Cadeia de Polimerase)
PHP	Heteroplasma de posição (em Inglês, <i>Pontual Heteroplasmy</i>)
Polimorfismo de ADN	Regiões da molécula de ADN que apresentam diferentes formas nos indivíduos de uma população. Também denominado polimorfismo genético
qPCR	<i>Real-Time PCR</i> (PCR em tempo real)
Primer	Pequeno segmento/fragmento de DNA que determina a posição a partir da qual a cadeia de DNA vai ser copiada.
rCRS	<i>Revised Cambridge Reference Sequence</i>
rpm	Rotações por minuto
SGBF-DS	Serviço de Genético e Biologia Forense – Delegação do Sul
SS	Sequenciação de Sanger
T	Timina
μL	Microlitro
°C	Graus Celsius

1. INTRODUÇÃO

1.1. Mitocôndria

A mitocôndria é um organelo celular, localizado no citoplasma, e presente em todas as células eucariotas [1]. Com origem em bactérias ancestrais [2], as mitocôndrias têm um papel crucial na produção de energia e na regulação da atividade celular, em processos como a respiração aeróbica, e apoptose, respectivamente [3], [4], [5]. São também um organelo que possui o seu próprio genoma - ADN mitocondrial – que está protegido por estruturas proteicas de ADN, conhecidas por nucleóides mitocondriais, espalhados pela matriz mitocondrial [6].

1.2. Genoma Mitocondrial

O genoma mitocondrial ou mitogenoma, é uma pequena molécula circular, de dupla cadeia, constituída por 16569 pares de bases (pbs) [6], [7]. As duas cadeias podem ser fisicamente distinguidas como a cadeia pesada ('heavy' (H) em inglês), rica em purinas (Guaninas e Adeninas), e a cadeia leve ('light' (L) em inglês), rica em pirimidinas (Citosinas e Timinas) [7].

Adicionalmente, o mitogenoma pode ser dividido em duas partes: a região controlo (CR) e a região codificante (codR). A codR é uma região codificante, que vai da posição 577 à 16023, que contém 37 genes (13 genes codificantes de proteínas e 24 genes que codificam para tRNA e rRNA) [7], [8].

A CR é uma pequena região não codificante e altamente variável, que vai da posição 1 à 576 e da 16024 à 16569. Esta região é também conhecida como *D-Loop* e tem uma extensão de 1122 pares de bases. É marcada pela presença dos segmentos de maior variabilidade genética no genoma mitocondrial, conhecidos como segmentos hipervariáveis – HVI, HVII e HVIII (*Hiper Variable Region* I, II e III). Estes segmentos, HVI (da posição 16024 à 16365), HVII (da posição 73 à 340) e HVIII (da posição 438 à 574), são locais onde ocorrem as variações genéticas mais proeminentes, contribuindo para a diversidade genética dentro da população [7], [9].

1.2.1. Características de ADN mitocondrial

O ADNmt apresenta quatro características que o tornam bastante importante no estudo da genética evolutiva: a transmissão (exclusivamente) materna; a ausência de recombinação; o elevado número de cópias, e a sua elevada taxa de mutação [10].

Aquando da fecundação, ocorre/dá-se a fusão dos núcleos dos gametas feminino e masculino, que tornam o genoma nuclear único para cada indivíduo, à exceção de gémeos homocigotos [11]. Contudo, o genoma mitocondrial é apenas transmitido pelas mitocôndrias dos oócitos originais [8], [10], [11], [12]. Este fenómeno pode ser explicado por dois processos [13]. O primeiro explica a existência de uma regulação negativa significativa do número de cópias do ADNmt durante a espermatogénese, contrariamente à oogénese [14]. No segundo, as mitocôndrias dos espermatozoides, que se encontram nas suas caudas, são degradadas por ativação de um mecanismo no citoplasma do oócito, após a fertilização [15], [16]. Esta característica torna o ADNmt útil em estudos de linhagens, ancestralidade e evolução de populações.

A segunda característica é a ausência de recombinação genética no ADNmt [12], permitindo que as regiões codificantes e não codificantes sejam combinadas para estabelecer relações entre indivíduos, através de uma árvore filogenética [10], onde as variações encontradas são o produto de mutações que se acumulam ao longo das gerações [17], [18].

O elevado número de cópias é outra característica vantajosa do ADNmt. Contrariamente ao ADN nuclear, que apenas apresenta 1 cópia por célula, cada mitocôndria possui em média entre 5 a 10 cópias e o número de mitocôndrias varia de acordo com o tipo de célula/tecido podendo chegar a existir centenas numa única célula [19], [20]. Assim, cada célula pode conter milhares de cópias de ADNmt, que se torna bastante vantajoso, quando se estuda amostras com pouca quantidade de ADN. Contudo, o número de cópias de ADNmt vai diminuindo com a idade [7].

A última característica é a elevada taxa de mutação do ADNmt, sendo que a maior parte da variação ocorre nos segmentos hipervariáveis (HVI, HVII e HVIII) [9]. Em comparação com o genoma nuclear, o ADNmt evolui 5 a 10 vezes mais rápido e a sua taxa de mutação é estimada em 1.3×10^{-8} a 1.89×10^{-8} por base por ano [7], sendo três vezes maior nas populações modernas [21]. Devido à elevada taxa de mutação, as células podem conter diferentes cópias de genoma mitocondrial, originando diversas “pools” de mitocôndrias, conhecidas por populações heteroplasmáticas de mitocôndrias [22].

1.2.2. Heteroplasmas

A heteroplasma é um estado de variabilidade das sequências das cópias de ADNmt, no mesmo indivíduo, podendo ser dentro da mesma mitocôndria ou em células diferentes [12]. Este acontecimento pode ser resultado de, primeiro o ócito ser heteroplasmático e, segundo, a ocorrência de mutações nas células somáticas, durante a replicação do ADNmt [23].

Com isto, os diferentes tecidos do organismo apresentam proporções variáveis de ADNmt, o que resulta num mosaicismo mitocondrial [24]. Existem dois tipos de heteroplasma: heteroplasma de comprimento (LHP); e heteroplasma de posição (PHP) [7], [12].

2.2.2.1. Heteroplasmas de comprimento (LHP):

As heteroplasmas de comprimento envolvem inserções ou deleções, que afetam no comprimento das sequências.

Estas variações têm implicações filogenéticas e podem ser associadas a grupos genéticos específicos [25]. As diferenças nas heteroplasmas de comprimento entre haplótipos idênticos de ADNmt não são consideradas critérios de exclusão, uma vez que é reconhecido que diferentes tecidos podem exibir padrões divergentes [7], [25].

2.2.2.2. Heteroplasma de posição (PHP):

A heteroplasma de posição, ou sequência, está relacionada com substituições em algumas sequências, e noutras não, de ADNmt [12], sendo detetada na sequenciação, pela presença de dois nucleótidos no mesmo local [26]. A observação de mais de duas posições de PHP, durante a análise de mitogenoma, pode indicar uma possível contaminação da amostra [26].

A ocorrência das heteroplasmas de posição depende do tipo de tecido, onde se encontra o ADNmt, já que a proporção de bases varia entre diferentes amostras, como sangue e cabelo, entre outros [26].

1.2.3. Aplicações do ADN mitocondrial

O mitogenoma, como já referido, apresenta diversas características que o diferenciam do genoma nuclear e o tornam uma ferramenta especial e útil na investigação.

A sua alta taxa de polimorfismo e mutação são vantajosas no estudo da genética antropológica e ciências forenses, entre outras áreas [27]. Os nucleótidos polimórficos presentes no ADNmt são úteis

para analisar variações dentro de uma espécie e identificar características específicas de populações, agrupadas em haplogrupos. Atualmente, os estudos do ADNmt, no âmbito de genética populacional, baseiam-se na sequenciação da região controlo, mais especificamente, nas três regiões hipervariáveis (HVI, HVII E HVIII) [28].

1.2.3.1. Genética Populacional

A herança estritamente materna e a ausência de recombinação por parte do ADNmt são características que contribuem para estudos de evolução e movimentos populacionais ao longo do tempo.

Existem duas teorias que explicam o mistério de como ocorreu a migração da humanidade e a sua diversificação pelo mundo: a hipótese de continuidade multirregional e o modelo “recentemente fora de África” [29]. A primeira propõe que o *Homo sapiens*, uma única espécie, saiu de África e começou a espalhar-se pelo mundo, há cerca de 1,5 milhões de anos. Esta teoria argumenta que as diferentes populações regionais estavam conectadas por fluxo génico contínuo, resultando na persistência de características esqueléticas em diversas regiões [29], [30]. Em contraste, a segunda, “*Out of Africa*” defende que o *Homo sapiens* surgiu em África há cerca de 100.000 anos e espalhou-se globalmente, substituindo outras espécies de *Homo* durante a sua expansão. De acordo com esta teoria, as morfologias regionais distintas fora de África desenvolveram-se após a saída do Homem do continente africano [29], [31].

Durante a evolução do Homem, as sequências de ADNmt de cada indivíduo acumularam mutações, resultando em diferentes haplótipos [32]. Quando os haplótipos são semelhantes entre si, compartilhando um ancestral comum, podem ser agrupados em haplogrupos [33]. Assim, o estudo das mutações no ADNmt possibilita a construção de árvores filogenéticas e a reconstrução dos padrões migratórios ao longo do tempo [7], [34]. A “*PhyloTree*” do ADNmt é a atual árvore filogenética de ADNmt e contém informações dos polimorfismos da região controlo e da região codificante do ADNmt [34]. Esta base de dados é permanentemente atualizada e já vai na versão 17 (<https://www.phylotree.org/>, consultado a 25-06-2024).

“*Mitochondrial Eve*” é, hipoteticamente, a mulher da qual todos os seres humanos são descendentes e o seu haplótipo é a raiz da árvore filogenética do ADNmt. O haplótipo da “*Mitochondrial Eve*”, haplótipo mtMRCA foi definido pela “*Reconstructed Sapiens Reference Sequence*” (RSRS) [7], [35], apartir de um grupo de indivíduos femininos, e estima-se que tenham surgido há aproximadamente 200 000 anos, em África [31].

Na origem desta árvore filogenética (Figura 1.1) está o macrohaplogrupo L, característico das populações de África-Subsariana, que se divide nos haplogrupos L0, L1, L2, L3, L4, L5 E L6 [34], [36], [37]. Foi a partir do haplogrupo L3, há aproximadamente 60 a 70 mil anos, que surgiram os macrohaplogrupos M e N, que estão presente nos continente Europeu e Asiático [38]. Os haplogrupos europeus, como H, I, J, K, T, U, V, W, e X, derivam do macrohaplogrupo N, representando cerca de 99% dos haplótipos europeus [17], [35].

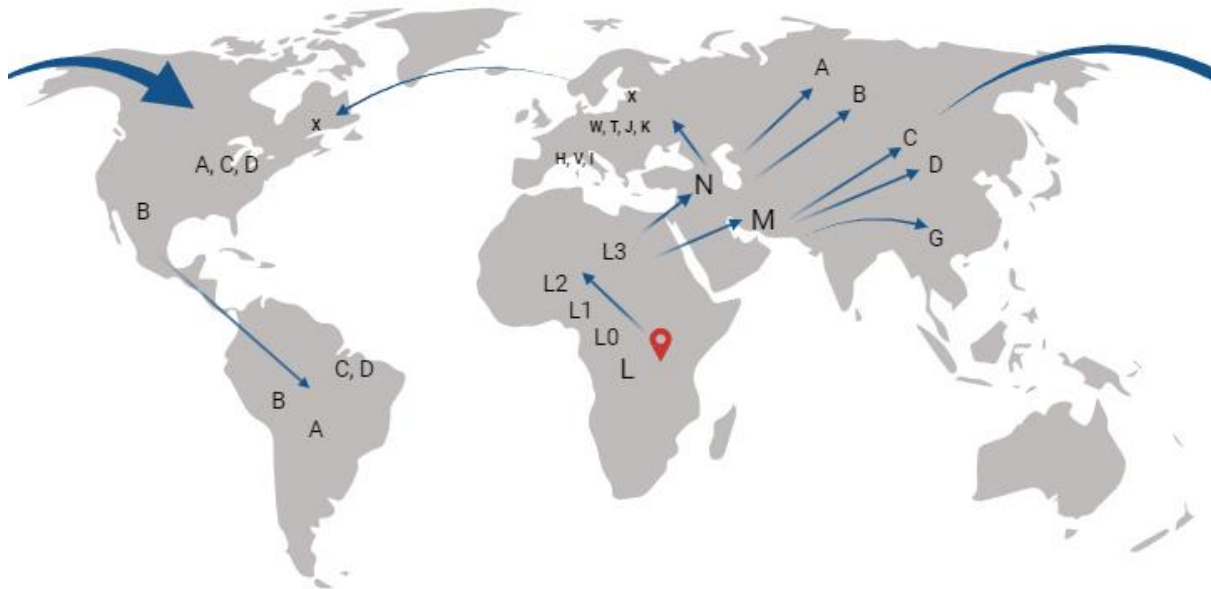


Figura 1.1 - Mapa da distribuição geográfica dos haplótipos do DNA mitocondrial humano (letras no mapa) deduzindo a origem dos humanos anatomicamente modernos e as possíveis rotas das migrações humanas “fora de África”. A localização a vermelho marca a origem da “Mitochondrial Eve” (Criada no BioRender)

Na região asiática, os haplogrupos A, B, F, e Y derivam do macrohaplogrupo N e os haplogrupos C, D, E, G e Z, derivam do macrohaplogrupo M [39]. Por fim, na América predominam os haplogrupos A, B, C, D e X, sugerindo que antes de ocorrer a colonização europeia, as populações asiáticas já haviam migrado para esta região [40].

1.2.3.2. Genética Forense

O ramo da genética forense é crucial na investigação criminal, empregando técnicas avançadas para analisar material genético com o objetivo de identificar e vincular indivíduos a cenas de crime. As amostras com ADN nuclear apresentam uma maior capacidade de discriminação, em comparação com o ADNmt e por isso, o uso de marcadores autossômicos na definição de perfis genéticos, como os *short tandem repeats* (STRs) e os *single nucleotide polymorphisms* (SNPs), é comum em investigação forense.

No entanto, nas amostras biológicas que apresentam o ADN nuclear degradado e/ou quantidade insuficiente, a análise do ADNmt torna-se uma ferramenta essencial. O elevado número de cópias aumenta a probabilidade de produzir dados para o *locus* do ADNmt em amostras comprometidas, como restos esqueléticos e cabelos soltos [41], [42]. A herança materna do ADNmt é outra característica particularmente útil em aplicações forenses, visto que indivíduos da mesma linhagem materna possuem haplótipos de ADNmt idênticos (excetuando-se eventos de mutação genética), até mesmo parentes distantes (da mesma linhagem materna), podem servir como referência adequada para comparação de ADNmt em casos de pessoas desaparecidas. Contudo, a análise de ADNmt também impossibilita a definição da identidade individual de um indivíduo [43].

Um dos casos forenses mais conhecidos é o caso da família Romanov. Os restos mortais da família imperial russa, assassinada em 1918, foram identificados através da análise de ADNmt, para conectar os restos da czarina a um parente materno distante, o príncipe Philip, do Reino Unido [44].

Em casos de investigação, na genética forense, a *revised Cambridge Reference Sequence* (rCRS) é uma sequência de referência do ADNmt utilizada para identificar variantes presentes nas amostras, com o objetivo de relatar novos haplótipos. Esta sequência é uma revisão da primeira sequência de ADNmt definida, a *Cambridge Reference Sequence* (CRS), que foi estabelecida a partir de uma mulher de ascendência europeia, e representa uma linhagem moderna na filogenia do ADNmt humano [45].

1.3. Base de dados

O aumento do uso do ADNmt em estudos científicos, especialmente em contextos forenses, destaca a importância de criar bases de dados mitocondriais para melhorar a qualidade e padronizar os resultados [46]. A criação das bases de dados ajuda a minimizar possíveis erros que aconteçam na metodologia e posteriormente na análise. Atualmente, existem várias bases de dados especializadas, como a EMPOP [47], o MITOMAP [48], o mtDB [49], a MitoVariome [50], o MitBase [51] e o MitoBreak [52].

A EMPOP é considerada a mais conceituada internacionalmente, no âmbito da genética forense, e procura aumentar a qualidade e confiança das sequências de ADNmt, submetendo todos os haplótipos a um rigoroso controlo de qualidade antes da sua inclusão na base de dados [47], [53]. Desde 2000, que a EMPOP (www.empop.org) tem tido um crescimento significativo no número de haplótipos armazenados, tendo registado 34 617 haplótipos em 2017 [47].

Em suma, a atualização desta base de dados mitocondriais é importante para avanços na análise de ADNmt em diferentes contextos.

1.4. Sequenciação de ADN mitocondrial

A análise do ADNmt é comumente realizada por meio da sequenciação. Atualmente, existem numerosas modificações e avanços na metodologia, que foram adaptados para o uso da análise de sequências, incluindo sequenciação de Sanger e sequenciação de segunda geração.

1.4.1. Sequenciação de Sanger

A sequenciação de Sanger (SS) é um método comumente utilizado para determinar sequências de ácidos nucleicos, desenvolvido por Frederick Sanger na década de 1970, tendo portanto mais de 40 anos [54].

Na técnica de sequenciação de Sanger, o ADN amplificado ou ADN complementar (ADNc) é emparelhado com um *primer* oligonucleotídico. Em seguida, a sequência é estendida pela enzima ADN polimerase, que adiciona desoxinucleotídeos trifosfatos (dNTPs: dATP, dGTP, dCTP, dTTP) ou didesoxinucleotídeos trifosfatos, que são terminadores de cadeia (ddNTPs: ddATP, ddGTP, ddCTP, ddTTP) [54], [55]. A diferença entre dNTPs e ddNTPs é que ao usar o segundo, a reação de extensão é interrompida, porque estes nucleotídeos não permitem que a cadeia continue a crescer, permitindo obter fragmentos de ADN de diferentes tamanhos [55].

A deteção de fragmentos marcados ocorre por meio de eletroforese capilar (CE), onde os fragmentos migram em velocidades distintas dependendo do seu tamanho (ou seja, fragmentos menores migram mais rápido do que fragmentos maiores). À medida que os fragmentos migram por um gel ou polímero, um laser estimula os fluoróforos ligados às moléculas de ddNTP, e a luz emitida é registada por um detetor. O comprimento de onda da emissão é então usado para identificar a base numa posição específica na sequência [7].

Contudo, existem algumas limitações na análise de sequências de baixa qualidade nos primeiros 15-40pb, devido à ligação do *primer* e à incapacidade de distinguir diferenças de pares de bases únicos em segmentos mais longos. Por isso, apesar do método de Sanger ser considerado padrão de ouro, os avanços genômicos permitiram uma nova sequenciação de alto rendimento, com aumento de sensibilidade e custos reduzidos [7], [56].

1.4.2. Sequenciação de Nova Geração

A Sequenciação de Nova Geração (NGS), também conhecida como sequenciação massiva em paralelo (SMP), ou sequenciação de segunda geração, é semelhante à sequenciação de Sanger. Ambos envolvem a sequenciação de fragmentos de ADN, embora, no NGS várias cadeias de ADN sejam sequenciadas em paralelo, tendo um maior rendimento e permitindo análise do genoma mitocondrial na sua totalidade na mesma reação (aprofundadas em níveis sem precedentes) [57].

Na análise de ADN, o NGS oferece a capacidade de interpretar amostras biológicas comprometidas, seja por misturas ou degradação, de maneira eficaz. Esta metodologia inovadora apresenta diversas vantagens, como a obtenção de informações genéticas mais detalhadas, multiplexação de diferentes tipos de marcadores, como SNPs e STRs, e a análise simultânea de múltiplas amostras [57]. Além disso, possibilita a análise de amostras forenses desafiadoras, utilizando fragmentos de leitura curtos que são adequados para amostras altamente degradadas e/ou danificadas, resultando num aumento na sensibilidade com centenas a milhares de leituras cobrindo a região alvo [7].

Existem diferentes sistemas nos métodos de detecção e geração de dados, após a sequenciação, que podem ser usados pela sequenciação de nova geração. Entre eles, a plataforma *Illumina* utiliza uma tecnologia de sequenciação por síntese com terminadores de corante reversível. [7] A Pirosequenciação, também, é um sistema de sequenciação por síntese, em que as cópias de sequências de ADN estão ligadas a esferas em uma placa *PicoTiter*. Nessa placa, nucleotídeos são adicionados ciclicamente, e cada vez que um nucleotídeo se liga é emitido um sinal luminoso [58].

Neste trabalho, é utilizada a tecnologia *Ion Torrent* que emprega sequenciação de semicondutores de iões, por meio de uma preparação química especializada chamada tecnologia *Ion AmpliSeq*, que é focada na sequenciação direcionada de fragmentos previamente amplificados. [59].

Considerando o ADNmt, estudos demonstraram que a NGS oferece uma abordagem custo-efetiva e de alto rendimento para gerar rapidamente dados completos de sequência do mitogenoma a partir de amostras de alta qualidade [60], [61]. Esta abordagem permite a detecção simultânea de mutações pontuais do ADNmt e grandes pontos de interrupção de deleção [62].

1.5. Validação dos métodos de sequenciação de primeira e segunda geração.

Na Delegação do Sul (DS) do Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF) são utilizados dois sequenciadores automáticos de primeira geração que empregam a tecnologia de eletroforese capilar: o *Applied Biosystems™ 3500 Genetic Analyser*, com 8 capilares e o *Applied Biosystems™ SeqStudio™ Genetic Analyzer*, com 4 capilares.

Atualmente, o INMLCF está em transição da metodologia de sequenciação de Sanger, que representa a sequenciação de primeira geração, para a sequenciação de nova geração, pertencente à segunda geração de sequenciação. Para essa mudança está a ser utilizado o sequenciador *Ion GeneStudio™ S5 System (Thermo Fisher Scientific)*, que realiza a sequenciação através da detecção de

alterações de pH provocadas pela libertação de iões de hidrogénio (H⁺) durante a polimerização do ADN e incorporação.

Cada um desses equipamentos tem características distintas e, antes de serem utilizados com segurança, devem ser previamente validados internamente. Esse processo de validação é essencial para determinar a eficácia e confiabilidade dos equipamentos na análise de dados.

2. OBJETIVOS

O presente estudo têm três objetivos. O primeiro é a validação da sequenciação da Sanger para verificar a sua especificidade na amplificação de amostras humanas e na análise de dados. O segundo é a validação do NGS, através da comparação de resultados entre NGS e a sequenciação de Sanger, da região controlo do ADNmt. Por último, o terceiro objetivo consiste na caracterização genética da população do Sul de Portugal e na definição de haplogrupo de cada amostra com mais precisão.

Desta forma, este trabalho constitui uma mais valia para o Serviço de Genética e Biologia Forense - Delegação do Sul (SGBF-DS) do INMLCF, sendo também importante para a sociedade, uma vez que, após a obtenção dos haplótipos de cada indivíduo, estes serão submetidos à EMPOP [47], a base de dados internacional de ADNmt mais relevante na área das ciências forenses, proporcionando resultados mais precisos e fiáveis para a análise de ancestralidade das amostras selecionadas. Esta análise fornecerá novos conhecimentos sobre as migrações das populações e sobre a história demográfica da região do Sul de Portugal.

3. METODOLOGIA

3.1. Seleção das Amostras

Neste estudo foram selecionadas 155 amostras de habitantes portugueses, residentes na região do Sul de Portugal. As amostras utilizadas para o estudo compreendem-se por zaragatoas bucais, provenientes de casos de investigação de parentesco biológico, que decorreram no SGBF-DS do INMLCF, entre 2012 e 2018. A Delegação do Sul do INMLCF, sediada em Lisboa, é responsável pela cobertura da população do Sul de Portugal, incluindo as regiões desde a Grande Lisboa até ao Algarve (Faro). Das amostras selecionadas, 75 pertenciam a indivíduos do género masculino e 80 a indivíduos do género feminino. Todas as amostras utilizadas foram previamente anonimizadas, impedindo a associação aos indivíduos.

Para a validação da especificidade da sequenciação de Sanger foram selecionadas 10 amostras todas de origem animal vertebrado não humano, anteriormente extraídas e purificadas

3.2. Sequenciação de Sanger

3.2.1. Amplificação:

Para a amplificação da região controlo total do ADNmt, que inclui as regiões HVI, HVII e HVIII, foram utilizados dois pares de *primers*. Para amplificar a região HVI, que se estende da posição 16024 até 16569, foi utilizado o par de *primers* L15971 (*forward*) e o H016 (*reverse*). O par de *primers* L16555 (*forward*) e H639 (*reverse*) foi utilizado para amplificar a região entre as posições 1 e 576, que corresponde às regiões HVII e HVIII.

Para cada amostra foram realizadas duas reações de amplificação, com um volume final de 10µL. Em cada reação, foram adicionados 5µL de *HotStarTaq Plus Master Mix*, 2,5µL de H₂O, 1µL de cada *primer* e 1,5µL de cada amostra. O programa do termociclador seguiu as seguintes condições cíclicas: 95°C durante 5 minutos; 35 ciclos de 94°C durante 30 segundos, 60°C durante 1 minuto e 30 segundos e 72°C durante 1 minuto, finalizando com uma extensão a 72°C durante 10 minutos.

3.2.2. Purificação dos produtos amplificados (*ExoSap*):

Após a amplificação, é realizada a purificação dos produtos amplificados para remover os *primers* e nucleótidos em excesso. Este passo envolve uma digestão enzimática utilizando o “*ExoSAP-IT™ PCR Product Cleanup*” (*Applied Biosystem™*) [63].

Para a purificação dos produtos amplificados foram adicionados 4µL de *ExoSAP-IT™* diretamente a cada 10µL de amostra amplificada. As amostras foram então colocadas no termociclador com o seguinte programa: 37°C durante 15 minutos, seguido de 80°C durante 15 minutos para inativação do reagente *ExoSAP™* [63].

3.2.3. Sequenciação:

No final da purificação procedeu-se a sequenciação direta dos fragmentos amplificados. Para este passo, foi utilizado o kit comercial *BigDye™ Terminator v3.1 Cycle Sequencing Kit* e *primers* de sequenciação específicos para os segmentos a serem analisados [64].

Para a sequenciação das amostras purificadas foram adicionados os seguintes componentes a cada poço de uma placa: 2µL de *Big Dye™ Terminator v3.1*, 1µL de tampão *BetterBuffer*, 1µL de *primer forward* ou *reverse*, e 1µL de amostra amplificada e purificada, ficando com o volume final de 5µL. De seguida, colocou-se a placa no termociclador, com a programação: 96°C durante 2 minutos; 35 ciclos, a 96°C durante 15 segundos, 50° durante 9 segundos, e 60°C durante 2 minutos; finalizando com 60°C durante 10 minutos.

3.2.4. Purificação BigDye Xterminator:

No final da sequenciação, foi realizada uma limpeza dos produtos sequenciados para remover todos os terminadores não incorporados e outros componentes que poderiam interferir nos resultados da eletroforese capilar, uma vez que estes terminadores estão marcados com fluorescência.

Este processo foi realizado com o *BigDye Xterminator™ Purification Kit* [65]. Em cada poço da placa de sequenciação foram adicionados 16,4µL de *SAM™ Solution*, 3,7µL de *Xterminator™ Solution Buffer* e 30µL de H₂O. Em seguida, a placa foi selada e colocada no vortéx a 18000 rpm durante 30 minutos.

Após esse período, a placa foi retirada do vortéx e centrifugada por 2 minutos. Finalmente, a placa foi colocada no *SeqStudio™ Genetic Analyzer* para a análise.

3.3. Sequenciação de Nova Geração (NGS)

3.3.1. Extração e Quantificação de ADN das amostras selecionadas

Na extração de ADN foi utilizado o kit *PrepFiler Express BTA™ Forensic DNA Extraction*. Previamente à extração foi preparada uma solução composta por 220µL de tampão, 3,5µL de proteinase k e 1,5µL de DTT, a ser aplicada em cada amostra. Durante a lise celular, as amostras foram incubadas

a 56°, numa rotação a 750 rotações por minuto (rpm) por 40 minutos. Após a incubação, as amostras foram transferidas para tubos de centrifugação e posteriormente colocadas no robô de extração, *AutoMate Express*.

Para a quantificação de ADNmt foi utilizado o kit *Quantifiler™ Trio DNA Quantification*, de acordo com as recomendações do fabricante (*Quantifiler Trio*, 2014) e o equipamento 7500 *Real-Time PCR*. A análise com *software HID Real-Time PCR Analysis v1.2*, permitiu determinar os valores de quantificação através da comparação dos valores C_t (*Cycle threshold*) com a reta padrão, construída com os valores de amostras *standard*.

3.3.2. Construção e Quantificação das Bibliotecas

Neste estudo foi utilizado o protocolo *Precision ID mtDNA Whole Genome Panel with HID Ion S5™ / HID Ion GeneStudio™ S5 System: Library Preparation on the HID Ion Chef™ System* [66], em que a preparação de bibliotecas da sequenciação de ADN em plataformas de NGS foi realizada de forma automática com o *Ion Chef™ System*.

No que toca à construção destas bibliotecas, foi utilizado o kit *Precision ID DL8*. Foram selecionadas 8 amostras de cada vez e pipetadas para uma placa *Precision ID IonCode Barcode adapters* – as primeiras 32 amostras com uma concentração final de 0,3ng/μL de ADNg (ADN genómico), as restantes com uma concentração final de 0,1ng/μL de ADNg e com um volume final de 15μL. De seguida, a placa foi centrifugada e colocada no *Ion Chef™*.

Posteriormente à inserção de todos os reagentes e consumíveis no *Ion Chef™*, representado na Figura 3.1 o equipamento foi programado para o *Precision ID mtDNA Whole Genome*, com duas “*pools*” de “*primers*”, vinte e dois ciclos de amplificação e quatro minutos de “*anneal*” e “*extension time*”. Após aproximadamente sete horas, terminada a construção de oito bibliotecas (com as amostras associadas ao seu IonCode), o tubo onde esta se encontra é retirado, identificado e reservado no congelador.

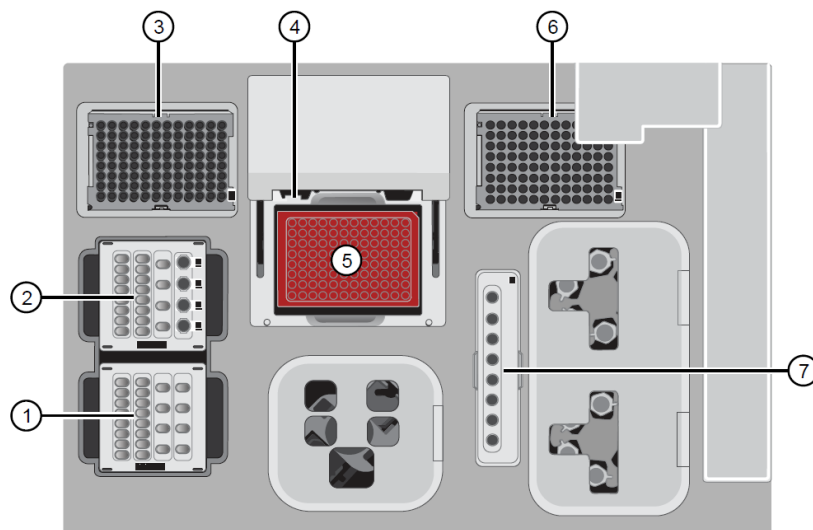


Figura 3.1 - Representação do equipamento *HID Ion Chef™* carregado para realização da pool de bibliotecas, retirada do manual “*Precision ID mtDNA Whole Genome Panel with HID Ion S5™*”. Utilizando o kit *Precision ID DL8*: 1- Embalagem de *Precision ID DL8 Solutions*; 2 – Embalagem de *Precision ID DL8 Reagents*; 3 – Embalagem de pontas *L8 Ion AmpliSeq™*; 4 – Selo de Alumínio Moldado para a *PCR*; 5 – Placa *Precision ID IonCode Barcode adapters* com 96 poços; 6 – Embalagem de pontas vazias; 7 – Cartucho de Enriquecimento.

A quantificação das bibliotecas, foi realizada por qPCR usando o kit *Ion Library TaqManTM Quantification*. Esta técnica permite determinar a quantidade de material genético presente na amostra, através da comparação do valor de C_t , ciclo em que a quantidade de material genético ultrapassa a *threshold line*. A comparação é feita com uma reta padrão, construída pelos valores de C_t , de diluições seriadas de uma amostra *standard* de quantidades conhecida.

3.3.3. Preparação do *Template* (carregamento do *chip*)

Antes da preparação do *template* (outra palavra para PCR), e uma vez que a qualidade de sequenciação de dados depende da concentração inicial correta da biblioteca, procedeu-se à sua diluição. Nesse sentido, algumas bibliotecas foram diluídas para uma concentração final de 30pM e com um volume final de 25 μ L e outras bibliotecas para uma concentração final de 20pM e com um volume final de 50 μ L.

Após a diluição foi utilizado o kit *Ion S5TM Precision ID Chef* para preparar o *template*. Todo o processo é automatizado no *Ion Chef System* e consiste, sucintamente, num PCR de emulsão (ePCR) onde ocorre a ligação das bibliotecas, das 32 amostras, a partículas (*Ion Spheres Particules* (ISP); de seguida, as partículas foram carregadas nos poços do *Ion 530TM Chip*. Por fim, o *chip* é retirado e reservado.

3.3.4. Sequenciação de Semicondutores de Iões (chip carregado)

Para a sequenciação foi utilizado o kit *Ion S5 Precision ID Sequencing*. O *Ion 530TM Chip* foi colocado no *Ion GeneStudioTM S5 System*, juntamente com o *Ion S5 Precision ID Sequencing Reagents* e o *Precision ID Wash Solutions* e o *Precision ID Cleaning Solutions*, para iniciar a sequenciação. No fim da sequenciação, o software *Torrent SuiteTM* foi usado para revisão de resultados.

3.3.5. Análise de dados (Bioinformática)

Durante a análise de dados em NGS, os resultados no sequenciador *Ion GeneStudioTM S5 System* são analisados com *software Torrent SuiteTM*. Estes resultados, provenientes do sequenciador, são condensados e resumidos para serem todos alinhados. De seguida, os resultados são exportados e analisados para o *software ConvergeTM*.

Na análise primária, o *software Torrent SuiteTM* permite fazer um pré-alinhamento à sequência de referência, rCRS, analisando a qualidade do *template* no *chip*, tal como a qualidade da sequenciação. Este *software* atua como um portal *web host* para revisão e distribuição de dados. É possível, ainda, obter um relatório síntese de toda a operação, sobre as ISPs e apresentar os resultados em ficheiros BAM (*Binary Alignment Map*).

Para detetar variações nas sequências, em comparação com a rCRS, e identificar com precisão a diversidade genética, através da sequenciação total do mitogenoma na análise secundária, retirou-se proveito da flexibilidade do *software ConvergeTM*, e dos seus programas *Applied BiosystemsTM Precision ID mtDNA Control Region* e *Applied BiosystemsTM Precision ID mtDNA Whole Genome*. Neste *software*, os ficheiros BAM são mapeados em *PhyloTree* e realinhados usando um algoritmo personalizado de alinhamento *Smith-Waterman* que integra informações de *PhyloTree* e EMPOP. As sequências incluídas neste programa estão em conformidade com as diretrizes do *European DNA Profiling Group* (EDNAP) [67] e as bases nucleotídicas foram codificadas seguindo o código apresentado no Quadro 3.1:

Quadro 3.1- Código IUPAC utilizado na nomenclatura das bases nucleotídicas de ADNmt.

Código IUPAC	Base Correspondente	Código IUPAC	Base Correspondente
A	Adenina	Y	Citosina ou Timina
C	Citosina	K	Guanina ou Timina
G	Guanina	M	Adenina ou Citosina
T	Timina	N	Adenina, Citosina, Guanina ou Timina
B	Citosina, Guanina ou Timina	V	Adenina, Citosina ou Guanina
D	Adenina, Guanina ou Timina	S	Guanina ou Citosina
H	Adenina, Citosina ou Timina	W	Adenina ou Timina
R	Adenina ou Timina		

3.4. Análise Filogenética

Para calcular a diversidade de sequência e o índice de diversidade nucleotídica da população do Sul, todos os haplótipos em estudo foram transferidos da sequência consenso determinada pelo *software ConvergeTM*, em formato FASTA. Foi selecionando apenas a região controlo (aproximadamente, 1122 pbs) de cada sequência. De seguida foram alinhados com o *software* MAFFT, com as definições “--retree 2 --maxiterate 1000”, as restantes definições foram deixadas no valor padrão. O ficheiro (.fasta) resultante do alinhamento foi convertido para o formato Arlequin, com o auxílio do *software PGD Spider v2.1.1.5*. No fim foi realizada a análise intrapopulacional no *software Arlequin v3.5.2.2*.

Foram selecionadas populações imigrantes em Lisboa, populações do Norte e Centro de Portugal e a população de Portugal (continental), para determinar as relações filogenéticas da população em estudo. Para estudar a variabilidade genética entre os grupos populacionais, realizou-se uma análise molecular da variância (AMOVA), utilizando o *software Arlequin v3.5.2.2*. Através deste *software* foram obtidos os valores do índice de fixação (F_{ST}) que avalia as distâncias genéticas interpopulacionais, e o valor de p , que indica o nível de significância (0,05).

Com base nos valores F_{ST} determinados e utilizando o programa *Neighbor* do *software Phylip v3.695*, foram calculadas as distâncias moleculares entre as sequências de ADNmt das populações analisadas. A árvore filogenética correspondente foi gerada e visualizada no programa *Treeview v1.6.6*.

4. RESULTADOS E DISCUSSÃO

4.1. Validação da Sequenciação de Sanger – Especificidade

No final da eletroforese capilar com o *SeqStudioTM Genetic Analyzer*, as 10 amostras de origem animal vertebrado não humano foram analisadas. A verificação da especificidade é crucial para garantir que os primers utilizados sejam específicos para o material genético que se pretende amplificar, neste caso, o ADNmt humano. Em contextos forenses onde, frequentemente, se recolhem amostras que podem conter misturas biológicas, a especificidade assegura que apenas as sequências de interesse sejam amplificadas e analisadas, evitando a inclusão de sequências indesejadas que poderiam comprometer a precisão dos resultados.

As amostras de animais foram submetidas aos processos de amplificação e sequenciação utilizando primers específicos para o ADNmt humano. Como esperado, nenhuma das amostras animais apresentou resultados, indicando que não ocorreu amplificação nem sequenciação. Este resultado confirma que os primers utilizados são específicos para o ADNmt humano, validando a sua aplicabilidade em genética forense com o instrumento *SeqStudio™ Genetic Analyzer*.

4.2. Análise das amostras de ADN mitocondrial – Sequenciação NGS

Após a sequenciação de semicondutores de iões, os resultados foram analisados no *software Torrent Suite™* que cria um relatório de pré-alinhamento sobre a análise da qualidade da sequenciação.

No *template* do chip 1, para a construção de bibliotecas as amostras foram preparadas para uma concentração final 0,3 ng/μL, para a preparação do *template* as bibliotecas foram diluídas para uma concentração final 30pM, com um volume final de 25μL, conforme indicado no protocolo [66]. O relatório de pré-alinhamento apresenta a densidade de carregamento das ISPs (Figura 4.1(A)), um resumo das condições dessas partículas (Figura 4.1(B)) e o comprimento das leituras geradas (Figura 4.1(C)). A Figura 4.1 (A) mostra a densidade de ISPs (91%), que depende da ligação de seqüências às ISPs, medida pela alteração de pH nos poços. Na Figura 4.1 (B), observa-se um gráfico cuja percentagem das barras inferiores (a azul e a cinzento) é correspondente aos valores da barra a azul imediatamente acima. De cima para baixo, a primeira barra, indica a densidade de poços ocupados e vazios. Na segunda barra, 100% das ISPs destes poços continham ADN. Na terceira barra, apesar de se pretender uma amplificação monoclonal do ADN em cada ISPs, 51% destas continham duas ou mais seqüências diferentes, ou seja, eram policlonais. E por fim, na quarta barra da percentagem de ISPs monoclonais (36% da biblioteca final) que passaram os filtros para detetar fragmentos policlonais, dímeros de primers, ou de baixa qualidade, foram utilizadas para o alinhamento. Ou seja, dos 91% poços do *chip* que podem conter ISPs (34,398,084 *poços*), apenas 18% correspondem a poços geradores de seqüências utilizáveis para análise (6,088,818 *reads*, ou leituras), sendo que o valor ideal para uma análise mais robusta é acima dos 30%. A Figura 4.1 (C) representa um histograma do comprimento da leitura das seqüências, apresentando uma média de 138 pbs de comprimento. De notar que, uma grande variação entre o número de leituras de cada amostra pode indicar um problema na qualidade da biblioteca, o que provavelmente está relacionado com o excesso de ADNmt [68].

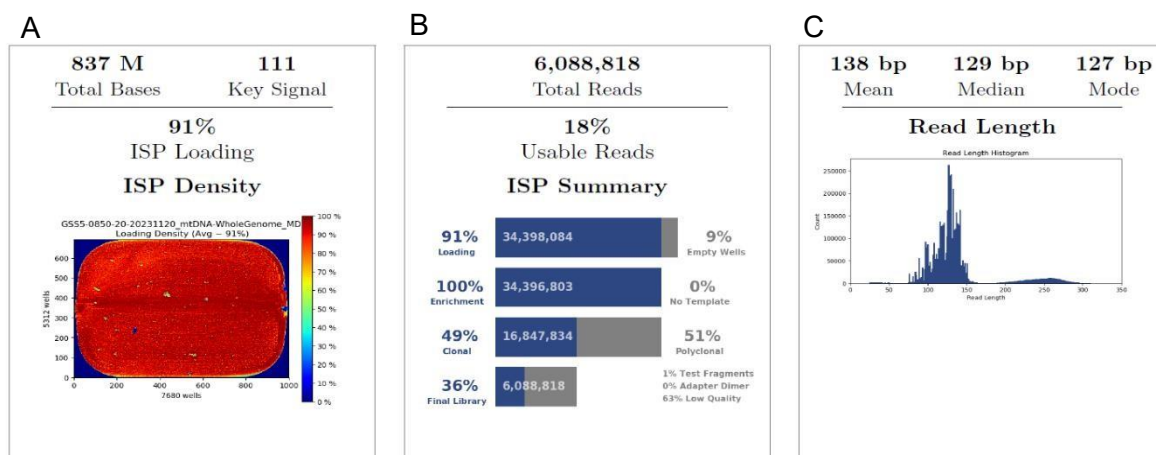


Figura 4.1 - Esquemas do relatório pré-alinhamento criado pelo *software Torrent Suite™* do *Ion 530™ Chip 1*. (A) Densidade de ISPs. (B) Resumo do estado das ISPs, seqüências utilizáveis. (C) Histograma do comprimento de leituras

O relatório de pré-alinhamento apresenta também um resumo do alinhamento das amostras com a rCRS, permitindo aferir a qualidade das leituras. Na Figura 4.2 (A) é visível o número total de bases alinhadas (732 milhões) e o número total de bases alinhadas em relação ao número de bases da sequência de referência (rCRS). Isto é, a cobertura da referência foi efetuada em média 43961,4 vezes

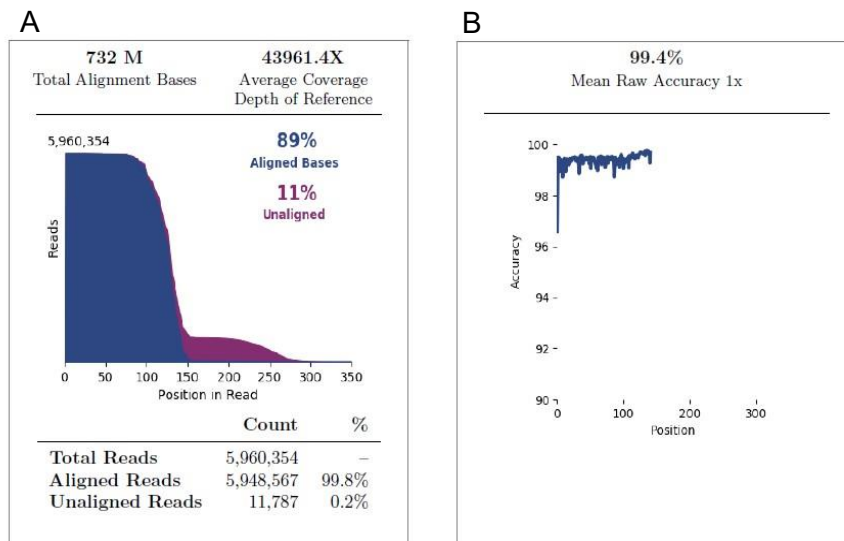


Figura 4.2 - Relatório do Alinhamento, relativo ao *Ion 530TM Chip 1* (alinhado a *PrecisionID_mtDNA-rCRS*). (A) Gráfico do alinhamento das leituras. (B) Precisão do alinhamento das leituras.

pelos bases nas leituras. Do número total de leituras geradas durante o processo de *basecalling* (atribuição de bases), 89% das bases encontram-se alinhadas com a sequência de referência. Na Figura 4.2 (B) evidencia-se a percentagem média de precisão do alinhamento para cada posição das leituras (99,4%). Este alinhamento é útil para verificar a qualidade da sequenciação e da biblioteca subjacente.

A criação do relatório supramencionado, ao final da sequenciação de cada *chip*, permite aperfeiçoar a definição dos parâmetros mais adequados à preparação de *templates* de ADNmt. Para tal, revelou-se essencial quantificar e controlar a concentração inicial das amostras.

No segundo *template*, as amostras foram preparadas com uma concentração final de 0.1ng/μL. Contudo, não tendo sido quantificadas com o kit *Ion Library TaqManTM Quantification*, as bibliotecas

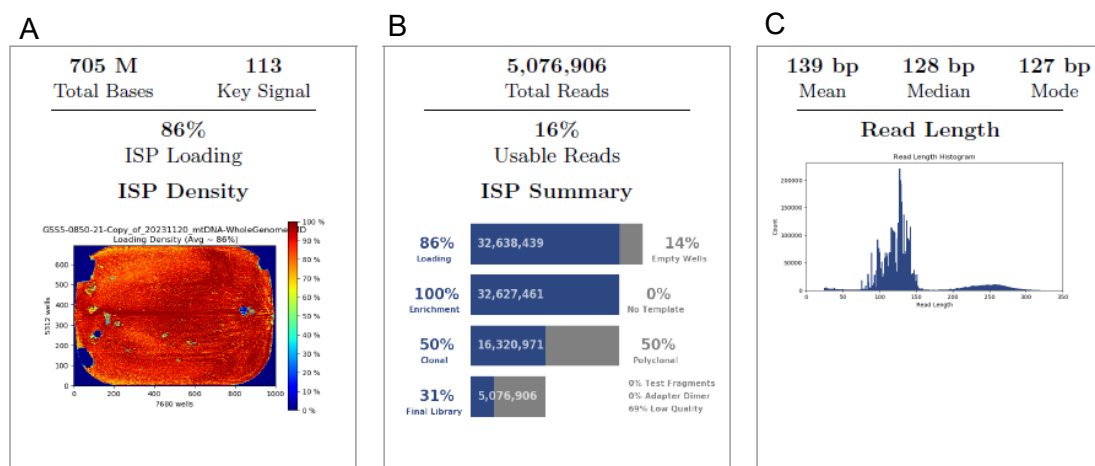


Figura 4.3 - Esquemas do relatório pré-alinhamento criado pelo *software Torrent SuiteTM do Ion 530TM Chip 2*. (A) Densidade de ISPs. (B) Resumo do estado das ISPs, sequências utilizáveis. (C) Histograma do comprimento de leituras.

revelaram-se detentoras de uma baixa percentagem de sequências utilizáveis (apenas 16%, correspondendo a 5,0760,906 *reads*), conforme mostrado na Figura 4.3 (B). Além disso, apenas 87% das sequências se alinharam à rCRS (Figura 4.4 (A)). Durante a análise com o *software Converge*TM, a

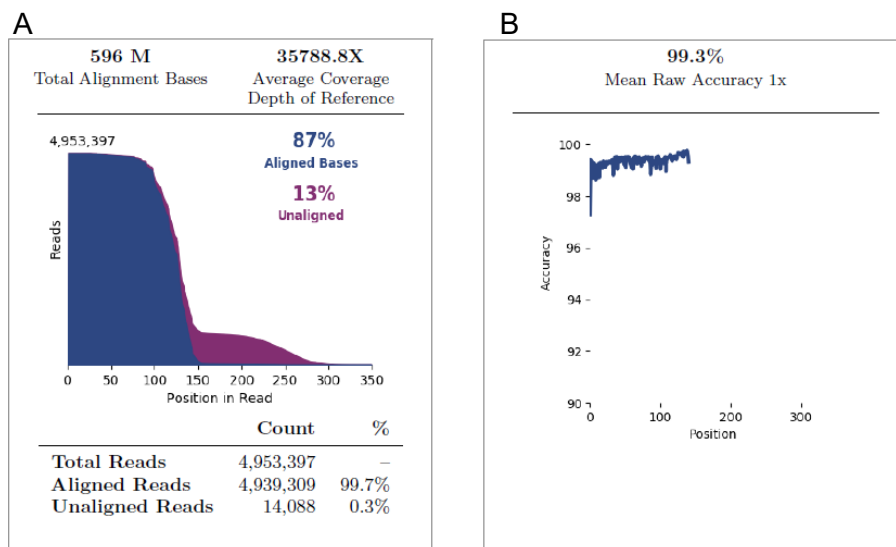


Figura 4.4 - Relatório do Alinhamento, relativo ao *Ion 530*TM *Chip 2* (alinhado a *PrecisionID_mtDNA-rCRS*). (A) Gráfico do alinhamento das leituras. (B) Precisão do alinhamento das leituras

ausência de quantificação das bibliotecas prejudicou os resultados provavelmente devido a concentrações desiguais entre bibliotecas, levando à sobreposição de resultados, à falta de cobertura e à sequenciação parcial da maioria das amostras.

Na preparação das bibliotecas seguintes, as amostras mantiveram uma concentração final 0,1ng/μL, e para a preparação dos *templates*, as bibliotecas, por estarem muito concentradas, foram diluídas até um volume final de 50μL e concentração final de 20pM.

A diluição das amostras e das bibliotecas melhorou significativamente a qualidade das sequências obtidas, em comparação aos dois *templates* anteriores. A diluição reduziu o número de sequência de baixa qualidade, aumentando o número de sequências utilizáveis (conforme demonstrado pela Figura 4.5). Além disso, as diluições permitiram obter um comprimento mais uniforme entre todas

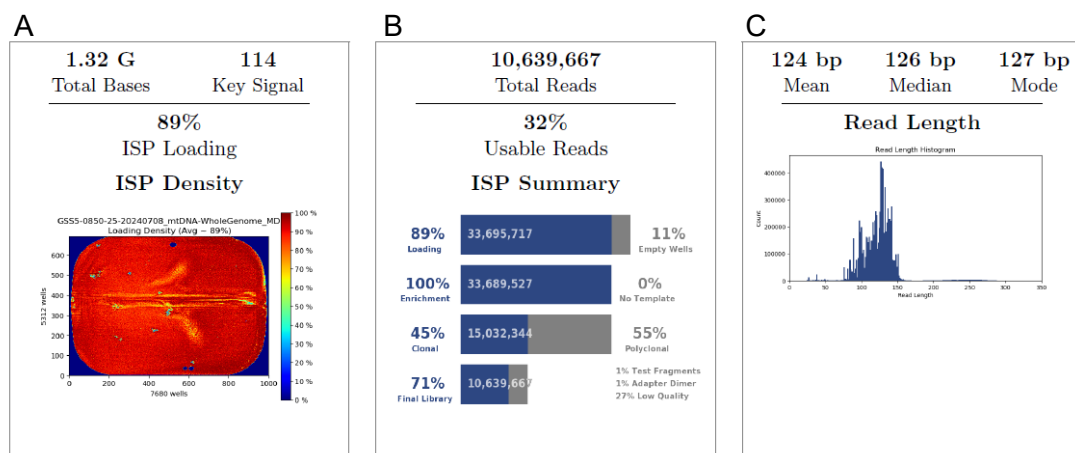


Figura 4.5 - Esquemas do relatório pré-alinhamento criado pelo *software Torrent Suite*TM do *Ion 530*TM *Chip 3*. (A) Densidade de ISPs. (B) Resumo do estado das ISPs, sequências utilizáveis. (C) Histograma do comprimento de leituras.

as sequências (Figura 4.6) o que foi crucial para aumentar a percentagem de alinhamento preciso com a rCRS.

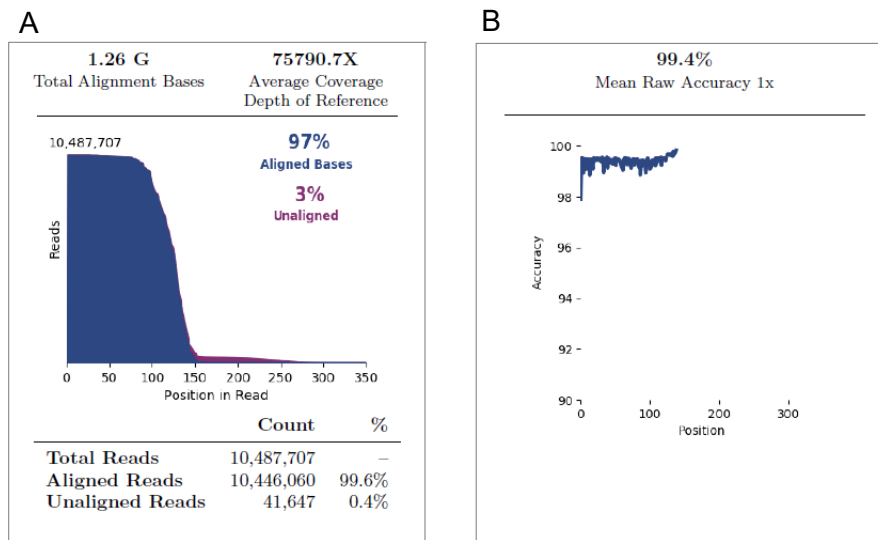


Figura 4.6 - Relatório do Alinhamento, relativo ao *Ion 530™ Chip 3* (alinhado a *PrecisionID_mtDNA-rCRS*). (A) Gráfico do alinhamento das leituras. (B) Precisão do alinhamento das leituras

Estes resultados estão de acordo com um estudo de sensibilidade do painel *Precision ID mtDNA Whole Genome*, que demonstrou que a diluição cuidadosa do ADN genómico pode, até certa medida, melhorar a cobertura completa do ADNmt e o alinhamento preciso com a rCRS. O estudo constatou que as amostras, com a diluição até 0.15pg de ADNg, alcançaram cobertura completa do ADNmt, e os replicados estavam em conformidade com a sequência conhecida, indicando que a diluição controlada pode minimizar os erros introduzidos por altas concentrações de ADN durante a amplificação [69].

4.3. Região controlo de ADN mitocondrial: Tecnologia NGS versus Sequenciação de Sanger

Para validar a metodologia NGS foi realizada uma comparação dos resultados da região controlo do ADNmt entre a sequenciação por NGS e a sequenciação de Sanger, das 155 amostras. Os resultados da região controlo do ADNmt por sequenciação de Sanger já haviam sido obtidos no SGBF-DS. Após a comparação, observou-se que os haplótipos na região controlo do ADNmt eram coincidentes entre as duas metodologias.

A comparação de oito resultados da região controlo por sequenciação de Sanger e NGS encontram-se descritos no Quadro 4.1. Neste quadro, observa-se uma uniformidade nos haplótipos, tendo em consideração apenas os polimorfismos da região controlo do ADNmt. No entanto, o estudo da região codificante do ADNmt permite detetar um maior número de polimorfismos característicos de haplogrupos mais refinados. Assim, é possível verificar na Quadro 4.1 que a mesma amostra pode ser classificada num haplogrupo mais detalhado e refinado quando analisados todos os seus polimorfismos. Por exemplo, a amostra GENPOP8603, com base apenas nos polimorfismos da região controlo, é classificada no haplogrupo K1a, mas ao incluir os polimorfismos da região codificante é classificada

no haplogrupo K1a4a1b1. Por outro lado, a amostra GENPOP8597 ao ser analisada na região controle ou em ambas as regiões, apresenta o mesmo haplogrupo.

Quadro 4.1 - Comparação do Haplogrupo nas amostras (GENPOP8594; GENPOP8595; GENPOP8596; GENPOP8597; GENPOP8599; GENPOP8600; GENPOP8601; GENPOP8603) de ADN mitocondrial com os dois métodos de sequenciação: Sanger VS NGS. (A azul estão as mutações localizadas na região controle e a preto estão as mutações localizadas na região codificante)

Amostras	Método	Haplogrupo	Haplótipo
GENPOP 8594	Sanger	R0	263G 315.1C 16189C 16519C
	NGS	H1	263G 315.1C 750G 1438G 3010A 4769G 8860G 9717A 15326G 16189C 16519C
GENPOP 8595	Sanger	U5b	73G 150T 195C 263G 309.1C 315.1C 16192T 16259T 16270T
	NGS	U5b1e	73G 150T 195C 263G 309.1C 315.1C 750G 1438G 2706G 3197C 4769G 5656G 7028T 7768G 8860G 9477A 10670T 11467G 11719A 12308G 12372A 13617C 14182C 14766T 15115C 15326G 16192T 16259T 16270T
GENPOP 8596	Sanger	R	73G 263G 315.1C 523del 524del
	NGS	H4a1a1a1a1	73G 263G 315.1C 523del 524del 750G 1438G 1738Y 3992T 4024G 4769G 5004C 8269A 8860G 9123A 10007C 10034C 10044G 10044G 10151del 13237a 14365T 14582G 15326G
GENPOP 8597	Sanger	L3b1a+@16124	73G 263G 315.1C 523del 524del 16223T 16278T 16362C 16519C
	NGS	L3b1a+@16124	73G 263G 315.1C 523del 524del 750G 1438G 2706G 3450T 4769G 5773A 6221C 7028T 8701G 8860G 9449T 9540C 10086G 10373A 10398G 10873C 11002G 11719A 12123T 12705T 13105G 13914A 14766T 15301A 15311G 15326G 15824G 15944del 16223T 16278T 16362C 16519C
GENPOP 8599	Sanger	H1ao	93G 146C 263G 315.1C 16278T 16519C
	NGS	H1ao	93G 146C 263G 315.1C 750G 1428G 3010A 4769G 8860G 11809C 11914A 15326G 16278T 16519C
GENPOP 8600	Sanger	U5a1	73G 263G 315.1C 16176T 16256T 16270T 16399G
	NGS	U5a1a1	73G 263G 315.1C 750G 1438G 1700C 2706G 16176T 16256T 16270T 16399G
GENPOP 8601	Sanger	T1a+152	73G 152C 195C 263G 309.1C 315.1C 16126C 16163G 16186T 16189C 16294T
	NGS	T1a1	73G 152C 195C 263G 309.1C 315.1C 709A 750G 1438G 1888A 2065G 2706G 4216C 4769G 4917G 7028T 7521A 8697A 8860G 9899C 10463C 11251G 11719A 12633A 13368A 14766T 14905A 15326G 15452A 15607G 15928A 16126C 16163G 16186T 16189C 16294T
GENPOP 8603	Sanger	K1a	73G 263G 309.1C 315.1C 497T 16224C 16311C 16519C
	NGS	K1a4a1b1	73G 263G 309.1C 315.1C 497T 750G 1189C 1438G 1811G 2706G 3480G 4769G 6260A 7028T 8098G 8860G 9055A 9698C 10398G 10550G 11299C 11467G 11485C 11719A 11840T 12308G 12372A 13740C 14167T 14766T 14798C 15217A 15326G 16224C 16311C 16519C

Durante a análise dos resultados, o *software Converge*TM demonstrou a capacidade de detetar com precisão heteroplasmias de posição (PHP) até 10% [70], algo que não era possível com a metodologia de Sanger. O haplótipo da GENPOP8919, determinado por sequenciação de Sanger, não apresenta qualquer substituição na posição 146. Com a metodologia NGS foi possível detetar uma substituição na posição 146 com uma frequência de 21%, passando a ser possível identificar a heteroplasmia de posição 146Y. De acordo com a notação do código IUPAC (Quadro 3.1), esta mutação mostra que o indivíduo, para esta posição pode apresentar o nucleótido T ou o nucleótido C.

Em outro caso, com a sequenciação de Sanger foi determinada a existência de uma substituição na posição 16183, na amostra GENPOP8542. Todavia, com a metodologia NGS, para a posição 16183, detetaram-se os polimorfismos 16183a e 16183M. Isto indica que para a posição 16183 há deleção do nucleótido A em algumas sequências e a presença do nucleótido C noutras sequências, verificando-se uma heteroplasmia de posição 16183M.

Estes resultados evidenciam que a sequenciação de NGS permite a sequenciação completa do mitogenoma e a deteção de todos os polimorfismos a diferentes frequências, resultando em haplótipos mais detalhados e numa determinação mais precisa e eficiente dos haplogrupos.

Os resultados da análise de dados da NGS, obtidos com o *software Converge*TM a partir dos arquivos BAM exportados pelo *software Torrent Suite*TM, foram mapeados no *PhyloTree* e, em seguida, realinhados utilizando um algoritmo de alinhamento personalizado *Smith-Waterman*, que integra informações de *PhyloTree* e EMPOP. Durante a avaliação dos haplótipos com a rCRS foram identificadas variantes nas sequências. A análise das variantes, com base na sua ocorrência no haplogrupo, permite a definição do haplogrupo. As variantes podem ser visualizadas no *software* em formato de tabela, linear ou circular [71].

4.4. Estudo Populacional

4.4.1. Estudo dos Haplogrupos

Com a metodologia NGS foi possível sequenciar o genoma mitocondrial total, obtendo haplótipos únicos. Isto evidencia 100% de variabilidade genética entre os indivíduos em estudo, da região do Sul de Portugal. Adicionalmente, na genética forense a sequenciação total da mitocôndria oferece um maior poder de discriminação da linhagem materna na atribuição de um haplogrupo mais específico [37].

Na análise de dados, o *software Converge*TM, que integra informações da árvore filogenética *PhyloTree* [34], permite a divisão dos haplótipos pelos respetivos haplogrupos. Cada haplogrupo é identificado por um carácter alfabético maiúsculo, seguido por um conjunto de caracteres numéricos e alfabéticos (minúsculos), permitindo assim uma subdivisão em subhaplogrupos. Para uma classificação precisa, a sequenciação total do genoma é fundamental. Adicionalmente, é também determinante que o haplótipo apresente todos os polimorfismos específicos que caracterizam o haplogrupo. Contudo, em algumas amostras foi observada a possibilidade de dois indivíduos pertencerem ao mesmo haplogrupo, embora apresentem haplótipos distintos. Por exemplo, as amostras GENPOP8594 e GENPOP8656 apesar de apresentarem haplótipos distintos pertencem ao haplogrupo H1, Quadro 4.2, devido à partilha do polimorfismo 3010A, característico desse haplogrupo.

Quadro 4.2 - Comparação dos haplótipos das amostras GENPOP8594 e GENPOP8656. (A negrito está marcado a variante característica do haplogrupo H1 e a sublinhado estão os polimorfismos partilhados pelas amostras)

Amostras	Haplótipo
GENPOP8594	16207G <u>16519C</u> 152C 263G 315.1C 750G 1438G 3010A 4769G 8860G 9717A 15326G
GENPOP8656	16189C <u>16519C</u> 263G 315.1C 750G 1438G 3010A 4769G 6755A 8860G <u>15326G</u>

As amostras GENPOP8694 e GENPOP8565 são outra demonstração, pois apesar de serem de indivíduos diferentes são classificadas no mesmo haplogrupo U4, caracterizado pelas variantes 195C, 499A, 1811G, 4646C, 5999C, 6047G, 11332T, 14620T, 15693C e 16356C, e apresentam haplótipos bastante semelhantes (Quadro 4.3):

Quadro 4.3 - Comparação dos haplótipos das amostras GENPOP8694 e GENPOP8565. (A negrito estão marcadas as variantes características do haplogrupo U4 e a vermelho estão os polimorfismos partilhados pelas amostras).

Amostras	Haplótipo
GENPOP8694	<u>16129A</u> 16356C <u>16519C</u> 73G 195C 196C 263G 315.1C 750G 1438G 1811G 2706G 4418C 4646C 4769G 5999C 6047G 7028T 7403G 7705C 8860G 11332T 11467G 11696A 11719A 12308G 12372A 14620T 14766T 15326G 15693C
GENPOP8565	<u>16129A</u> 16293G 16356C <u>16519C</u> 73G 195C 196C 263G 315.1C 499A 750G 1438G 1811G 2706G 4418C 4646C 4769G 5999C 6047G 7028T 7403G 7705C 8860G 11332T 11467G 11696A 11719A 12308G 12372A 13695T 14620T 14766T 15326G 15693C

Assim, a análise dos dados recolhidos com o *software Converge™* permitiu definir o haplogrupo de todas as amostras com precisão. A Figura 4.7 apresenta a frequência de cada haplogrupo presente no grupo de amostras, da população do Sul de Portugal. Em alguns casos, não foi possível obter cobertura de todo o genoma e, provavelmente, em algumas amostras o haplogrupo não é específico, porque não foi possível detetar todas as mutações.

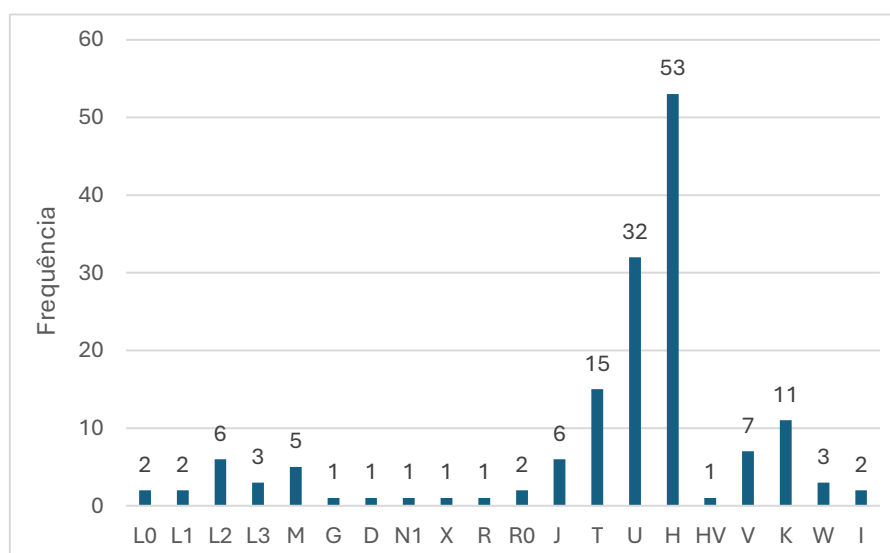


Figura 4.7 - Gráfico com os diferentes haplogrupos e respetivas frequências identificados nas 155 sequências de ADNmt de indivíduos da população do sul de Portugal

Como já foi mencionado, os haplogrupos característicos de populações na Europa são o H, I, J, K, T, U, V, W, e X, que derivam do macrohaplogrupo N [17, 39]. Neste estudo, o haplogrupo H está presente em 34,19% das amostras; o haplogrupo U está presente em 20,65%; o haplogrupo T em 9,68%; o K em 7,10%; os haplogrupos V em 4,52%; e o J em 3,87% das amostras. O haplogrupo W encontra-se em 1,94%; o haplogrupo I em 1,29%; e por último, o haplogrupo X está presente em 0,65% das amostras. Em conjunto, 83,87% das amostras são haplogrupos característicos da população europeia, o que reflete a elevada contribuição de linhagens europeias na composição genética da população analisada.

Neste estudo foram também encontradas 13 amostras (8,39%) classificadas no haplogrupo L (L0, L1, L2 e L3), característico de populações africanas. No haplogrupo M foram classificadas 5 amostras (3,23%) e no haplogrupo R0 foram classificadas 2 amostras (1,29%). As restantes amostras foram classificadas nos haplogrupos G, D, N1, R e HV, cada um com uma percentagem de 0,65%.

Cada haplogrupo foi analisado individualmente para identificar os polimorfismos característicos, tanto da região controlo, como da região codificante, e verificar a sua concordância com os polimorfismos definidos pela *PhyloTree*.

O haplogrupo mais ancestral da árvore filogenética de ADNmt é o haplogrupo L, que se subdivide em sete subhaplogrupos (L0-L6)[72]. Este haplogrupo apresenta uma alta frequência no continente africano. No entanto, devido à migração das populações africanas ao longo dos séculos, também é encontrado entre indivíduos de populações europeias, asiáticas e americanas [73]. Na *PhyloTree* o haplogrupo L0 é caracterizado pelas mutações 263G, 1048C, 3516a, 5442C, 6185C, 9042T, 9347G, 10589A, 12007A e 12720G. Contudo, neste estudo, as amostras classificadas no haplogrupo L0 exibiram apenas os polimorfismos 263G, 3516a, 5442C, 6185C, 9042T, 9347G, 10589A, 12007A e 12720G. As amostras classificadas no haplogrupo L1, apresentavam todas as mutações características do haplogrupo, que é definido pelos polimorfismos 3666A, 7055G, 13789C, 14178C e 14560A. O haplogrupo L2 é descrito pelas mutações 146C, 150T, 152C, 2416C, 8206A, 9221G, 10115C e 13590A, 16311T e 16390A. Todas as amostras classificadas no haplogrupo L2 mostraram ser coincidentes com os polimorfismos definidos na *PhyloTree*, exceto na posição 16311, onde não foi detetada nenhuma alteração em nenhuma amostra. O haplogrupo L3, na *PhyloTree*, apresenta os polimorfismos 769G, 1018G e 16311T, e entre os seus diversos subhaplogrupos, neste estudo, destacaram-se os subhaplogrupos L3f, L3b e L3e, que apresentam respetivamente as seguintes mutações: 3396C, 4218C, 15514C e 16209C; 3450T, 5773A, 6221C, 9449T, 10086G, 13105G, 13914a, 15311G, 15824G, 16124C, 16278T e 16362C e; 2352C e 14212C. As amostras classificadas no haplogrupo L3, apresentaram os polimorfismos que caracterizam os subhaplogrupos L3f, L3b e L3e, exceto o 16124C na amostra do subhaplogrupo L3b. Particularmente na Península Ibérica, a presença dos haplogrupos L em populações europeias já foi documentada em estudo anteriores, o que pode ser explicado pelo tráfico de escravos de africanos desde meados do século XV até ao século XIX [73], [74].

Os macrohaplogrupos M e N, presentes nas populações euroasiáticas, surgiram há aproximadamente 65 mil anos do haplogrupo L3 [38]. O haplogrupo M é composto por diversos subhaplogrupos característicos das populações da Ásia Meridional e Oriental [75] e uma das explicações da sua presença nas populações europeias é a migração, no final do Pleistoceno, dos humanos modernos da Índia [76]. Na *PhyloTree*, o macrohaplogrupo M é caracterizado pelos polimorfismos 489C, 10400T, 14783C e 15043A. Contudo, foi realizado um estudo com o objetivo de estudar a variação e distribuição do macrohaplogrupo M e foi observado que este se divide no haplogrupo M asiático, caracterizado pela *PhyloTree*, e o haplogrupo M africano, que apresenta mais

uma variante (10873C) [38]. No presente estudo, 5 amostras foram classificadas nos subhaplogrupos M1, M3 e M27. Um estudo revela a presença do haplogrupo M em indivíduos de etnia cigana nas populações ibéricas, o que pode indicar uma origem cigana, embora o haplogrupo seja predominante nas populações da Ásia Meridional e Oriental [77].

Os haplogrupos G e D, que divergiram do haplogrupo M, são predominantes nas populações do leste asiático. Entre as amostras existe um indivíduo, para cada um destes haplogrupos. O haplogrupo G, na *PhyloTree*, é determinado pelos polimorfismos 709A, 4833G, 5108C e 16362C e neste estudo a amostra classificada no subhaplogrupo G1 apresenta apenas a mutação 4793G. Outros polimorfismos característicos deste haplogrupo não foram identificados, provavelmente devido à sequenciação parcial da amostra. O haplogrupo D é caracterizado pelos polimorfismos 5178A e 16362C, que também são observados na amostra classificada neste haplogrupo. A presença dos haplogrupos G e D em indivíduos da população europeia sugere ascendência do leste asiático [78].

O macrohaplogrupo N teve origem no haplogrupo L3 e divergiu para os haplogrupos A, I, S, W, X, Y e macrohaplogrupo R, estando estes distribuídos pelas populações europeias, bem como nos continentes asiático, americano e na Oceânia. [17], [31]. Neste estudo, 7 indivíduos foram classificados nos haplogrupos N1, I, X e W e apresentaram os seguintes polimorfismos: 10238C e 12501A; 10034C e 16129A; 6221C, 6371T, 13966G, 14470C 16189C e 16278T; 195C, 204C 207A, 1243C, 3505G, 5460A, 8251A, 8994A, 11947G, 15884c e 16292T. Estas mutações estão de acordo com as variações dispostas na *PhyloTree*. Estes haplogrupos encontram-se presentes na Euroásia Ocidental e sul da Ásia, com exceção dos haplogrupos N1 e X que podem ser encontrados, respetivamente, no continente africano e americano [79].

O haplogrupo R, originado a partir do macrohaplogrupo N e vinculado às populações euroasiáticas, está amplamente distribuído na Europa, Ásia e no Norte de África, incluindo áreas como a Tunísia e Marrocos. Este haplogrupo originou os subhaplogrupos R0, HV e V, entre outros [80]. Das amostras analisadas, 11 indivíduos foram classificados nos haplogrupos R, R0, HV e V, que na *PhyloTree* apresentam as respetivas variantes: 73A e 11719G; 64T, 2442C e 16362C; 2706G e 7028T; e 4580A. Neste estudo, todas as amostras apresentavam as variantes características dos haplogrupos, dispostas na *PhyloTree*.

O haplogrupo H, o mais comum na Europa, é um descendente do haplogrupo HV [81], representando quase 50% das linhagens mitocondriais europeias. Acredita-se que este haplogrupo tenha surgido no Médio Oriente há cerca de 30-35 mil anos e se tenha expandido para o sul da Europa há cerca de 20 mil anos. Este haplogrupo também é encontrado no Norte de África, Médio Oriente e em menor frequência no Nordeste da Índia e Ásia Central [82]. A sua distribuição geográfica sugere uma conexão com a expansão dos agricultores neolíticos que repovoaram a Europa após o último máximo glacial, corroborada pela sua presença em populações no Norte de África e Médio Oriente. [81], [83]. Este haplogrupo é um marcador genético importante para compreender a história europeia e a expansão humana. O haplogrupo H possui vários subhaplogrupos, como H1, H2, H3, entre outros, sendo o H1 o mais prevalente na Europa.

No presente estudo, o haplogrupo H mostrou ser o mais frequente, tendo-se identificado 53 indivíduos divididos entre os subhaplogrupos H1, H2, H3, H4, H5, H6, H7, H10, H17, H30 e H31, que apresentam os respetivos polimorfismos: 3010A; 1438A; 152C e 6776C; 3992T, 5004C, 9123A; 456T e 16304C; 239C, 16362C e 16482G; 4793G; 14470A; 3915A e 16129A; 8200C; 146C, 195C, 7930T e 10771G.

Na observação da *PhyloTree*, na árvore de ADNmt, há aglomerados de haplogrupos como JT e UK [84]. Os haplogrupos J e T são caracterizados pelos polimorfismos 4216C, 11251G, 15452A e 16126C, e são encontrados principalmente em população europeia [85]. Os haplogrupos U e K são caracterizados pelos respetivos polimorfismos 11467G, 12308G e 12372A, e 10550G, 11299C, 14798C, 16224C e 16311C, correspondendo a 20% de todas as sequências ADNmt europeu [86].

O aglomerado TJ subdivide-se em dois haplogrupos: o T e J. O haplogrupo T, que surgiu há aproximadamente 25.000 anos no Médio Oriente, espalhou-se pela Europa, Índia e África Oriental, nos últimos 15 milénios, dividindo-se nos subgrupos T1-T5. Neste estudo, foram identificados os subgrupos T1 e T2, em 15 amostras, predominantes na Europa e no Médio Oriente [87]. Estes subhaplogrupos apresentam as respetivas variantes: 12633A, 16163G e 16189C; e 11812G e 14233G. Algumas amostras classificadas no subhaplogrupo T2 podem ainda apresentar a mutação 16296T, visto que a sua presença é considerada recorrente/instável no subhaplogrupo, pela *PhyloTree*. Todas as amostras classificadas nestes subhaplogrupos apresentavam as suas variantes e, como esperado em algumas amostras do subhaplogrupo T2, observou-se a presença da variante 16296T.

O haplogrupo J tem uma distribuição uniforme na Eurásia Ocidental, representando entre de 6 a 22% dos ADNmt nas populações modernas [84] e derivou do haplogrupo JT há aproximadamente 45 000 anos [88], [89]. Na *PhyloTree*, o haplogrupo J é caracterizado pelos polimorfismos 295T, 489C, 10398G, 12612G, 13708A e 16069T, dividindo-se nos subhaplogrupos J1 e J2. O subhaplogrupo J1 surgiu há cerca de 27 000 anos, sendo encontrado em populações europeias e caracterizado pelas variantes 462T e 3010A. Já o subhaplogrupo J2 apareceu há cerca de 19.000 anos, com as mutações 150T, 152C 7476T e 15257A, estando recorrentemente presente no Mediterrâneo, incluindo Grécia, Itália e Espanha [88].

Neste estudo ambos os subhaplogrupos (J1 e J2) foram identificados em 6 amostras. Todas as amostras apresentam as variantes que caracterizam o haplogrupo J e os subhaplogrupos para os quais tinham sido classificadas.

O haplogrupo U surgiu na Eurásia Ocidental há cerca de 38 000 anos [90], sendo caracterizado pelas variantes 11467G, 12308G e 12372A. Este haplogrupo, amplamente presente nas populações europeias, subdividiu-se em nove subhaplogrupos (U1 a U9) cada um com mutações e distribuições geográficas específicas [82], [90]. No presente trabalho, o haplogrupo U foi o segundo mais frequente na população em estudo, com 19 amostras distribuídas entre os subhaplogrupos U2, U3, U4, U5, U6 e U8. Um estudo anterior também confirma a presença do haplogrupo U na Península Ibérica como o segundo mais frequente [91].

Todas as variantes das amostras classificadas dentro do haplogrupo U coincidiram com as variantes expostas na *PhyloTree*, exceto duas amostras que apresentaram falta de cobertura em algumas regiões. O subhaplogrupo U5 foi o mais frequente (16 amostras), sendo caracterizado pelas variantes 16192T e 16270T. Contudo, nem todas apresentaram a mutação 16192T, mas não foi por falta de cobertura, podendo ser isto explicado pela presença de heteroplasmias (coexistência de diferentes sequências mitocondriais no mesmo indivíduo) ou reversões de mutações. Este subhaplogrupo é frequente no Norte da Europa e Ásia, tendo dispersado pela Europa Ocidental e Oriental, e a sua presença neste estudo sugere que alguns indivíduos tenham ascendência de populações europeias nórdicas [92].

O subhaplogrupo U6 está presente em 5 amostras e é caracterizado pelas variantes 3348G e 16172C. Todas as amostras apresentaram ambas as mutações. Embora o subhaplogrupo U6 seja mais característico das populações do Norte de África, a sua presença em Portugal já tinha sido documentada,

em outros estudos. Em particular, uma investigação, sugere que este subhaplogrupo pode ter sido introduzido, durante o período de domínio muçulmano, no centro e sul da Península Ibérica [93].

Os subhaplogrupos U2, U3, U4 e U8 divergiram do mesmo ramo, na *PhyloTree*, e apresentam em comum a variante 1811G. Individualmente, o subhaplogrupo U2 apresenta a variante 16051G, o subhaplogrupo U3 as variantes 150T, 14139G, 15454C e 16343G, o subhaplogrupo U4 apresenta as variantes 195C, 499A, 4646C, 5999C, 6047G, 11332T, 14620T, 15693C e 16356C, e o subhaplogrupo U8 apresenta a variante 9698C. Nas 11 amostras agrupadas nestes subhaplogrupos foram identificadas todas as mutações características de cada subhaplogrupo. Estes subhaplogrupos encontram-se distribuídos pela Europa com uma frequência inferior comparativamente ao subhaplogrupo U5 [90].

O haplogrupo K é derivado do subhaplogrupo U8 e é considerado frequente entre as populações da Europa e Ásia ocidental [86]. É caracterizado pelos polimorfismos 10550G, 11299C, 14798C, 16224C e 16311C, tendo evoluído nos subhaplogrupos K1, K2 e K3. Neste estudo foram detetadas 11 amostras pertencentes ao haplogrupo K e todas coincidiam com as variantes expostas na *PhyloTree*. Das amostras, 10 foram classificadas no subhaplogrupo K1 e apresentaram as mutações 1189C e 10398G, características deste subhaplogrupo. A restante amostra foi classificada no subhaplogrupo K2 e as suas mutações foram coincidentes com as variantes, 146C e 9716C, presentes na *PhyloTree*.

Os haplótipos, da região controlo, destas amostras encontram-se submetidos na base de dados internacional EMPOP, com o código de acesso EMPOP00879, sendo agora necessário atualizar com a nova informação da região codificante.

4.4.2. Diversidade Genética

Através do *software Arlequin*, os resultados forneceram informação sobre a diversidade de sequência e o índice de diversidade nucleotídica da população do Sul de Portugal. A diversidade de sequência (*H*, *gene diversity*) representa a probabilidade de que dois haplótipos escolhidos aleatoriamente sejam diferentes, enquanto a diversidade nucleotídica (π , *nucleotide diversity*) considera as diferenças médias entre nucleotídeos em todos os pares de indivíduos.

A análise da diversidade genética da população do Sul revelou um valor de diversidade nucleotídica de 0.038772 +/- 0.018712 e uma diversidade de sequência de 1,0000 +/- 0.0007. Em comparação com resultados de outro estudo [94], este valor é relativamente elevado, indicando que em média 3,88% das posições nucleotídicas diferem entre os indivíduos. O valor de *H* próximo de 1,0000 indica que a maioria das sequências são únicas, o que ressalta a diversidade genética. Esta diversidade na região do Sul de Portugal pode ter beneficiado de eventos históricos que permitiram o fluxo de populações, promovendo a variabilidade.

4.5. Análise filogenética

Para estudar a variabilidade genética entre os grupos populacionais descritos no Quadro 4.4, realizou-se uma análise molecular da variância (AMOVA), que permitiu quantificar a variabilidade genética tanto entre indivíduos dentro da mesma população, como entre populações diferentes.

Quadro 4.4- População do presente estudo e populações selecionadas da literatura (populações africanas imigrantes e população de Portugal), tendo por base a região controlo total do ADNmt

Populações	Tamanho da Amostra	Método de Sequenciação	Referência
Sul de Portugal	155	NGS	Presente estudo
Cabo Verde	103	Sanger	[95]
Guiné-Bissau	80	Sanger	[96]
Angola	173	Sanger	[28]
Moçambique	83	Sanger	[97]
Portugal	292	Sanger	[94]

Foram calculados os valores de F_{ST} e os respetivos valores de p , utilizados para quantificar a diferenciação genética entre a população do Sul de Portugal e as populações em estudo, apresentados na Tabela 4.1.

Tabela 4.1 - Valores de F_{ST} e o p relativos às populações do estudo filogenético conseguidos através do programa Arlequin para a totalidade da região controlo do ADNmt: Sul de Portugal; Imigrantes de Cabo Verde; Imigrantes de Guiné-Bissau; Imigrantes de Angola; Imigrantes de Moçambique; Portugal

Populações	Sul de Portugal	Cabo Verde	Guiné-Bissau	Angola	Moçambique	Portugal
Sul de Portugal	-	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000
Cabo Verde	0.10012	-	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000
Guiné-Bissau	0.33280	0.48755	-	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000
Angola	0.08870	0.04388	0.46320	-	0.00901+- 0.0091	0.00000+- 0.0000
Moçambique	0.06454	0.06223	0.48102	0.01273	-	0.00000+- 0.0000
Portugal	0.02353	0.21911	0.60170	0.16161	0.14944	-

A comparação entre o Sul de Portugal e a população de Portugal, resultaram em valores baixos de F_{ST} (0,02), sugerindo uma proximidade genética e a ausência de barreiras significativas ao fluxo génico entre estas regiões. Em contraste, a população imigrante da Guiné-Bissau apresentou o valor de F_{ST} mais elevado (0,33), estatisticamente significativa ($p < 0.05$), sugerindo uma diferenciação genética relevante com a população do Sul de Portugal.

Na árvore filogenética (Figura 4.8), construída a partir dos valores de F_{ST} a proximidade entre o Sul e Portugal é evidente, o que indica uma baixa diferenciação genética entre estas duas populações, evidenciando uma proximidade genética esperada, visto que a população do Sul está inserida na população portuguesa.

As populações africanas imigrantes (Angola, Moçambique e Cabo Verde) mostram uma relação próxima entre si, observado em outros estudos [97], embora estejam também relativamente distantes da população do Sul de Portugal. A disposição dessas populações na árvore pode ser explicada pela presença de haplogrupos africanos compartilhados, observados no estudo populacional, refletindo uma ancestralidade genética comum e a migração histórica dessas populações para Portugal, especialmente para a área do Sul de Portugal [28], [95].

A população da Guiné-Bissau, no entanto, aparece mais afastada na árvore filogenética, indicando uma maior diferenciação genética, relativamente às outras populações africanas imigrantes estudadas e à população do Sul. Este distanciamento sugere que, embora existam conexões genéticas entre as populações africanas, a população da Guiné-Bissau possui uma maior variabilidade genética em comparação com as populações de Angola, Moçambique e Cabo Verde. Essa observação está em consonância com a diversidade genética dentro do continente africano e com as diferenças históricas e geográficas que podem ter influenciado os padrões de variabilidade genética entre Guiné-Bissau e outras regiões africanas que tiveram maior fluxo migratório com a região do Sul de Portugal [96].

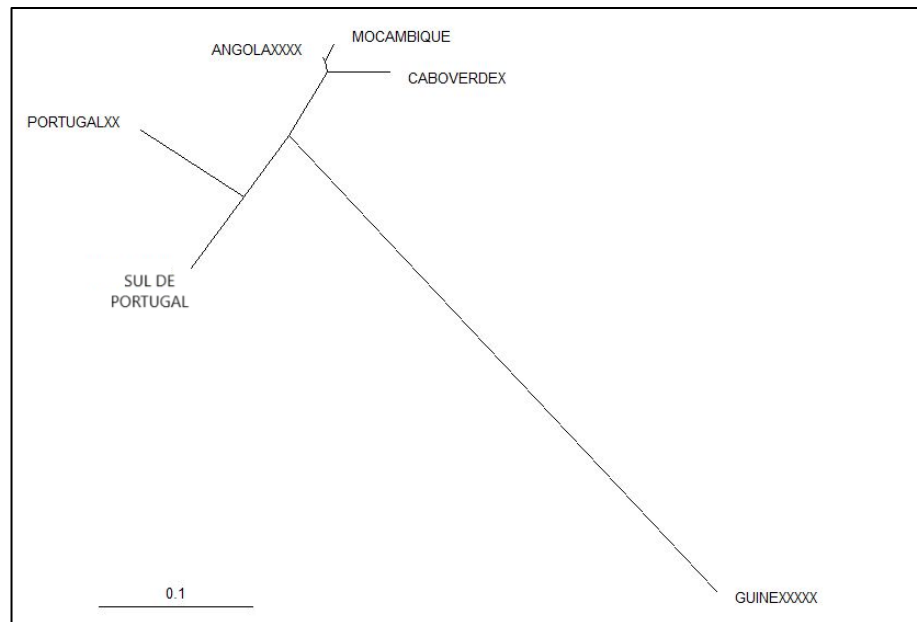


Figura 4.8 - Representação filogenética das distâncias genéticas entre a população do Sul, as populações africanas imigrantes em Lisboa e a população de Portugal, selecionadas da literatura.

Foi também calculada a AMOVA, da região total do ADNmt dos haplótipos de indivíduos da população do Sul, Norte e Centro de Portugal, obtidos com a metodologia NGS, que estão descritos no Quadro 4.5:

Quadro 4.5 - População do presente estudo e populações portuguesas das regiões do Centro e Norte, com amostras que foram sequenciadas com NGS para análise filogenética, tendo por base a região controlo do ADNmt.

População	Tamanho da amostra	Método de Sequenciação	Referências
Centro	201	NGS	[98]
Norte	63	NGS	[99]
Sul	155	NGS	Presente estudo

Os valores de F_{ST} e os respetivos valores de p entre as diferentes populações regionais de Portugal, estão apresentados na Tabela 4.2. Estes valores sugerem uma ligeira distância genética (0.03058) entre o Centro e Norte, enquanto estas duas regiões comparadas com o Sul apresentam uma distância genética maior (0.16249 e 0.18664, respetivamente).

Tabela 4.2 - Valores de F_{st} e os respectivos valores de p , recolhidos através do software Arlequin para a totalidade da região controlo do ADNmt, relativos aos haplótipos das regiões portuguesas (Norte, Centro e Sul de Portugal), obtidos com a sequenciação de nova geração. (* significa que o valor de p é inferior a 0,01)

Populações	Sul	Norte	Centro
Sul	-	*	*
Norte	0.18664	-	*
Centro	0.16249	0.03058	-

Através destas distâncias moleculares foi construída uma árvore filogenética, Figura 4.9, com as populações das 3 regiões de Portugal, onde se observou a distinção marcante entre o Sul e as outras duas regiões. A longa distância filogenética que separa o Sul do Norte e Centro sugere uma diferenciação genética significativa, o que pode estar associado ao espaço geográfico. O Sul de Portugal, apresenta uma maior área geográfica, na recolha de amostras, em comparação com o Norte e Centro. Estas distâncias já tinham sido documentadas [94], mostrando que a população do Sul de Portugal é caracterizada por populações mais dispersas, que formaram um perfil genético distinto na população do Sul.

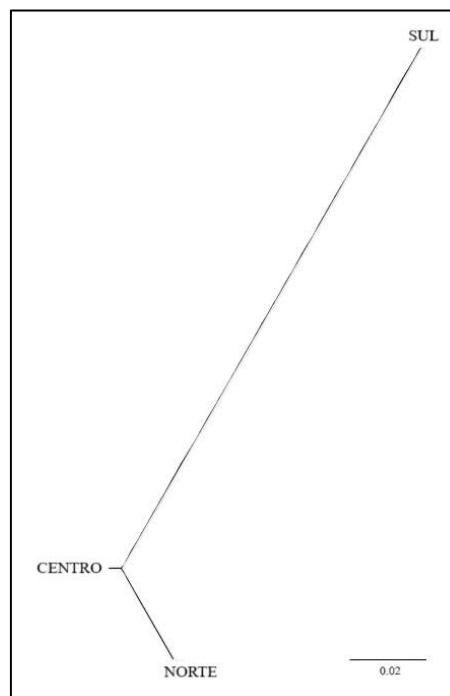


Figura 4.9 - Representação filogenética das distâncias genéticas entre as populações portuguesas (população das regiões Norte, Centro e Sul) estudadas com a metodologia NGS

Em contraste, o Norte e o Centro apresentam uma proximidade filogenéticas muito maior entre si, indicando uma maior semelhança genética. Essa proximidade sugere que essas regiões mantiveram um fluxo génico mais consistente ao longo do tempo, talvez devido à proximidade geográfica. Esta relação próxima entre as regiões apoia a hipótese de uma história evolutiva compartilhada, enquanto a distância da população do Sul destaca a diversidade genética interna de Portugal.

4.6. Aplicações Clínicas e Genética Médica da Sequenciação do ADN mitocondrial

4.6.1. Genoma mitocondrial – função e aplicação

No INMLCF, no Serviço de Genética e Biologia Forense, a sequenciação do ADNmt é realizada com fins específicos, como em estudos de parentesco ou na associação de indivíduos à mesma linhagem materna. Contudo, o estudo do genoma mitocondrial também apresenta vantagens em outras áreas, como na genética médica, especialmente no estudo de doenças hereditárias, distúrbios oncológicos e até no envelhecimento [100], [101].

A mitocôndria, como referido, é um organelo sito dentro da célula, que possui o seu próprio genoma. O genoma mitocondrial codifica 37 genes e participa ativamente, em conjunto com o genoma nuclear, em diversos processos da regulação celular, como a apoptose, o metabolismo energético, entre outros [7], [8]. É através do processo de fosforilação oxidativa que as mitocôndrias produzem 90% da energia celular [5], [100].

4.6.2. Mutações no ADN mitocondrial associadas a doenças

Para além do mencionado no ponto anterior, o ADNmt apresenta uma taxa de mutação elevada quando comparado com o ADN nuclear, o que pode resultar no aparecimento de variantes que levam a disfunções mitocondriais [101], [102]. Quando ocorrem distúrbios mitocondriais, os primeiros órgãos afetados são aqueles que apresentam grandes necessidades energéticas, como o cérebro e os músculos esqueléticos [100].

Já foram relatadas mais de 250 mutações diferentes no ADNmt associadas a variantes patogénicas. Estas mutações podem ser pontuais ou resultar de rearranjos em larga escala (duplicações ou deleções) [100]. Quando uma mutação é detetada, a célula pode apresentar 100% do ADNmt mutado, condição denominada homoplasmia, ou pode conter uma mistura de ADNmt mutado e normal, conhecida como heteroplasmia [103].

As mutações pontuais mais comuns e bem estudadas são a A3243G, a A8344G e a T8993G [101]. A mutação A3243G, que envolve a substituição de um A por um G, está associada a indivíduos com diabetes e surdez (MIDD) [104]. A mutação A8344G, uma substituição de A por G, provoca alterações no metabolismo energético e miopatias, resultando na síndrome de epilepsia mioclônica com fibras rotas vermelhas (MERRF) [105]. Já a mutação T8993C, uma troca de um T por um C, é reportada em indivíduos com a Síndrome de Leigh, uma doença que provoca destruição progressiva do sistema central nervoso [106]. Todas estas variantes são hereditárias, transmitidas pela via materna, e geralmente estão no estado de heteroplasmia [101].

A neuropatia ótica hereditária de Leber (LHON) é um exemplo de doença causada por uma mutação pontual homoplásmica, sendo frequentemente ligada à variante G11778A, que representa uma substituição de um G por um A. No entanto, a variante T14484C, embora menos comum, também pode causar a LHON, sendo uma mutação pontual heteroplásmica [101], [107]. Por fim, o síndrome de Pearson resulta de uma deleção de vários nucleótidos no genoma mitocondrial, causando desregulações no pâncreas e na medula óssea [108].

4.6.3. Relação entre Haplogrupos Mitocondriais e Doenças

A evolução do ADNmt resultou na formação de haplogrupos específicos, definidos por distribuições geográficas, permitindo a classificação de sequências em haplogrupos europeus, africanos, asiáticos ou americanos, caracterizados por um conjunto de polimorfismos [35], [36], [37], [38], [39], [40]. Alguns estudos sugerem que há associações entre certos haplogrupos e o risco aumentado de diversas doenças, bem como uma possível ligação com o processo de envelhecimento [100], [101].

Um estudo do genoma mitocondrial revelou que 75% dos indivíduos com LHON e com a mutação T14484C pertenciam ao haplogrupo europeu J. Isto sugere, que a presença desta variante genética, associada a este haplogrupo, pode aumentar a chance de a mutação causar LHON [101], [108].

Outro estudo, realizado com doentes franco-canadenses, sugeriu uma associação entre a doença de Alzheimer (AD) e o haplogrupo J. Notavelmente, um haplótipo composto por três variantes (nas posições 5633, 7476 e 15812) foi identificado em 10,5% dos doentes com AD. Por outro lado, foi levantada a hipótese de que o haplogrupo T contém variantes que podem prevenir ou reduzir a incidência de AD. No entanto, a falta de confirmação dessa associação em diferentes populações reforça a possibilidade de que a contribuição mitocondrial esteja limitada a determinados grupos [109].

No caso da doença de Parkinson (PD), os estudos apresentaram resultados contraditórios. Enquanto alguns sugerem que os haplogrupos J e K, que contêm a variante 10398G, conferem proteção contra a PD, outros associam os haplogrupos J e T à mutação 4216C, que está relacionada a um maior risco de desenvolvimento da doença [110], [111].

Num estudo sobre fertilidade masculina conduzido entre um grupo de homens europeus com problemas de fertilidade foi observado que os indivíduos pertencentes ao haplogrupo H apresentavam uma maior motilidade espermática, sugerindo uma possível relação entre este haplogrupo e a eficiência da fosforilação oxidativa mitocondrial [112].

O haplogrupo J, além de aumentar a penetrância de mutações associadas à LHON, também parece estar relacionado com a longevidade. Num estudo realizado com três grupos da população finlandesa (crianças, adultos de meia-idade e idosos com 90 anos ou mais) verificou-se que os haplogrupos J e U eram predominantes entre os indivíduos com 90 ou mais anos [113]. Por outro lado, uma investigação realizada no Japão indicou que o haplogrupo D foi o mais comum entre centenários japoneses, sugerindo uma associação da mutação C5178A com a longevidade [114].

Contudo, é importante salientar que muitas dessas associações foram observadas apenas em estudos isolados, o que limita a formulação de conclusões amplas e definitivas [115]. Assim, embora existam indícios de que os polimorfismos relacionados aos haplogrupos possam atuar como fatores de risco, ainda é necessário realizar investigações mais abrangentes, com maior rigor metodológico e envolvendo múltiplas subpopulações ou grandes levantamentos populacionais, para que essas correlações possam ser consideradas robustas e generalizáveis.

5. CONCLUSÃO

Este projeto teve três objetivos: validar a sequenciação de Sanger para verificar a sua especificidade em ADNmt humano; validar a sequenciação de nova geração, comparando os resultados da região controlo do ADNmt com os da sequenciação de Sanger; e, com a metodologia NGS, sequenciar o mitogenoma na sua totalidade e realizar a caracterização genética da população da região do Sul de Portugal.

Para a validação da especificidade da Sequenciação de Sanger, com o instrumento *SeqStudio™ Genetic Analyzer*, verificou-se que os primers utilizados amplificam exclusivamente o ADNmt humano. Este resultado foi confirmado pela ausência de amplificação das amostras de origem animal vertebrado não humano.

A comparação da sequenciação da região controlo do ADNmt entre Sanger e NGS mostrou que os haplótipos coincidem, mas houve divergências na atribuição de haplogrupos em algumas amostras. Estas diferenças devem-se à maior sensibilidade do NGS, que permitiu a deteção de mutações adicionais na região codificante do mitogenoma, que não tinham sido identificadas pela metodologia de Sanger. Assim, a metodologia NGS demonstrou ser mais precisa e informativa, fornecendo haplótipos mais detalhados e uma definição mais exata dos haplogrupos.

A metodologia NGS revelou-se uma técnica fiável para a sequenciação do ADNmt. No entanto, para garantir que as sequências apresentem alta qualidade, é importante uma preparação adequada das bibliotecas, sendo necessário controlar a concentração do ADN. Este trabalho mostrou que a concentração ideal de amostras e bibliotecas, para a preparação de bibliotecas e templates, respetivamente, não corresponde exatamente ao indicado pelo fabricante. Nos ensaios realizados, verificou-se que a concentração inicial inferior das amostras e também de bibliotecas geram melhores resultados finais.

A sequenciação completa do mitogenoma utilizando a metodologia NGS permitiu a identificação de haplogrupos com maior precisão. A análise dos haplogrupos forneceu informações sobre a diversidade genética da região do Sul e as possíveis origens dos indivíduos da população. Assim:

- A população em estudo apresentou 100% de haplótipos únicos, o que revelou uma grande variabilidade genética na população do Sul de Portugal. Analisando apenas a região controlo do ADNmt, só teria sido possível obter 79,4% de haplótipos únicos.
- De acordo com os haplogrupos obtidos, 83,87% são característicos da Europa, o que demonstra o contributo de linhagens europeias na população do Sul.
- O haplogrupo H foi o mais comum na região do Sul, presente em 34,19%, seguido pelos haplogrupos U (20,65%), T (9,68%) e K (7,10), o que está de acordo com outros estudos sobre a composição genética de populações europeias.
- A deteção do haplogrupo L (8,39%), característico das populações africanas, já era esperada e veio reforçar a história de migração e tráfico de escravos entre os séculos XV-XIX.
- Os haplogrupos M, D e G, de origem asiática, também foram detetados entre as amostras, contribuindo para a diversidade genética.

A sequenciação de nova geração demonstrou ser uma ferramenta valiosa para a caracterização genética de populações, fornecendo informação sobre a diversidade e origem genética da população da região do Sul, e os seus resultados vão contribuir para a base de dados EMPOP.

Na análise filogenética, a comparação da população de Sul com a população portuguesa e as populações imigrantes residentes em Lisboa, e com as populações portuguesas do Norte e Centro do país permitiram observar padrões de relacionamento genético entre elas.

Na primeira árvore, que inclui a população do Sul de Portugal, a população de Portugal e populações imigrantes de Angola, Moçambique, Cabo Verde e Guiné-Bissau, verifica-se que o Sul e Portugal estão muito próximos geneticamente, formando um clade. As populações imigrantes africanas encontram-se mais afastadas, especialmente a Guiné-Bissau, que se distancia consideravelmente das outras populações. Esse afastamento da população da Guiné-Bissau reflete a diversidade genética, dentro do continente africano, e as diferentes origens geográficas. No entanto, a relativa proximidade entre o Sul de Portugal e as outras populações africanas imigrantes era esperada pelo intercâmbio genético ao longo da história.

Na segunda árvore, que compara as populações do Norte, Centro e Sul de Portugal, constatou-se uma clara distinção entre a população do Sul e as demais regiões. O Sul apresenta maior divergência genética em relação ao Norte e Centro, que estão mais próximos entre si. Este padrão pode refletir variações históricas e demográficas dentro de Portugal, apontando possíveis influências genéticas distintas ao longo do tempo em diferentes regiões do país.

No entanto, o estudo das mutações do ADNmt transcende a genética forense, e mostrou-se vantajoso noutras áreas, como a genética médica. Algumas investigações sugerem uma relação entre haplogrupos mitocondriais e o risco de doenças como o Alzheimer e Parkinson, entre outras. Contudo, com o objetivo de compreender melhor esta relação, são necessárias mais investigações abrangentes, com metodologias rigorosas e envolvendo diversas populações, para validar e expandir essas associações de forma robusta e generalizável.

6. REFERÊNCIAS

- [1] J. Cummins, «Mitochondrial DNA in mammalian reproduction», set. 1998, doi: 10.1530/revreprod/3.3.172.
- [2] E. Braschi e H. M. McBride, «Mitochondria and the culture of the Borg: Understanding the integration of mitochondrial function within the reticulum, the cell, and the organism», *BioEssays*, vol. 32, n.º 11, pp. 958–966, nov. 2010, doi: 10.1002/bies.201000073.
- [3] D. Grasso, L. X. Zampieri, T. Capelôa, J. A. Van De Velde, e P. Sonveaux, «Mitochondria in cancer», *Cell Stress*, vol. 4, n.º 6, pp. 114–146, jun. 2020, doi: 10.15698/cst2020.06.221.
- [4] F. J. Bock e S. W. G. Tait, «Mitochondria as multifaceted regulators of cell death», *Nat. Rev. Mol. Cell Biol.*, vol. 21, n.º 2, pp. 85–100, fev. 2020, doi: 10.1038/s41580-019-0173-8.
- [5] M. Picard, O. S. Shirihai, B. J. Gentil, e Y. Buelle, «Mitochondrial morphology transitions and functions: implications for retrograde signaling?», *Am. J. Physiol.-Regul. Integr. Comp. Physiol.*, vol. 304, n.º 6, pp. R393–R406, mar. 2013, doi: 10.1152/ajpregu.00584.2012.
- [6] R. Filograna, M. Mennuni, D. Alsina, e N. Larsson, «Mitochondrial DNA copy number in human disease: the more the better?», *FEBS Lett.*, vol. 595, n.º 8, pp. 976–1002, abr. 2021, doi: 10.1002/1873-3468.14021.
- [7] K. Sturk-Andreaggi, *Mitochondrial Genome Analysis Using Next Generation Sequencing for Forensic Applications*. Uppsala: Acta Universitatis Upsaliensis, 2022.
- [8] J.-W. Taanman, «The mitochondrial genome: structure, transcription, translation and replication», *Biochim. Biophys. Acta BBA - Bioenerg.*, vol. 1410, n.º 2, pp. 103–123, fev. 1999, doi: 10.1016/S0005-2728(98)00161-3.
- [9] A. Amorim, T. Fernandes, e N. Taveira, «Mitochondrial DNA in human identification: a review», *PeerJ*, vol. 7, p. e7314, ago. 2019, doi: 10.7717/peerj.7314.
- [10] T. Kivisild, «Maternal ancestry and population history from whole mitochondrial genomes», *Investig. Genet.*, vol. 6, n.º 1, p. 3, 2015, doi: 10.1186/s13323-015-0022-2.
- [11] M. Sato e K. Sato, «Maternal inheritance of mitochondrial DNA: Degradation of paternal mitochondria by allogeneic organelle autophagy, allophagy», *Autophagy*, vol. 8, n.º 3, pp. 424–425, mar. 2012, doi: 10.4161/auto.19243.
- [12] W. Parson *et al.*, «DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing»,

Forensic Sci. Int. Genet., vol. 13, pp. 134–142, nov. 2014, doi: 10.1016/j.fsigen.2014.07.010.

[13] C. M. Gustafsson, M. Falkenberg, e N.-G. Larsson, «Maintenance and Expression of Mammalian Mitochondrial DNA», *Annu. Rev. Biochem.*, vol. 85, n.º 1, pp. 133–160, jun. 2016, doi: 10.1146/annurev-biochem-060815-014402.

[14] N.-G. Larsson, A. Oldfors, J. David Garman, G. S. Barsh, e D. A. Clayton, «Down-Regulation of Mitochondrial Transcription Factor a During Spermatogenesis in Humans», *Hum. Mol. Genet.*, vol. 6, n.º 2, pp. 185–1991, fev. 1997, doi: 10.1093/hmg/6.2.185.

[15] H. Kaneda, J. Hayashi, S. Takahama, C. Taya, K. F. Lindahl, e H. Yonekawa, «Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis.», *Proc. Natl. Acad. Sci.*, vol. 92, n.º 10, pp. 4542–4546, mai. 1995, doi: 10.1073/pnas.92.10.4542.

[16] P. Sutovsky, R. D. Moreno, J. Ramalho-Santos, T. Dominko, C. Simerly, e G. Schatten, «Ubiquitin tag for sperm mitochondria», *Nature*, vol. 402, n.º 6760, pp. 371–372, nov. 1999, doi: 10.1038/46466.

[17] B. Budowle, M. W. Allard, M. R. Wilson, e R. Chakraborty, «Forensics and Mitochondrial DNA: Applications, Debates, and Foundations», *Annu. Rev. Genomics Hum. Genet.*, vol. 4, n.º 1, pp. 119–141, set. 2003, doi: 10.1146/annurev.genom.4.070802.110352.

[18] J. L. Elson, R. M. Andrews, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, e N. Howell, «Analysis of European mtDNAs for Recombination», *Am. J. Hum. Genet.*, vol. 68, n.º 1, pp. 145–153, jan. 2001, doi: 10.1086/316938.

[19] F. J. Iborra, H. Kimura, e P. R. Cook, «The functional organization of mitochondrial genomes in human cells», *BMC Biol.*, vol. 2, n.º 1, p. 9, mai. 2004, doi: 10.1186/1741-7007-2-9.

[20] P. Silva-Pinheiro e M. Minczuk, «The potential of mitochondrial genome engineering», *Nat. Rev. Genet.*, vol. 23, n.º 4, pp. 199–214, abr. 2022, doi: 10.1038/s41576-021-00432-x.

[21] V. M. Cabrera, «Human molecular evolutionary rate, time dependency and transient polymorphism effects viewed through ancient and modern mitochondrial DNA genomes», *Sci. Rep.*, vol. 11, n.º 1, p. 5036, mar. 2021, doi: 10.1038/s41598-021-84583-1.

[22] A. Konrad *et al.*, «Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in *Caenorhabditis elegans* Spontaneous Mutation Accumulation Lines of Differing Population Size», *Mol. Biol. Evol.*, p. msx051, jan. 2017, doi: 10.1093/molbev/msx051.

- [23] N. D. Beuzen, M. J. Stear, e K. C. Chang, «Molecular markers and their use in animal breeding», *Vet. J.*, vol. 160, n.º 1, pp. 42–52, jul. 2000, doi: 10.1053/tvj.2000.0468.
- [24] C. B. Jackson, D. M. Turnbull, M. Minczuk, e P. A. Gammage, «Therapeutic Manipulation of mtDNA Heteroplasmy: A Shifting Perspective», *Trends Mol. Med.*, vol. 26, n.º 7, pp. 698–709, jul. 2020, doi: 10.1016/j.molmed.2020.02.006.
- [25] A. W. Röck, A. Dür, M. Van Oven, e W. Parson, «Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA)», *Forensic Sci. Int. Genet.*, vol. 7, n.º 6, pp. 601–609, dez. 2013, doi: 10.1016/j.fsigen.2013.07.005.
- [26] A. Tagliabracci e C. Turchi, «mtDNA exploitation in forensics», em *The Human Mitochondrial Genome*, Elsevier, 2020, pp. 145–169. doi: 10.1016/B978-0-12-819656-4.00007-3.
- [27] R. K.-B. Atig, S. Hsouna, E. Beraud-Colomb, e S. Abdelhak, «[Mitochondrial DNA: properties and applications]», *Arch. Inst. Pasteur Tunis*, vol. 86, n.º 1–4, pp. 3–14, 2009.
- [28] F. Simão *et al.*, «Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA», *Forensic Sci. Int. Genet.*, vol. 15, pp. 33–38, mar. 2015, doi: 10.1016/j.fsigen.2014.09.013.
- [29] P. Guha, S. K. Srivastava, S. Bhattacharjee, e T. K. Chaudhuri, «Human migration, diversity and disease association: a convergent role of established and emerging DNA markers», *Front. Genet.*, vol. 4, 2013, doi: 10.3389/fgene.2013.00155.
- [30] A. G. Thorne e M. H. Wolpoff, «The Multiregional Evolution of Humans», *Sci. Am.*, vol. 266, n.º 4, pp. 76–83, abr. 1992, doi: 10.1038/scientificamerican0492-76.
- [31] A. C. Wilson e R. L. Cann, «The recent African genesis of humans», *Sci. Am.*, vol. 266, n.º 4, pp. 68–73, abr. 1992, doi: 10.1038/scientificamerican0492-68.
- [32] M. T. Lott *et al.*, «mtDNA Variation and Analysis Using Mitomap and Mitomaster», *Curr. Protoc. Bioinforma.*, vol. 44, n.º 1, dez. 2013, doi: 10.1002/0471250953.bi0123s44.
- [33] C. J. Pérez-Amado, A. Bazan-Cordoba, A. Hidalgo-Miranda, e S. Jiménez-Morales, «Mitochondrial Heteroplasmy Shifting as a Potential Biomarker of Cancer Progression», *Int. J. Mol. Sci.*, vol. 22, n.º 14, p. 7369, jul. 2021, doi: 10.3390/ijms22147369.
- [34] M. Van Oven e M. Kayser, «Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation», *Hum. Mutat.*, vol. 30, n.º 2, pp. E386–E394, fev. 2009, doi: 10.1002/humu.20921.

- [35] D. M. Behar *et al.*, «A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root», *Am. J. Hum. Genet.*, vol. 90, n.° 4, pp. 675–684, abr. 2012, doi: 10.1016/j.ajhg.2012.03.002.
- [36] A. Brehm e A. Rosa, «African human mtDNA phylogeography at-a-glance», *J. Anthropol. Sci.*, n.° 89, pp. 1–34, 2011, doi: 10.4436/jass.89006.
- [37] D. M. Behar *et al.*, «The Dawn of Human Matrilineal Diversity», *Am. J. Hum. Genet.*, vol. 82, n.° 5, pp. 1130–1140, mai. 2008, doi: 10.1016/j.ajhg.2008.04.002.
- [38] L. Quintana-Murci, O. Semino, H.-J. Bandelt, G. Passarino, K. McElreavey, e A. S. Santachiara-Benerecetti, «Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa», *Nat. Genet.*, vol. 23, n.° 4, pp. 437–441, dez. 1999, doi: 10.1038/70550.
- [39] T. Kivisild *et al.*, «The Emerging Limbs and Twigs of the East Asian mtDNA Tree», *Mol. Biol. Evol.*, vol. 19, n.° 10, pp. 1737–1751, out. 2002, doi: 10.1093/oxfordjournals.molbev.a003996.
- [40] D. A. (Weiss) Bolnick e D. G. Smith, «Unexpected patterns of mitochondrial DNA variation among Native Americans from the Southeastern United States», *Am. J. Phys. Anthropol.*, vol. 122, n.° 4, pp. 336–354, dez. 2003, doi: 10.1002/ajpa.10284.
- [41] T. D. Anderson, J. P. Ross, R. K. Roby, D. A. Lee, e M. M. Holland, «A validation study for the extraction and analysis of DNA from human nail material and its application to forensic casework», *J. Forensic Sci.*, vol. 44, n.° 5, pp. 1053–1056, set. 1999.
- [42] T. Melton, G. Dimick, B. Higgins, L. Lindstrom, e K. Nelson, «Forensic mitochondrial DNA analysis of 691 casework hairs», *J. Forensic Sci.*, vol. 50, n.° 1, pp. 73–80, jan. 2005.
- [43] W. Goodwin, A. Linacre, e P. Vanezis, «The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims», *Electrophoresis*, vol. 20, n.° 8, pp. 1707–1711, jun. 1999, doi: 10.1002/(SICI)1522-2683(19990101)20:8<1707::AID-ELPS1707>3.0.CO;2-F.
- [44] D. Syndercombe Court, «Mitochondrial DNA in forensic use», *Emerg. Top. Life Sci.*, vol. 5, n.° 3, pp. 415–426, set. 2021, doi: 10.1042/ETLS20210204.
- [45] R. M. Andrews, I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, e N. Howell, «Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA», *Nat. Genet.*, vol. 23, n.° 2, pp. 147–147, out. 1999, doi: 10.1038/13779.

- [46] A. Röhl, B. Brinkmann, L. Forster, e P. Forster, «An annotated mtDNA database», *Int. J. Legal Med.*, vol. 115, n.º 1, pp. 29–39, ago. 2001, doi: 10.1007/s004140100217.
- [47] W. Parson e A. Dür, «EMPOP—A forensic mtDNA database», *Forensic Sci. Int. Genet.*, vol. 1, n.º 2, pp. 88–92, jun. 2007, doi: 10.1016/j.fsigen.2007.01.018.
- [48] E. Ruiz-Pesini *et al.*, «An enhanced MITOMAP with a global mtDNA mutational phylogeny», *Nucleic Acids Res.*, vol. 35, n.º Database, pp. D823–D828, jan. 2007, doi: 10.1093/nar/gkl927.
- [49] M. Ingman, «mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences», *Nucleic Acids Res.*, vol. 34, n.º 90001, pp. D749–D751, jan. 2006, doi: 10.1093/nar/gkj010.
- [50] Y. S. Lee, W.-Y. Kim, M. Ji, J. H. Kim, e J. Bhak, «MitoVariome: a variome database of human mitochondrial DNA», *BMC Genomics*, vol. 10, n.º S3, p. S12, dez. 2009, doi: 10.1186/1471-2164-10-S3-S12.
- [51] M. Attimonelli, «MitBASE: a comprehensive and integrated mitochondrial DNA database. The present status», *Nucleic Acids Res.*, vol. 28, n.º 1, pp. 148–152, jan. 2000, doi: 10.1093/nar/28.1.148.
- [52] J. Damas, J. Carneiro, A. Amorim, e F. Pereira, «MitoBreak: the mitochondrial DNA breakpoints database», *Nucleic Acids Res.*, vol. 42, n.º D1, pp. D1261–D1268, jan. 2014, doi: 10.1093/nar/gkt982.
- [53] B. Zimmermann, A. W. Röck, A. Dür, e W. Parson, «Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software», *Croat. Med. J.*, vol. 55, n.º 2, pp. 115–120, abr. 2014, doi: 10.3325/cmj.2014.55.115.
- [54] B. M. Crossley *et al.*, «Guidelines for Sanger sequencing and molecular assay monitoring», *J. Vet. Diagn. Invest.*, vol. 32, n.º 6, pp. 767–775, nov. 2020, doi: 10.1177/1040638720905833.
- [55] F. Sanger, S. Nicklen, e A. R. Coulson, «DNA sequencing with chain-terminating inhibitors», *Proc. Natl. Acad. Sci.*, vol. 74, n.º 12, pp. 5463–5467, dez. 1977, doi: 10.1073/pnas.74.12.5463.
- [56] N. Gupta e V. K. Verma, «Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality», em *Microbial Technology for the Welfare of Society*, vol. 17, P. K. Arora, Ed., em *Microorganisms for Sustainability*, vol. 17. , Singapore: Springer Singapore, 2019, pp. 313–341. doi: 10.1007/978-981-13-8844-6_15.
- [57] B. E. Slatko, A. F. Gardner, e F. M. Ausubel, «Overview of Next-Generation Sequencing Technologies», *Curr. Protoc. Mol. Biol.*, vol. 122, n.º 1, p. e59, abr. 2018, doi: 10.1002/cpmb.59.

- [58] S. Balzer, K. Malde, A. Lanzén, A. Sharma, e I. Jonassen, «Characteristics of 454 pyrosequencing data--enabling realistic simulation with flowsim», *Bioinforma. Oxf. Engl.*, vol. 26, n.º 18, pp. i420-425, set. 2010, doi: 10.1093/bioinformatics/btq365.
- [59] C. Marshall, K. Sturk-Andreaggi, J. Daniels-Higginbotham, R. S. Oliver, S. Barritt-Ross, e T. P. McMahon, «Performance evaluation of a mitogenome capture and Illumina sequencing protocol using non-probative, case-type skeletal samples: Implications for the use of a positive control in a next-generation sequencing procedure», *Forensic Sci. Int. Genet.*, vol. 31, pp. 198–206, nov. 2017, doi: 10.1016/j.fsigen.2017.09.001.
- [60] W. Parson *et al.*, «Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples», *Forensic Sci. Int. Genet.*, vol. 15, pp. 8–15, mar. 2015, doi: 10.1016/j.fsigen.2014.11.009.
- [61] J. C. B. Nunez e M. F. Oleksiak, «A Cost-Effective Approach to Sequence Hundreds of Complete Mitochondrial Genomes», *PLOS ONE*, vol. 11, n.º 8, p. e0160958, ago. 2016, doi: 10.1371/journal.pone.0160958.
- [62] A. Legati *et al.*, «Current and New Next-Generation Sequencing Approaches to Study Mitochondrial DNA», *J. Mol. Diagn.*, vol. 23, n.º 6, pp. 732–741, jun. 2021, doi: 10.1016/j.jmoldx.2021.03.002.
- [63] Applied Biosystems, «ExoSAP-IT™ PCR Product Cleanup», 2017.
- [64] Applied Biosystems, «BigDye® Terminator v3.1 Cycle Sequencing Kit Protocol», 2002.
- [65] Applied Biosystems, «BigDye XTerminator™ Purification Kit», 2020.
- [66] Applied Biosystems, *Precision ID mtDNA Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System*. Thermo Fisher, 2018. [Online]. Disponível em: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017770_PrecisionID_mtDNA_Panels_S5_UG.pdf
- [67] W. Parson e H.-J. Bandelt, «Extended guidelines for mtDNA typing of population data in forensic science», *Forensic Sci. Int. Genet.*, vol. 1, n.º 1, pp. 13–19, mar. 2007, doi: 10.1016/j.fsigen.2006.11.003.
- [68] S. D. L. Magalhães, «Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias».
- [69] C. Faccineto *et al.*, «Internal validation and improvement of mitochondrial genome sequencing using the Precision ID mtDNA Whole Genome Panel», *Int. J. Legal Med.*, vol. 135, n.º 6, pp. 2295–2306, nov. 2021, doi: 10.1007/s00414-021-02686-w.

- [70] Thermo Fisher Scientific, «Solutions for mtDNA analysis». 2019. [Online]. Disponível em: <https://assets.thermofisher.com/TFS-Assets/GSD/Flyers/precisionid-mtDNA-panel-flyer.pdf>
- [71] *Converge Forensic Analysis Software*.
- [72] K. Kaneva *et al.*, «Mitochondrial DNA haplogroup, genetic ancestry, and susceptibility to Ewing sarcoma», *Mitochondrion*, vol. 67, pp. 6–14, nov. 2022, doi: 10.1016/j.mito.2022.09.002.
- [73] A. Salas *et al.*, «The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade», *Am. J. Hum. Genet.*, vol. 74, n.º 3, pp. 454–465, mar. 2004, doi: 10.1086/382194.
- [74] Q. Mairal *et al.*, «Linguistic isolates in Portugal: Insights from the mitochondrial DNA pattern», *Forensic Sci. Int. Genet.*, vol. 7, n.º 6, pp. 618–623, dez. 2013, doi: 10.1016/j.fsigen.2013.08.009.
- [75] P. Marrero, K. K. Abu-Amero, J. M. Larruga, e V. M. Cabrera, «Carriers of human mitochondrial DNA macrohaplogroup M colonized India from southeastern Asia», *BMC Evol. Biol.*, vol. 16, n.º 1, p. 246, dez. 2016, doi: 10.1186/s12862-016-0816-8.
- [76] T. Kivisild *et al.*, «Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages», *Curr. Biol.*, vol. 9, n.º 22, pp. 1331–1334, nov. 1999, doi: 10.1016/S0960-9822(00)80057-3.
- [77] I. Mendizabal *et al.*, «Reconstructing the Indian Origin and Dispersal of the European Roma: A Maternal Genetic Perspective», *PLoS ONE*, vol. 6, n.º 1, p. e15988, jan. 2011, doi: 10.1371/journal.pone.0015988.
- [78] M. Derenko *et al.*, «Origin and Post-Glacial Dispersal of Mitochondrial DNA Haplogroups C and D in Northern Asia», *PLoS ONE*, vol. 5, n.º 12, p. e15214, dez. 2010, doi: 10.1371/journal.pone.0015214.
- [79] M. Metspalu *et al.*, «Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans», *BMC Genet.*, vol. 5, n.º 1, p. 26, 2004, doi: 10.1186/1471-2156-5-26.
- [80] J. M. Larruga, P. Marrero, K. K. Abu-Amero, M. V. Golubenko, e V. M. Cabrera, «Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area», *BMC Evol. Biol.*, vol. 17, n.º 1, p. 115, dez. 2017, doi: 10.1186/s12862-017-0964-5.
- [81] U. Roostalu *et al.*, «Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and

Caucasian Perspective», *Mol. Biol. Evol.*, vol. 24, n.º 2, pp. 436–448, fev. 2007, doi: 10.1093/molbev/msl173.

[82] M. Richards *et al.*, «Tracing European founder lineages in the Near Eastern mtDNA pool», *Am. J. Hum. Genet.*, vol. 67, n.º 5, pp. 1251–1276, nov. 2000.

[83] A. W. R. Whittle, V. Cummings, e British Academy, Eds., *Going over: the Mesolithic-Neolithic transition in north-west Europe*. em Proceedings of the British Academy, no. no. 144. Oxford: Published for the British Academy by Oxford University Press, 2007.

[84] C. Fraumene *et al.*, «High Resolution Analysis and Phylogenetic Network Construction Using Complete mtDNA Sequences in Sardinian Genetic Isolates», *Mol. Biol. Evol.*, vol. 23, n.º 11, pp. 2101–2111, nov. 2006, doi: 10.1093/molbev/msl084.

[85] S. Finnilä e K. Majamaa, «Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population», *J. Hum. Genet.*, vol. 46, n.º 2, pp. 64–69, fev. 2001, doi: 10.1007/s100380170110.

[86] C. Herrnstadt *et al.*, «Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian, and European Haplogroups», *Am. J. Hum. Genet.*, vol. 70, n.º 5, pp. 1152–1171, mai. 2002, doi: 10.1086/339933.

[87] M. Pala *et al.*, «Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia», *Am. J. Hum. Genet.*, vol. 90, n.º 5, pp. 915–924, mai. 2012, doi: 10.1016/j.ajhg.2012.04.003.

[88] R. Villems, «HUMAN MITOCHONDRIAL DNA HAPLOGROUP J IN EUROPE AND NEAR EAST».

[89] D. Kristjansson, T. G. Schurr, J. Bohlin, e A. Jugessur, «Phylogeographic history of mitochondrial haplogroup J in SCANDINAVIA», *Am. J. Biol. Anthropol.*, vol. 180, n.º 2, pp. 298–315, fev. 2023, doi: 10.1002/ajpa.24666.

[90] S. Översti e J. U. Palo, «Variation in the Substitution Rates among the Human Mitochondrial Haplogroup U Sublineages», *Genome Biol. Evol.*, vol. 14, n.º 7, p. evac097, jul. 2022, doi: 10.1093/gbe/evac097.

[91] R. Barral-Arca *et al.*, «Meta-Analysis of Mitochondrial DNA Variation in the Iberian Peninsula», *PLOS ONE*, vol. 11, n.º 7, p. e0159735, jul. 2016, doi: 10.1371/journal.pone.0159735.

[92] D. Kristjansson, J. Bohlin, T. T. Nguyen, A. Jugessur, e T. G. Schurr, «Evolution and dispersal of mitochondrial DNA haplogroup U5 in Northern Europe: insights from an unsupervised learning approach to phylogeography», *BMC Genomics*, vol. 23, n.º 1, p. 354, mai. 2022, doi: 10.1186/s12864-022-08572-y.

- [93] L. Pereira, M. J. Prata, e A. Amorim, «Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation», *Ann. Hum. Genet.*, vol. 64, n.º 6, pp. 491–506, nov. 2000, doi: 10.1046/j.1469-1809.2000.6460491.x.
- [94] S. L. Marques *et al.*, «Portuguese mitochondrial DNA genetic diversity—An update and a phylogenetic revision», *Forensic Sci. Int. Genet.*, vol. 15, pp. 27–32, mar. 2015, doi: 10.1016/j.fsigen.2014.10.004.
- [95] P. Morais, A. Amorim, C. Vieira Da Silva, T. Ribeiro, J. Costa Santos, e H. Afonso Costa, «Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis», *J. Genet.*, vol. 94, n.º 3, pp. 509–512, set. 2015, doi: 10.1007/s12041-015-0552-7.
- [96] A. Amorim *et al.*, «Updated mtDNA study of Guinea-Bissau immigrant population living in Lisbon», *Forensic Sci. Int. Genet. Suppl. Ser.*, vol. 6, pp. e329–e331, dez. 2017, doi: 10.1016/j.fsigss.2017.09.122.
- [97] M. Proença De Campos *et al.*, «The immigrant population from Mozambique in Lisbon: Updated mitochondrial DNA portrait», *Forensic Sci. Int. Genet. Suppl. Ser.*, vol. 6, pp. e298–e300, dez. 2017, doi: 10.1016/j.fsigss.2017.09.134.
- [98] B. A. C. Hilário, «Genetic characterisation of the Central Portugal population through analysis of the mtDNA control region using NGS», de Aveiro, Aveiro, 2024. [Online]. Disponível em: <http://hdl.handle.net/10773/43724>
- [99] B. A. F. Rego, «Caracterização genética da população do Norte de Portugal através do estudo da Região Controlo do mtDNA com metodologia NGS», de Aveiro, Aveiro, 2024. [Online]. Disponível em: <http://hdl.handle.net/10773/44260>
- [100] S. Mahmud *et al.*, «Use of Next-Generation Sequencing for Identifying Mitochondrial Disorders», *Curr. Issues Mol. Biol.*, vol. 44, n.º 3, pp. 1127–1148, fev. 2022, doi: 10.3390/cimb44030074.
- [101] R. W. Taylor e D. M. Turnbull, «Mitochondrial DNA mutations in human disease», *Nat. Rev. Genet.*, vol. 6, n.º 5, pp. 389–402, mai. 2005, doi: 10.1038/nrg1606.
- [102] C. Chen, Y. Chen, e M.-X. Guan, «A peep into mitochondrial disorder: multifaceted from mitochondrial DNA mutations to nuclear gene modulation», *Protein Cell*, vol. 6, n.º 12, pp. 862–870, dez. 2015, doi: 10.1007/s13238-015-0175-z.
- [103] I. E. Nasseh, C. H. Tengan, B. H. Kiyomoto, e A. A. Gabbai, «Doenças Mitocondriais», *Rev. Neurociências*, vol. 9, n.º 2, pp. 60–69, jan. 2019, doi: 10.34024/rnc.2001.v9.8921.
- [104] J. M. W. Van Den Ouweland *et al.*, «Mutation in mitochondrial tRNA^{Leu}(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness», *Nat. Genet.*, vol. 1, n.º 5, pp. 368–371, ago. 1992, doi: 10.1038/ng0892-368.

[105] J. M. Shoffner, M. T. Lott, A. M. S. Lezza, P. Seibel, S. W. Ballinger, e D. C. Wallace, «Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA^{Lys} mutation», *Cell*, vol. 61, n.º 6, pp. 931–937, jun. 1990, doi: 10.1016/0092-8674(90)90059-N.

[106] G. M. Pastores *et al.*, «Leigh syndrome and hypertrophic cardiomyopathy in an infant with a mitochondrial DNA point mutation (T8993G)», *Am. J. Med. Genet.*, vol. 50, n.º 3, pp. 265–271, abr. 1994, doi: 10.1002/ajmg.1320500310.

[107] A. M. Cunha, O. Faria, e F. Falcão-Reis, «Neuropatia Ótica Hereditária de Leber: caracterização clínica e genética», *Rev. Soc. Port. Oftalmol.*, p. Vol. 44 N.º 4 (2020), dez. 2020, doi: 10.48560/RSPO.20305.

[108] A. Rötig *et al.*, «Pearson's marrow-pancreas syndrome. A multisystem mitochondrial disorder in infancy.», *J. Clin. Invest.*, vol. 86, n.º 5, pp. 1601–1608, nov. 1990, doi: 10.1172/JCI114881.

[109] P. Chagnon, M. Gee, M. Filion, Y. Robitaille, M. Belouchi, e D. Gauvreau, «Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population», *Am. J. Med. Genet.*, vol. 85, n.º 1, pp. 20–30, jul. 1999, doi: 10.1002/(sici)1096-8628(19990702)85:1<20::aid-ajmg6>3.0.co;2-k.

[110] J. M. Van Der Walt *et al.*, «Mitochondrial Polymorphisms Significantly Reduce the Risk of Parkinson Disease», *Am. J. Hum. Genet.*, vol. 72, n.º 4, pp. 804–811, abr. 2003, doi: 10.1086/373937.

[111] O. Ross, «mt4216C variant in linkage with the mtDNA TJ cluster may confer a susceptibility to mitochondrial dysfunction resulting in an increased risk of Parkinson's disease in the Irish», *Exp. Gerontol.*, vol. 38, n.º 4, pp. 397–405, abr. 2003, doi: 10.1016/S0531-5565(02)00266-8.

[112] E. Ruiz-Pesini *et al.*, «Human mtDNA Haplogroups Associated with High or Reduced Spermatozoa Motility», *Am. J. Hum. Genet.*, vol. 67, n.º 3, pp. 682–696, set. 2000, doi: 10.1086/303040.

[113] A.-K. Niemi, A. Hervonen, M. Hurme, P. Karhunen, M. Jylhä, e K. Majamaa, «Mitochondrial DNA polymorphisms associated with longevity in a Finnish population», *Hum. Genet.*, vol. 112, n.º 1, pp. 29–33, jan. 2003, doi: 10.1007/s00439-002-0843-y.

[114] M. Tanaka, J.-S. Gong, J. Zhang, Y. Yamada, H.-J. Borgeld, e K. Yagi, «Mitochondrial genotype associated with longevity and its inhibitory effect on mutagenesis», *Mech. Ageing Dev.*, vol. 116, n.º 2–3, pp. 65–76, jul. 2000, doi: 10.1016/S0047-6374(00)00149-4.

[115] C. Herrnstadt e N. Howell, «An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders», *Mitochondrion*, vol. 4, n.º 5–6, pp. 791–798, set. 2004, doi: 10.1016/j.mito.2004.07.041.

7. ANEXOS

Anexo 7.1 – Estudo aceite para comunicação oral no 22º Congresso Nacional de Medicina Legal e Ciências Forenses, intitulado “ADN mitocondrial: Estudo de validação interna de metodologia de sequenciação”



Anexo 7.2 – Tabela com as amostras de animais vertebrados não humano utilizadas nos testes de especificidade.

Espécie	Espécie
<i>Equus caballus</i>	<i>Felis catus</i>
<i>Sus scrofa domesticus</i>	M9 – 2017
<i>Bos taurus</i>	M9 – 2019
<i>Ovis aries</i>	M9 – 2022
<i>Canis lupus familiaris</i>	M9 – 2023

Anexo 7.3 – Tabela com todas as amostras e os respetivos haplogrupos e haplótipos da região controlo.

Amostra	Haplogrupo	Haplótipo
GENPOP 8472	N1a1a1a2	16086C 16147A 16164G 16172C 16223T 16248T 16261T 16320T 16355T 16519C 73G 152C 199C 204C 207A 263G 315.1C 573.1C
GENPOP 8474	H	16519C 263G 315.1C
GENPOP 8490	V3a	16240G 16298C 72C 263G 315.1C

GENPOP 8499	T1a1	16126C 16163G 16186T 16189C 16294T 16519C 73G 195C 263G 309.1C 315.1C
GENPOP 8680	L3e4a	16051G 16223T 16264T 16519C 73G 150T 263G 315.1C 523del 524del
GENPOP 8681	M	16519C 263G 309.1C 315.1C
GENPOP 8682	K1a3a	16224C 16311C 16519C 73G 263G 309.1C 315.1C 497T
GENPOP 8683	K1a4a1b1	16224C 16311C 16519C 73G 263G 309.1C 315.1C 497T
GENPOP 8684	J1c2a1	16069T 16126C 16519C 73G 185A 188G 228A 263G 295T 309.1C 315.1C 462T 489C
GENPOP 8691	T2b3+151	16126C 16294T 16296T 16304C 16519C 73G 151T 263G 309.1C 315.1C
GENPOP 8693	U5b2b	16148T 16261T 16390A 73G 94A 150T 263G 315.1C
GENPOP 8694	U4a1	16129A 16356C 16519C 73G 195C 196C 263G 315.1C 499A
GENPOP 8695	K1a1a1	16311C 263G 309.1C 315.1C
GENPOP 8892	U5b1b1	16189C 16193del 16270T 73G 150T 263G 315.1C
GENPOP 8893	T2b3c	16126C 16294T 16296T 16304C 73G 151T 200G 263G 309.1C 315.1C
GENPOP 8894	T2b	16126C 16294T 16296T 16304C 16519C 73G 195C 263G 309.1C 315.1C 321C
GENPOP 8895	H1	16519C 263G 309.1C 315.1C 524.AC
GENPOP 8896	H6a1b2e	16192T 16274A 16362C 16482G 239C 263G 309.1C 315.1C
GENPOP 8897	U4c1	16179T 16356C 16519C 73G 150T 195C 263G 315.1C 499A 524.AC
GENPOP 8898	H1	16519C 263G 309.1C 315.1C
GENPOP 8899	U5a2a1	16114A 16192T 16256T 16270T 16294T 16526A 73G 263G 309.1C 315.1C
GENPOP 8900	H1c3a	16176T 16219G 16519C 257G 263G 309.1C 315.1C 477C
GENPOP 8901	H1p	16519C 263G 315.1C
GENPOP 8902	U4a1b	16134T 16288C 16356C 16519C 73G 152C 195C 263G 315.1C 499A
GENPOP 8903	H3k	16519C 152C 263G 309.1C 315.1C 523del 524del
GENPOP 8904	H	16519C 263G 309.1C 315.1C 502T
GENPOP 8905	V3a	16519C 72C 151T 152Y 263G 309.1C 315.1C

GENPOP 8907	T2b3d	16126C 16294T 16296T 16304C 16344T 16519C 73G 151T 152C 263G 309.1C 315.1C
GENPOP 8908	H1a6	16183M 16183a 16189C 16519C 73G 263G 315.1C
GENPOP 8909	H4a1a1a1a	73G 152C 263G 309.1C 315.1C 523del 524del
GENPOP 8910	H1e1a	16519C 263G 315.1C 523del 524del
GENPOP 8913	U6a3b1	16172C 16183C 16189C 16219G 16278T 16311C 73G 146C 152C 185A 188G 263G 309.1C 315.1C
GENPOP 8914	H1	16316G 16519C 263G 315.1C
GENPOP 8915	U5a1c2a1	16227G 16256T 16270T 16311C 16320T 16399G 16519C 60.lt 71del 73G 152C 263G 309.1C 315.1C 573.1C
GENPOP 8916	H1j2	16519C 263G 315.1C
GENPOP 8917	M	16519C 195C 263G 309.1C 315.1C
GENPOP 8918	U5b2b3a1	16224C 16270T 16T 73G 150T 199C 263G 279C 315.1C 517T
GENPOP 8919	H1e1a1	16311C 16519C 146Y 150T 263G 315.1C
GENPOP 8920	L1c2b1c	16187T 16189C 16223T 16265C 16278T 16286A 16292T 16294T 16311C 16360T 16519C 16527T 73G 152C 182T 186A 189C 195C 198T 204C 247A 263G 297G 315.1C 316A
GENPOP 8921	H5a1j	16209C 16304C 263G 315.1C 456T 513A
GENPOP 8923	L2a1l2a	16093C 16189C 16192Y 16223T 16260T 16278T 16294T 16309G 16390A 16519C 73G 143A 146C 152C 195C 263G 309.1C 315.1C
GENPOP 8924	K1a3a2	16093C 16153A 16224C 16311C 16320T 16519C 73G 263G 309.1C 315.1C 497T
GENPOP 8925	H1c4b	16519C 263G 315.1C 477C 523del 524del
GENPOP 8931	H1j	16189C 16519C 263G 315.1C
GENPOP 8932	H31a	16319A 16519C 72G 146C 195C 263G 315.1C 533G
GENPOP 8933	W1	16223T 16292T 16320T 16519C 73G 119C 189G 195C 198T 204C 207A 263G 309.1C 315.1C
GENPOP 8934	L2a1l2	16189C 16223T 16278T 16294T 16390A 73G 146C 152C 195C 263G 309.1C 315.1C 534T
GENPOP 8939	H	16519C 263G 309.1C 315.1C 480C
GENPOP 8941	H	16519C 263G 315.1C
GENPOP 8942	L3flb1a1	16129A 16209C 16223T 16292T 16295T 16311C 16519C 73G 189G 200G 263G 272G 309.1C 315.1C
GENPOP 8943	U4c1	16179T 16356C 16519C 73G 150T 195C 263G 315.1C 499A 524.AC

GENPOP 8944	H	16519C 263G 315.1C
GENPOP 8945	R0a	16126C 16362C 60.1T 64T 263G 309c 315.1C 451G
GENPOP 8946	T2b	16126C 16294T 16296T 16304C 16519C 73G 263G 309.1C 315.1C
GENPOP 8947	V14	16093C 16298C 72C 263G 309.1C 315.1C
GENPOP 89848	M1a3b1	16183C 16189C 16249C 16265C 16274A 16311C 16519C 73G 195C 263G 309.1C 315.1C 489C
GENPOP 8949	U3a1c	16301T 16343G 16356C 16390A 16519C 73G 150T 263G 315.1C
GENPOP 8951	X2+225	16172C 16193del 16189C 16223T 16255A 16278T 16519C 73G 146C 153G 195C 225A 226C 263G 315.1C
GENPOP 8952	T1a2a	16126C 16186T 16189C 16261T 16294T 16519C 73G 152C 195C 196C 263G 309.1C 315.1C 384G
GENPOP 8953	L2a1fla	16519C 263G 315.1C
GENPOP 8954	U5b2b3a1	16152C 16224C 16270T 16T 73G 150T 199C 263G 279C 315.1C 517T
GENPOP 8956	R	16256T 16274A 16519C 73G 195C 214G 263G 309.1C 315.1C
GENPOP 8957	K1a4a1a2b	16224C 16311C 16519C 73G 189G 263G 309.1G 315.1G 497T 524.AC
GENPOP 8958	H1b1b	16129A 16183M 16183a 16189C 16355T 16356C 16362C 16519C 263G 309.2C 315.1C 523del 524del
GENPOP 8961	U4b1a4	16356C 16519C 73G 195C 263G 315.1c 499A
GENPOP 8967	J1c7	16069T 16126C 16261T 73G 146C 185A 263G 295T 309.1C 315.1C 462T 489C 513A
GENPOP 8968	H2a1e1	16354T 263G 310del 311c 315c 575T
GENPOP 8969	T2c1d+152	16126C 16292T 16294T 16519C 73G 146C 152C 263G 279C 309.1C 315.1C
GENPOP 8970	H1k1a	16189C 16290T 16519C 263G 309.2C 315.1C
GENPOP 8971	U6a1a1	16172C 16183C 16189C 16193del 16219G 16239T 16278T 16320T 73G 263G 309.1C 315.1C
GENPOP 8972	L2c	16126C 16183M 16183del 16189C 16223T 16269G 16278T 16293G 16390A 73G 93G 146C 150T 152C 182T 195C 198T 263G 309.1C 315.1C 325T 523del 524del
GENPOP 8973	H1	16519C 263G 309c 315.1C
GENPOP 8975	U5a1a1	16256T 16270T 16399G 73G 263G 309c 315.1C
GENPOP 8500	H	16235G 16291T 263G 315.1C
GENPOP 8501	J2b1c1	16069T 16126C 16193T 16319A 16360T 73G 150T 152C 263G 295T 309.1C 315.1C 489C 523del 524del

GENPOP 8503	V2a1	16298C 72C 235G 263G 309.1C 315.1C 390G
GENPOP 8504	H	16169T 16519C 263G 315.1C
GENPOP 8505	W1h	16192Y 16223T 16292T 16346A 16519C 73G 189G 195C 198T 204C 207A 263G 309.1C 315.1C
GENPOP 8508	U5b2a4a	16189C 16270T 73G 150T 263G 315.1C 524.AC
GENPOP 8510	U2e1a1	16051G 16129C 16183C 16189C 16362C 16519C 73G 152C 217C 263G 309.1C 315.1C 340T 508G 524.AC
GENPOP 8512	HV0b	16298C 72C 195C 198T 263G 315.1C
GENPOP 8516	H2a2	16235G 16291T 263G 309.1C 315.1C
GENPOP 8520	G1a1a	16519C 73G 263G 315.1C
GENPOP 8525	H7d3	16153A 16519C 263G 315.1C
GENPOP 8529	T2c1d	16126C 16292T 16294T 16296T 16519C 73G 146C 263G 315.1C
GENPOP 8532	M3d	16344T 16519C 93G 263G 315.1C
GENPOP 8533	D4	16093C 16192T 16223T 16271C 16316G 16362C 73G 184A 263G 309.1C 315.1C 489C
GENPOP 8535	H1c3	16519C 195C 257G 263G 309.1C 315.1C 376G 477C
GENPOP 8539	U5a2a1	16114A 16192T 16256T 16270T 16294T 16526A 73G 195C 263G 309.1C 315.1C
GENPOP 8542	U8b1b1	16094C 16183M 16183a 16189C 16234T 16324C 16519C 73G 152C 195C 263G 315.1C 573.4C
GENPOP 8546	U5a1i1	16192T 16256T 16270T 16399G 73G 207A 263G 315.1C
GENPOP 8547	M27b1	16145A 16519C 64T 263G 315.1C
GENPOP 8552	H6a1b2e	16362C 16482G 239C 263G 309.1C 315.1C
GENPOP 8557	H10a1b	16114T 16344T 16519C 263G 309.1C 315.1C
GENPOP 8558	U6a6b1	16092C 16172C 16219G 16278T 73G 263G 315.1C
GENPOP 8559	T2	16126C 16248T 16294T 16519C 73G 263G 309.1C 315.1C
GENPOP 8560	L2a112	16189C 16223T 16278T 16294T 16390A 73G 146C 152C 195C 263G 315.1C 534T
GENPOP 8561	R0a	16126C 16362C 60.1T 64T 263G 309.1C 315.1C 451G
GENPOP 8562	U4d3	16257T 16356C 16519C 73G 195C 263G 315.1C 499A 573.2C

GENPOP 8563	H10a	16519C 146C 263G 315.1C
GENPOP 8564	U6a7b1	16172C 16278T 73G 195C 198T 263G 309.1C 315.1C 455.1T
GENPOP 8565	U4a1	16129A 16293G 16356C 16519C 73G 195C 196C 263G 315.1C 499A
GENPOP 8566	K1a4a1g	16196A 16224C 16311C 16362C 16519C 73G 152C 263G 309.1C 315.1C 497T 524.ACAC
GENPOP 8567	H1c3	16519C 195C 257G 263G 309.1C 315.1C 477C
GENPOP 8570	H5a1c1a	16304C 195C 263G 315.1C 456T 523del 524del
GENPOP 8573	H3q1	16362C 16519C 152C 263G 315.1C
GENPOP 8574	L0a2a2a2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523del 524del
GENPOP 8575	K1a3a3	16224C 16234T 16311C 16356C 16519C 73G 263G 309.1C 315.1C 497T
GENPOP 8582	H5a4a1	16294T 16304C 16320T 263G 315.1C 456T
GENPOP 8584	K1a2c	16224C 16311C 16519C 73G 146C 152C 263G 315.1C 497T 524.ACAC
GENPOP 8586	T1a1	16126C 16163G 16186T 16187T 16189C 16294T 16519C 73G 152C 195C 263G 315.1C
GENPOP 8587	L1b1a12a	16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16400T 16519C 73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523del 524del
GENPOP 8589	H	16519C 263G 315.1C
GENPOP 8590	V	16298C 16519C 72C 263G 309.1C 315.1C
GENPOP 8591	H1a1a	16162G 16519C 73G 93G 263G 315.1C
GENPOP 8594	H1	16189C 16519C 263G 315.1C
GENPOP 8595	U5b1e	16192T 16259T 16270T 73G 150T 195C 263G 309.1C 315.1C
GENPOP 8596	H4a1a1a1a1	73G 263G 315.1C 523del 524del
GENPOP 8597	L3b1a+@16124	16223T 16278T 16362C 16519C 73G 263G 315.1C 523del 524del
GENPOP 8599	H1ao	16278T 16519C 93G 146C 263G 315.1C
GENPOP 8600	U5a1a1	16176T 16256T 16270T 16399G 73G 263G 315.1C 374R
GENPOP 8601	T1a1	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 309.1C 315.1C
GENPOP 8603	K1a4a4a1b1	16224C 16311C 16519C 73G 263G 309.1C 315.1C 497T

GENPOP 8604	U5b2a1a2	16278T 16519C 73G 150T 263G 315.1C
GENPOP 8610	I4a	16129A 16223T 16391A 16519 73G 199C 204C 250C 263G 315.1C 573.CCC
GENPOP 8613	V	16298C 16320T 72C 131C 263G 309.1C 315.1C
GENPOP 8618	U4c1	16179T 16356C 16519C 73G 195C 263G 309.1C 315.1 499A
GENPOP 8619	H2a2	16235G 16291T 263G 309.1C 315.1C
GENPOP 8624	U6a7a1b	16172C 16219G 16278T 73G 150T 151A 152C 263G 294C 315.1C
GENPOP 8625	H30b1	16519C 152C 263G 309.1C 315.1C
GENPOP 8626	H3q1	16519C 152C 263G 315.1C
GENPOP 8627	L0d2	16519C 263G 309.1C 315.1C
GENPOP 8628	H3g4	16242T 16519C 152C 263G 309.1C 315.1C
GENPOP 8633	H3	16519C 152C 263G 309c 315.1C
GENPOP 8638	J1c2	16069T 16126C 16519C 73G 185A 188G 228A 263G 295T 315.1C 462T 489C
GENPOP 8639	U5a1b	16192T 16256T 16270T 16298C 16399G 73G 263G 315.1C
GENPOP 8642	U5b1g	16192T 16270T 16519C 73G 150T 151T 228A 263G 315.1C 573.CC
GENPOP 8643	K2b1a	16224C 16311C 16362C 16519C 73G 146C 263G 315.1C 572T
GENPOP 8644	U5b2b3a1	16224C 16270T 16T 73G 150T 199C 263G 279C 315.1C 517T
GENPOP 8648	H3ap	16519C 263G 309.1C 315.1C
GENPOP 8651	T1a2a	16126C 16163G 16172C 16186T 16189C 16261T 16283G 16294T 16519C 73G 152C 263G 309c 315.1C 384G
GENPOP 8654	V	16298C 16311C 72C 263G 315.1C
GENPOP 8655	W3a1c	16182C 16183C 16188T 16189C 16223T 16292T 16519C 73G 189G 194T 195C 204C 207A 263G 309.1C 315.1C
GENPOP 8656	H1	16207G 16519C 152C 263G 315.1C
GENPOP 8657	T2b	16126C 16294T 16296T 16304C 16519C 73G 263G 315.1C
GENPOP 8658	L2'3'4'5'6	73G 152C 263G 315.1C
GENPOP 8660	J1d5	16183M 16183a 16189C 16519C 73G 263G 315.1C

GENPOP 8661	T2b4	16126C 16192T 16294T 16304C 16519C 73G 263G 309.1C 315.1C
GENPOP 8662	J1c2b1	16069T 16126C 16179T 73G 185A 228A 263G 295T 315.1C 462T 489C
GENPOP 8664	K1b1a1d1	16223T 16224C 16311C 16319A 16463G 16519C 73G 152C 263G 309.1C 315.1C
GENPOP 8666	H1c3	16519C 195C 257G 263G 309.1C 315.1C 477C
GENPOP 8668	H5a2	16304C 263G 309.1C 315.1C 456T
GENPOP 8669	I1a1c	16129A 16172C 16223T 16311C 16391A 16519C 73G 199C 203A 204C 250C 263G 315.1C 455.1T 573.3C
GENPOP 8674	H2a2	16235G 16291T 93G 263G 309.1C 315.1C
FIA151	H30b1	16519C 152C 263G 309.1C 315.1C