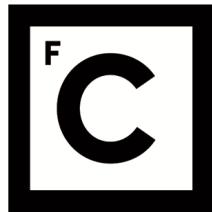


UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE BIOLOGIA VEGETAL



**Ciências**  
**ULisboa**

## **Caracterização genética da população do Norte de Portugal através do estudo da região controlo de mtDNA**

Inês do Rosário Colaço Balsemão Barbosa

**Mestrado em Biologia Molecular e Genética**

Dissertação orientada por:  
Doutora Benedita Abreu Ferreira da Silva  
Professor Doutor António Amorim Santos

## Agradecimentos

Ao longo deste projeto fui acompanhada por várias pessoas que, de forma indelével, com as suas sugestões, contribuíram e permitiram a realização deste trabalho. A todos o meu muito obrigado.

Deixo um agradecimento especial,

À minha orientadora, Doutora Benedita Silva, pelo apoio incondicional, orientação, disponibilidade, ajuda, alegria, ensinamentos e pertinência das suas críticas e sugestões concedidas. Bem como a sua amizade desde o primeiro dia.

Ao Professor Doutor António Amorim, por ter aceitado orientar a minha tese, na qualidade de orientador interno, pelo apoio prestado e pelas valiosas recomendações que permitiram reformular a tese enriquecendo-a com mais informação.

À Dra. Gabriela Lima e à Dra. Heloísa Costa, pelo apoio incondicional. As suas relevantes informações, disponibilidade e conhecimentos permitiram que a minha tarefa tenha sido mais produtiva e tornaram possível a realização desta tese.

À restante equipa do Serviço de Genética e Biologia Forense do INMLCF da delegação do Norte e Sul, por me terem recebido de braços abertos.

À Ana, pela amizade que surgiu no desenvolvimento deste projeto e por todo o apoio e momentos que passámos juntas.

Aos meus amigos, pelo apoio especial que me prestaram ao longo deste trabalho e anos.

Aos meus pais, irmão e avós, que sempre me apoiaram, incentivaram e por me terem acompanhado em mais uma etapa da minha vida.

Ao Vasco pela companhia, carinho e pela constante motivação que me transmite.

À faculdade, aos seus docentes e funcionários, por tudo o que me deram ao longo destes 2 anos tanto a nível académico como pessoal.

Finalmente, ao curso que hoje, fruto dos recentes acontecimentos, mais do que nunca, assume um papel relevante no desenvolvimento de um mundo mais bem preparado para os desafios que se avizinham.

## Resumo

O DNA mitocondrial (mtDNA) apresenta características como o elevado número de cópias por célula, herança uniparental materna, elevada taxa de mutação e ausência de recombinação, que o tornam útil em estudos de origem e evolução humana, assim como na área forense.

Dada a sua relevância, o objetivo principal deste trabalho prende-se com a caracterização genética da população da região do Norte de Portugal através do estudo do mtDNA. Foram estudados 87 indivíduos, residentes nesta região e não relacionados entre si. No âmbito da realização de ensaios de sequenciação de mtDNA, foi ainda realizada a validação interna no laboratório da Delegação do Norte, do Serviço de Genética e Biologia Forense, do Instituto Nacional de Medicina Legal e Ciências Forenses, do novo sequenciador automático *Seqstudio™ Genetic Analyzer Instrument*. Os ensaios de sensibilidade e especificidade foram concluídos com sucesso, ficando definida a concentração mínima detetável em 0,001ng. Foram validados os ensaios para as amostras de cabelos, manchas de sangue e zaragatoas bucais com resultados satisfatórios.

No estudo populacional, a região controlo total do mtDNA foi amplificada e sequenciada, utilizando dois pares de primers – L15971/H016 e L16555/H639. A sua análise revelou elevada variabilidade genética, com elevada frequência de haplótipo únicos. Estes, corresponderam, na sua maioria (59,52%) a haplogrupos característicos e típicos de populações europeias. A partir dos estudos filogenéticos comparativos, quanto às populações estudadas, verificou-se que a população do Norte de Portugal se encontra geneticamente mais próxima das populações de Portugal e Alemanha e mais distante da população de imigrantes de Cabo Verde.

Palavras-Chave: DNA mitocondrial, *Seqstudio™ Genetic Analyzer*, Validação interna, Região do Norte de Portugal, Genética Populacional

## Abstract

Mitochondrial DNA presents several characteristics such as a high number of copies per cell, maternal uniparental inheritance, high mutation rate, and absence of recombination, which make it useful in studies of human origin and evolution, as well as in forensics.

Given its relevance, the main objective of this work concerns the genetic characterization of the population of the Northern region of Portugal through the study of mtDNA. A total of 87 unrelated individuals living in this locality were studied. As part of carrying out mtDNA sequencing tests, internal validation was also carried out in the laboratory of the North Delegation, of the Genetics and Forensic Biology Service, of the National Institute of Legal Medicine and Forensic Sciences, of the new automatic sequencer *Seqstudio™ Genetic Analyzer Instrument*. Sensitivity and specificity tests were successfully followed, with the minimum detectable concentration being defined at 0.001ng. Assays for hair samples, bloodstains, and buccal swabs with satisfactory results have been validated.

In the populational study, the total mtDNA control region was amplified and sequenced using two pairs of primers – L15971/H016 and L16555/H639. Their analysis revealed a high genetic variability, with a high frequency of unique haplotypes. Most of these (59.52%) corresponded to haplogroups characteristic and typical of European populations. From the comparative phylogenetic studies, it was found that the population of Northern Portugal is genetically closer to the populations of Portugal and Germany and more distant from the immigrant population of Cape Verde.

Keywords: Mitochondrial DNA, *Seqstudio™ Genetic Analyzer*, Internal validation, Northern Portugal Region, Population Genetics

# Índice

<b>Agradecimentos</b> .....	<b>i</b>
<b>Resumo</b> .....	<b>ii</b>
<b>Abstract</b> .....	<b>iii</b>
<b>Lista de Quadro e Figuras</b> .....	<b>vi</b>
<b>Lista de abreviaturas, siglas e símbolos</b> .....	<b>vii</b>
<b>1. Introdução</b> .....	<b>1</b>
<b>1.1 DNA Mitocondrial</b> .....	<b>1</b>
1.1.1 Características do mtDNA .....	2
1.1.2 Heteroplasmia .....	2
1.1.3 Aplicações do mtDNA.....	3
1.1.3.1 Genética Populacional .....	3
1.1.3.2 Aplicações Forenses.....	4
1.1.4 Bases de Dados de mtDNA .....	4
<b>1.2 Validação do método</b> .....	<b>5</b>
<b>1.3 Estudo Populacional - Região do Norte de Portugal</b> .....	<b>5</b>
<b>1.4 Objetivos</b> .....	<b>6</b>
<b>2. Materiais e Métodos</b> .....	<b>6</b>
<b>2.1 Amostras Biológicas para os ensaios de validação</b> .....	<b>6</b>
<b>2.2 Amostras Biológicas para o estudo populacional</b> .....	<b>6</b>
<b>2.3 Preparação de Amostras Biológicas</b> .....	<b>7</b>
<b>2.4 Extração do DNA</b> .....	<b>7</b>
• Extração por <i>Prep-n-Go™ Buffer</i> .....	7
• Extração por <i>PrepFiler Express™</i> e <i>PrepFiler Express™ BTA</i> .....	7
• Extração por <i>Chelex®100 (Bio-Rad)</i> .....	8
<b>2.5 Amplificação por PCR</b> .....	<b>8</b>
<b>2.6 Verificação da Amplificação</b> .....	<b>9</b>
<b>2.7 Purificação dos produtos amplificados</b> .....	<b>9</b>
<b>2.8 Sequenciação direta</b> .....	<b>10</b>
<b>2.9 Limpeza dos terminadores</b> .....	<b>10</b>
<b>2.10 Eletroforese capilar no sequenciador automático e Análise dos produtos sequenciados</b> ....	<b>11</b>
<b>2.11 Análise estatística e filogenética</b> .....	<b>12</b>
<b>3. Resultados e Discussão</b> .....	<b>13</b>

<b>3.1 Validação da análise dos polimorfismos do DNA mitocondrial por sequenciação no <i>SeqStudio™ Genetic Analyzer Instrument</i></b> .....	<b>13</b>
3.1.1 Validação de resultados dos controlos positivos e amostras controlo .....	13
3.1.2 Estudos de Sensibilidade .....	14
3.1.3 Estudos de Especificidade.....	14
3.1.4 Validação do estudo de mtDNA de cabelos extraídos por <i>PrepFiler Express™ BTA</i> .....	15
3.1.5 Validação do estudo da região controlo total do mtDNA.....	16
3.1.6 Validação do estudo de mtDNA nas manchas de sangue FTA.....	16
3.1.7 Validação do estudo de mtDNA de zaragoas bucais extraídas por <i>Prep-n-Go™</i> .....	17
<b>3.2 Estudo Populacional da Região do Norte.....</b>	<b>17</b>
3.2.1 Estudo dos Haplótipos.....	18
3.2.2 Estudo das heteroplasmias.....	21
3.2.3 Estudo dos haplogrupos .....	21
• Haplogrupo R, R0 e HV .....	22
• Haplogrupo H .....	23
• Haplogrupo L .....	23
• Haplogrupo M .....	23
• Haplogrupo T.....	24
• Haplogrupo U .....	24
• Haplogrupo J.....	24
• Haplogrupo K.....	25
• Haplogrupo W, X e V .....	25
3.2.4 Análise Filogenética.....	25
<b>4. Conclusão.....</b>	<b>28</b>
<b>5. Bibliografia .....</b>	<b>30</b>
<b>6. Anexos .....</b>	<b>41</b>

## Lista de Quadro e Figuras

**Figura 1.1.1** – Esquema representativo da molécula de mtDNA humana.

**Figura 1.1.3.1.1** – Mapa com representação esquemática das migrações do Homem pelo mundo e distribuição geográfica dos haplogrupos.

**Figura 2.8.1** – Sequenciação direta do DNA com ddNTPs fluorescentes.

**Figura 2.10.1** – Exemplo de um segmento de um eletroferograma obtido no programa *Sequencing Analysis 7*, após sequenciação com o *primer* H016.

**Figura 2.10.2** – Representação, através de um eletroferograma, de um segmento (a) com heteroplasmia de comprimento e (b) com heteroplasmia de posição.

**Figura 3.1.1.1** – Gel de eletroforese capilar (*QIAxcel® ScreenGel*) para verificação da amplificação dos controlos positivos com os *primers* HV1, HV2, HV3 e RV.

**Figura 3.1.2.1** – Gel de eletroforese capilar (*QIAxcel® ScreenGel*) para verificação da amplificação do teste de sensibilidade com os *primers* HV1, HV2, HV3 e RV.

**Figura 3.2.3.1** – Representação esquemática dos diferentes haplogrupos e respetivas frequências identificados nas 84 sequências de mtDNA de indivíduos da região do norte de Portugal.

**Figura 3.2.4.1** - Representação filogenética das distâncias genéticas entre a população do Norte de Portugal e as restantes selecionadas da literatura.

**Tabela 1.1.1** – *Primers* utilizados para análise da região controlo do mtDNA.

**Tabela 2.10.1** – Código IUPAC utilizado na análise do mtDNA.

**Tabela 3.2.1.1** – Haplótipos mais frequentes nas 84 sequências de indivíduos da região do norte de Portugal, estudadas no presente estudo, bem como a sua frequência.

**Tabela 3.2.1.2** – Número e tipo de alterações polimórficas identificadas na região controlo total do mtDNA das 84 amostras de indivíduos da região do norte de Portugal.

**Tabela 3.2.1.3** – Polimorfismos observados apenas uma única vez na região HV1 do mtDNA.

**Tabela 3.2.1.4** – Polimorfismos observados apenas uma única vez na região HV2 do mtDNA.

**Tabela 3.2.1.5** - Polimorfismos observados apenas uma única vez na região HV3 do mtDNA.

**Tabela 3.2.1.6** - Polimorfismos observados apenas uma única vez nas regiões não compreendidas por HV1, HV2 e HV3 do mtDNA.

**Tabela 3.2.2.1** – Frequência de haplótipos homoplásmicos e heteroplásmicos nas 84 sequências de mtDNA analisadas e região hipervariável de ocorrência.

**Tabela 1.2.4.1** - População do presente estudo e populações selecionadas da literatura, utilizadas no estudo filogenético, tendo por base a região controlo total do mtDNA.

**Tabela 3.2.4.2** - Valores de  $F_{st}$  e  $p$  relativos às populações do estudo filogenético obtidos através do programa *Arlequin* para a totalidade da região controlo do mtDNA.

## Lista de abreviaturas, siglas e símbolos

<b>A</b>	Adenina
<b>AMOVA</b>	Análise de variância molecular
<b>C</b>	Citosina
<b>CE</b>	Eletroforese capilar (Inglês, <i>capillary electrophoresis</i> )
<b>CRS</b>	<i>Cambridge Reference Sequence</i>
<b>ddNTPs</b>	Didesoxinucleótidos Trifosfatados
<b>DEL</b>	Deleção
<b>DNA</b>	<i>Deoxyribonucleic acid</i> (Ácido Desoxirribonucléico)
<b>dNTPs</b>	Desoxinucleótidos Trifosfatos
<b>Electroferograma</b>	Diagrama de resultados obtidos na análise feita por eletroforese em sequenciador automático. No caso da sequenciação, o electroferograma é constituído por uma série de picos adjacentes, sendo que, cada pico possui uma das quatro cores possíveis que representam cada uma das diferentes bases que constituem o ADN.
<b>EDNAP</b>	<i>European DNA Profiling Group</i>
<b>EMPOP</b>	<i>EDNAP mitochondrial DNA population database</i>
<b>G</b>	Guanina
<b>HV1</b>	Região hipervariável 1
<b>HV2</b>	Região hipervariável 2
<b>HV3</b>	Região hipervariável 3
<b>INMLCF</b>	Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.
<b>IUPAC</b>	<i>International Union of Pure &amp; Applied Chemistry</i>
<b>mtDNA</b>	DNA mitocondrial
<b>Pb</b>	Par de bases (Inglês, <i>Base pair</i> )
<b>PCR</b>	<i>Polymerase chain reaction</i> (Reação em Cadeia de Polimerase)
<b>Perfil Genético</b>	Conjunto de características hereditárias ou padrões genotípicos que cada indivíduo possui para um determinado número de marcadores genéticos.
<b>Polimorfismo de DNA</b>	Regiões da molécula de DNA que apresentam diferentes formas nos indivíduos de uma população. Também denominado polimorfismo genético.
<b>Primer</b>	Pequeno segmento/fragmento de DNA que determina a posição a partir da qual a cadeia de DNA vai ser copiada.
<b>rCRS</b>	<i>Revised Cambridge Reference Sequence</i>
<b>RFLP</b>	<i>Restriction fragment length polymorphism</i>
<b>rpm</b>	Rotações por minuto
<b>RV1</b>	Região hipervariável localizada entre as regiões HV1 e HV2
<b>SGBF-N</b>	Serviço de Genético e Biologia Forense – Delegação do Norte
<b>T</b>	Timina
<b>Taq Polimerase</b>	Enzima Polimerase do ADN da bactéria <i>Thermus aquaticus</i>

# 1. Introdução

## 1.1 DNA Mitocondrial

A utilização de DNA humano em genética forense percorreu um longo caminho desde que o “*DNA fingerprinting*” foi descrito em 1985 [1]. Após esta descoberta, que teve um impacto significativo no sistema de justiça e na sociedade, em geral, as sequências de DNA humano repetidas em tandem tornaram-se a escolha preferida para estudos forenses. No entanto, quando se está perante casos nos quais o DNA nuclear está demasiado degradado ou não se consegue colher em quantidade suficiente para ser analisado com resultados reportáveis, pode ter de se recorrer ao mtDNA (DNA mitocondrial).

As mitocôndrias são organelos semiautónomos e dinâmicos encontrados na maioria das células eucarióticas. Uma vez que apresentam um genoma próprio, mtDNA, são capazes de manter a capacidade de replicação e síntese proteica independentemente do resto da célula [2]. Têm como principal papel a produção de energia da célula, na síntese de ATP através da fosforilação oxidativa. Adicionalmente, também estão envolvidas noutros processos metabólicos como a sinalização celular para transporte de cálcio, apoptose, biossíntese de ácidos gordos e ciclo de Krebs ([3–5]).

O mtDNA humano descoberto em 1963 por M. Nass e S. Nass, consiste numa molécula circular de cadeia dupla com cerca de 16569 pares de bases [6,7]. É constituída maioritariamente por uma região codificante, com a exceção da região controlo (D-loop). Esta região controlo tem um comprimento de aproximadamente 1100 pb e está localizada entre as posições 16024-576. As posições dos nucleótidos são numeradas de acordo com a sequência Anderson revista [8]. A primeira posição situa-se no meio da região controlo e separa as duas regiões hipervariáveis (HV1 e HV2), a região HV1 situa-se nas posições 16024 a 16365, enquanto a HV2 engloba as posições 73 a 340 (figura 1.1.1) [9,10]. A pequena dimensão de cada região permite a amplificação por PCR e, portanto, HV1 e HV2 são sequenciadas para fins forenses. No entanto, se as amostras forem sujeitas a grandes processos de degradação, os cerca de 300pb que constituem HV1 e HV2 podem não estar intactos. Desta maneira, para além dos *primers forward* e *reverse* HV1 e HV2, é comum, na amplificação, utilizar primers para a região hipervariável HV3 situada entre as posições 438 a 574 [11] (figura 1.1.1). Na tabela 1.1.1 apresentamos um dos conjuntos de primers que possibilita o estudo da região controlo do mtDNA.

Em 1981 foi anunciada a sequência do DNA mitocondrial humano que ficou determinada como a sequência de referência de Cambridge (CRS) [8]. Em 1999 esta sequência foi revista e foi criada a sequência rCRS, uma versão corrigida da original [12], sendo a utilizada como base universal de comparação. As sequências de uma determinada amostra são comparadas com a rCRS por forma a obter o haplótipo de cada indivíduo. Este haplótipo é constituído pelo conjunto de alterações observadas relativamente à rCRS e é descrito pela posição onde ocorreu a mutação bem como pela base nucleotídica alterada [13].

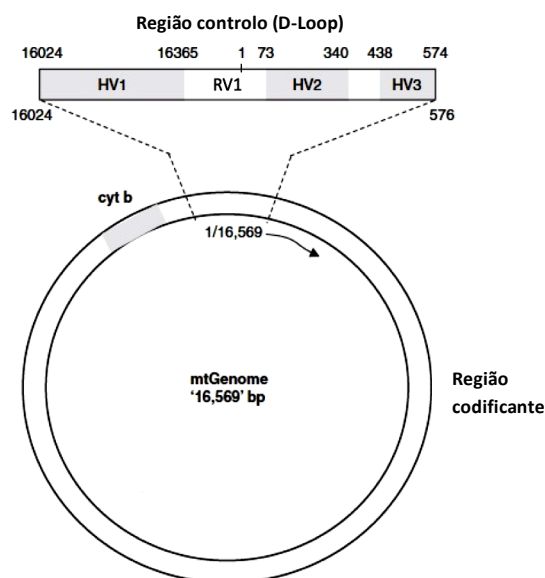


Figura 1.1.1 - Esquema representativo da molécula de mtDNA humana.

Fonte: Adaptado de Butler, J. M. 2009.

Tabela 1.1.1 - Primers utilizados para análise da região controlo do mtDNA.

Região	Primer Forward	Sequência 5'-3'	Primer Reverse	Sequência 5'-3'	Região alvo
HV1	F15989	CCCAAAGCTAAGATTCTAAT	R16433	TTGTGCGGGATATTGATTTCAC	16024 - 16365
HV2	F16533	TAGCCACACGTTCCCCTT	R460	AGGGGAAAATAATGTGTAGT	73 - 340
HV3	F317	CTTCTGGCCACAGCACTTAAAC	R619	AACATTTTCAGTGTATTGCTTTGAG	438 - 574
RV1	F16363	CCCCATGGATGACCCCCC	R293	AATTTTTTGTATGATGTCTGTGTGG	16366 - 72
Total	F15989	CCCAAAGCTAAGATTCTAAT	R52	AAATGCATGGAGAGCTCC	16024 - 16569
	F16533	TAGCCACACGTTCCCCTT	R619	AACATTTTCAGTGTATTGCTTTGAG	1-576
	L15971	TTA ACT CCA CCA TTA GCA CC	H016	TGA TAG ACC TGT GAT CCA TCG TGA	16024 - 16569
	L16555	CCC ACA CGT TCC CCT TAA AT	H639	GGG TGA TGT GAG CCC GTC TA	1 - 576

### 1.1.1 Características do mtDNA

Uma das principais vantagens do mtDNA relativamente ao DNA nuclear, consiste no facto de haver um maior número de cópias por células, o que torna mais fácil a obtenção de resultados em amostras muito degradadas e/ou antigas. Enquanto o DNA nuclear contém apenas 2 cópias por células, o mtDNA oscila entre 2 a 10 cópias por mitocôndria o que corresponde a cerca de  $10^3$  a  $10^4$  cópias por células [14,15].

Uma outra vantagem reside no facto do mtDNA ter uma transmissão maioritariamente materna, ou seja, todos os indivíduos que tenham a mesma herança materna possuem o mesmo mtDNA (herança haplotípica uniparental materna), com exceção da ocorrência de mutações [16,17]. Esta característica torna o mtDNA útil em estudos de linhagens, ancestralidade e evoluções de populações. No entanto, esta vantagem acarreta a desvantagem de não permitir uma identificação individual, tal como pretendida na maioria dos casos forenses.

Outra das características que faz do mtDNA um bom marcador genético é ser uma molécula bastante polimórfica, devido à elevada taxa de mutação. A aparente falta de mecanismos de reparação bem como a baixa fidelidade da mtDNA polimerase fazem com que o mtDNA evolua 5 a 10 vezes mais rapidamente que o genoma nuclear [18,19]. A maior parte da variação encontra-se nos segmentos hipervariáveis, HV1, HV2 e HV3[18]. Em média, existem 8 diferenças nucleotídicas no mtDNA entre indivíduos caucasianos e 15 entre indivíduos de origem africana quando pertencentes a diferentes linhagens maternas [20].

A ausência de recombinação genética é outra das características presentes no mtDNA [21,22]. Assim, as variações encontradas no genoma mitocondrial resultam de mutações que se vão acumulando ao longo de várias gerações, permitindo estabelecer relações filogenéticas entre indivíduos que partilhem a mesma linhagem materna [10].

### 1.1.2 Heteroplasmia

Uma dificuldade técnica frequentemente encontrada é a presença de heteroplasmia em alguns indivíduos. A heteroplasmia é caracterizada pela deteção de mais do que uma sequência de mtDNA num único indivíduo [23] que, embora possa ser utilizada como um fator de identificação [24], na maioria das vezes dificulta-a.

Podem ser classificadas em heteroplasmias de posição e de comprimento, sendo a de comprimento a forma mais comum encontrada na região controladora do mtDNA, nomeadamente nos segmentos HV1 e HV2. Este tipo caracteriza-se pela coexistência de pelo menos duas moléculas de mtDNA que diferem em número de nucleótidos, geralmente citosina, traduzindo-se numa diferença de comprimento [23].

As heteroplasmias de posição estão associadas a mutações pontuais em apenas algumas sequências de mtDNA de um indivíduo e são caracterizadas pela sobreposição de dois picos de nucleótidos diferentes no electroferograma [25].

### 1.1.3 Aplicações do mtDNA

Devido às suas características, o estudo do genoma mitocondrial tem aplicações nas mais diversas áreas, incluindo na genética populacional e forense. Os primeiros estudos efetuados tinham por base o uso de técnicas de restrição enzimática que utilizavam a molécula de mtDNA total, - RFLP (*Restriction fragment length polymorphism*) - no entanto, apresentavam uma baixa capacidade de discriminação e resolução [6,17]. Com o avanço tecnológico dos métodos de amplificação - Polymerase Chain Reaction (PCR) - e de sequenciação de Sanger, foi possível estudar mais detalhadamente a molécula do mtDNA, aumentando a qualidade e a credibilidade dos resultados obtidos. Inicialmente, começou-se por estudar o segmento HV1 [26,27], incluindo-se mais tarde o estudo do segmento HV2 [28,29]. Mais recentemente, tornou-se comum o estudo das três regiões hipervariáveis (HV1, HV2 e HV3) [30] bem como o estudo da totalidade da região controladora do mtDNA [31–33], o que veio aumentar o poder de individualização entre as amostras.

#### 1.1.3.1 Genética Populacional

Devido às características particulares do mtDNA, como a herança estritamente materna e a ausência de recombinação, é possível determinar a composição genética da linhagem materna de uma população. Tal permite reconstruir a sua origem e evolução.

Existem duas teorias relativamente à origem do homem moderno: a teoria *Out of Africa* [6,34] e a teoria Multirregional [35,36]. Atualmente, a teoria mais aceite é a *Out of Africa* que postula que o homem moderno teve origem em África, há aproximadamente 200 mil anos, propagando-se pelo resto do mundo, substituindo as populações humanas mais arcaicas, como os Neandertais (figura 1.1.3.1.1). O ancestral feminino comum é conhecido como Eva Mitocondrial [6,37].

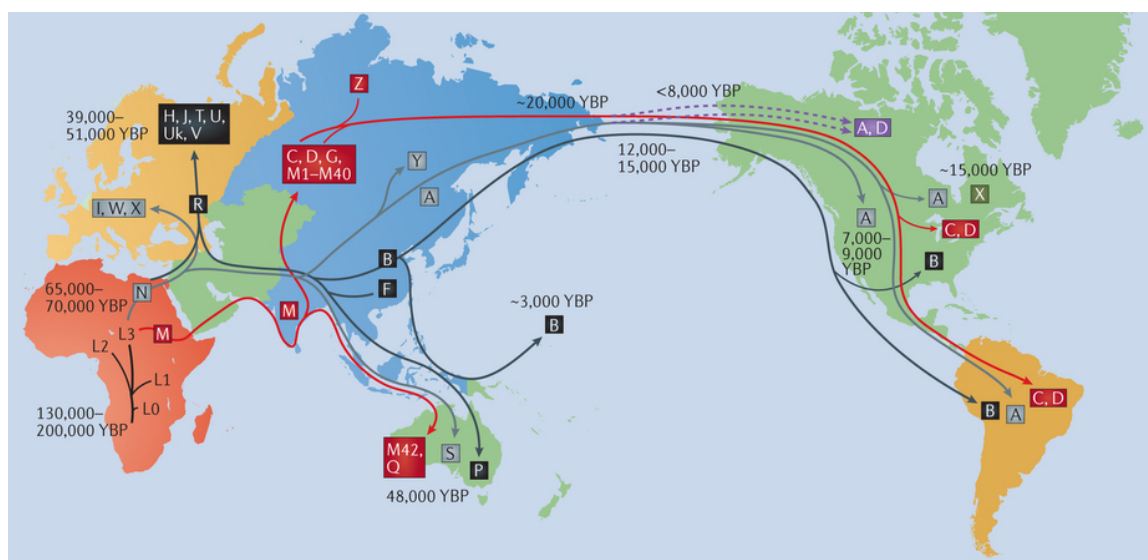


Figura 1.1.3.1.1 – Mapa com representação esquemática das migrações do Homem pelo mundo e distribuição geográfica dos haplogrupos. (Adaptado de [38])

A evolução dos indivíduos ao longo das gerações é refletida pela acumulação de mutações na sequência de mtDNA. Estas mutações definem o haplótipo de cada indivíduo que, por sua vez, é agrupado em haplogrupos. Cada haplogrupo contém haplótipos similares, com polimorfismos específicos, que partilham um ancestral comum. Desta forma, através do estudo das mutações do mtDNA de cada indivíduo é possível contruir árvores filogenéticas, bem como reconstruir os processos migratórios que ocorreram ao longo dos anos [39]. Atualmente, a árvore filogenética mais utilizada é a Phylotree [40]. Esta engloba os polimorfismos presentes tanto na região controladora como na região codificante do mtDNA estando permanentemente a ser atualizada, tendo a última revisão sido feita em Fevereiro de 2016. A raiz desta árvore engloba o haplogrupo característico da África Subsariana, o macrohaplogrupo L que se ramifica nos haplogrupos L0, L1, L2, L3, L4, L5 e L6, característicos de populações africanas [40–42]. Foi há cerca de 60 a 70 mil anos que, a partir do haplogrupo L3, surgiram os macrohaplogrupos M e N, presentes na Ásia e na Europa [43]. A partir do macrohaplogrupo N surgiram os haplogrupos H, I, J, K, T, U, V, W e X, constituintes de aproximadamente 99% dos haplótipos europeus [10,44]. As linhagens asiáticas derivam dos haplogrupos A, B, F e Y, que se ramificaram a partir do macrohaplogrupo N, e dos haplogrupos C, D, E, G e Z, que se ramificaram a partir do macrohaplogrupo M [45,46]. O continente americano é constituído, na grande maioria pelos haplogrupos A, B, C, D e X, sugerindo que, antes da colonização europeia, as populações asiáticas haviam entrado no continente americano [47].

#### 1.1.3.2 Aplicações Forenses

Ao longo dos tempos, o ramo da genética forense tem-se concentrado na utilização de DNA, mais especificamente de marcadores genéticos também conhecidos como polimorfismos de DNA, para permitir a identificação genética em perícias forenses.

Quando as amostras biológicas não apresentam DNA nuclear em quantidade e qualidade desejáveis é necessário usar o DNA mitocondrial. Embora este possua um menor poder de discriminação tem-se demonstrado eficiente em casos onde não se consegue utilizar o DNA nuclear [48–50]. Em casos forenses que tem como objetivo a comparação entre indivíduos que pertencem à mesma linhagem materna, como por exemplo a identificação de um familiar desaparecido, a utilização do mtDNA é vantajosa, uma vez que estes indivíduos apresentam o mesmo haplótipo, salvo mutações que são raras [10]. Contudo, a análise do mtDNA também apresenta desvantagens sendo a principal, a impossibilidade de definir a identidade individual de um indivíduo [51].

A primeira utilização do mtDNA para resolver um caso forense ocorreu em 1996, utilizado para confirmar o suspeito de um crime de violação e homicídio em Tenesse, nos Estados Unidos [52]. Um dos casos mais conhecidos efetuado em restos cadavéricos, que envolveu a utilização mtDNA, foi a identificação de indivíduos pertencentes à família Romanov, último Czar da Rússia [53].

#### 1.1.4 Bases de Dados de mtDNA

Com o uso crescente do mtDNA para resolver casos forenses, surgiu a necessidade de se criar bases de dados de mtDNA. Estas vieram permitir uma comparação dos haplótipos bem como a determinação da sua incidência e da frequência de mutações numa determinada população [54]. De entre as inúmeras bases de dados existentes, destacam-se a EMPOP [55], MITOMAP [56] e mtDB [57].

A EMPOP (EDNAP mitochondrial DNA population database), base de dados estabelecida no Instituto de Medicina Legal de Innsbruck, é a mais utilizada a nível mundial para estudos de mtDNA. Esta tem como objetivo aumentar a qualidade e a confiança das sequências de mtDNA permitindo obter informação estatística mais acertada de parâmetros populacionais importantes para estudos forenses e

populacionais [55]. Todas as sequências submetidas na EMPOP são submetidas a um controlo de qualidade rigoroso antes de serem integradas na base de dados [58].

## 1.2 Validação do método

No Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. (INMLCF) - Delegação do Norte, existem três modelos de sequenciadores automáticos que utilizam a tecnologia de eletroforese capilar: *Applied Biosystems™ 3130xl Genetic Analyzer*, com 16 capilares; *Applied Biosystems™ 3500 Genetic Analyzer*, com 8 capilares e *Applied Biosystems™ SeqStudio™ Genetic Analyzer*, com 4 capilares. Estes equipamentos foram lançados em momentos distintos e diferem em vários fatores, tais como: sensibilidade, robustez, rendimento máximo, número de *dyes* detetáveis, polímeros recomendados e manutenção, entre outros [59,60]. Atualmente, toda a sequenciação do mtDNA na Delegação do Norte, é feita no *3130xl Genetic Analyzer* porque o *3500 Genetic Analyzer* têm instalado o capilar para a análise de fragmentos. Como é preciso trocar o capilar sempre que se verifique necessário a análise de fragmentos por eletroforese capilar ou a sequenciação, o sequenciador *3130xl Genetic Analyzer* é utilizado para a sequenciação do DNA mitocondrial, enquanto o *3500 Genetic Analyzer* para a análise de fragmentos de DNA nuclear.

Mais recentemente, foi adquirido o *Seqstudio™ Genetic Analyzer Instrument*. Este apresenta uma maior facilidade de utilização, um tempo de preparação de amostras mais curto e uma maior flexibilidade de aplicação para realizar sequenciação de *Sanger* e análise de fragmentos na mesma execução, sem requerer mudar de capilar ou polímero [61]. Trata-se assim de uma tecnologia bastante útil para sequenciar o mtDNA, muitas vezes essencial em certos casos periciais. No entanto, carece de validação.

Para que sejam introduzidos na rotina pericial de um laboratório forense, todos os reagentes e equipamentos precisam de ser validados internamente, uma vez que é necessário assegurar que funcionam eficazmente e que se adaptam às condições do laboratório em causa. Esta validação tem de estar bem documentada e precisa de garantir que, numa elevada percentagem de amostras analisadas, os resultados são satisfatórios. Desta maneira, esta validação interna é indispensável pois permite avaliar a segurança, robustez, fiabilidade e reprodutibilidade do procedimento laboratorial, bem como a integridade dos reagentes, equipamentos e protocolos envolvidos no processo de determinação de perfis genéticos de DNA.

## 1.3 Estudo Populacional - Região do Norte de Portugal

A região do Norte de Portugal é composta pelas sub-regiões do litoral Norte (distritos de Viana do Castelo, Braga e Porto) e do interior Norte (distritos de Vila Real e Bragança). Esta tem uma história rica e diversa, tendo sido sucessivamente ocupada desde a pré-história por vários povos, desde celtas, romanos, visigodos a mouros [62,63].

Em meados do século 500 a.c., os Celtas, originários da Europa Ocidental e Central, estabeleceram-se na região e construíram várias fortificações, sendo a língua Galaico-Portuguesa considerada uma herança sua. Mais tarde, na época romana, a região do Norte foi conquistada e incorporada na província romana de *Gallaecia Bracarense*. Posteriormente aos romanos, vieram os Suevos e durante os séculos V e VI, o povo germânico dos visigodos invadiram o território estabeleceram-se na Europa Ocidental, principalmente na Península Ibérica [63]. Já no século VII, os muçulmanos, invadiram a península Ibérica e fixaram-se na região do Norte de Portugal, construindo cidades como o Porto e Guimarães. Foi a partir do norte de Portugal que D. Afonso Henriques constituiu Portugal como reino independente [62].

A região do norte de Portugal tem assim uma história complexa de migrações e invasões. Estudos genéticos que têm sido realizados em diferentes populações portuguesas, demonstram que há uma variedade de linhagens mitocondriais, nomeadamente com origens europeias e africanas. Além disso, também houve influências de outros grupos étnicos, como os romanos, mouros e árabes. Portanto é provável que a região do norte de Portugal tenha uma mistura de ancestrais de diferentes origens [30,64,65].

## 1.4 Objetivos

O presente estudo tem, assim, como um dos objetivos fazer a validação interna da análise dos polimorfismos do DNA mitocondrial por sequenciação no *SeqStudio™ Genetic Analyzer Instrument*. Desta forma, este trabalho constitui uma mais-valia para o SGBF-N do INMLCF assim refletindo num benefício para a sociedade, uma vez que disponibiliza resultados mais precisos e, portanto, mais fiáveis.

Outro dos objetivos do trabalho consiste na realização de um estudo genético da população da região do norte de Portugal, através da caracterização da totalidade da região controlo do seu mtDNA, para obter a diversidade e heterogeneidade genética. Para tal, após a obtenção dos haplótipos de cada indivíduo, os mesmos são submetidos à base de dados internacional de mtDNA com maior relevância na área das Ciências forenses, a EMPOP [55]. Posteriormente, visa-se determinar os haplogrupos em que cada haplótipo se insere, bem como realizar uma comparação interpopulacional, em particular, com outras populações europeias e populações imigrantes em Portugal. Esta comparação tem por finalidade determinar qual o grau de proximidade entre as populações, fornecendo novos conhecimentos sobre o impacto das migrações das populações do passado sobre a história demográfica do País.

## 2. Materiais e Métodos

### 2.1 Amostras Biológicas para os ensaios de validação

No presente estudo foram utilizadas amostras de DNA controlo provenientes de casas comerciais com quantidades e perfis/sequências conhecidas. Foram, ainda, estudadas amostras controlo provenientes de ensaios interlaboratoriais, portanto, com sequências bem definidas e conhecidas.

Estudamos também zaragoas bucais e manchas de sangue FTA colhidas a 20 voluntários e amostras de cabelos recolhidos com e sem raiz a 10 indivíduos voluntários.

Antes da sua inclusão no estudo, foi obtido o consentimento escrito de todos os voluntários, após prévia informação sobre o propósito do mesmo. As amostras a utilizar estão todas devidamente codificadas, ficando assegurada a proteção da identidade dos indivíduos e o correspondente sigilo.

### 2.2 Amostras Biológicas para o estudo populacional

Na caracterização genética da população do Norte de Portugal foram analisadas 87 amostras de sangue de indivíduos que residem nesta região, não relacionados entre si. As amostras estavam armazenadas no SGBF-N e eram referentes a processos de investigação de parentesco biológico de 2019.

Os cartões FTA são tratados quimicamente permitindo a lise das células, a desnaturação das proteínas, a proteção dos ácidos nucleicos e a inativação de determinados microrganismos. As principais vantagens da utilização destes cartões são o fácil transporte das amostras sem preocupações quanto a riscos biológicos e o estável armazenamento por longo prazo à temperatura ambiente [66].

## 2.3 Preparação de Amostras Biológicas

No caso da análise dos cabelos, antes da extração do DNA, é necessário realizar uma observação micro e macroscópica, proceder à descrição e consequente seleção dos mesmos, a que se segue uma lavagem. Este procedimento é efetuado para garantir que qualquer material biológico contaminante estranho ao cabelo seja removido.

Numa primeira etapa fez-se uma observação macroscópica onde se regista o comprimento e a cor do cabelo. Esta é complementada com uma observação microscópica que permite identificar a presença ou não de bolbo e de medula bem como das eventuais características da mesma - contínua, descontínua ou inexistente.

Após a observação, os cabelos foram cortados e a amostra com cerca de 4 centímetros foi colocada num Tubo *Eppendorf*, devidamente identificado, para prosseguir com a lavagem e extração. Sempre que for possível, o fragmento a utilizar deve conter o bolbo.

Para a lavagem dos cabelos, a cada tubo, deve ser adicionado etanol a 70%. Após a lavagem e centrifugação, o etanol é retirado e adicionada água destilada, seguindo-se uma nova lavagem e centrifugação. A lavagem é feita no termobloco a 20°C durante 30 minutos com agitação, sendo que a centrifugação dura 1 minuto a velocidade máxima, 12000 rpm (rotações por minuto).

## 2.4 Extração do DNA

- Extração por *Prep-n-Go™ Buffer*

O DNA das amostras de zaragoas bucais e manchas de sangue, utilizadas na validação do equipamento, foi extraído utilizando o reagente tampão *Prep-n-Go™ Buffer*.

Nas zaragoas bucais, cortou-se um dos dentes das mesmas colocando-o num tubo de *ependorf*. No caso das manchas de sangue, foi feito um punch de 2 mm. De seguida, adicionou-se aos tubos 100 µl do tampão *Prep-n-Go™ Buffer* e 100 µl de água. Após agitar os tubos de *ependorf*, 5-10 segundos, no vortex, colocaram-se os mesmos no termobloco durante 20 minutos a 90°.

Posteriormente, deixou-se as amostras arrefecer e de seguida foram colocadas no congelador a -20°C ou então prosseguiu-se para o passo de amplificação.

- Extração por *PrepFiler Express™* e *PrepFiler Express™ BTA*

Na validação o kit *PrepFiler Express™ Forensic DNA Extraction* foi utilizado na extração do DNA das manchas de sangue FTA enquanto o kit *PrepFiler Express™ BTA Forensic DNA Extraction* foi na extração do DNA dos cabelos.

No *PrepFiler Express™* foram utilizadas colunas *PrepFiler LySep* devidamente colocadas em tubos de amostra. A cada coluna foi adicionado um punch de 2 mm de mancha de sangue e 500µl de uma solução de lise. Esta solução foi preparada através de 5µl de DTT 1M (0,154g/ml) e 500µl *PrepFiler Lysis Buffer*, por amostra. De seguida, as amostras foram incubadas a 70° e 750 rpm durante 40 minutos e centrifugadas durante 2 minutos a 10000g, permitindo desta maneira a lise das células e a sua separação do substrato [67].

No caso do *PrepFiler Express™ BTA* a cada coluna foi adicionada uma amostra e 220µl tampão, 7 µl *proteinase k* e 3µl DTT. A incubação das amostras foi feita a 56° e 750rpm durante 3 horas e a centrifugação durante 2 minutos a 10000g.

Os tubos de amostra, com a solução de lise, foram colocados no instrumento *Applied Biosystems™ AutoMate Express™ Forensic DNA Extraction System* que realiza a extração do DNA.

- Extração por *Chelex®100 (Bio-Rad)*

O método *Chelex®* consiste em adicionar uma resina quelante à amostra. A resina *Chelex®* liga-se aos íons metálicos polivalentes, como os íons de magnésio, retendo as enzimas responsáveis pela destruição do DNA, assim protegendo o DNA [68].

Para cada amostra, utilizada no estudo populacional, foi colocado um punch de 2mm num tubo de eppendorf de 1,5mL, devidamente identificado. Posteriormente, foi adicionado a cada tubo 1 mL de água, deixando cerca de 30 minutos a incubar à temperatura ambiente. De seguida, as amostras foram agitadas 5-10 segundos num vortex e centrifugadas a 14000 rpm e 20° durante 3 minutos. Depois de se remover cerca de 930µl de sobrenadante a cada amostra, foram adicionados 180 µl de solução de *Chelex®* 100 a 5%, previamente preparada. Os tubos de *eppendorf* com a amostra e a solução de *Chelex®* foram colocados, num primeiro momento, num termobloco a 56° e 500 rpm durante 15 minutos. Seguidamente, após agitação no *vortex*, as amostras foram colocadas no termobloco a 100° e 300 rpm durante 8 minutos. Para finalizar, as amostras foram agitadas, uma última vez no vortex e colocadas na centrifuga a 14000 rpm durante 5 minutos. Por fim foram armazenadas num congelador a -20°C.

## 2.5 Amplificação por PCR

A amplificação é uma técnica que tem como objetivo replicar uma determinada região do DNA, delimitada por *primers*, aumentando exponencialmente o número de cópias dessa região [69]. A técnica de PCR é constituída por três fases que se podem repetir por um número variável de vezes: Desnaturação - as cadeias complementares da molécula de DNA são separadas; Annealing – os primers ligam-se à região flanqueadora; e Extensão – A DNA polimerase começa a sintetizar novas cadeias de DNA complementares à sequência alvo [69].

Para a validação do *Seqstudio™ Genetic Analyzer Instrument*, na amplificação do mtDNA, por cada amostra, foram adicionados 10µl de *HotStarTaq Plus Master Mix*; 0,5µl de *primer forward* (10 µM); 0,5µl de *primer reverse* (10µM); e valores variáveis de água e de DNA que, habitualmente, são 1 µl de DNA e 8 µl de água; perfazendo um volume total de 20µl. Para a amplificação da região HV1 são utilizados, normalmente, o par de *primers* F15989 e R16433; para a região HV2 o par de *primers* F16533 e R460 e na região HV3 utiliza-se o par F317 e R619. De modo a ser possível estudar a região presente entre HV1 e HV2, foi amplificada e sequenciada uma região denominada de RV1 com os *primers* F16363 e R293 (figura 1.1.1). No caso de se querer amplificar a região controlo total recorre-se aos *primers* F15989 e R619, sendo que na sequenciação a região total é dividida em duas através da utilização dos *primers* R52 e F16533. O programa no termociclador utilizado tem as condições cíclicas de: 95°C durante 5 minutos; 40C de 95°C durante 45s, 58°C durante 1 minuto e 72°C durante 1 minuto. Para finalizar assume uma extensão de 72°C durante 10 minutos.

No estudo populacional foram utilizadas condições de amplificação ligeiramente diferentes. O volume final é de 10 µl. Para tanto, por cada amostra são adicionados 5 µl de *HotStarTaq Plus Master Mix*, 1 µl de *primers* e, consoante o volume de DNA adicionado à reação, pode-se juntar água para perfazer o volume final.

Para a amplificação da região controlo total do mtDNA, que inclui as regiões HV1, HV2 e HV3, foram utilizados dois pares de *primers*. O par L15971 (direto) e H016 (reverso) permitiu a amplificação entre as posições 16024 e 16569, enquanto o par L16555/H639 amplificou a região entre as posições 1 a 576. O programa no termociclador utilizado tem as condições cíclicas de: 94°C durante 5 minutos; 35C de 94°C durante 30s, 60°C durante 90s e 72°C durante 60s, finalizando com uma extensão a 72° durante 10 minutos.

A cada amplificação são adicionados um controlo positivo (padrão de perfil genético conhecido) e um negativo (reagentes de amplificação sem amostra). Como controlo positivo podem-se utilizar o 2800M, K562 e 9947A, de 10 ng/ $\mu$ l; e 007 e 09948 de 0,1 ng/ $\mu$ l. Estes controlos deverão ser sujeitos ao mesmo procedimento seguido para as amostras amplificadas nas fases posteriores à amplificação. No final, o controlo positivo deverá apresentar a sequência conhecida e no controlo negativo não se deverá observar qualquer sequência.

## 2.6 Verificação da Amplificação

Após a amplificação, foi essencial fazer a verificação da mesma, utilizando o equipamento *QIAxcel® Advanced* (Qiagen®). Este passo tem como objetivo confirmar a amplificação das amostras a analisar para se poder prosseguir com o percurso laboratorial. As amostras que não amplificaram devem ser repetidas e/ou descartadas, não devendo avançar-se para os passos seguintes.

O sistema *QIAxcel® advanced* utiliza a eletroforese capilar em gel para permitir uma rápida separação dos nucleótidos com base no seu tamanho. Quando comparado com o sistema tradicional de eletroforese em gel de agarose, este sistema apresenta diversas vantagens tais como: uma maior sensibilidade de deteção; uma melhor resolução dos fragmentos; uma análise rápida; um menor desperdício de amostra; e um sistema de carregamento e análise automatizados. Após a corrida, o *software QIAxcel® ScreenGel* fornece tanto um electroferograma e ainda imagens do gel [70].

Às amostras tem de ser adicionado um Marcador de Alinhamento e um Marcador de Tamanho. O Marcador de Alinhamento é injetado a partir da posição *MARKER1* do tabuleiro tampão e co-migra com as amostras de DNA para análise. No SGBF-N estão disponíveis os Marcadores de Alinhamento QX 15 pb/1 kb e QX 15 pb/3 kb. O Marcador de Tamanho é utilizado para uma determinação ótima do tamanho do DNA. Quando se utiliza o Marcador de Alinhamento QX 15 pb/1 kb tem de se utilizar o marcador de tamanho de DNA QX 50-800 pb, se for o marcador de alinhamento QX 15 pb/3 kb o utilizado, o marcador de tamanho de DNA QX a utilizar é o 100 pb - 2,5 kb [70].

Este passo foi apenas aplicado para validação interna da análise dos polimorfismos do DNA mitocondrial por sequenciação no *SeqStudio™ Genetic Analyzer Instrument*, devido à inexistência do equipamento no SGBF-S.

## 2.7 Purificação dos produtos amplificados

A purificação dos produtos amplificados é um processo realizado após a verificação da amplificação, por forma a remover os *primers* e nucleótidos em excesso. Esta purificação foi feita por digestão enzimática utilizando o “*ExoSAP-IT™ PCR Product Cleanup*” (*Applied Biosystems™*) [71].

Na validação dos polimorfismos do DNA, adicionou-se em primeiro lugar 5 $\mu$ l do reagente *ExoSAP-IT™* diretamente a cada produto da reação. De seguida, as amostras foram colocadas no termociclador no programa de purificação que consiste em: 40 minutos a 37°C seguidos de 20 minutos a 80°C.

No estudo populacional, foram adicionados 4 $\mu$ l do reagente *ExoSAP-IT™* a cada 10 $\mu$ l de amostra amplificada e estas foram colocadas no termociclador no programa: 37°C durante 15 minutos, seguidos de mais 15 minutos a 80°C.

O reagente *ExoSAP-IT™* foi inativado por aquecimento a 80°C durante 15 minutos [71].

## 2.8 Sequenciação direta

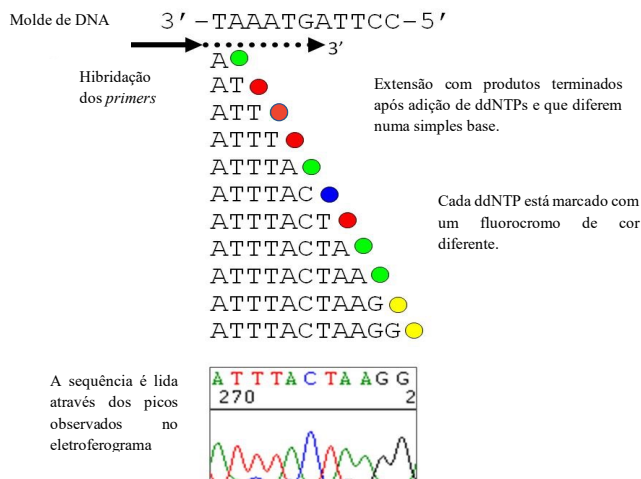


Figura 2.8.1 - Sequenciação direta do DNA com ddNTPs fluorescentes. O primer liga-se a uma região específica do DNA e é prolongado com uma polimerase. Como existe uma mistura de dNTPs e ddNTPs para cada um dos quatro nucleótidos possíveis, alguns dos produtos da extensão são interrompidos pela incorporação de um ddNTP. Cada ddNTP tem uma cor diferente o que permite a distinção de cada produto de extensão.

Fonte: Adaptado de Butler, J. M. 2009.

síntese *in vitro* (figura 2.8.1). No final, obtém-se um conjunto de cadeias simples de DNA com diferentes tamanhos, com terminadores incorporados marcados com fluorescência, o que, todas juntas, permitem construir a sequência alvo. Para a sequenciação direta das amostras utilizadas na validação, por cada amostra, foram adicionados 1,5µl de *Master Mix*; 1µl de *Primer Forward* ou *Reverse* (3,2 µM); e um volume variável de água e de DNA, sendo habitual utilizar-se 1µl de amostra e 6,5µl de água; perfazendo um volume total de 10µl. De seguida, as amostras foram colocadas no termociclador com as condições cíclicas de: 25°C de 96°C durante 10 segundos, 50°C durante 5 segundos e 60°C durante 4 minutos. No caso de amostras utilizadas para o estudo populacional, foram adicionadas por amostra 2µl *BigDye™ Terminator v3.1*, 1µl de tampão *Better Buffer®*, 1µl de *primer* e 1µl de produto amplificado, perfazendo um volume final de 5µl. De seguida, as amostras foram colocadas no termociclador com as condições cíclicas de: 96°C durante 2 minutos, seguido de 35°C a 96°C durante 15 segundos, 50°C durante 9 segundos e 60°C durante 2 minutos, finalizando com 10 minutos a 60°C. Um iniciador que foi concebido para reconhecer uma região específica do DNA, é ligado e prolongado com uma polimerase. Como existe uma mistura de dNTPs e ddNTPs para cada um dos quatro nucleótidos possíveis, alguns dos produtos de extensão são interrompidos pela incorporação de um ddNTP enquanto outras moléculas continuam a ser estendidas. Cada ddNTP é rotulado com um *dye* diferente o que permite que cada produto de extensão seja distinguido por cor. Uma separação por tamanho dos produtos de extensão permite a leitura da sequência de DNA desde uma resolução suficiente esteja presente para ver cada base com clareza.

## 2.9 Limpeza dos terminadores

Antes de se colocar as amostras no sequenciador automático, deve-se proceder à limpeza dos terminadores que não foram incorporados na sequenciação direta. Esta limpeza é fundamental para evitar interferências nos resultados, uma vez que os terminadores estão marcados com fluorescência. Este processo foi realizado com o *BigDye® XTerminator Purification Kit* (AB) que contém dois reagentes: *Xterminator solution*, responsável por capturar os dNTP, ddNTPs e primers não

Posteriormente à purificação das amostras amplificadas, prosseguiu-se para a sequenciação direta dos fragmentos amplificados. Esta sequenciação foi feita utilizando um *kit* comercial para sequenciação, *BigDye™ Terminator v3.1 Cycle Sequencing Kit*, e *primers* de sequenciação específicos para os segmentos a analisar [72]. A sequenciação direta, ou sequenciação pelo método enzimático direto ou método de Sanger, é, tal como o PCR, um processo de amplificação que envolve a utilização de *Taq* polimerase, dNTPs, ddNTPs (terminadores marcados com diferentes fluorescências) e de um único *primer*. Os terminadores quando são incorporados na cadeia de DNA em construção, impedem a continuação do alongamento da nova cadeia durante a sua

incorporados; e *SAM TM Solution* que melhora a atuação do reagente *X Terminator* e estabiliza a amostra após a limpeza [73].

Nas placas de sequenciação para a validação do *SeqStudio™*, para efeitos de limpeza, foram adicionados 45µl de SAM e 10µl de *X-terminator* a cada amostra. Seguidamente, a placa foi colocada no agitador a 2500 rotações por minuto, durante 30 minutos. Tratando-se de placas de sequenciação para o estudo populacional, foram adicionados a cada amostra 50µl de uma mix constituída por 30µl de água, 16,4µl de SAM e 3,7µl de *X-terminator*. De seguida, a placa foi colocada a agitar a 2000rpm durante 20 minutos. Em ambos os casos, posteriormente à agitação, a placa é centrifugada durante 2 minutos a 1000g, permitindo que os componentes capturados fiquem agregados no fundo da placa, visto não serem aspirados pelos capilares durante a eletroforese capilar.

## 2.10 Eletroforese capilar no sequenciador automático e Análise dos produtos sequenciados

Após a sequenciação e respetiva limpeza dos terminadores, as amostras foram colocadas no sequenciador automático e submetidas a eletroforese capilar com deteção automática da fluorescência. Durante este processo os diferentes fragmentos migram, através dos capilares do sequenciador, desde o polo negativo para o positivo, consoante o seu peso molecular. Os fragmentos de menor tamanho migram a uma maior velocidade que os maiores. A deteção é feita pela emissão da fluorescência, num determinado comprimento de onda, que os fluorocromos dos ddNTPs emitem quando são excitados por um feixe de laser.

Como um dos objetivos do estudo é validar o analisador genético *SeqStudio™ Genetic Analyzer*, as amostras a analisar são colocadas concomitantemente no *3130xl Genetic Analyzer* e no *SeqStudio™ Genetic Analyzer*, de modo a poder comparar-se os resultados obtidos em ambos.

Os perfis de DNA obtidos foram, em primeiro lugar, analisados no *Sequencing Analysis Software 7*, onde é possível obter um electroferograma (figura 2.10.1). A avaliação da qualidade dos electroferograma permite selecionar as amostras que podem prosseguir para análise no *Software SeqScape™ 4* e as que têm de voltar a ser amplificadas e/ou sequenciadas. De seguida, as amostras com qualidade são analisadas no *SeqScape™* onde se faz o alinhamento das amostras e a comparação com a sequência padrão do mtDNA. Este identifica apenas as diferenças entre as duas sequências discriminando todas as bases nucleotídicas alteradas e as respetivas posições.

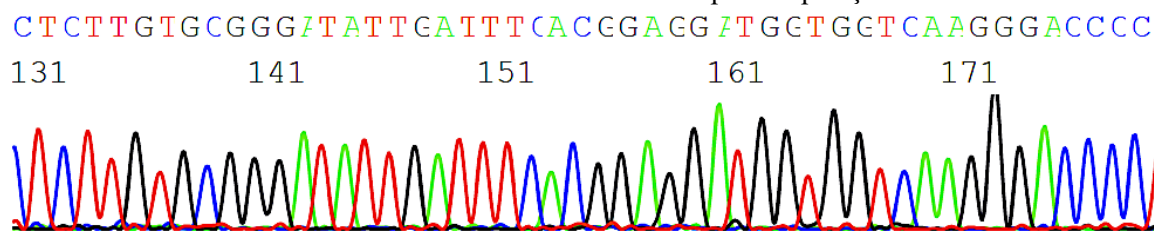


Figura 2.10.1 - Exemplo de um segmento de um electroferograma obtido no programa *Sequencing Analysis 7*, após sequenciação com o primer H016.

Para todas as amostras do estudo populacional, bem como algumas das amostras da validação, foi sequenciada a região controlo total, compreendida entre as posições 16024 a 16569. No entanto, para outras amostras utilizadas na validação foram estudadas as regiões HV1, HV2, HV3 e RV. Em ambos os casos foram sempre sequenciadas duas cadeias, direta e reversa, de forma a comparar os segmentos e confirmar os resultados.

Tabela 2.10.1 - Código IUPAC utilizado na análise do mtDNA.

Código IUPAC	Base(s) nucleotídica(s)	Código IUPAC	Base(s) nucleotídica(s)
A	Adenina	Y	Citosina ou Timina
C	Citosina	K	Guanina ou Timina
G	Guanina	M	Adenina ou Citosina
T	Timina	N	Qualquer base
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 Guanina		

A caracterização dos haplótipos é realizada segundo as considerações da *European DNA Profiling Group* (EDNAP) e as regras de nomenclatura da *International Union of Pure & Applied Chemistry* (IUPAC) (tabela 2.10.1) [13,74].

As mutações por substituição podem ser transições, entre purinas (A e G) ou entre pirimidinas (C e T), ou transversões, entre uma purina e uma pirimidina ou vice-versa. Estas são indicadas pela posição onde ocorre a mutação e pelo nucleótido que originou essa alteração.

Outros tipos de alterações que podem ser encontradas são as inserções ou as deleções. As primeiras são identificadas indicando a posição nucleotídica, imediatamente anterior à alteração, seguida de um ponto e um (.1) (para a primeira inserção), .2, na segunda inserção e assim sucessivamente, acrescentando-se ainda o nucleótido inserido. As deleções são representadas pela posição do nucleótido juntamente com o sufixo “DEL” ou “-”. No caso particular de inserções ou deleções que ocorram em regiões homopoliméricas, são sempre identificadas na última posição [13].

Em amostras onde estava presente heteroplasmia de comprimento, foi apenas possível ler a sequência até à posição heteroplásmica, devido a uma sobreposição de bases nucleotídicas (figura 2.10.2A). Nestes casos, foi necessário efetuar a repetição da sequenciação das amostras, para confirmar a sequência obtida. Outra característica que se observou em algumas amostras foi heteroplasmia de posição que é representada pela sobreposição de nucleótidos numa mesma posição, traduzindo-se em dois picos no eletroferograma (figura 2.10.2B). A sua identificação é realizada pela posição em que ocorre a alteração, seguida da letra correspondente, de acordo com o código descrito pela IUPAC [13].

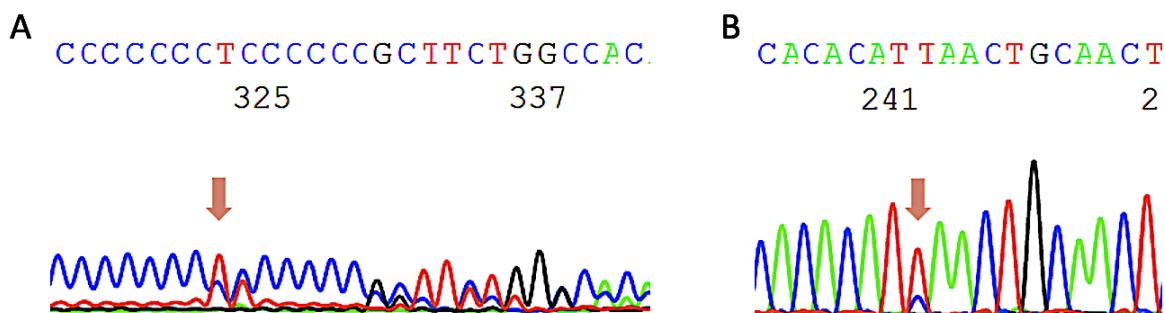


Figura 2.10.2 – Representação, através de um eletroferograma, de um segmento (a) com heteroplasmia de comprimento e (b) com heteroplasmia de posição. As heteroplasmias estão identificadas por uma seta.

Para a determinação dos haplogrupos a que cada haplótipo pertence recorreu-se à árvore filogenética *PhyloTree*, desenvolvida por van Oven e Kayser [40].

## 2.11 Análise estatística e filogenética

Numa primeira instância, através do *software Arlequin v.3.5.2.2* foram calculados os valores da diversidade nucleotídica da região controlo total das amostras estudadas. Este *software* também foi utilizado para realizar a análise de variância molecular (AMOVA), que permite determinar as relações

filogenéticas entre a população em estudo e as selecionadas da literatura, e daí obter medidas de distâncias genéticas interpopulacionais [75]. As distâncias genéticas foram alcançadas com base no valor do índice de fixação (FST), sendo também calculado o *p-value*, indicativo do nível de significância (0.05).

As distâncias moleculares entre populações foram obtidas, tendo por base os valores de Fst e recorrendo ao programa *Neighbor* do *software Phylip v.3.698* [76]. A sua representação gráfica, em forma de árvore filogenética, foi alcançada recorrendo ao *software* programa *TreeView v. 2.0.8* [77].

### 3. Resultados e Discussão

#### 3.1 Validação da análise dos polimorfismos do DNA mitocondrial por sequenciação no *SeqStudio™ Genetic Analyzer Instrument*

##### 3.1.1 Validação de resultados dos controlos positivos e amostras controlo

O primeiro passo de validação do *SeqStudio™ Genetic Analyzer* foi o estudo dos controlos positivos. Para tal, foram amplificados os controlos positivos 2800M (10 ng/μl), K562 (10 ng/μl), 9947A (10 ng/μl), 007 (0,1 ng/μl) e 9948 (0,1 ng/μl) com os pares de *primers* HV1, HV2, HV3 e RV.

Os resultados obtidos no *Qiaxcel®* (figura 3.1.1.1- A e B) demonstraram a amplificação correta de todos os controlos com os *primers* HV1 e HV2. A amplificação com os *primers* HV3 falhou na totalidade (figura 3.1.1.1C); já com os *primers* RV houve amplificação dos controlos positivos 2800M, K562, 9947A e 9948 (figura 3.1.1.1D).

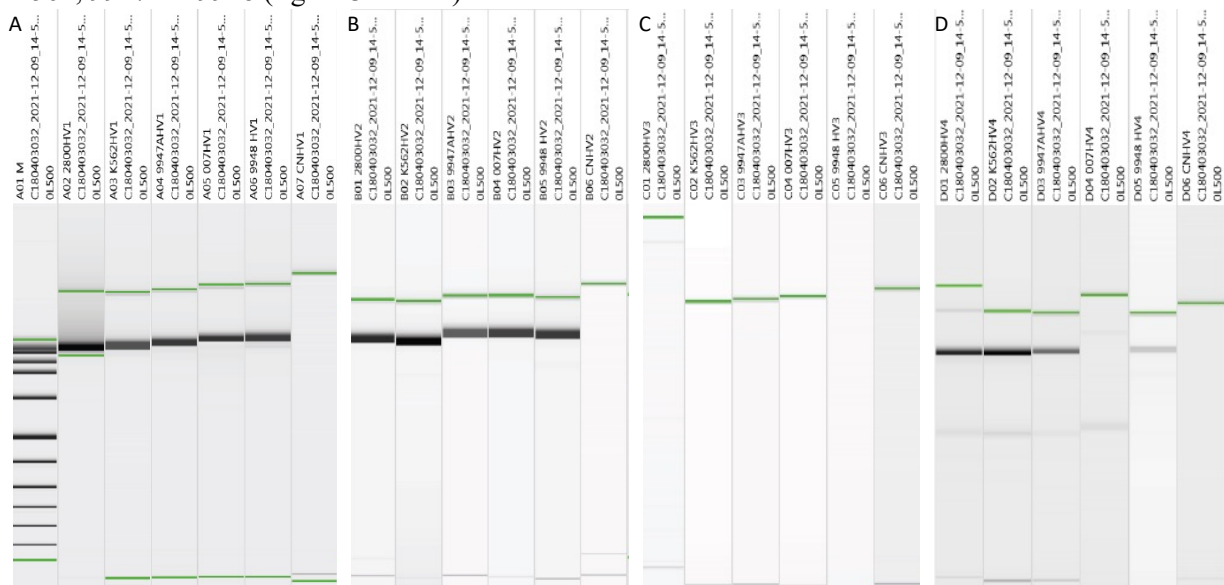


Figura 3.1.1.1 - Gel de eletroforese capilar (*QIAXCEL® ScreenGel*) para verificação da amplificação dos controlos positivos com os *primers* HV1, HV2, HV3 e RV. A – Verificação da amplificação com os *primers* HV1. Existe uma amplificação satisfatória de todos os controlos positivos. B – Verificação da amplificação com os *primers* HV2. Existe uma amplificação satisfatória de todos os controlos positivos. C – Verificação da amplificação com os *primers* HV3. Não houve amplificação de qualquer controlo positivo. D – Verificação da amplificação com os *primers* RV. Existe uma amplificação dos controlos positivos 2800M, K562, 9947 e 9948.

Prosseguindo com a sequenciação das amostras amplificadas com os *primers* HV1 e HV2, tanto no *3130xl Genetic Analyzer* como no *SeqStudio™ Genetic Analyzer* e comparando os resultados obtidos, foi possível aferir que o *SeqStudio™ Genetic Analyzer*, tal como os *softwares* do mesmo, estão em conformidade com o *3130xl Genetic Analyzer*, já que os resultados são iguais e com boa qualidade.

Posteriormente, foram amplificadas amostras controle, com os *primers* HV3 e RV, que tinham falhado. Seguiu-se a sequenciação das amostras amplificadas com estes *primers*, a qual correu satisfatoriamente. Ficaram assim validados os controles positivos/amostras controle quando utilizados os *primers* HV1, HV2, HV3 e RV, com a respectiva sequência descrita no anexo 6.1.

### 3.1.2 Estudos de Sensibilidade

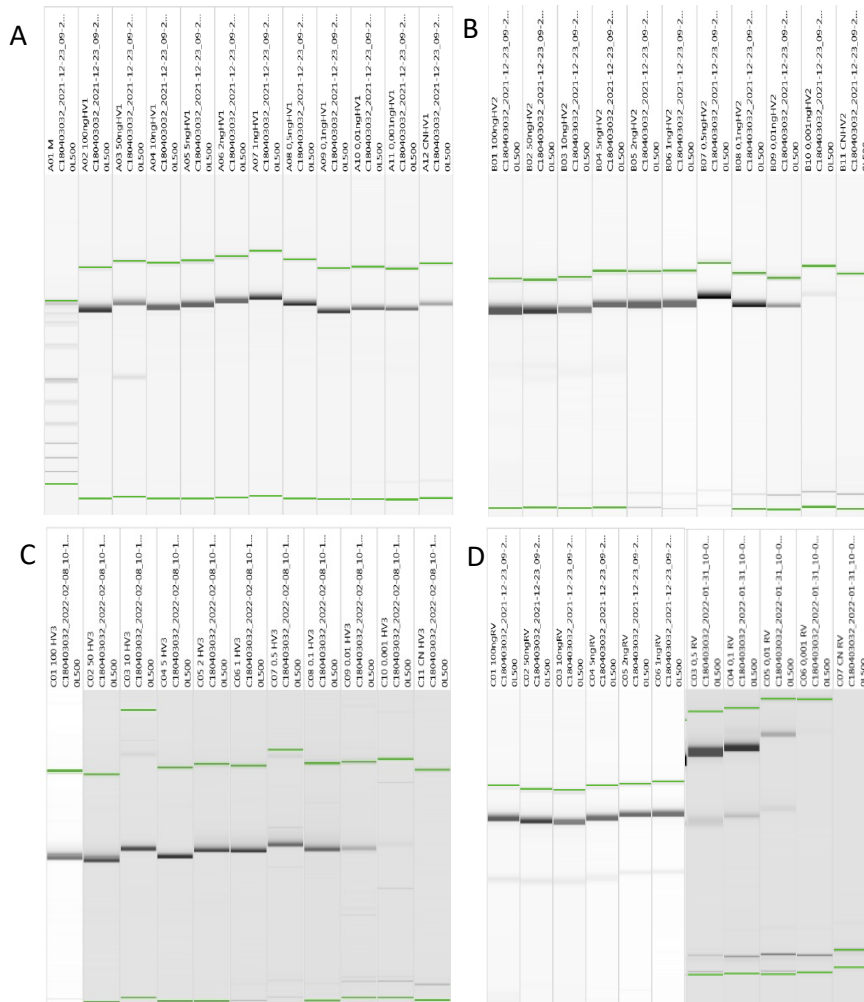


Figura 3.1.2.1- Gel de eletroforese capilar (QIAXcel® ScreenGel) para verificação da amplificação dos testes de sensibilidade com os primers das regiões HV1, HV2, HV3 e RV. A – Verificação da amplificação com os primers HV1. Existe uma amplificação satisfatória de todas as concentrações. B – Verificação da amplificação com os primers HV2. Existe uma amplificação satisfatória de todas as concentrações. C – Verificação da amplificação com os primers HV3. Existe uma amplificação satisfatória de todas as concentrações. D - Verificação da amplificação com os primers RV. Existe uma amplificação satisfatória de todas as concentrações.

Um teste de grande importância é o da sensibilidade. Neste, são analisadas diferentes concentrações do DNA (0.00001ng; 0.0001ng; 0.001ng; 0.01ng; 0.1ng; 0.5ng; 1ng; 2ng; 5ng; 10ng; 50ng; 100ng), por forma a saber qual a gama de quantidades capazes de produzir resultados de tipagem fiáveis.

Na verificação da amplificação conseguiu-se perceber que, à medida que as concentrações vão diminuindo, a sua deteção vai-se revelando com menor intensidade (figura 3.1.2.1), chegando mesmo a não ser detetável. Numa primeira fase, comparando os resultados obtidos no Qiaxcel® com os do SeqStudio™ Genetic Analyzer verifica-se que, com a utilização dos primers correspondentes à região HV1, HV2, HV3 e RV, existe sequenciação desde os 100ng até aos 0.001ng (figura 3.1.2.1); daí que seja necessário fazer mais diluições para testar a partir de que ponto de concentração é possível obter resultados.

Consequentemente, foram testadas mais duas concentrações, 0.0001ng e 0.00001ng, não se tendo registado qualquer amplificação e sequenciação das amostras. Foi assim definida a concentração mínima detetável de 0.001ng quando utilizados os primers das regiões HV1, HV2, HV3 e RV.

### 3.1.3 Estudos de Especificidade

A verificação da especificidade é um passo crucial para garantir que os primers são específicos do mtDNA humano. Foram assim amplificadas e sequenciadas amostras animais (anexo 6.2) com os primers humanos HV1, HV2, HV3 e RV.

Como esperado, tanto na amplificação, como na sequenciação, não foram obtidos resultados, confirmando que os mesmos são específicos para o mtDNA humano.

A verificação da especificação também foi útil para confirmar que os resultados negativos do *Qiaxcel*® estão em conformidade com os resultados negativos no sequenciador *SeqStudio*™ *Genetic Analyzer*. Nenhuma das amostras animais revelou resultados, quer na verificação da amplificação quer na sequenciação, o que não se verificou no controlo positivo como esperado.

#### 3.1.4 Validação do estudo de mtDNA de cabelos extraídos por *PrepFiler Express*™ *BTA*

Em determinados casos forenses o material biológico disponível são cabelos. Nestes casos, devido à pouca quantidade de DNA nuclear que normalmente está disponível, é frequentemente utilizado o mtDNA para se obter resultados.

Inicialmente, os cabelos são observados ao microscópio para avaliar as suas características morfológicas e selecionar a zona da extração do DNA, preferencialmente o local da raiz, caso esta esteja presente. Após a extração, embora apenas seja detetado o DNA nuclear, as amostras são quantificadas para se conseguir alcançar uma perceção da quantidade de DNA existente. Assim, na amplificação é adicionada 1  $\mu$ l das amostras que apresentam mais de 1  $ng/\mu$ l de DNA na quantificação, enquanto nas restantes são adicionados 5  $\mu$ l.

Posteriormente à amplificação faz-se a respetiva verificação. Este passo permite selecionar quais as amostras que seguem para a sequenciação.

No presente estudo, como anteriormente descrito, foram extraídas e amplificadas 20 amostras de 10 indivíduos. Em primeiro lugar, tentou-se compreender se existe uma diferença significativa entre a quantidade de DNA existente em cabelos arrancados ou cabelos caídos. Como tal, fez-se uma análise estatística às amostras de cabelo recorrendo ao *IBM*® *SPSS*® *software*.

Primeiro, para perceber se as amostras seguiam uma distribuição normal, foi realizado um teste de normalidade, onde se considerou  $H_0$ : A quantidade de DNA presente nos cabelos segue uma distribuição normal e  $H_1$ : A quantidade de DNA presente nos cabelos não segue uma distribuição normal. Uma vez que estamos perante um estudo com uma amostra pequena, inferior a 30, são analisados os resultados obtidos a partir do teste *Shapiro-wilk*. Neste, obteve-se um *p-value* igual a 0,142, ou seja 14,2 % (anexo 6.3). Como esse valor é superior a 5%, aceita-se a hipótese nula [78]. Visto que a distribuição é normal e que os dois tipos de cabelos são do mesmo indivíduo, a análise estatística prosseguiu através do teste T com duas amostras emparelhadas [79]. Neste, considerou-se  $H_0$ : A quantidade de DNA presente em cabelos arrancados é igual ao dos cabelos caídos e  $H_1$ : A quantidade de DNA presente em cabelos arrancados é diferente ao dos cabelos caídos. Uma vez que o valor da significância bilateral é de 0,016, 1,6%, sendo inferior aos 5%, rejeita-se  $H_0$  (anexo 6.4).

Assim, podemos assegurar que há evidência estatística suficiente para afirmar que a quantidade de DNA presente em cabelos arrancados é diferente da que está presente em cabelos caídos. Neste caso, a quantidade dos cabelos arrancados é significativamente maior que a de cabelos caídos pois, a média de quantidade de DNA nos cabelos arrancados é de 1,6615453, um valor bem superior à média de 0,2199382, apresentada nos cabelos caídos (anexo 6.5). Esta conclusão foi tomada assente num nível de confiança de 95%. Note-se que os limites do intervalo de confiança a 95% são ambos positivos (0,366 e 2,517) ou seja, não contém o 0, não podendo ser as quantidades iguais (anexo 6.4). Assim, ficou provado que os cabelos caídos apresentam uma menor quantidade de DNA, provavelmente devido a uma maior degradação do mesmo por exposição a diversos fatores, o que se traduziu num maior insucesso na amplificação. Apenas as amostras B1R, B2R, B4R, B5R e B8R, todas extraídas de raízes de cabelos arrancados, apresentaram mais de 1  $ng/\mu$ l de DNA na quantificação, daí resultando numa amplificação satisfatória de todas as regiões (anexo 6.6). No entanto, a maior parte das amostras, independentemente da quantidade de DNA obtida na quantificação, conseguiu amplificar com sucesso

em duas ou mais regiões, o que os torna numa amostragem possível de ser utilizada (anexo 6.6). Os haplótipos obtidos com as amostras de cabelo estão descritos no anexo 6.6.

Assim, ficou validada a utilização de cabelos para a obtenção de perfis de mtDNA no sequenciador genético *SeqStudio™*.

### 3.1.5 Validação do estudo da região controlo total do mtDNA

Quando o mtDNA é extraído de cabelos, como se trata de amostras expostas a diversos processos de degradação, a amplificação e a sequenciação são divididas em quatro regiões diferentes, de forma a ser possível a obtenção de sequências mais curtas capazes de serem amplificadas. No entanto, o mtDNA também pode ser extraído de manchas de sangue ou zaragatoas bucais, sendo nestes casos provável que toda a região controlo esteja intacta. Assim, amplificar a região controlo total reduz o tempo de trabalho laboratorial e os custos da análise, otimizando o processo.

Para começar, é necessário fazer a validação da região controlo total dos controlos positivos, já que em todos os estudos estes têm de ser incluídos para comparação. Como tal, foram testados três destes controlos (007 – 0,1 ng/μl, 2800M – 10 ng/μl e 9947A – 10 ng/μl) com três quantidades diferentes de amostra adicionada (1μl, 2,5μl, 4μl) e o volume final de amplificação reduzido a metade, 10 μl. Como o estudo da região controlo total reduz a utilização de *primers* da sequenciação, de 8 para 4, os 20μl de volume final da amplificação anteriormente necessários, deixam assim de ser precisos, fazendo então sentido o ajuste do volume total final para 10μl. Como nunca se tinha feito a amplificação da região controlo total, não se sabendo por isso qual a quantidade ideal de amostra a adicionar na amplificação, optou-se por testar três quantidades diferentes para se chegar à melhor solução.

Efetuada a análise, foi observado que, independentemente da quantidade inicial de amostra adicionada, existem sempre resultados detetáveis, sendo para tanto suficiente a utilização de 1μl de amostra. Para além disso, constatou-se que as mutações ora detetadas na validação da região controlo total, são iguais às mutações da validação das regiões dos controlos positivos, como era expectável. Como se esperava, também todas as sequências de um mesmo controlo, independentemente da quantidade inicial de amostra, são iguais. No controlo 9947A 1μl houve uma falha na sequenciação da região HV1, por faltarem as mutações 16311C e 16519C; no entanto a restante sequência foi igual. Esta falha pode ter sido originada por um problema na incorporação dos *primers* na sequenciação ou uma falha de leitura do equipamento.

### 3.1.6 Validação do estudo de mtDNA nas manchas de sangue FTA

Quando são utilizadas manchas de sangue em casos forenses, faz-se normalmente uso do DNA nuclear. No entanto, este pode estar degradado, sendo então necessário recorrer ao DNA mitocondrial.

As manchas de sangue podem ser trabalhadas de diversas maneiras nos casos forenses. O DNA pode ser extraído por *PrepFiler Express™*, por *Prep-n-Go™* ou, no caso de manchas de sangue obtidas em cartões FTA, utilizar-se diretamente um punch na amplificação.

No presente estudo foi extraído DNA, através do *kit PrepFiler Express™*, em 20 amostras de diferentes indivíduos. Foi posteriormente amplificada e sequenciada a região controlo total das amostras e assim obtida a respetiva sequência (anexo 6.7). Todas as amostras amplificaram e sequenciaram com resultados reportáveis. Tal indica que é um método com uma grande qualidade para a obtenção de resultados.

De seguida, na tentativa de otimizar o processo, foi testada a amplificação com punch direto; no entanto, a grande maioria das amostras não amplificou. Este resultado pode ser devido a certos compostos sanguíneos, como por exemplo, a hemoglobina e a imunoglobina G [80], que funcionam

como inibidores de PCR, afetando a amplificação quando ao baixar ou bloquear a atividade da DNA polimerase ou quando interage com ácidos nucleicos [81].

Uma vez que o tempo disponível para realizar o trabalho laboratorial estava a terminar, optou-se por testar apenas a região HV1 com 3 amostras, por amplificação direta, mas com diluição prévia. Desta forma, caso tornasse a não registar resultados não seria necessário continuar a testar a abordagem dos punches de sangue; porém, caso estes fossem positivos, o laboratório sempre poderia posteriormente continuar a validação dos mesmos. Os punches foram assim colocados em água, durante cerca de 20 minutos, visando diluir as amostras e reduzir o número de inibidores. À amplificação foram adicionados 1  $\mu$ l e 2  $\mu$ l de amostra. Embora os resultados da amplificação fossem fracos, a grande maioria das amostras amplificou e sequenciou corretamente, não se tendo detetado diferença relevante entre a adição de 1  $\mu$ l ou 2  $\mu$ l. A amostra que já tinha sido previamente sequenciada na validação das manchas de sangue extraídas por *PrepFiler*<sup>TM</sup>, A2, deu uma sequência igual na zona correspondente à região HV1. Embora os resultados tenham sido positivos, a quantidade amostral não é suficiente para a validação do método. Continua a faltar analisar um maior número de amostras, testar todas as regiões do mtDNA e a região controlo total.

Atendendo que a utilização do punch direto perde em rapidez dada a necessidade de um pré-tratamento, foi testada a amplificação e sequenciação das 4 regiões do mtDNA através das 20 manchas de sangue extraídas pelo método *Prep-n-Go*<sup>TM</sup>. O rebenamento das membranas celulares provocado por este método pode servir de pré-tratamento para as manchas de sangue. É, no entanto, necessária uma diluição de 1:4 prévia à adição da amostra na amplificação, a fim de diminuir a concentração dos inibidores presentes. A grande maioria das amostras amplificou, tendo, porém, havido uma falha generalizada na região RV. Esta pode ter sido originada por um problema nos *primers*, ou da adição dos mesmos na amplificação, ou de uma outra causa não definida. Assim, tendo sido validado o método nas regiões HV1, HV2 e HV3, falta apenas realizar uma repetição da região RV e a análise da região controlo total.

### 3.1.7 Validação do estudo de mtDNA de zangaratoas bucais extraídas por *Prep-n-Go*<sup>TM</sup>

Tal como acontece com as manchas de sangue, também pode ter de se recorrer ao mtDNA nas zangaratoas bucais. Neste caso o DNA extrai-se por *PrepFiler Express*<sup>TM</sup> ou *Prep-n-Go*<sup>TM</sup>.

Este método não foi muito abordado no estudo, devido ao pouco tempo disponível para terminar o trabalho laboratorial. Foi apenas feita uma pequena experiência para ter uma perspectiva dos resultados a que se pode chegar. Com efeito, tal viabilizaria, a perceção sobre se valeria a pena continuar a validação das amostras extraídas por *Prep-n-Go*<sup>TM</sup> com diluição de 1:4 ou se teria de se recorrer a um método diferente, futuramente.

Assim, foram estudadas 4 amostras extraídas por *Prep-n-Go*<sup>TM</sup>. Estas foram diluídas na proporção de 1:4, e apenas estudada a região HV1. Houve uma amplificação e sequenciação de todas as amostras, sendo as suas sequências iguais às das manchas de sangue, uma vez que se trata dos mesmos indivíduos.

Embora os resultados fossem um bom presságio para a validação do método, ficou a faltar a utilização de um maior número de amostras, a amplificação e conseqüente sequenciação de todas as regiões e da região controlo total, bem como a análise das amostras extraídas por *PrepFiler Express*<sup>TM</sup>.

## 3.2 Estudo Populacional da Região do Norte

No presente trabalho, determinaram-se os haplótipos de 84 indivíduos e os respetivos haplogrupos. A amostragem inicial continha 87 indivíduos, tendo, no entanto, sido retiradas 3 amostras dada a impossibilidade de obtenção de resultados com a qualidade desejada.

Para que fosse assegurada a qualidade das sequências e dos haplogrupos obtidos, os haplótipos descritos no estudo foram submetidos a uma base de dados internacional, a EMPOP [12,55], e posteriormente inseridos na mesma e disponibilizados com um código de acesso.

### 3.2.1 Estudo dos Haplótipos

A partir dos dados da totalidade da região controlo do mtDNA obtidos no estudo, foi possível construir uma base de dados que apresenta os haplótipos de cada indivíduo, o número de alterações existentes e o haplogrupo correspondente (anexo 6.8).

Foram identificados 79 haplótipos diferentes nas 84 amostras analisadas, dos quais 77 (91,7%) estavam presentes apenas uma única vez. A elevada frequência de haplótipos únicos identificada neste estudo, vai de acordo com outros estudos de mtDNA da população portuguesa, especificamente da região do norte de Portugal [64,65].

Os restantes 7 haplótipos foram identificados em mais do que uma amostra, estando o haplótipo mais frequente presente em 5 sequências (tabela 3.2.1.1). A existência de haplótipo coincidentes entre indivíduos não relacionados demonstra o baixo poder de discriminação e individualização do mtDNA como marcador genético, tornando dúbida a sua utilização em situações de identificação individual. Desta maneira, a aplicação do mtDNA na rotina forense depende da frequência dos haplótipos obtidos na população. Se o haplótipo for raro, o estudo deste marcador genético pode contribuir para restringir a identidade de um indivíduo. Assim, bases de dados, como a EMPOP, são cruciais, uma vez que permitem realizar uma pesquisa e averiguar a coincidência entre os haplótipos. Neste estudo, verificou-se que 19 amostras não foram ainda descritas na base de dados. O haplótipo encontrado em 5 amostras do estudo coincide com 791 sequências de 38361 existentes na EMPOP, o que indica que é um haplótipo frequente. No entanto, o haplótipo presente nas amostras NOR032 e NOR069, apenas é encontrado em 1 sequência na EMPOP, sugerindo que não é um haplótipo frequente.

*Tabela 3.2.1.1 – Haplótipos mais frequentes nas 84 sequências de indivíduos da região do norte de Portugal, estudadas no presente estudo, bem como a sua frequência.*

Amostras	Frequência	Haplótipos
NOR003; NOR041; NOR053; NOR066; NOR076	5	16519C 263G 315.1C
NOR032; NOR069	2	16320A 16519C 263G 315.1C

Comparando a região controlo total do mtDNA com a rCRS [12], foram identificadas 757 alterações nucleotídicas em 135 posições polimórficas. Através da análise da tabela 3.2.1.2, verificou-se que o tipo de polimorfismos mais frequente são as substituições, com 572 alterações (75,6%). De forma detalhada, 554 substituições correspondem a transições (96,9%), sendo 351 entre pirimidinas (63,4%) e 203 entre purina (36,6%), demonstrando-se assim que o tipo de transições predominante foi entre pirimidinas. Por outro lado, 15 substituições correspondem a transversões (2,6%) e 3 a heteroplasmias de posição (0,5%). Foram também identificadas 160 inserções (21,1%) e 25 deleções (3,3%). Estes resultados foram consistentes com os obtidos em estudos anteriores feitos na população portuguesa [64,65]. Dentro da região HV1, localizada entre as posições 16024 e 16353, verificou-se a existência de uma zona poli-C, entre as posições 16183 e 16193 que apresenta uma alta variabilidade nucleotídica. Esta zona é caracterizada pela sequência ACCCCCTCCCC que possui uma sequência de citosinas interrompida por uma timina, na posição 16189. Nesta região foram observados 31 polimorfismos, o que corresponde a 4,1% de todas as alterações existentes. Na posição 16183, 5 sequências apresentavam uma substituição de uma adenina (A) por uma citosina (C). Tanto na posição 16186 como na posição 16187, verificou-se que 3 sequências continham uma mutação de uma C para uma timina (T). A posição 16189, constituída por uma T na sequência de referência, sofreu uma substituição para uma C em 15 amostras. Por último, na posição 16192 a C é alterado para uma T em 5 sequências.

Tabela 3.2.1.2 – Número e tipo de alterações polimórficas identificadas na região controlo total do mtDNA das 84 amostras de indivíduos da região do norte de Portugal.

Alterações	Nucleótido	Número de Polimorfismos
Transições entre pirimidinas	T → C	228
	C → T	123
Transições entre Purinas	A → G	171
	G → A	32
Transversões	A → C	6
	A → T	1
	C → A	3
	G → T	2
	G → C	3
Inserções	C	148
	A	10
	T	2
Deleções	A	12
	C	13
Heteroplasmia de posição	G → R	1
	C → Y	1
	T → Y	1

Através da análise dos resultados da região HV1 constatou-se que a posição 16126, tal como a posição 161311, apresentavam uma substituição de T para C em 20 amostras (23,8%), sendo estas alterações as mais comuns em HV1. Constatou-se também que, em 12 amostras (14,3%), a posição 16223 apresentava uma T em vez de uma C. Observou-se ainda que a transição de um C por um T estava presente em 9 sequências (10,7%), na posição 16294.

Nesta região foi possível detetar também 27 polimorfismos (32,1%) que aparecem apenas uma vez nas 84 sequências estudadas (tabela 3.2.1.3). Não foram detetadas deleções e inserções no conjunto de resultados obtidos para a região HV1.

Tabela 3.2.1.3 – Polimorfismos observados apenas uma única vez na região HV1 do mtDNA.

Posição	Alteração	Posição	Alteração	Posição	Alteração	Posição	Alteração
16037	A → G	16179	C → T	16235	A → G	16290	C → T
16075	T → C	16182	A → C	16239	C → T	16295	C → T
16145	G → A	16213	G → R	16240	A → G	16301	C → T
16148	C → T	16215	A → T	16263	T → C	16316	A → G
16153	G → A	16218	C → T	16265	A → G	16318	A → G
16174	C → T	16231	T → C	16266	C → T	16325	T → C
16178	T → C	16234	C → T	16271	T → C		

Localizada entre as posições 73 a 340, temos a região HV2. Nesta estão presentes duas mutações extremamente comuns, uma na posição 263 e outra na posição 315. Na rCRS, na posição 263, está presente uma A e, entre as posições 311 e 315, estão presentes 5 C; no entanto, estes polimorfismos foram considerados como raros, não estando presentes na maioria dos seres humanos [12]. Desta maneira, como era de esperar, no presente estudos todas as amostras apresentam as duas mutações, a transição de uma A para uma guanina (G) na posição 263 e a inserção de uma C na posição 315.

A alteração de uma A por uma G na posição 73, foi encontrada em 46 sequências (54,8%) sendo considerada como a quarta mutação mais frequente em todo o estudo. Nesta região, outras posições polimórficas apresentam uma elevada taxa de substituição, tal como a posição 152, onde se observa a alteração de uma T por uma C em 16 amostras (19%) e na posição 195, onde em 13 amostras (15,5%), existe também a alteração de uma T por uma C.

Tal como na região HV1, a região HV2 também apresenta uma zona poli-C, compreendida entre as posições 303 e 316 e caracterizada pela sequência CCCCCCTCCCCG. Durante o estudo foram detetados 135 polimorfismos (17,9%) neste local. Para além das 84 inserções observadas em 315.1, como descrito anteriormente, na posição 309, 44 amostras apresentavam uma inserção de uma C e 7 amostras a inserção de duas C.

Nesta região foram detetados 12 polimorfismos de baixa frequência, que apenas aparecem uma única vez como descritos na tabela 3.2.1.4. Para além das inserções já descritas não foram detetadas mais nenhuma, tal como não foram detetadas deleções.

Tabela 3.2.1.4 - Polimorfismos observados apenas uma única vez na região HV2 do mtDNA.

Posição	Alteração	Posição	Alteração	Posição	Alteração	Posição	Alteração
93	A → G	215	A → G	226	T → C	256	C → T
94	G → A	222	C → T	227	A → G	294	T → C
199	T → C	225	G → A	235	A → G	319	T → C

Por último, temos a região HV3, compreendida entre as posições 438 e 574. Esta apresenta uma zona de repetição nucleotídica AC, entre as posições 515 e 524, que concentra uma grande quantidade de polimorfismos. Esta zona, caracterizada pela sequência ACACACACAC, integra um total de 44 polimorfismos (5,8%). Em 12 sequências (14,3%) verificou-se a deleção dos dímeros AC, sendo este polimorfismo assinalado como 523DEL e 524DEL, segundo a nomenclatura [82]. Foram também identificadas 7 sequências com a inserção de um dinucleótido AC na posição 524 e 1 sequência com a inserção de 4 dinucleótidos AC (sendo esta sequência constituída por 9 repetições desse dinucleótido). Para além desta zona de repetição nucleotídica, à semelhança do que acontece na região HV1 e HV2, a região HV3 também apresenta uma zona poli-C, caracterizada pela sequência ACCCCCCA, entre as posições 567 e 574. No presente estudo, verificou-se a inserção de 3 citosinas (573.1C 573.2C 573.3C), em uma das amostras. Na região HV3 também foram encontrados 3 polimorfismos de baixa frequência (tabela 3.2.1.5). Não foram detetadas mais inserções que as mencionadas. Relativamente a deleções, para além das indicadas, foi detetada também a deleção de uma C na posição 498.

Para além das regiões HV1, HV2 e HV3, podem ocorrer mutações em posições não compreendidas por estas regiões. Desta maneira entre as posições 16365 a 16569, 1 a 72 e 341 a 437 foram observadas 75 alterações (8,2%) em 12 posições polimórficas. A posição 16519 apresenta a substituição de uma T para uma C em 62 amostras (73,8%). Em 2 amostras, identificou-se, na posição 16390, uma mutação de G para A e, na posição 357, uma mutação de A para G. As restantes 9 posições mutadas aparecem apenas uma única vez nas sequências analisadas (tabela 3.2.1.6).

Tabela 3.2.1.5 - Polimorfismos observados apenas uma única vez na região HV3 do mtDNA.

Posição	Alteração
458	C → T
498	DEL C
513	G → A

Tabela 3.2.1.6 - Polimorfismos observados apenas uma única vez nas regiões não compreendidas por HV1, HV2 e HV3 do mtDNA.

Posição	Alteração	Posição	Alteração	Posição	Alteração
16399	A → G	57	T → C	385	A → G
16400	C → T	58	T → C	390	A → G
16456	G → A	64	C → T	417	G → C

Nenhuma das sequências estudadas apresenta concordância total com a rCRS. O haplótipo que apresentou um menor número de alterações polimórficas, em relação à rCRS, foi de 3, num total de 5 sequências (NOR003, NOR041, NOR053, NOR066 e NOR076). Contrariamente, a sequência que contém o maior número de polimorfismos é a NOR046 que apresenta 24 substituições: 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16362C 16400T 16519C 73G 152C 182T 185T 189G 195C 247A 263G 315.1C 357G 523DEL 524DEL. O número médio de alterações nucleotídicas encontradas relativamente à rCRS foi de 9 por amostra.

A partir do *software Arlequin* foi possível obter o valor do índice de diversidade nucleotídica que indica o número médio de diferenças (por site) entre duas sequências retiradas ao acaso de uma amostra. O valor do mesmo foi de  $0.008657 \pm 0.004438$  (SD), o que se assemelha aos valores obtidos noutros estudos sobre Portugal e da região do Norte de Portugal [64,65].

### 3.2.2 Estudo das heteroplasmias

Nas 84 amostras analisadas para o estudo populacional, 35 continham heteroplasmia, o que corresponde a cerca de 41,7% das sequências. No estudo efetuado por Marques *et al.* (2014), foi observada uma taxa de heteroplasmias de 42,8%, pelo que os valores encontrados no presente estudo se enquadram nos anteriormente referenciados [65].

Ao longo do estudo, foram detetados apenas 3 indivíduos (3,6%) com heteroplasmia de posição. Tal ocorreu nas amostras NOR008, NOR016 e NOR030 e, respetivamente, nas posições 16213, 16260 e 16362. Na posição 16213 é visível uma adenina e uma guanina, sendo esta classificada, de acordo com o código estabelecido pela IUPAC, como 16213. Já nas posições 16260 e 16362, pode-se observar uma citosina e uma timina, classificadas como 16260Y e 16362Y.

Relativamente às heteroplasmias de comprimento (tabela 3.2.2.1), foram identificadas 37 heteroplasmias distribuídas por 32 amostras (38,1%). Destas, 4 apresentavam heteroplasmia apenas associada à zona poli-C da região HV1, compreendida entre as posições 16182 e 16191; 23 apresentavam heteroplasmia apenas na zona poli-C da região HV2, compreendida entre as posições 303 e 316; e 5 apresentavam heteroplasmia em ambas as regiões hipervariáveis de HV1 e HV2. Não foi encontrada nenhuma heteroplasmia de comprimento associada à região HV3.

Valores similares foram detetados noutros estudos, entre eles por Marques *et al.* (2014), onde foi verificada uma taxa de heteroplasmia de posição de 4,1% e uma taxa de heteroplasmia de comprimento de 39,7%. O estudo, também referenciou que a taxa de heteroplasmia de comprimento de HV2 foi superior à de HV1, rondando os 28,8%. Todos os resultados estão de acordo com os resultados alcançados no presente trabalho [65].

Todos os casos de heteroplasmia de comprimento foram confirmados por uma sequenciação em duplicado das duas cadeias de mtDNA.

Tabela 3.2.2.1 – Frequência de haplótipos homoplásmicos e heteroplásmicos nas 84 sequências de mtDNA analisadas e região hipervariável de ocorrência.

	Tipo de Heteroplasmia	Segmento hipervariável	Frequência	%
Amostras homoplásmicas			49	58,3
Amostras heteroplásmicas	Posição		3	3,6
	Comprimento	HV1	4	4,8
		HV2	23	27,4
		HV1 + HV2	5	6

### 3.2.3 Estudo dos haplogrupos

Os 79 haplótipos obtidos no estudo foram divididos em 52 haplogrupos distintos. Esta classificação das amostras em haplogrupos é feita, tendo por base, a árvore filogenética *PhyloTree*, desenhada por van Over e Kayser, em 2009 [40]. Cada haplogrupo é identificado por letras do alfabeto seguidas de números, que permite uma subdivisão em subhaplogrupos. Para que um haplótipo seja corretamente identificado e inserido num haplogrupo, precisa de ter todos os polimorfismos específicos que caracterizam esse haplogrupo, embora possa apresentar outros para além dos específicos. Desta forma, dois indivíduos com haplótipos semelhantes, mas com alguns polimorfismos diferentes, podem pertencer ao mesmo haplogrupo. Por exemplo, os haplótipos das amostras NOR034 e NOR071 são, respetivamente, 16304C 16519C 263G 309.1C 315.1C 456T e 16304C 263G 309.1C 315.1C 456T

523DEL 524DEL. Embora apresentem algumas diferenças ambas as amostras são inseridas no haplogrupo H5 que é caracterizado pelos polimorfismos 16304C e 456T, presentes em ambas as amostras.

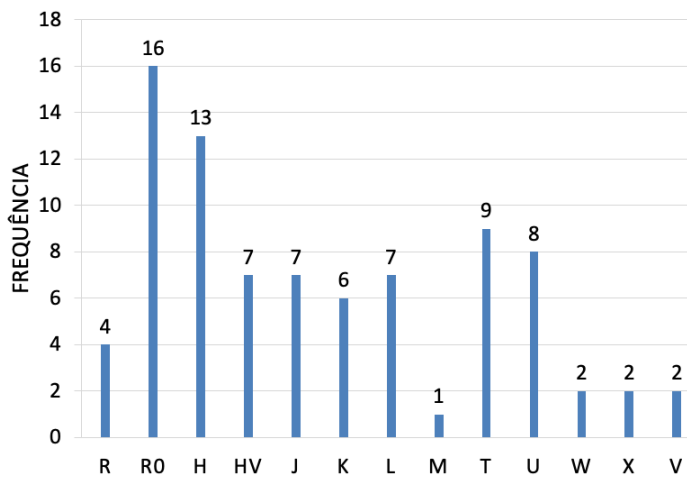


Figura 3.2.3.1 – Representação esquemática dos diferentes haplogrupos e respectivas frequências identificados nas 84 seqüências de mtDNA de indivíduos da região do norte de Portugal.

Através da análise dos resultados obtidos foi possível elaborar a figura 3.2.3.1 onde estão presentes as frequências dos haplogrupos na população em estudo.

Como haplogrupos característicos de populações europeias temos os haplogrupos H, I, J, K, M, T, U, V, W e X [10,83]. Neste estudo, o haplogrupo H está presente em 15,48% das amostras, o haplogrupo J em 8,33%, K em 7,14%, M em 1,19%, T em 10,71%, U em 9,52%, V em 2,38%, W em 2,38% e X em 2,38%, o que juntos, perfazem 59,52%. A alta frequência destes haplogrupos revela a grande contribuição de linhagens europeias na composição genética da população em estudo.

No presente estudo também se constataram 7 amostras (8,33%), percentagem ao haplogrupo L característico de populações africanas. A frequência destes haplogrupos revela a presença de linhagens maternas africanas na composição genética da população do norte de Portugal. A sua presença pode refletir o passado histórico de tráfico de escravos de África [84].

Os restantes 31,14% estão distribuídos pelos haplogrupos R (4,76%), R0 (19,05%) e HV (8,33%). O haplogrupo R e HV são encontrados na Euroásia e no Norte de África [85,86] enquanto o haplogrupo R0 é maioritário na Península Arábica e também no Norte de África [87].

- Haplogrupo R, R0 e HV

O haplogrupo R é um descendente do macrohaplogrupo N, e que está associado ao povoamento da Euroásia há cerca de 70.000 anos. Este é considerado um dos haplogrupos mais comuns do mtDNA que está amplamente distribuído [85]. Neste estudo, 4 amostras (4,76%) pertencem a este haplogrupo.

Dos subhaplogrupos de R, o mais comum é o R0. Este haplogrupo é um dos mais antigos e considerado como o ancestral comum de todos os haplogrupos R. Estima-se que tenha surgido há cerca de 200.000 anos. É comum especialmente entre os povos no Norte de África, incluindo Marrocos e Tunísia [88,89]. No entanto, embora com frequências geralmente mais reduzidas, também é encontrado em indivíduos de outras etnias, incluindo europeus, asiáticos e americanos [31,90,91]. Este haplogrupo divide-se em dois sub-ramos, um maior, conhecido como HV e outro menor conhecido por R0a. No presente estudo, o haplogrupo R0 foi encontrado em 19% das amostras, sendo o haplogrupo a registar o maior número de ocorrências. As frequências de R0 em Portugal, referidos noutros estudos, variam um pouco. Trata-se, no entanto, de um haplogrupo frequentemente detetado [65,92].

Como subhaplogrupo R temos ainda o haplogrupo HV. Como possível origem deste haplogrupo, aponta-se a região do oeste do Irão, Mesopotâmia e sul do Cáucaso, uma vez que foi neste local que foi encontrada a maior prevalência de HV. Na Europa, esta linhagem é rara no norte e oeste, mas comum no sul e leste. Assim sendo, este haplogrupo pode ser encontrado na Ásia Central e Ocidental, Europa do Sul e Oriental e Norte de África [86]. Neste estudo, o haplogrupo HV foi detetado em 7 amostras (8,33%), sendo três amostras pertencentes ao subhaplogrupo HV0, valores que estão de acordo com os apresentados na literatura [65].

- Haplogrupo H

O haplogrupo H é o mais comum na Europa, representando cerca de 40% a 50% de todas as linhagens mitocondriais europeias. Pensa-se que este haplogrupo teve origem no médio oriente, há cerca de 30-35 mil anos, tendo expandindo-se para o sul da Europa há cerca de 20.000 anos. Também é comumente encontrado em áreas do Norte de África e Médio Oriente e, embora em menor frequência, em locais do Nordeste da Índia e Ásia Central [93–95]. A distribuição geográfica do haplogrupo H sugere que a sua origem está relacionada com a expansão dos agricultores neolíticos que repovoaram a Europa após o último máximo glacial. A presença do haplogrupo H em áreas do Norte da África e do Médio oriente suporta essa teoria, fornecendo informações valiosas sobre as conexões históricas entre essas regiões [95,96]. É geralmente considerado um importante marcador genético para compreender a história da Europa e da expansão humana.

Este haplogrupo é constituído por vários subhaplogrupos, como H1, H2, H3, entre outros. H1 é o mais comum na Europa, enquanto o subtipo H3 é o mais comum no Norte da África e no Oriente Médio.

Identificaram-se 13 indivíduos (15%) pertencentes a este haplogrupo, o que está conforme com a distribuição generalizada das linhagens H pela Europa [94]. Dentro do haplogrupo H, 53,85% das amostras estão inseridas no subhaplogrupo H1, dividindo-se as restantes pelos subhaplogrupos H5 (15,38%) e H14 (7,69%). Os resultados são consistentes com estudos anteriores que mostram picos de frequência do subhaplogrupo H1 na Península Ibérica e áreas circundantes [30,65].

- Haplogrupo L

O haplogrupo L é um dos mais antigos e diversificados do genoma humano. Este é originário de África e é encontrado com alta frequência em populações africanas [40,97]. Além disso, o haplogrupo L é encontrado em populações europeias, asiáticas e americanas, devido à migração humana ao longo dos anos [97]. Em geral, o haplogrupo L é considerado um indicador importante da diversidade genética humana e da história evolutiva dos seres humanos.

O haplogrupo L é dividido em vários subgrupos, cada um com sua própria história evolutiva e geográfica. Os subgrupos do haplogrupo L incluem L1, L2 e L3, que são os mais antigos e estreitamente relacionados entre si. Eles são encontrados principalmente em populações africanas e são considerados os ancestrais dos outros subgrupos do haplogrupo L [98,99]. Outros subgrupos incluem L4, L5 e L6, que são encontrados principalmente em populações afro-descendentes fora da África. L4 e L5 são encontrados com maior frequência em populações do leste de África, enquanto L6 é encontrado com maior frequência em populações da África Central e Oriental [41,98].

Neste estudo, embora não sejam tipicamente encontradas em populações europeias, algumas linhagens L (8%), L3 (60%) e L1b (40%) foram detetadas [64,100]. Valores semelhantes foram registados em estudos anteriores feitos na região de Portugal [30,64,65]. A presença destas linhagens na Península Ibérica tem sido explicada pelo tráfico de escravos de África, desde meados do século XV até o século XIX [84].

- Haplogrupo M

O haplogrupo M é um dos principais haplogrupos encontrados na Ásia, especificamente na Ásia Meridional e Oriental. É o haplogrupo mais comum na população indiana e também encontrado em altas frequências em outras populações da Ásia Meridional e Oriental. Alguns estudos sugerem que o haplogrupo M pode ter surgido há cerca de 60.000 anos, durante a expansão humana na Ásia [43,101]. É composto por diversos subgrupos conhecidos como os haplogrupos M1-M7, descobertos posteriormente.

O haplogrupo M5a1b foi detetado em uma única amostra (1%) do estudo. Como já foi referido este haplogrupo é encontrado normalmente em populações indianas [101] no entanto, mais recente mente foi detetado em grupos de ciganos ibéricos [102], indicativo de que o indivíduo poderá ser de ascendência cigana.

- Haplogrupo T

O haplogrupo T é bastante comum na Europa e Próximo Oriente, compondo sensivelmente 10% dos mtDNAs na Europa e aproximadamente 8% no Próximo Oriente. Este pode ser dividido em vários subgrupos T1-T5, sendo que no presente estudo apenas os subhaplogrupos T1 e T2 foram detetados. Ambos os subhaplogrupos tiveram origem, há sensivelmente 21.000 anos no Próximo Oriente, dispersando-se mais recentemente para a Europa. O T2 está normalmente presente em maior percentagem na Europa Ocidental e Oriente Próximo, sendo que T1 também se encontra presente na Europa oriental [103].

No estudo, o haplogrupo T é detetado em 9 amostras (10,73%), 4 pertencentes a T1 e 5 a T2. Estes valores são concordantes com os apresentados em outros estudos sobre a região de Portugal [30,64,65,92].

- Haplogrupo U

Estima-se que o haplogrupo U surgiu há mais de 50 mil anos, sendo específico de populações europeias, embora tivesse também sido identificado na Índia e África, nomeadamente no Norte de África. Este haplogrupo é dividido em vários subhaplogrupos, de U1 a U9, cada um com as suas características e distribuição geográfica próprias [94,104,105].

Este haplogrupo está presente em 9,52% das amostras do estudo, observado pela presença dos subhaplogrupos U2, U4 e U5. O subhaplogrupo U2 foi identificado em 2 indivíduos. É encontrado principalmente em populações indianas, sendo raro na Europa. Na Europa foram encontrados os clusters U4 e U5, cada um presente em 3 indivíduos. É de admitir encontrar o haplogrupo U5 na Ásia [94,106]. Estes valores encontrados nos subhaplogrupos do haplogrupo U não destoam dos previamente detetadas na população portuguesa [30,64,65].

O subhaplogrupo U6, embora em baixas concentrações, foi registado em todos os estudos da região do Norte de Portugal. Este é também característico de populações do norte e noroeste africano e a sua presença em Portugal é justificada com o movimento populacional associado ao tráfico de escravos ou durante as invasões berbere-árabe da Península [30,64,65,84]. No entanto, no presente estudo não foram encontradas amostras com este haplogrupo o que pode ser justificado pela baixa frequência do mesmo em Portugal ou pela necessidade de uma maior amostragem.

- Haplogrupo J

O haplogrupo J representa quase 9% dos haplogrupos Europeus e aproximadamente 13% dos haplogrupos do Próximo Oriente. No estudo realizado esteve presente em 7 das amostras, numa proporção de 8,33% [103].

Datado de aproximadamente 33.000 anos o subhaplogrupo J1 é o mais comum e está dividido em três subclados, sendo o J1c principalmente encontrado na Europa, nomeadamente no Centro, Balcãs e na Ucrânia. No estudo, 6 dos haplogrupos J1 pertencem ao subhaplogrupo J1c, e o restante pertence ao subhaplogrupo J2a1. O haplogrupo J2, muito mais raro, apresenta dois subclados europeus: J2a1, presente na Europa central e J2b1 mais focado na Europa mediterrânea central e atlântica [103]. Comparando os resultados obtidos com os apresentados noutros papers de estudo do mtDNA em Portugal, observamos que os valores estão dentro do esperado [30,64,65].

- Haplogrupo K

Considerado um subgrupo do haplogrupo U, temos o haplogrupo K, comum em toda a Europa. Este divide-se nos subhaplogrupos K1, K2 e K3 [107]. Nas 6 amostras (7,14%) correspondentes ao haplogrupo K, identificaram-se 5 que se enquadram no subhaplogrupo K1 e 1 no subhaplogrupo K2. O mais o comum, subhaplogrupo K1, chegou ao continente europeu por migrações provenientes do Próximo Oriente e é frequentemente encontrado nas áreas do mediterrâneo [108,109]. As concentrações encontradas relativamente a este haplogrupo, são semelhantes às detetadas nos estudos de Mairal *et al.* (2013) e Marques *et al.* (2014).

- Haplogrupo W, X e V

Os haplogrupos W, X e V encontram-se presentes no estudo em concentrações muito baixas. Cada um apresenta duas amostras, correspondendo a uma frequência de 2,38% por haplogrupo.

Estudos genéticos sugerem que o haplogrupo W é originário da região do Cáucaso, tendo-se espalhando posteriormente pela Europa, Ásia ocidental e Sul da Ásia. Este divide-se em 7 subhaplogrupos, sendo que os presentes no estudo W e W6 são encontrados, na sua maioria, na Europa [110].

O haplogrupo X, considerado como um dos mais raros no mundo, é geralmente encontrado em baixas frequências na Europa e na América. Este representa cerca de aproximadamente 4% dos haplogrupos detetados na Europa [111,112].

O haplogrupo V, deriva do subclado HV0a do haplogrupo HV. É um haplogrupo relativamente raro, sendo encontrado em cerca de 4% dos europeus nativos [113,114], podendo também pode ser encontrado em algumas partes do noroeste de África [115].

As baixas frequências encontradas no estudo, destes três haplogrupos, estão de acordo com os dados, constantes da literatura referente a estes haplogrupos, bem como as determinadas noutros estudos [30,65].

### 3.2.4 Análise Filogenética

Tabela 2.2.4.1 - População do presente estudo e populações selecionadas da literatura, utilizadas no estudo filogenético, tendo por base a região controlo total do mtDNA.

Populações	Tamanho da Amostra	Referência
Imigrantes cabo-verdianos	103	[32]
Imigrantes angolanos	173	[33]
Imigrantes moçambicanos	83	[116]
Imigrantes brasileiros	75	[117]
Norte de Portugal	84	Presente estudo
Portugal	292	[65]
Alemanha	199	[118]
Marrocos	509	[89]
Miranda do Douro	121	[30]
Zamora	113	[30]
Euroásia Ocidental	31	[119]

Para a estimar a variabilidade genética entre grupos populacionais, recorreu-se à análise molecular da variância (AMOVA). Para tal foram selecionadas, da literatura, populações previamente estudadas, que estão descritas na tabela 3.2.4.1.

Com referência a cada população foram calculados os valores das distâncias genéticas (Fst) e os respetivos valores de p (tabela 3.2.4.2). Os Fst entre as diferentes populações permitiu

obter uma árvore onde são visíveis as distâncias genéticas entre as populações (figura 3.2.4.1). Valores de p iguais ou superiores a 0,05 indicam semelhança entre duas populações. Verificou-se, todavia, que os valores de p obtidos, para algumas populações eram inferiores ao valor indicado e superior ao valor indicado para outras. Desta maneira, e com referência à população em estudo as populações de imigrantes que habitam em Portugal, bem como Marrocos e Mirando do Douro apresentam diferenças

significativas, já o mesmo não sucedendo com as populações de Portugal, Alemanha, Zamora e os Países da Euroásia Ocidental.

A partir da figura 3.2.4.1 foi possível verificar a formação de dois grupos: um composto pelas populações de imigrantes do continente africano em Portugal e outro composto pelas populações europeias. Este resultado já era previsível, uma vez que, em termos genéticos, as populações africanas são mais semelhantes entre si do que relativamente a populações de outros continentes, bem como as populações europeias são mais semelhantes entre si do que relativamente a populações de outros continentes.

Tabela 3.2.4.2 - Valores de *Fst* e *p* relativos às populações do estudo filogenético obtidos através do programa Arlequin para a totalidade da região controlo do mtDNA. CV – Imigrantes Cabo-Verdianos; Angola - Imigrantes Angolanos; Moçam - Imigrantes Moçambicanos; Brasil – Imigrantes Brasileiros; Norte – Norte de Portugal; PT – Portugal; Alem – Alemanha; Marro – Marrocos; MiDou – Mirando do Douro; EuroOci – Euroásia ocidental

Populações	CV	Angola	Moçam	Brasil	Norte	PT	Alem	Marro	MiDou	Zamora	EuroOci
CV	-	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000
Angola	0.04388	-	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000
Moçam	0.06223	0.06223	-	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000
Brasil	0.06915	0.04857	0.03946	-	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000
Norte	0.20646	0.14193	0.13466	0.06577	-	0.92793+-0.0196	0.54054+-0.0470	0.0000+-0.0000	0.0000+-0.0000	0.19820+-0.0466	0.39640+-0.0680
PT	0.21911	0.16161	0.14944	0.06867	-0.00278	-	0.25225+-0.0353	0.0000+-0.0000	0.0000+-0.0000	0.03604+-0.0148	0.31532+-0.0365
Alem	0.23908	0.17697	0.16836	0.08385	-0.00040	0.00062	-	0.0000+-0.0000	0.0000+-0.0000	0.09009+-0.0271	0.11712+-0.0333
Marro	0.12289	0.09709	0.08260	0.02211	0.02111	0.02034	0.03155	-	0.0000+-0.0000	0.0000+-0.0000	0.0000+-0.0000
MiDou	0.24984	0.18027	0.17499	0.09887	0.00909	0.01076	0.01010	0.03992	-	0.0000+-0.0000	0.05405+-0.0201
Zamora	0.22148	0.16303	0.15735	0.07980	0.00223	0.00319	0.00339	0.02773	0.00633	-	0.04505+-0.0279
EuroOci	0.20463	0.13463	0.12457	0.06611	0.00116	0.00157	0.00537	0.02582	0.01342	0.01400	-

Como era de esperar, através da análise da matriz de distâncias interpopulacionais observou-se que a população do norte de Portugal encontra-se geneticamente mais próxima da população portuguesa, o que faz todo o sentido uma vez que a população do norte de Portugal está inserida na população portuguesa. As características evidenciadas neste estudo, como haplótipos, haplogrupos e diversas frequências estavam todas de acordo com as obtidas noutros estudos relativos à população portuguesa no seu todo, o que justifica esta proximidade de valor de *Fst* (-0.00278) e um valor tão alto de *p* (0.92793+-0.0196) [30,64,65]. A seguir à população portuguesa, a população da Alemanha foi a que apresentou uma menor distância genética face à população do Norte de Portugal (*Fst* = -0.00040 e  $\rho$  = 0.54054+-0.0470). Esta proximidade genética não é de estranhar uma vez que a população do norte de Portugal, como vem demonstrado no estudo, compartilha as características da diversidade do mtDNA geralmente registadas em europeus (todos os haplogrupos europeus foram detetados). No entanto, a população com as amostras dos Países da Europa Ocidental é a que acusa uma proximidade maior. No estudo de Zimmermann *et al.* (2011) foram analisadas apenas 31 amostras e, embora diversos países estivessem representados (Alemanha, França, Polónia, Bósnia, Roménia, Suíça, Itália, Áustria, Turquia, Cazaquistão, Dinamarca, Portugal, Espanha e Rússia), 14 amostras (45%) pertenciam à Alemanha, estando os outros países representados com apenas 1 ou 2 amostras. Devido à supremacia de amostras da Alemanha, é normal que a proximidade entre o norte de Portugal e os Países da Euroásia Ocidental

seja alta ( $F_{st} = 0.00116$  e  $0.39640 \pm 0.0680$ ), embora não tão alta como a verificada com a população apenas alemã. No que diz respeito à população de Mirando do Douro, dentro das populações europeias, está é a que apresenta uma maior distância genética da população do Norte de Portugal. Embora, Miranda do Douro pertença à região do norte de Portugal, esta população é caracterizada por um nível relativamente baixo de diversidade genética e por um padrão de distribuição de haplogrupos que possui algumas singularidades comparativamente ao apresentado em Portugal. Com efeito, estes haplogrupos apresentam uma elevada frequência de alguns haplogrupos tipicamente encontrados em populações judaicas, o que justifica estes resultados [30].

Fora das populações europeias, Marrocos é a população que apresenta uma maior proximidade com a população em estudo ( $F_{st} = 0.02111$ ). Uma vez que Portugal recebeu influências significativas do Norte de África, devido à colonização árabe, as frequências de haplogrupos específicos destas regiões são relatadas com uma frequência superior às descritas para outras populações europeias [120]. Desta maneira, este povo é geneticamente mais próximo da população portuguesa do que das populações africanas [121,122].

Comparando as diferentes comunidades de imigrantes a viver em Lisboa, a que se encontra geneticamente mais próxima da população portuguesa é a brasileira.

Dentro das populações imigrantes africanas a mais próxima geneticamente é a moçambicana enquanto a cabo-verdiana é a que se encontra mais distante. Estes resultados já foram detetados anteriormente em estudos com outros marcadores genéticos [123]. Como explicação para estes resultados existem duas hipóteses: a modulação diferencial dos padrões genéticos, devido aos diferentes momentos de colonização pelo povo português; e as diferentes rotas estabelecidas na migração do povo Bantu ao longo da África subsariana, uma para a zona este e outra para oeste, que originou os padrões genéticos distintos destas regiões [124].

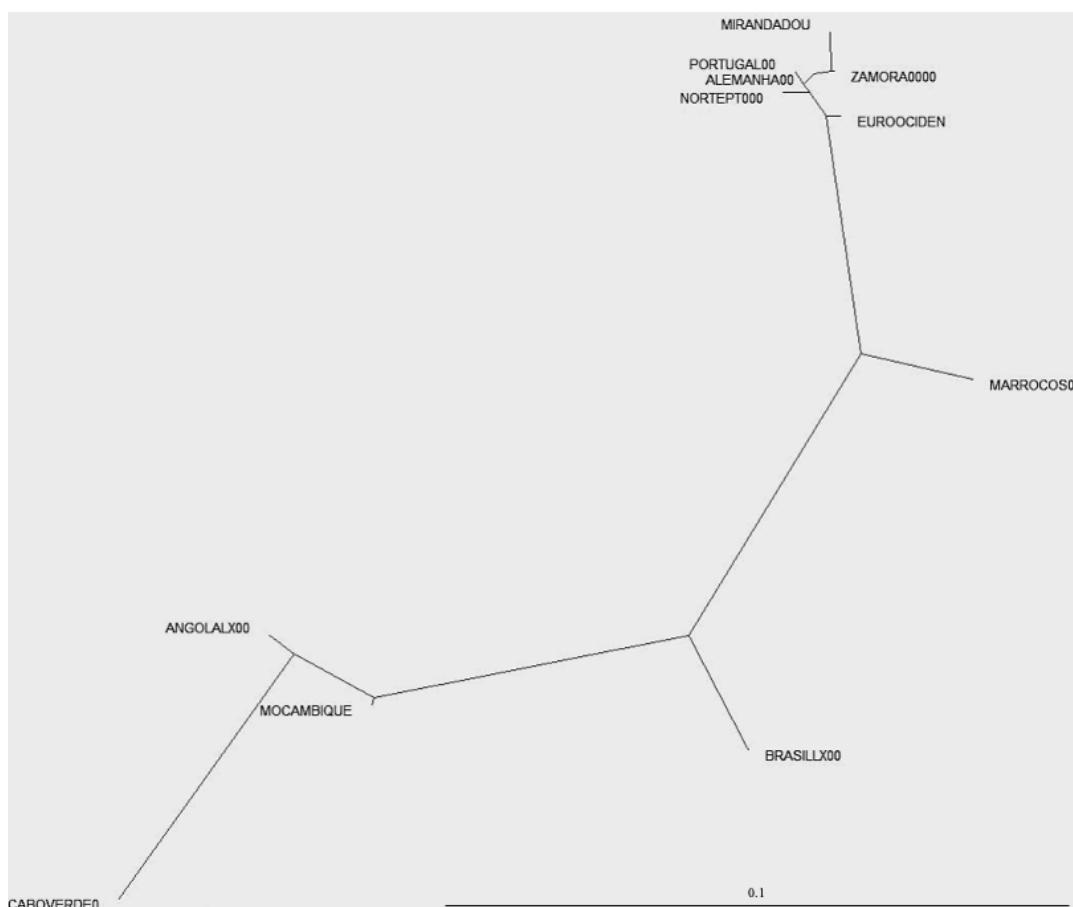


Figura 3.2.4.1 - Representação filogenética das distâncias genéticas entre a população do Norte de Portugal e as restantes selecionadas da literatura.

## 4. Conclusão

Um processo de validação envolve sempre o objetivo de melhor alcançar a integração de elementos na rotina laboratorial. Insere-se neste propósito a validação interna da análise dos polimorfismos do DNA mitocondrial por sequenciação no *SeqStudio*<sup>TM</sup> Genetic Analyzer Instrument, a partir da qual se concluiu que:

- O *SeqStudio*<sup>TM</sup> Genetic Analyzer ficou validado, aquando da validação dos controlos positivos, ao comparar os resultados obtidos por este com os obtidos pelo *3130xl Genetic Analyzer*.
- No estudo da sensibilidade, a concentração mínima detetável, aquando da utilização dos *primers* HV1, HV2, HV3 e RV, foi definida como sendo de 0,001ng.
- No estudo de especificidade confirmou-se que os *primers* são específicos para o mtDNA humano.
- Os cabelos arrancados apresentam uma maior quantidade de DNA, quando comparados com os cabelos caídos, o que se traduz numa melhor qualidade de resultados obtidos.
- A amplificação e sequenciação da região controlo total das amostras de manchas de sangue extraídas por *PrepFiler Express*<sup>TM</sup>, apresentaram resultados com qualidade.
- A amplificação de 3 amostras de manchas de sangue com punch direto, após uma diluição das mesmas apresentou resultados promissores, para a região de HV1.
- Nas manchas de sangue extraídas por *Prep-n-Go*<sup>TM</sup>, foram amplificadas e sequenciadas as 4 regiões do mtDNA. A grande maioria das amostras sequenciou, ficando apenas a faltar a validação da região RV devido a uma falha geral na amplificação;
- Todas as amostras de zaragoas bucais extraídas por *Prep-n-Go*<sup>TM</sup> obtiveram bons resultados na validação deste método.
- Para terminar a validação é necessário utilizar mais amostras e estudar todas as regiões, incluindo a região controlo total, tanto com a extração por *Prep-n-Go*<sup>TM</sup> como por *PrepFiler Express*<sup>TM</sup>.

Um estudo populacional tem por objetivo obter informação sobre a diversidade e, eventualmente, origem genética de uma determinada população. O estudo populacional da região controlo do mtDNA de 87 amostras da região do Norte de Portugal insere-se neste objetivo. Assim:

- A população apresenta uma elevada percentagem de haplótipo únicos, com 77 haplótipos diferentes em 87 amostras, revelando uma grande variabilidade genética, já descrita por outros autores, para a população portuguesa.
- Conforme identificado em 5 amostra o haplótipo mais comum determinado no estudo é o 16519C 263G 315.1C.
- A análise das sequências de mtDNA permitiu detetar um total de 757 alterações nucleotídicas em 153 posições polimórficas. A alteração mais frequente foi a substituição (75,6%) e dentro desta, as transições entre pirimidinas foram as que tiveram uma maior incidência.
- O estudo das sequências obtidas confirmou a existência de uma zona muito polimórfica na região HV1, entre as posições 16183 e 16193, e na região HV2, entre as posições 303 e 316. Foi na região polimórfica de HV2 que se verificaram as alterações mais frequentes do estudo: a inserção de uma citosina na posição 315 (315.1C) e a alteração de uma citosina por uma guanina na posição 263, ambas com uma taxa de 100%.
- O maior número de alterações polimórficas encontradas numa sequência, em relação à rCRS foi de 24 e o menor, de 3. O número médio de alterações nucleotídicas por amostra de é 9.

- A taxa de heteroplasmia no estudo foi de 41,7%. Relativamente às heteroplasmias de posição, foram detetadas 3, enquanto heteroplasmias de comprimento foram encontradas 37, sendo que a grande maioria destas ocorreu na região HV2.
- Entre os 79 haplótipos obtidos, 59,52% foram classificados em haplogrupos europeus H, I, J, K, M, T, U, V, W e X. O haplótipo mais abundante foi R0 (19%), seguido do haplogrupo H (15%). De um modo geral, na população em estudo, foi encontrada uma composição de haplogrupo de mtDNA típica da Europa Ocidental.
- O haplogrupo L, haplogrupo tipicamente africano, foi encontrado numa taxa de 8%. Estas altas frequências já tinham sido detetadas anteriormente em estudos da região de Portugal e podem ser explicadas pelo tráfico de escravos, provenientes de África, em que Portugal era interveniente.
- A análise filogenética permitiu verificar a formação de dois grupos: um, composto pelas populações europeias e outro, mais distante, composto pelas populações do continente africano.
- Através da obtenção das distâncias genéticas entre a população do Norte de Portugal com outras populações selecionadas da literatura, confirmou-se a proximidade desta região com Portugal e com as populações europeias.
- Relativamente às diferentes comunidades de imigrantes que residem em Portugal, constatou-se que a população em estudo se encontra geneticamente mais próxima da população de imigrantes brasileiros, seguindo-se a população de imigrantes moçambicanos.

Com o presente estudo, ficou concluída uma grande parte da validação do estudo dos polimorfismos do mtDNA no analisador genético *SeqStudio*<sup>TM</sup> viabilizando assim a introdução dos métodos validados na rotina pericial do laboratório forense do Serviço de Genética e Biologia Forenses, Delegação do Norte do INMLCF. Por outro lado, este trabalho também contribuiu para uma atualização da informação sobre a composição genética da população do Norte de Portugal, bem como para reforçar a importância da realização de estudos populacionais e da análise do mtDNA em contexto de estudos históricos relativos a trajetos migratórios, e aferir sobre as relações filogenéticas entre diferentes comunidades de populações humanas.

## 5. Bibliografia

- [1] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific 'fingerprints' of human DNA, *Nature*. 316 (1985) 76–79. <https://doi.org/10.1038/316076a0>.
- [2] J. Cummins, Mitochondrial DNA in mammalian reproduction, *Rev Reprod*. 3 (1998) 172–182. <https://doi.org/10.1530/ror.0.0030172>.
- [3] D.C. Logan, The mitochondrial compartment, *J Exp Bot*. 58 (2007) 1225–1243. <https://doi.org/10.1093/jxb/erj151>.
- [4] F. Legros, F. Malka, P. Frachon, A. Lombès, M. Rojo, Organization and dynamics of human mitochondrial DNA, *J Cell Sci*. 117 (2004) 2653–2662. <https://doi.org/10.1242/jcs.01134>.
- [5] H. Lodish, A. Berk, C.A. Kaiser, M. Krieger, M.P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira, *Molecular Cell Biology*, W. H. Freeman, 2006.
- [6] R.L. Cann, M. Stoneking, A.C. Wilson, Mitochondrial DNA and human evolution, *Nature*. 325 (1987) 31–36. <https://doi.org/10.1146/annurev.genom.6.080604.162249>.
- [7] M.M.K. Nass, S. Nass, Intramitochondrial fibers with DNA characteristics I. Fixation and Electron Staining Reactions, *J Cell Biol*. 19 (1963) 593–611. <https://doi.org/10.1083/jcb.19.3.593>.
- [8] A. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijin, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, *Nature*. 290 (1981) 457–465. <https://doi.org/10.1038/290457a0>.
- [9] J.M. Butler, B.C. Levin, Forensic applications of mitochondrial DNA, *Trends Biotechnol*. 16 (1998) 158–162. [https://doi.org/doi.org/10.1016/S0167-7799\(98\)01173-1](https://doi.org/doi.org/10.1016/S0167-7799(98)01173-1).
- [10] B. Budowle, M.W. Allard, M.R. Wilson, R. Chakraborty, Forensics and mitochondrial DNA: applications, debates, and foundations, *Annual Review of Genomics and Human Genetics*. 4 (2003) 119–141. <https://doi.org/10.1146/annurev.genom.4.070802.110352>.
- [11] S. Lutz, H.J. Weisser, J. Heizmann, S. Pollak, A third hypervariable region in the human mitochondrial D-loop, *Hum Genet*. 101 (1997) 384.
- [12] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat Genet*. 23 (1999) 147. <https://doi.org/10.1038/13779>.
- [13] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, E. Pokorak, M. Prinz, A. Salas, P.M. Schneider, T.J. Parsons, DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing,

Forensic Sci Int Genet. 13 (2014) 134–142.  
<https://doi.org/10.1016/j.fsigen.2014.07.010>.

- [14] P. Fernández-Silva, J.A. Enriquez, J. Montoya, Replication and transcription of mammalian mitochondrial DNA, *Exp Physiol.* 88 (2003) 41–56. <https://doi.org/10.1113/eph8802514>.
- [15] P. Borst, A.M. Kroon, Mitochondrial DNA: Physicochemical Properties, Replication, and Genetic Function, *Int Rev Cytol.* 26 (1969) 107–190. [https://doi.org/10.1016/s0074-7696\(08\)61635-6](https://doi.org/10.1016/s0074-7696(08)61635-6).
- [16] J.T. Case, D.C. Wallace, Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts, *Somatic Cell Genet.* 7 (1981) 103–108. <https://doi.org/10.1007/BF01544751>.
- [17] R.E. Giles, H. Blanc, H.M. Cann, D.C. Wallace, Maternal inheritance of human mitochondrial DNA, *Proceedings of the National Academy of Sciences.* 77 (1980) 6715–6719. <https://doi.org/10.1073/pnas.77.11.6715>.
- [18] A. Amorim, T. Fernandes, N. Taveira, Mitochondrial DNA in human identification: A review, *PeerJ.* 7 (2019) 7314. <https://doi.org/10.7717/peerj.7314>.
- [19] W.M. Brown, M. George, JR, A.C. Wilson, Rapid evolution of animal mitochondrial DNA, *Proceedings of the National Academy of Sciences of the United States of America.* 76 (1979) 1967–1971. <https://doi.org/10.1073/pnas.76.4.1967>.
- [20] B. Budowle, M.R. Wilson, J.A. DiZinno, C. Stauffer, M.A. Fasano, M.M. Holland, K.L. Monson, Mitochondrial DNA regions HVI and HVII population data, *Forensic Sci Int.* 103 (1999) 23–35. [https://doi.org/10.1016/s0379-0738\(99\)00042-0](https://doi.org/10.1016/s0379-0738(99)00042-0).
- [21] J.L. Elson, R.M. Andrews, P.F. Chinnery, R.N. Lightowers, D.M. Turnbull, N. Howell, Analysis of European mtDNAs for Recombination, *Am. J. Hum. Genet.* 68 (2001) 145–153. <https://doi.org/10.1086/316938>.
- [22] Y. Kraytsberg, M. Schwartz, T.A. Brown, K. Ebralidse, W.S. Kunz, D.A. Clayton, J. Vissing, K. Khrapko, Recombination of Human Mitochondrial DNA, *Science* (1979). 304 (2004) 981. <https://doi.org/10.1126/science.1096342>.
- [23] T. Melton, Mitochondrial DNA Heteroplasmy, *Forensic Sci Rev.* 16 (2004) 1–20.
- [24] P.L. Ivanov, M.J. Wadhams, R.K. Roby, M.M. Holland, V.W. Weedn, T.J. Parsons, Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II, *Nat Genet.* 12 (1996) 417–420. <https://doi.org/10.1038/ng0496-417>.
- [25] K.A. Roberts, C. Calloway, Characterization of Mitochondrial DNA Sequence Heteroplasmy in Blood Tissue and Hair as a Function of Hair Morphology, *J Forensic Sci.* 56 (2011) 46–60. <https://doi.org/10.1111/j.1556-4029.2010.01540.x>.

- [26] A. Brandstätter, W. Parson, Mitochondrial DNA heteroplasmy or artefacts - A matter of the amplification strategy?, *Int J Legal Med.* 117 (2003) 180–184. <https://doi.org/10.1007/s00414-002-0350-7>.
- [27] A. Nagai, I. Nakamura, F. Shiraki, Y. Bunai, I. Ohya, Sequence polymorphism of mitochondrial DNA in Japanese individuals from Gifu Prefecture, *Leg Med.* 5 (2003) S210–S213. [https://doi.org/10.1016/s1344-6223\(02\)00114-1](https://doi.org/10.1016/s1344-6223(02)00114-1).
- [28] H. Afonso Costa, M. Carvalho, V. Lopes, F. Balsa, A.M. Bento, A. Serra, L. Andrade, M.J. Anjos, M.C. Vide, S. Pantoja, D.N. Vieira, F. Corte-Real, Mitochondrial DNA sequence analysis of a native Bolivian population, *J Forensic Leg Med.* 17 (2010) 247–253. <https://doi.org/10.1016/j.jflm.2010.02.011>.
- [29] J.C. Alvarez, D.L.E. Johnson, J.A. Lorente, E. Martinez-espín, L.J. Martinez-Gonzalez, M. Allard, M.R. Wilson, B. Budowle, Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set, *Leg Med.* 9 (2007) 293–304. <https://doi.org/10.1016/j.legalmed.2007.05.005>.
- [30] Q. Mairal, C. Santos, M. Silva, S.L. Marques, A. Ramos, M.P. Aluja, A. Amorim, M.J. Prata, L. Alvarez, Linguistic isolates in Portugal: Insights from the mitochondrial DNA pattern, *Forensic Sci Int Genet.* 7 (2013) 618–623. <https://doi.org/10.1016/j.fsigen.2013.08.009>.
- [31] L. Palencia, L. Valverde, A. Álvarez, L.M. Cainé, S. Cardoso, M.A. Alfonso-Sánchez, M.F. Pinheiro, M.M. de Pancorbo, Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): Predominance of the European input, *Int J Legal Med.* 124 (2010) 331–336. <https://doi.org/10.1007/s00414-010-0464-2>.
- [32] P. Morais, A. Amorim, C. da Vieira Silva, T. Ribeiro, J. Costa Santos, H. Afonso Costa, V.C. da Silva, C.J. Santos, A.H. Costa, Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis, *J Genet.* 94 (2015) 509–512. <https://doi.org/10.1007/s12041-015-0552-7>.
- [33] F. Simão, H.A. Costa, C.V. da Silva, T. Ribeiro, M.J. Porto, J.C. Santos, A. Amorim, Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA, *Forensic Sci Int Genet.* 15 (2015) 33–38. <https://doi.org/10.1016/j.fsigen.2014.09.013>.
- [34] L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A.C. Wilson, African Populations and the Evolution of Human Mitochondrial DNA, *Science.* 253 (1991) 1503–1507. <https://doi.org/10.1126/science.1840702>.
- [35] M.H. Wolpoff, J. Hawks, R. Caspari, Multiregional, Not Multiple Origins, *Am J Phys Anthropol.* 112 (2000) 129–136. [https://doi.org/10.1002/\(SICI\)1096-8644\(200005\)112:1<129::AID-AJPA11>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1096-8644(200005)112:1<129::AID-AJPA11>3.0.CO;2-K).
- [36] I. Tattersall, Out of Africa Again... and Again?, *Sci Am.* 276 (1997) 60–67. <https://doi.org/10.1038/scientificamerican0503-38sp>.
- [37] M. Stoneking, H. Soodyall, Human evolution and the mitochondrial genome, *Curr Opin Genet Dev.* 6 (1996) 731–736. [https://doi.org/10.1016/s0959-437x\(96\)80028-1](https://doi.org/10.1016/s0959-437x(96)80028-1).

- [38] J.B. Stewart, P.F. Chinnery, The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease, *Nat Rev Genet.* 16 (2015) 530–542. <https://doi.org/10.1038/nrg3966>.
- [39] N. Howell, I. Kubacka, D.A. Mackey, How Rapidly Does the Human Mitochondrial Genome Evolve?, *Am. J. Hum. Genet.* 59 (1996) 501–509. <https://doi.org/8751850>.
- [40] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum Mutat.* 30 (2009) E386–394. <https://doi.org/10.1002/humu.20921>.
- [41] A. Rosa, A. Brehm, African human mtDNA phylogeography at-a-glance, *Journal of Anthropological Sciences.* 89 (2011) 25–58. <https://doi.org/10.4436/jass.89006>.
- [42] D.M. Behar, R. Villems, H. Soodyall, J. Blue-Smith, L. Pereira, E. Metspalu, R. Scozzari, H. Makkan, S. Tzur, D. Comas, J. Bertranpetit, L. Quintana-Murci, C. Tyler-Smith, R.S. Wells, S. Rosset, The Dawn of Human Matrilineal Diversity, *Am J Hum Genet.* 82 (2008) 1130–1140. <https://doi.org/10.1016/j.ajhg.2008.04.002>.
- [43] L. Quintana-Murci, O. Semino, H.-J. Bandelt, G. Passarino, K. Mcelreavey, & A.S. Santachiara-Benerecetti, Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa, *Nat Genet.* 23 (1999) 437–441. <https://doi.org/10.1038/70550>.
- [44] D.M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N.M. Silva, T. Kivisild, A. Torroni, R. Villems, A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root, *The American Journal of Human Genetics.* 90 (2012) 675–684. <https://doi.org/10.1016/j.ajhg.2012.03.002>.
- [45] A. Torroni, T.G. Schurr, M.F. Cabell, M.D. Brown, J. v Neel, M. Larsen, D.G. Smith, C.M. Vullo, D.C. Wallace, Asian Affinities and Continental Radiation of the Four Founding Native American mtDNAs, *Am. J. Hum. Genet.* 53 (1993) 563–590.
- [46] T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S.S. Papiha, H.-J. Bandelt, R. Villems, The emerging limbs and twigs of the East Asian mtDNA tree, *Mol. Biol. Evol.* 19 (2002) 1737–1751. <https://doi.org/10.1093/oxfordjournals.molbev.a003996>.
- [47] D.A. Bolnick, D.G. Smith, Unexpected Patterns of Mitochondrial DNA Variation among Native Americans from the Southeastern United States, *Am J Phys Anthropol.* 122 (2003) 336–354. <https://doi.org/10.1002/ajpa.10284>.
- [48] S. Lutz, H.J. Weisser, J. Heizmann, S. Pollak, mtDNA as a tool for identification of human remains Identification using mtDNA, *Int J Legal Med.* 109 (1996) 205–209. <https://doi.org/10.1007/BF01225519>.
- [49] M.R. Wilson, J.A. Dizinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, *Int J Legal Med.* 108 (1995) 68–74. <https://doi.org/10.1007/BF01369907>.

- [50] J.M. Butler, *Forensic DNA Typing: biology, technology, and genetics of STR markers.*, 2nd ed., Elsevier, 2005.
- [51] W. Goodwin, A. Linacre, P. Vanezis, The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims, *Electrophoresis*. 20 (1999) 1707–1711. [https://doi.org/10.1002/\(SICI\)1522-2683\(19990101\)20:8<1707::AID-ELPS1707>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1522-2683(19990101)20:8<1707::AID-ELPS1707>3.0.CO;2-F).
- [52] B.C. Leland Davis, Mitochondrial DNA: State of Tennessee v. Paul Ware, *GenePrint*. (1998) 6–7.
- [53] M.D. Coble, The identification of the Romanovs: Can we (finally) put the controversies to rest?, *Investig Genet*. 2 (2011). <https://doi.org/10.1186/2041-2223-2-20>.
- [54] A. Röhl, B. Brinkmann, L. Forster, P. Forster, An annotated mtDNA database, *Int J Legal Med*. 115 (2001) 29–39. <https://doi.org/10.1007/s004140100217>.
- [55] W. Parson, A. Dür, EMPOP-A forensic mtDNA database, *Forensic Sci Int Genet*. 1 (2007) 88–92. <https://doi.org/10.1016/j.fsigen.2007.01.018>.
- [56] E. Ruiz-Pesini, M.T. Lott, V. Procaccio, J.C. Poole, M.C. Brandon, D. Mishmar, C. Yi, J. Kreuziger, P. Baldi, D.C. Wallace, An enhanced MITOMAP with a global mtDNA mutational phylogeny, *Nucleic Acids Res*. 35 (2007) D823–D828. <https://doi.org/10.1093/nar/gkl927>.
- [57] M. Ingman, U. Gyllensten, mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences., *Nucleic Acids Res*. 34 (2006) D749–D751. <https://doi.org/10.1093/nar/gkj010>.
- [58] B. Zimmermann, A.W. Röck, A. Dür, W. Parson, Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software, *Croat Med J*. 55 (2014) 115–120. <https://doi.org/10.3325/cmj.2014.55.115>.
- [59] Applied Biosystems, Applied Biosystems® 3130/3130xl Genetic Analyzers, (2012).
- [60] Applied Biosystems, Applied Biosystems™ 3500/3500xl Genetic Analyzer, (2021).
- [61] AppliedBiosystems, SeqStudio™ Genetic Analyzer Instrument and Software User Guide, B.0 (2019).
- [62] D. Campos, F. Ribeiro, J. Guerreiro, L. Gabriel, R. Miranda, S. Machado, *História*, (2014). <https://projetonorte.wordpress.com/norte/historia/> (accessed January 26, 2023).
- [63] X. Jordana, A. Malgosa, B. Casté, C. Tornero, Lost in transition: the dietary shifts from Late Antiquity to the Early Middle Ages in the North Eastern Iberian Peninsula, *Archaeol Anthropol Sci*. 11 (2019) 3751–3763. <https://doi.org/10.1007/s12520-019-00777-9>.
- [64] L. Pereira, M.J. Prata, A. Amorim, Diversity of mtDNA lineages in Portugal : not a genetic edge of European variation, *Ann. Hum. Genet*. 64 (2000) 491–506.

- [65] S.L. Marques, A. Goios, A.M. Rocha, M.J. Prata, A. Amorim, L. Gusmão, C. Alves, L. Alvarez, Portuguese mitochondrial DNA genetic diversity - An update and a phylogenetic revision, *Forensic Sci Int Genet.* 15 (2015) 27–32. <https://doi.org/10.1016/j.fsigen.2014.10.004>.
- [66] E. Picard-Meyer, J. Barrat, F. Cliquet, Use of filter paper (FTA) technology for sampling, recovery and molecular characterisation of rabies viruses, *J Virol Methods.* 140 (2007) 174–182. <https://doi.org/10.1016/j.jviromet.2006.11.011>.
- [67] AppliedBiosystems, PrepFiler Express and AutoMate Express DNA extraction systems, n.d.
- [68] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *Biotechniques.* 10 (1991) 506–513.
- [69] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction, *Cold Spring Harb Symp Quant Biol.* 51 (1986) 263–273. <https://doi.org/10.1101/sqb.1986.051.01.032>.
- [70] Qiagen, QIAxcel<sup>®</sup> DNA Handbook. 5th edn, (2014). [www.qiagen.com](http://www.qiagen.com).
- [71] Applied Biosystems, ExoSAP-IT<sup>™</sup> PCR Product Cleanup, (2017).
- [72] Applied Biosystems, BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing Kit Protocol, (2002).
- [73] Applied Biosystems, BigDye XTerminator<sup>™</sup> Purification Kit USER GUIDE, (2020).
- [74] G. Tully, W. Bär, B. Brinkmann, A. Carracedo, P. Gill, N. Morling, W. Parson, P. Schneider, Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles, *Forensic Sci Int.* 124 (2001) 83–91. [https://doi.org/10.1016/s0379-0738\(01\)00573-4](https://doi.org/10.1016/s0379-0738(01)00573-4).
- [75] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol Ecol Resour.* 10 (2010) 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
- [76] J. Felsenstein, PHYLIP (Phylogeny Inference Package) version 3.695, Distributed by the Author, Department of Genome Sciences, University of Washington, Seattle. (2013). <http://evolution.gs.washington.edu/phylip/newicktree.html>.
- [77] R.D.M. Page, TreeView: an application to display phylogenetic trees on personal computers, *Comput. Appl. Biosci.* 12 (1996) 357–358. <https://doi.org/10.1093/bioinformatics/12.4.357>.
- [78] S.S. Shapiro, M.B. Wilk, H.J. Chen, A Comparative Study of Various Tests for Normality, *J Am Stat Assoc.* 63 (1968) 1343–1372. <https://doi.org/10.2307/2285889>.
- [79] A. Ross, V.L. Willson, Paired Samples T-Test, in: *Basic and Advanced Statistical Tests*, 2017: pp. 17–19. [https://doi.org/10.1007/978-94-6351-086-8\\_4](https://doi.org/10.1007/978-94-6351-086-8_4).

- [80] M. Sidstedt, J. Hedman, E.L. Romsos, L. Waitara, L. Wadsö, C.R. Steffen, P.M. Vallone, P. Rådström, Inhibition mechanisms of hemoglobin, immunoglobulin G, and whole blood in digital and real-time PCR, *Anal Bioanal Chem.* 410 (2018) 2569–2583. <https://doi.org/10.1007/s00216-018-0931-z>.
- [81] I.G. Wilson, Inhibition and facilitation of nucleic acid amplification, *Appl Environ Microbiol.* 63 (1997) 3741–3751. <https://doi.org/10.1128/aem.63.10.3741-3751.1997>.
- [82] B. Bär, B. Brinkmann, A. Budowle, P. Carracedo, M. Gill, P. Holland, W. Lincoln, N. Mayr, Morling Bjørnar, P.M. Olaisen, G. Schneider, M. Tully, W.J. Wilson, M. Wolfgang, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing., *Vox Sang.* 119 (2000) 79–85. [https://doi.org/10.1016/s0379-0738\(00\)00161-4](https://doi.org/10.1016/s0379-0738(00)00161-4).
- [83] A. Torroni, K. Huoponen, P. Francalacci, M. Petrozzi, L. Morelli, R. Scozzari, D. Obinu, M.-L. Savontaus, D.C. Wallace, Classification of European mtDNAs from an analysis of three European populations, *Genetics Society of America.* 144 (1996) 1835–1850. <https://doi.org/10.1093/genetics/144.4.1835>.
- [84] L. Alvarez, C. Santos, A. Ramos, R. Pratdesaba, P. Francalacci, M.P. Aluja, Mitochondrial DNA patterns in the Iberian Northern plateau: population dynamics and substructure of the Zamora province, *Am J Phys Anthropol.* 142 (2010) 531–539. <https://doi.org/10.1002/ajpa.21252>.
- [85] J.M. Larruga, P. Marrero, K.K. Abu-Amero, M. v. Golubenko, V.M. Cabrera, Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area, *BMC Evol Biol.* 17 (2017). <https://doi.org/10.1186/s12862-017-0964-5>.
- [86] M. Shamooun-Pour, M. Li, D.A. Merriwether, Rare human mitochondrial HV lineages spread from the Near East and Caucasus during post-LGM and Neolithic expansions, *Sci Rep.* 9 (2019). <https://doi.org/10.1038/s41598-019-48596-1>.
- [87] A. Non, Analyses of genetic data within an interdisciplinary framework to investigate recent human evolutionary history and complex disease, 2010.
- [88] C. Turchi, L. Buscemi, E. Giacchino, V. Onofri, L. Fendt, W. Parson, A. Tagliabracci, Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: An enrichment of forensic mtDNA databases with Northern Africa data, *Forensic Sci Int Genet.* 3 (2009) 166–172. <https://doi.org/10.1016/j.fsigen.2009.01.014>.
- [89] R. Aboukhalid, K. Sturk-Andreaggi, M. Bouabdellah, D. Squalli, J.A. Irwin, S. Amzazi, Mitochondrial DNA control region variation from samples of the Moroccan population, *Int J Legal Med.* 127 (2013) 757–759. <https://doi.org/10.1007/s00414-012-0813-4>.
- [90] A. Brandstätter, B. Zimmermann, J. Wagner, T. Göbel, A.W. Röck, A. Salas, A. Carracedo, W. Parson, Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability

- in Central Europe and Middle East, *BMC Evol Biol.* 8 (2008).  
<https://doi.org/10.1186/1471-2148-8-191>.
- [91] A. Rakha, Fatima, M.S. Peng, A. Adan, R. Bi, M. Yasmin, Y.G. Yao, mtDNA sequence diversity of Hazara ethnic group from Pakistan, *Forensic Sci Int Genet.* 30 (2017) e1–e5.  
<https://doi.org/10.1016/j.fsigen.2017.07.004>.
- [92] L. Prieto, B. Zimmermann, A. Goios, A. Rodriguez-Monge, G.G. Paneto, C. Alves, A. Alonso, C. Fridman, S. Cardoso, G. Lima, M.J. Anjos, M.R. Whittle, M. Montesino, R.M.B. Cicarelli, A.M. Rocha, C. Albarrán, M.M. de Pancorbo, M.F. Pinheiro, M. Carvalho, D.R. Sumita, W. Parson, The GHEP-EMPOP collaboration on mtDNA population data - A new resource for forensic casework, *Forensic Sci Int Genet.* 5 (2011) 146–151.  
<https://doi.org/10.1016/j.fsigen.2010.10.013>.
- [93] P. Brotherton, W. Haak, J. Templeton, G. Brandt, J. Soubrier, C. Jane Adler, S.M. Richards, C. der Sarkissian, R. Ganslmeier, S. Friederich, V. Dresely, M. van Oven, R. Kenyon, M.B. van der Hoek, J. Korlach, K. Luong, S.Y.W. Ho, L. Quintana-Murci, D.M. Behar, H. Meller, K.W. Alt, A. Cooper, S. Adhikarla, A.K. Ganesh Prasad, R. Pitchappan, A. Varatharajan Santhakumari, E. Balanovska, O. Balanovsky, J. Bertranpetit, D. Comas, B. Martínez-Cruz, M. Melé, A.C. Clarke, E.A. Matisoo-Smith, M.C. Dulik, J.B. Gaieski, A.C. Owings, T.G. Schurr, M.G. Vilar, A. Hobbs, H. Soodyall, A. Javed, L. Parida, D.E. Platt, A.K. Royyuru, L. Jin, S. Li, M.E. Kaplan, N.C. Merchant, R. John Mitchell, C. Renfrew, D.R. Lacerda, F.R. Santos, D.F. Soria Hernanz, R. Spencer Wells, P. Swamikrishnan, C. Tyler-Smith, P. Paulo Vieira, J.S. Ziegler, Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans, *Nat Commun.* 4 (2013) 1764. <https://doi.org/10.1038/ncomms2656>.
- [94] M. Richards, V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F. Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Gölge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F.C. Calí, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. di Rienzo, A. Novelletto, A. Oppenheim, S. Nørby, N. Al-Zaheri, S. Santachiara-Benerecetti, R. Scozzari, A. Torroni, H.-J. Bandelt, Tracing European Founder Lineages in the Near Eastern mtDNA Pool, *Am. J. Hum. Genet.* 67 (2000) 1251–1276.
- [95] A. Achilli, C. Rengo, C. Magri, V. Battaglia, A. Olivieri, R. Scozzari, F. Cruciani, M. Zeviani, E. Briem, V. Carelli, P. Moral, J.-M. Dugoujon, U. Roostalu, E.-L. Loogväli, T. Kivisild, H.-J. Bandelt, M. Richards, R. Villems, A.S. Santachiara-Benerecetti, O. Semino, A. Torroni, The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool, *Am. J. Hum. Genet.* 75 (2004) 910–918. <https://doi.org/10.1086/425590>.
- [96] A.W.R. Whittle, Vicki. Cummings, British Academy., *Going over: the Mesolithic-Neolithic transition in north-west Europe*, Published for the British Academy by Oxford University Press, 2007.

- [97] A. Salas, M. Richards, M.-V. Lareu, R. Scozzari, A. Coppa, A. Torroni, V. Macaulay, A. Ngel Carracedo, The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade, *Am. J. Hum. Genet.* 74 (2004) 454–465. <https://doi.org/10.1086/382194>.
- [98] M.K. Gonder, H.M. Mortensen, F.A. Reed, A. de Sousa, S.A. Tishkoff, Whole-mtDNA genome sequence analysis of ancient african lineages, *Mol Biol Evol.* 24 (2007) 757–768. <https://doi.org/10.1093/molbev/msl209>.
- [99] M. Silva, F. Alshamali, P. Silva, C. Carrilho, F. Mandlate, M. Jesus Trovoadá, V. Cerny, L. Pereira, P. Soares, 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2, *Sci Rep.* 5 (2015). <https://doi.org/10.1038/srep12526>.
- [100] A. Salas, D. Comas, M.V. Lareu, J. Bertranpetit, A. Carracedo, mtDNA analysis of the Galician population: a genetic edge of European variation, *European Journal of Human Genetics.* 6 (1998) 365–375. <https://doi.org/10.1038/sj.ejhg.5200202>.
- [101] K. Thangaraj, G. Chaubey, V.K. Singh, A. Vanniarajan, I. Thanseem, A.G. Reddy, L. Singh, In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup “M” in India, *BMC Genomics.* 7 (2006) 151. <https://doi.org/10.1186/1471-2164-7-151>.
- [102] I. Mendizabal, C. Valente, A. Gusmão, C. Alves, V. Gomes, A. Goios, W. Parson, F. Calafell, L. Alvarez, A. Amorim, L. Gusmão, D. Comas, M.J. Prata, Reconstructing the Indian origin and dispersal of the european Roma: A maternal genetic perspective, *PLoS One.* 6 (2011) e15988. <https://doi.org/10.1371/journal.pone.0015988>.
- [103] M. Pala, A. Olivieri, A. Achilli, M. Accetturo, E. Metspalu, M. Reidla, E. Tamm, M. Karmin, T. Reisberg, B.H. Kashani, U.A. Perego, V. Carossa, F. Gandini, J.B. Pereira, P. Soares, N. Angerhofer, S. Rychkov, N. Al-Zahery, V. Carelli, M.H. Sanati, M. Houshmand, J. Hatina, V. MacAulay, L. Pereira, S.R. Woodward, W. Davies, C. Gamble, D. Baird, O. Semino, R. Villems, A. Torroni, M.B. Richards, Mitochondrial DNA signals of late glacial recolonization of europe from near eastern refugia, *Am J Hum Genet.* 90 (2012) 915–924. <https://doi.org/10.1016/j.ajhg.2012.04.003>.
- [104] J.C. Rando, F. Pinto, A.M. Gonzalez, M. Hernandez, J.M. Larruga, V.M. Cabrera, H.J. Bandelt, Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations, *Ann Hum Genet.* 62 (1998) 531–550. <https://doi.org/10.1046/j.1469-1809.1998.6260531.x>.
- [105] S. Finnilä, L.E. Hassinen, L. Ala-Kokko, K. Majamaa, Phylogenetic network of the mtDNA haplogroup U in Northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis, *Am. J. Hum. Genet.* 66 (2000) 1017–1026. <https://doi.org/10.1086/302802>.
- [106] T. Kivisild, M.J. Bamshad, K. Kaldma, M. Metspalu, E. Metspalu, M. Reidla, S. Laos, J. Parik, W.S. Watkins, M.E. Dixon, S.S. Papiha, S.S. Mastana, M.R. Mir, V. Ferak, R. Villems, Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages, *Current Biology.* 9 (1999) 1331–1334. [https://doi.org/10.1016/s0960-9822\(00\)80057-3](https://doi.org/10.1016/s0960-9822(00)80057-3).

- [107] A. Torroni, M. Richards, V. Macaulay, P. Forster, R. Villems, S. Nørby, M.-L. Savontaus, K. Huoponen, R. Scozzari, H.-J. Bandelt, mtDNA Haplogroups and Frequency Patterns in Europe, *Am J Hum Genet.* 66 (2000) 1173–1177. <https://doi.org/10.1086/302789>.
- [108] P. Endicott, J.J. Sanchez, E. Metspalu, D.M. Behar, T. Kivisild, The unresolved location of Ötzi's mtDNA within haplogroup K, *Am J Phys Anthropol.* 132 (2007) 590–591. <https://doi.org/10.1002/ajpa.20501>.
- [109] V. Coia, G. Cipollini, P. Anagnostou, F. Maixner, C. Battaglia, F. Brisighelli, A. Gómez-Carballa, G. Destro Bisol, A. Salas, A. Zink, Whole mitochondrial DNA sequencing in Alpine populations and the genetic history of the Neolithic Tyrolean Iceman, *Sci Rep.* 6 (2016) 18932. <https://doi.org/10.1038/srep18932>.
- [110] M. Hay, Haplogroup W (mtDNA), (2020). [https://www.eupedia.com/europe/Haplogroup\\_W\\_mtDNA.shtml](https://www.eupedia.com/europe/Haplogroup_W_mtDNA.shtml) (accessed February 24, 2023).
- [111] M.D. Brown, S.H. Hosseini, A. Torroni, H.J. Bandelt, J.C. Allen, T.G. Schurr, R. Scozzari, F. Cruciani, D.C. Wallace, mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America?, *Am. J. Hum. Genet.* 63 (1998) 1852–1861. <https://doi.org/10.1086/302155>.
- [112] M. Reidla, T. Kivisild, E. Metspalu, K. Kaldma, K. Tambets, H.-V. Tolk, J. Parik, E.-L. Loogväli, M. Derenko, B. Malyarchuk, M. Bermisheva, S. Zhadanov, E. Pennarun, M. Gubina, M. Golubenko, L. Damba, S. Fedorova, V. Gusar, E. Grechanina, I. Mikerezi, J.-P. Moisan, A. Chaventré, E. Khusnutdinova, L. Osipova, V. Stepanov, M. Voevoda, A. Achilli, C. Rengo, O. Rickards, G.F. de Stefano, S. Papiha, L. Beckman, N. Howell, A. Torroni, R. Villems, Origin and Diffusion of mtDNA Haplogroup X, *Am. J. Hum. Genet.* 73 (2003) 1178–1190. <https://doi.org/10.1086/379380>.
- [113] N. Maca-Meyer, P. Sánchez-Velasco, C. Flores, J.-M. Larruga, A.-M. González, A. Oterino, F. Leyva-Cobán, Y Chromosome and Mitochondrial DNA Characterization of Pasiegos, a Human Isolate from Cantabria (Spain), *Ann Hum Genet.* 67 (2003) 329–339. <https://doi.org/10.1046/j.1469-1809.2003.00045.x>.
- [114] L. Pereira, V. Černý, M. Cerezo, N.M. Silva, M. Hájek, A. Vašíková, M. Kujanová, R. Brdika, A. Salas, Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel, *European Journal of Human Genetics.* 18 (2010) 915–923. <https://doi.org/10.1038/ejhg.2010.21>.
- [115] K. Fadhlou-Zid, S. Plaza, F. Calafell, M. ben Amor, D. Comas, A.B. el Gaaied, Mitochondrial DNA heterogeneity in Tunisian Berbers, *Ann Hum Genet.* 68 (2004) 222–233. <https://doi.org/10.1046/j.1529-8817.2004.00096.x>.
- [116] M. Proença de Campos, H. Afonso Costa, C. Vieira da Silva, V. Bogas, T. Ribeiro, M.J. Porto, A. Amorim, The immigrant population from Mozambique in Lisbon: Updated mitochondrial DNA portrait, *Forensic Sci Int Genet Suppl Ser.* 6 (2017) 298–300. <https://doi.org/10.1016/j.fsigss.2017.09.134>.

- [117] M. Marcelino, Mitochondrial DNA characterization of Brazilian immigrant population living in Lisboa. (Poster), III Internacional Meeting of the Portuguese Society of Genetics. (2022).
- [118] M. Eduardoff, G. Huber, B. Bayer, D. Schmid, K. Anslinger, T. Göbel, B. Zimmermann, P.M. Schneider, A.W. Röck, W. Parson, Mass spectrometric base composition profiling: Implications for forensic mtDNA databasing, *Forensic Sci Int Genet.* 7 (2013) 587–592. <https://doi.org/10.1016/j.fsigen.2013.05.007>.
- [119] B. Zimmermann, A. Röck, G. Huber, T. Krämer, P.M. Schneider, W. Parson, Application of a west Eurasian-specific filter for quasi-median network analysis: Sharpening the blade for mtDNA error detection, *Forensic Sci Int Genet.* 5 (2011) 133–137. <https://doi.org/10.1016/j.fsigen.2010.10.003>.
- [120] R. Gonçalves, A. Freitas, M. Branco, A. Rosa, A.T. Fernandes, A. Brehm, Y-chromosome lineages from Portugal, Madeira and Açores record elements of Sephardim and Berber ancestry, *Annals Human Genetics.* 69 (2005) 443–456. <https://doi.org/10.1016/j.ics.2005.09.099>.
- [121] J.L. Rodriguez-Flores, K. Fakhro, F. Agosto-Perez, M.D. Ramstetter, L. Arbiza, T.L. Vincent, A. Robay, J.A. Malek, K. Suhre, L. Chouchane, R. Badii, A.A.N. Al-Marri, C.A. Khalil, M. Zirie, A. Jayyousi, J. Salit, A. Keinan, A.G. Clark, R.G. Crystal, J.G. Mezey, Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations, *Genome Res.* 26 (2016) 151–162. <https://doi.org/10.1101/gr.191478.115>.
- [122] K.K. Abu-Amero, J.M. Larruga, V.M. Cabrera, A.M. González, Mitochondrial DNA structure in the Arabian Peninsula, *BMC Evol Biol.* 8 (2008). <https://doi.org/10.1186/1471-2148-8-45>.
- [123] A. Inácio, H.A. Costa, C.V. da Silva, T. Ribeiro, M.J. Porto, J.C. Santos, G. Igrejas, A. Amorim, Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa, *Int J Legal Med.* 131 (2017) 657–660. <https://doi.org/10.1007/s00414-016-1484-3>.
- [124] S. Belezã, L. Gusmão, A. Amorim, C. Angel, A. Salas, The genetic legacy of western Bantu migrations, *Hum Genet.* 117 (2005) 366–375. <https://doi.org/10.1007/s00439-005>.

## 6. Anexos

Anexo 6.1 - Sequência de cada uma das regiões do mtDNA dos controlos positivos. rCRS - Sem alterações relativamente à sequência padrão.

	HV1	HV2	HV3	RV
007	16129A; 16223T; 16391A	73G; 152C; 199C; 204C; 207A; 250C; 263G; 315.1C	460C	16519C; 73G; 152C; 199C; 204C; 207A
2800M	rCRS	152C; 263G; 315.1C	477C	16519C; 152C; 263G
K562	16126C; 16294T; 16296T; 16324C	73G; 263G; 310D; 311D; 312D; 313D	rCRS	16519C; 73G; 263G
9974A	16311C	93G; 195C; 214G; 263G; 309.1C; 309.2C; 315.1C	rCRS	16519C; 152C; 263G
9948	16126C; 16163G; 16186T; 16189C; 16294T	73G; 152C; 195C; 263G; 309.1C; 315.1C	rCRS	16519C; 73G; 152C; 195C; 263G

Anexo 6.2 - Amostras animais utilizadas nos testes de especificidade.

E1 e E2	<i>Equus caballus</i> , sexo feminino	M9-2018	<i>Ovis aries</i>
E3 a E8	<i>Equus caballus</i> , sexo masculino	M9-2017-1 (915-1)	<i>Mustela putorius</i>
S1 a S4	<i>Sus scrofa domesticus</i> , sexo masculino	M9-2017-2 (915-1-C1)	<i>Mustela putorius</i>
S5 a S8	<i>Sus scrofa domesticus</i> , sexo feminino	M9-2016-1 (674-1)	<i>Oryctolagus cuniculus</i>
C1	<i>Canis lupus familiaris</i> , sexo masculino	M9-2016-2 (674-1-C1)	<i>Oryctolagus cuniculus</i>
C2	<i>Canis lupus familiaris</i> , sexo feminino	Coelho 1 e 2	<i>Oryctolagus cuniculus</i>
G1	<i>Felis catus</i> , sexo feminino	Galinha 1 e 2	
G2 e G3	<i>Felis catus</i> , sexo masculino	Frango 1	
M9-2020	<i>Sus scrofa</i>		

### Testes de Normalidade

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
dif_Quantidade	,247	8	,166	,867	8	,142

a. Correlação de Significância de Lilliefors

Anexo 6.3 - Tabela retirada do software IBM® SPSS® com os valores obtidos no teste de normalidade.

### Teste de amostras emparelhadas

Par	Diferenças emparelhadas	Diferenças emparelhadas							Significância	
		Média	Desvio Padrão	Erro de média padrão	95% Intervalo de Confiança da Diferença		t	df	Unilateral p	Bilateral p
					Inferior	Superior				
Par 1	QuantidadeDNACabeloArrancado - QuantidadeDNACabeloNã oArrancado	1,44160716	1,28655566	,45486611	,36601971	2,51719461	3,169	7	,008	,016

Anexo 6.3 - Tabela retirada do software IBM® SPSS® com os valores obtidos no teste T para amostras emparelhadas.

### Estatísticas de amostras emparelhadas

		Média	N	Desvio Padrão	Erro de média padrão
Par 1	QuantidadeDNACabeloArrancado	1,6615453	8	1,63347523	,57752070
	QuantidadeDNACabeloNã oArrancado	,2199382	8	,52296079	,18489456

Anexo 6.5 - Tabela retirada do software IBM® SPSS® com os valores obtidos para cada uma das amostras de cabelos estudadas.

Anexo 4.6 - Sequência das regiões HV1, HV2, HV3 e RV de amostras de cabelo. rCRS - Sem alterações relativamente à sequência padrão

Amostras	HV1	HV2	HV3	RV
B1R/B1	16239T	152C; 263G; 309.1C; 315.1C	rCRS	16519C; 152C; 263G
B2R/B2	16239T	152C; 263G; 309.1C; 315.1C	rCRS	165C; 152C
B3R/B3	16239T			16519C; 152C; 263G
B4R/B4	16239T	152C; 263G; 309.1C; 315.1C	rCRS	16519C; 152C; 263G
B5R/B5	16239T	152C; 263G; 309.1C; 315.1C	rCRS	16519C; 152C; 263G
B6R/B6	16111T; 16180G; 16192T; 16223T; 16311C	73G; 150T; 263G; 315.1C		
B7R/B7	16362C	239C; 263G; 315.1C		16482G
B8R/B8	16239T	152C; 263G; 309.1C; 315.1C	rCRS	16519C; 152C; 263G
B9R/B9	16069T; 16126C; 16145A; 16231C; 16261T	73G; 150T; 152C; 215G; 263G; 295T; 315.1C; 319C		73G; 150T; 152C; 215G; 263G
B10R/B10	16069T; 16126C; 16145A; 16231C; 16261T	73G; 150T; 152C; 215G; 263G; 295T; 315.1C; 319C;		73G; 150T; 152C; 215G; 263G

Anexo 6.7 - Sequência da região controlo total de amostras de manchas de sangue extraídas por PrepFiler™.

	Região controlo total
A1	16239T; 16519C; 152C; 263G; 309.1C; 315.1C
A2	16239T; 16519C; 152C; 263G; 309.1C; 315.1C
A3	16239T; 16519C; 152C; 263G; 309.1C; 315.1C
A4	16362C; 16482G; 239C; 263G; 315.1C
A5	16093C; 16126C; 16239T; 16294T; 16296T; 16304C; 16335G; 16519C; 73G; 151T; 263G; 309.1C; 315.1C
A6	16126C; 16172C; 16294T; 16296T; 16304C; 16519C; 73G; 309.1C; 315.1C
A7	16356C; 16519C; 73G; 195C; 263G; 315.1C; 499A
A8	16189C; 16223T; 16255A; 16278T; 16519C; 73G; 146C; 153G; 195C; 225A; 226C; 263G; 309.1C; 315.1C
A9	16224C; 16311C; 16519C; 73G; 263G; 315.1C; 497T; 523.1C; 523.2A; 523.3C; 523.4A
A10	16169T; 16519C; 263G; 315.1C
A11	16304C; 16356C; 16519C; 73G; 195C; 263G; 315.1C; 499A
A12	16239T; 16519C; 152C; 263G; 309.1C; 315.1C
A13	16111T; 16180G; 16192T; 16223T; 16311C; 16519C; 73G; 150T; 263G; 315.1C; 489C
A14	16304C; 16519C; 263G; 309.1C; 315.1C; 456T
A15	16069T; 16126C; 16145A; 16231C; 16261T; 73G; 150T; 152C; 215G; 263G; 295T; 315.1C; 319C; 489C; 513A
A16	16192T; 16270T; 73G; 150T; 195C; 199C; 263G; 315.1C; 523.1C; 523.2A
A17	16183C; 16189C; 16270T; 16362C; 16519C; 73G; 150T; 263G; 315.1C
A18	16126C; 16519C; 152C; 195C; 263G; 309.1C; 315.1C
A19	16069T; 16126C; 16145A; 16231C; 16261T; 73G; 150T; 152C; 215G; 263G; 295T; 315.1C; 319C; 489C; 513A
A20	16069T; 16126C; 16145A; 16231C; 16261T; 73G; 150T; 152C; 215G; 263G; 295T; 315.1C; 319C; 489C; 513A

Anexo 6.8 - Resultados obtidos para os 87 indivíduos da Região do Norte de Portugal. Código do haplótipo, haplogrupo a que pertence, número de alterações polimórficas relativamente à rCRS e haplótipo.

Amostra	Haplogrupo	Polimorfismos	Haplótipo
NOR001	J1c1	12	16069T 16126C 16256T 73G 185A 228A 263G 295T 315.1C 462T 482C 489C
NOR002	L3h1b2	19	16129A 16172C 16174T 16192T 16218T 16223T 16256A 16311C 16362C 73G 151T 152C 189C 195C 263G 294C 315.1C 523DEL 524DEL
NOR003	R0	3	16519C 263G 315.1C
NOR004	T2b3+151	11	16126C 16271C 16294T 16296T 16304C 16519C 73G 151T 263G 309.1C 315.1C
NOR005	T1	9	16126C 16163G 16189C 16294T 16519C 73G 263G 309.1C 315.1C
NOR006	T2b6	10	16126C 16294T 16296T 16304C 16519C 73G 263G 309.1C 315.1C 458T
NOR007	V3a	6	16240G 16298C 72C 263G 309.1C 315.1C
NOR008	H14b1	7	16126C 16213R 16519C 152C 263G 309.1C 315.1C
NOR009	J1c	11	16069T 16126C 16311C 73G 185A 228A 263G 295T 315.1C 462T 489C
NOR010	H1a3	6	16051G 16162G 16519C 73G 263G 315.1C
NOR011	K1c2	11	16224C 16311C 16320T 16519C 73G 146C 152C 263G 309.1C 315.1C 498DEL
NOR012	R0	4	16519C 263G 309.1C 315.1C
NOR013	H	6	16239T 16519C 152C 263G 309.1C 315.1C
NOR014	L3f1b1	11	16209C 16223T 16292T 16295T 16311C 16519C 73G 189G 200G 263G 315.1C
NOR015	U2e1	16	16051G 16129C 16183C 16189C 16362C 16519C 73G 152C 217C 263G 309.1C 315.1C 340T 508G 524.1A 524.2C
NOR016	H	5	16129A 16260Y 16519C 263G 315.1C
NOR017	HV0	4	16298C 72C 263G 315.1C
NOR018	HV	4	16172C 16311C 263G 315.1C
NOR019	H1aj1a	9	16172C 16192T 16456A 16519C 207A 263G 309.1C 309.2C 315.1C
NOR020	H	6	16235G 16291T 93G 263G 309.1C 315.1C
NOR021	K1a+150	10	16224C 16311C 16519C 73G 150T 263G 315.1C 497T 524.1A 524.2C
NOR022	L3f	9	16209C 16223T 16311C 16519C 73G 195C 200G 263G 315.1C
NOR023	W6	14	16192T 16223T 16292T 16325C 16519C 73G 189G 194T 195C 204C 207A 263G 309.1C 315.1C
NOR024	U2e1	15	16129C 16183C 16189C 16362C 16519C 73G 152C 200G 217C 263G 315.1C 340T 508G 524.1A 524.2C
NOR025	K1a	13	16093C 16224C 16311C 16519C 73G 150T 195C 263G 309.1C 315.1C 497T 524.1A 524.2C
NOR026	U4c1	9	16179T 16356C 16519C 73G 195C 263G 309.1C 315.1C 499A
NOR027	R0	4	16093C 16519C 263G 315.1C
NOR028	L3f	13	16209C 16223T 16311C 16519C 73G 152C 189G 200G 263G 309.1C 315.1C 523DEL 524DEL
NOR029	K1a	11	16224C 16234T 16311C 16356C 16519C 73G 263G 292.1A 292.2T 315.1C 497T
NOR030	R0	4	16362Y 16519C 263G 315.1C
NOR031	U5b1+16189	12	16093C 16183C 16187T 16189C 16192T 16270T 16519C 73G 150T 195C 263G 315.1C
NOR032	R0	4	16320A 16519C 263G 315.1C

Anexo 6.8 –(Continuação)

Amostra	Haplogrupo	Polimorfismos	Haplótipo
NOR033	R0	6	16519C 185A 263G 309.1C 309.2C 315.1C
NOR034	H5	6	16304C 16519C 263G 309.1C 315.1C 456T
NOR035	R0	4	16189C 16519C 263G 315.1C
NOR036	T1a9	12	16126C 16163G 16186T 16189C 16249C 16294T 16311C 16519C 73G 263G 309.1C 315.1C
NOR038	R30a1b	6	16209C 16519C 73G 263G 309.1C 315.1C
NOR039	W+194	15	16223T 16292T 16318G 16519C 73G 189G 194T 195C 204C 207A 263G 309.1C 315.1C 523DEL 524DEL
NOR040	H1C	6	16519C 263G 315.1C 477C 523DEL 524DEL
NOR041	R0	3	16519C 263G 315.1C
NOR042	R0a+60.1T	13	16126C 16264T 16362C 58C 60.1T 64T 263G 309.1C 309.2C 315.1C 573.1C 573.2C 573.3C
NOR043	HV	5	146C 152C 263G 309.1C 315.1C
NOR044	T2e	12	16126C 16153A 16294T 16296T 16519C 73G 150T 199C 263G 315.1C 523DEL 524DEL
NOR046	L1b1a+189	24	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16362C 16400T 16519C 73G 152C 182T 185T 189G 195C 247A 263G 315.1C 357G 523DEL 524DEL
NOR047	L1b1a16	22	16126C 16187T 16189C 16215T 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185C 195C 247A 263G 315.1C 357G 523DEL 524DEL
NOR048	J1c	13	16069T 16126C 16192T 16519C 73G 185A 188G 228A 263G 295T 315.1C 462T 489C
NOR049	K1a	10	16093C 16224C 16290T 16311C 16519C 73G 263G 309.1C 315.1C 497T
NOR050	T1a+152	12	16126C 16163G 16186T 16189C 16294T 16519C 73G 152C 195C 263G 309.1C 315.1C
NOR051	L3f1b	11	16209C 16223T 16311C 16519C 73G 189G 200G 263G 315.1C 523DEL 524DEL
NOR052	K2	8	16224C 16311C 16519C 73G 146C 152C 263G 315.1C
NOR053	R0	3	16519C 263G 315.1C
NOR054	R0	5	16519C 152C 263G 309.1C 315.1C
NOR055	T2b33	10	16126C 16266T 16294T 16304C 16519C 73G 263G 309.1C 315.1C 385G
NOR056	R	11	16278T 16311C 16519C 73G 263G 309.1C 309.2C 315.1C 497T 523DEL 524DEL
NOR057	U4a3a	18	16265G 16356C 16362C 16519C 73G 195C 247A 263G 315.1C 499A 524.1A 524.2C 524.3A 524.4C 524.5A 524.6C 524.7A 524.8C
NOR058	V16	8	16298C 16301T 72C 227G 263G 309.1C 309.2C 315.1C
NOR059	X3a	14	16182C 16183C 16189C 16223T 16278T 16519C 73G 146C 153G 256T 263G 309.1C 309.2C 315.1C
NOR060	R	9	16148T 16261T 16390A 73G 94A 150T 263G 309.1C 315.1C
NOR061	U5a1	7	16189C 16256T 16270T 16399G 73G 263G 315.1C
NOR062	R0	4	16316G 16519C 263G 315.1C
NOR063	J1c1	12	16069T 16126C 16172C 73G 185A 228A 263G 295T 315.1C 462T 482C 489C
NOR064	HV	10	16311C 16356C 57C 60.1T 152C 263G 309.1C 315.1C 523DEL 524DEL
NOR065	M5a1b	12	16129A 16223T 16291T 16298C 16519C 73G 263G 309.1C 315.1C 489C 524.1A 524.2C
NOR066	R0	3	16519C 263G 315.1C
NOR067	H1q1	5	16209C 16519C 263G 309.1C 315.1C
NOR069	R0	4	16320A 16519C 263G 315.1C
NOR070	H1q3	5	16037G 16519C 263G 315.1C 390G

Anexo 6.8 –(Continuação)

Amostra	Haplogrupo	Polimorfismos	Haplótipo
NOR071	H5	7	16304C 263G 309.1C 315.1C 456T 523DEL 524DEL
NOR072	J1c3e	13	16069T 16126C 16256T 16390A 73G 185A 228A 263G 295T 309.1C 315.1C 462T 489C
NOR073	U5b	8	16189C 16270T 73G 150T 263G 315.1C 524.1A 524.2C
NOR074	R	4	73G 150T 263G 315.1C
NOR075	T2	9	16126C 16256T 16294T 16296T 16519C 73G 263G 309.1C 315.1C
NOR076	R0	3	16519C 263G 315.1C
NOR077	H1c1	6	16263C 16519C 263G 309.1C 315.1C 477C
NOR078	HV	5	16519C 150T 263G 309.1C 315.1C
NOR079	J1c2	16	16069T 16126C 16291T 16519C 73G 146C 185A 188G 222T 228A 235G 263G 295T 315.1C 462T 489C
NOR080	R0	5	16519C 146C 263G 309.1C 315.1C
NOR081	HV0	7	16298C 72C 195C 263G 309.1C 315.1C 417C
NOR082	T1a9	10	16126C 16163G 16186T 16189C 16249C 16294T 16311C 16519C 73G 263G
NOR083	X2+225	14	16183C 16189C 16223T 16260T 16278T 16519C 73G 153G 195C 225A 226C 263G 309.1C 315.1C
NOR084	HV0	5	16298C 72C 263G 309.1C 315.1C
NOR085	H1b3	9	16075C 16178C 16189C 16356C 16519C 263G 309.1C 309.2C 315.1C
NOR086	HV	7	16311C 152C 263G 309.1C 315.1C 523DEL 524DEL
NOR087	J2a1a1	15	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 215G 263G 295T 315.1C 319C 489C 513A