



Escola Superior de Saúde Egas Moniz

**Estudo do DNA mitocondrial de um grupo de
imigrantes oriundos de Cabo Verde residentes
em Lisboa**

**Contributo para o conhecimento da origem e evolução da população de um
território integrado no ex-Império Colonial Português durante 500 anos**

Paulo Jorge Sampaio Morais

**Dissertação de Mestrado
Lisboa, 2013**



Mestrado em Biologia Molecular em Saúde

Título: Estudo do DNA mitocondrial de um grupo de imigrantes oriundos de Cabo Verde residentes em Lisboa: Contributo para o conhecimento da origem e evolução da população de um território integrado no ex-Império Colonial Português durante 500 anos

Dissertação apresentada à Escola Superior de Saúde Egas Moniz com vista à obtenção do grau de Mestre em Biologia Molecular em Saúde

Autor: Paulo Jorge Sampaio Morais

Orientadores: Especialista Superior de Medicina Legal António Amorim e Especialista Superior de Medicina Legal Heloísa Afonso Costa

Investigação laboratorial realizada no Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.

Data: Setembro 2013

Agradecimentos

Ao meu orientador, Professor António Amorim, por ter aceite a orientação da presente Dissertação de Mestrado, por estar sempre disponível para qualquer assunto, dúvida e correção, e pela enorme ajuda na elaboração da tese.

À minha orientadora, Mestre Heloísa Afonso Costa, por ter aceite a coorientação da presente Dissertação de Mestrado, por me ter transmitido todo o conhecimento da área e por contribuir sempre com uma boa dose de apoio e otimismo.

Ao Rodolfo Santos, por me ter ambientado ao contexto laboratorial e auxiliado nos primeiros passos a nível técnico. À Dr.^a Manuela Marques, pela ajuda na obtenção de muitos dos artigos utilizados na elaboração da tese.

Ao Sr. Prof. Doutor Jorge Costa Santos, por permitir que realizasse o meu projeto nas instalações do INMLCF.

À Dr.^a. Rosa Espinheira e a todo o pessoal do serviço de Genética e Biologia Forense do INMLCF, por toda a ajuda técnica, simpatia e disponibilidade.

À coordenação do Mestrado em Biologia Molecular em Saúde, Prof. Dr.^a. Alexandra Maia e Silva, por ter aceite este projeto.

À Sara, por me fazer acreditar e lutar pelos meus objetivos, obrigada pelo amor e carinho!

Aos meus pais, pelo incentivo e apoio constante, pelo suporte financeiro e emocional.

Aos restantes colegas da 3^a Edição do Mestrado de Biologia Molecular em Saúde, pelo convívio e amizade.

Resumo

O DNA mitocondrial apresenta características particulares, tais como um elevado número de cópias por célula, herança uniparental materna, elevada taxa de mutação e ausência de recombinação, que permitem a sua utilização na clarificação da origem e evolução das populações humanas. O objetivo deste estudo foi obter informação sobre a origem, em termos genéticos, da população imigrante de Cabo Verde residente atualmente em Lisboa. Para tal procedeu-se à construção de uma base de dados de DNA mitocondrial de forma a caracterizar a diversidade genética de 103 imigrantes oriundos de Cabo Verde, residentes em Lisboa.

A região controlo do DNA mitocondrial foi amplificada e sequenciada, entre as posições 16024 e 576, utilizando dois pares de primers (L15997/H016 e L1655/H599). As sequências obtidas foram inseridas na base de dados de DNA mitocondrial com maior relevância na área das Ciências Forenses, a *EDNAP Forensic mtDNA Population Database* - EMPOP.

A análise da região controlo revelou elevada variabilidade genética, com elevada frequência de haplótipos únicos. A maioria das sequências de DNA mitocondrial corresponde a haplogrupos característicos de populações africanas. Uma pequena minoria corresponde a haplogrupos euroasiáticos. Os resultados obtidos são semelhantes aos resultados alcançados em estudos anteriores sobre a origem e evolução da população de Cabo Verde, o que sugere que a composição genética da população imigrante de Cabo Verde a residir atualmente em Lisboa é representativa da composição genética do arquipélago. Os resultados obtidos não contradizem a versão histórica da colonização do arquipélago de Cabo Verde, segundo a qual a referida colonização envolveu escravas africanas e indivíduos portugueses do sexo masculino, mobilizados à altura para a ex-colónia.

Palavras chave: DNA Mitocondrial; Cabo Verde; Lisboa; Genética populacional

Abstract

Mitochondrial DNA presents peculiar characteristics that promote its use in the clarification of the origin and evolution of human populations, such as maternal inheritance, high mutation rate and absence of recombination. The aim of this study was to obtain information about the origin of the immigrant population of Cabo Verde currently living in Lisbon. As such, a database for mtDNA was built to analyze the genetic diversity of 103 immigrants from Cabo Verde, living in Lisboa.

Mitochondrial DNA control region was amplified and sequenced between positions 16024 and 576, using two sets of primers (L15997/H016 and L16555/H599). Obtained haplotypes were submitted to the most important mitochondrial DNA database in Forensic Sciences, *EDNAP Forensic mtDNA Population Database* – EMPOP.

The analysis of the control region revealed high genetic diversity, with a high frequency of unique haplotypes. Most mitochondrial DNA sequences correspond to African haplogroups. A minority of the sequences were linked to Euroasian haplogroups. Obtained results are similar to results achieved in previous studies about the origin and evolution of Cabo Verde's population, which suggests that the genetic composition of Cabo Verde's immigrant population living in Lisbon is representative of the genetic makeup of the archipelago. The results do not contradict the historical version of the colonization of the archipelago, which states the colonization involved female African slaves and male Portuguese, mobilized by that time to the ex-colony.

Keywords: Mitochondrial DNA; Cabo Verde; Lisboa; Population genetics

Índice Geral

•	Resumo	2
•	Abstract	3
•	Índice Geral	4
•	Índice de Figuras	6
•	Índice de Tabelas.....	7
•	Lista de abreviaturas	8
•	Introdução	10
1.	Cabo Verde.....	10
1.1.	Generalidades.....	10
1.2.	O Império Colonial Português e a descoberta de Cabo Verde	12
1.3.	Colonização e crescimento de Cabo Verde.....	12
1.4.	Destabilização do estado económico e Independência de Cabo Verde .	13
1.5.	Migração em Cabo Verde.....	14
1.6.	Atualidade de Cabo Verde	15
2.	Mitocôndria	16
2.1.	Estrutura e funções.....	16
2.2.	Origem da mitocôndria e semi-autonomia.....	17
3.	DNA Mitocondrial	18
3.1.	Genoma mitocondrial	19
3.2.	Características do genoma mitocondrial	20
3.2.1.	Código genético	21
3.2.2.	Número de cópias.....	21
3.2.3.	Taxa de mutação	21
3.2.4.	Herança materna	23
3.2.5.	Ausência de recombinação	24
3.2.6.	Heteroplasmia	25
4.	Estudo do genoma mitocondrial	27
4.1.	DNA mitocondrial e Genética Populacional.....	29
4.1.1.	Origem do homem moderno	30

4.1.2.	Migrações humanas e dispersão da população africana.....	31
4.2.	DNA mitocondrial e Ciências Forenses	33
4.3.	Outras aplicações do genoma mitocondrial	34
5.	Bases de dados genéticos de DNA mitocondrial.....	35
6.	Cabo Verde – composição genética atual.....	35
•	Objetivos.....	36
•	Materiais e Métodos	37
7.	Procedimento experimental	37
7.1.	Técnicas utilizadas	37
7.2.	Seleção de amostras	38
7.3.	Extração de DNA	38
7.4.	Amplificação	39
7.5.	Purificação dos produtos da reação de amplificação.....	40
7.6.	Sequenciação.....	41
7.7.	Purificação dos produtos da reação de sequenciação	43
7.8.	Deteção dos produtos sequenciados	44
7.9.	Análise dos produtos sequenciados	45
7.10.	Análise estatística e filogenética.....	47
•	Resultados e Discussão.....	48
8.	Resultados	48
8.1.	Estudo dos haplótipos representados	48
8.2.	Estudo de heteroplasmias.....	55
8.3.	Cálculo dos parâmetros de diversidade.....	56
8.4.	Estudo dos haplogrupos representados	60
8.5.	Análise filogenética.....	62
9.	Discussão.....	64
9.1.	Estudo da totalidade da região controlo do genoma mitocondrial	64
9.2.	Análise dos resultados obtidos.....	69
9.3.	Significado filogenético	77
•	Conclusões	79
•	Bibliografia	81

Figura 1.	Mapa com a localização das ilhas pertencentes ao grupo da Macaronésia	12
Figura 2.	Mapa de Cabo Verde	13
Figura 3.	Fotografia da estátua de Diogo Gomes, um dos navegadores envolvidos na descoberta de Cabo Verde, em Praia	14
Figura 4.	Representação esquemática referente aos emigrantes cabo-verdianos por país, em %, segundo dados do Censo de 2000	17
Figura 5.	Representação esquemática da estrutura da mitocôndria	18
Figura 6.	Representação esquemática do genoma mitocondrial	21
Figura 7.	Demonstração esquemática de um exemplo de herança materna de mtDNA para 18 indivíduos	25
Figura 8.	Mapa com uma representação esquemática da expansão do ser humano pelas diversas regiões do globo e respetiva distribuição geográfica de haplogrupos	32
Figura 9.	Técnica de sequenciação com ddNTPs marcados com fluorocromos de diferentes cores ...	42
Figura 10.	Esquema de purificação por <i>Xterminator</i> e aspiração automática pelo capilar	46
Figura 11.	Imagem de electroferograma obtido após a eletroforese capilar no <i>3130 Genetic Analyzer</i>	48
Figura 12.	Imagem de um electroferograma de sequência da região controlo do mtDNA com heteroplasmia de comprimento	50
Figura 13.	Representação esquemática das percentagens observadas por cada haplogrupo para os 103 indivíduos com ascendência cabo-verdiana	64
Figura 14.	Árvore resultante da análise das distâncias genéticas obtidas entre as populações referidas na tabela 18	66

Tabela 1.	Características discriminantes entre o DNA nuclear e o DNA mitocondrial	20
Tabela 2.	Primers utilizados na amplificação da região controlo total do DNA mitocondrial humano	42
Tabela 3.	Reagentes da reação de amplificação da região controlo total do DNA mitocondrial humano e respetivo volume por amostra	43
Tabela 4.	Condições da reação em cadeia da polimerase para a região controlo total do DNA mitocondrial humano, efetuada no termociclador <i>GeneAmp PCR system 9700</i>	43
Tabela 5.	Reagentes da reação de sequenciação e volume por amostra	45
Tabela 6.	Condições da reação de sequenciação efetuada no termociclador <i>GeneAmp PCR system 9700</i>	45
Tabela 7.	Condições da corrida electroforética realizada no <i>3130 Genetic Analyzer</i>	47
Tabela 8.	Nomenclatura de bases nucleotídicas de acordo com os códigos da IUPAC	50
Tabela 9.	Código do haplótipo, haplogrupo a que pertence, número de indivíduos que apresentam o haplótipo, número de alterações polimórficas em relação à rCRS e haplótipo	52
Tabela 10.	Número de haplótipos únicos, repetidos e diferentes, considerando a região controlo total do mtDNA dos 103 indivíduos estudados	54
Tabela 11.	Número e tipo de alterações polimórficas da região controlo total do mtDNA detetadas nas 103 sequências de mtDNA analisadas	54
Tabela 12.	Diferentes polimorfismos observados na zona de poli-C da região HV1 do mtDNA, entre as posições 16183 e 16193 e número de sequências onde ocorreu heteroplasmia de comprimento	56
Tabela 13.	Frequência absoluta de cada nucleótido na região controlo do mtDNA, considerando as alterações polimórficas relativamente à rCRS	57
Tabela 14.	Frequência de haplótipos homoplásmicos e heteroplásmicos nas 103 sequências analisadas e região hipervariável de ocorrência	59
Tabela 15.	Frequência absoluta e frequência relativa de cada nucleótido nas posições polimórficas da região controlo do mtDNA, considerando as alterações relativamente à rCRS, apresentadas na tabela 9	60
Tabela 16.	Parâmetros de diversidade da região HV1 do mtDNA de populações africanas e portuguesas	62
Tabela 17.	Parâmetros de diversidade da região controlo do mtDNA de populações africanas	62
Tabela 18.	População do presente estudo e populações selecionadas da literatura, utilizadas no estudo filogenético	65
Tabela 19.	Valores de F_{st} entre as populações do estudo filogenético	65
Tabela 20.	Valores de p relativos às populações comparadas no estudo	66

Lista de abreviaturas

°C: Graus Celsius

μAmp: Microampere

μm: Micrometro

μL: Microlitro

A: Adenina

ATP: Adenosina trifosfato

C: Citosina

CA: California

CRS: *Cambridge Reference Sequence* (Sequência de Referência de Cambridge)

dNTP: Desoxinucleótidos trifosfatados

ddATP: Didesoxiadenina trifosfatada

ddCTP: Didesoxicitosina trifosfatada

ddGTP: Didesoxiguanina trifosfatada

ddNTP: Didesoxinucleótidos trifosfatados

ddTTP: Didesoxitimina trifosfatada

D-loop: *Displacement Loop* (Laço de deslocamento)

Da: Dalton

DNA: Ácido desoxirribonucleíco

EDNAP: *European DNA Profiling Group*

EDTA: Ácido etilenodiaminotetracético

EMPOP: *EDNAP Forensic mtDNA Population Database*

F_{ST}: *Fixation index* (Índice de fixação)

g: Força gravitacional

G: Guanina

H: *Heavy* (pesado)

HV1: Hipervariável 1

HV2: Hipervariável 2

HV3: Hipervariável 3

INE: Instituto Nacional de Estatística

IUPAC: *International Union of Pure and Applied Chemistry*

Km: Quilómetro

L: *Light* (leve)

LHON: *Leber hereditary optic neuropathy* (Neuropatia óptica hereditária de Leber)

mL: Mililitro

mtDNA: DNA mitocondrial

N: *North* (Norte)

nDNA: DNA nuclear

NUMT: *Nuclear mitochondrial DNA* (Pseudogene mitocondrial no genoma nuclear)

O_H: Origem de replicação

Pb: Pares de bases

PCR: *Polymerase Chain Reaction* (Reação em cadeia da polimerase)

rCRS: *Revised Cambridge Reference Sequence* (Sequência de Referência de Cambridge revista)

RFLP: *Restriction Fragment Length Polymorphism*

ROI: Intermediário reativo de oxigênio

rSAP: *Recombinant Shrimp Alkaline Phosphatase* (Fosfatase Alcalina Recombinante de Camarão)

SNP: *Single nucleotide polymorphism* (Polimorfismo de nucleótido único)

T: Timina

tRNA: RNA de transferência

UK: *United Kingdom* (Reino Unido)

VNTR: Repetições em *tandem* de variado número

W: *West* (oeste)

Introdução

1. Cabo Verde

1.1. Generalidades

Cabo Verde é um país insular africano que em 2010 tinha 491,683 habitantes (Instituto Nacional de Estatística de Cabo Verde, 2010). Está situado ao largo da costa Africana Ocidental, a 450-600 km do Senegal, na latitude de 14°40' e 17°30' N e na longitude de 21°30' e 25°30' W. O seu nome tem origem na península de Cabo Verde, no Senegal, o ponto mais oeste do continente Africano. Está integrado na Macaronésia (Figura 1), um grupo de 4 arquipélagos situados no Oceano Atlântico (Açores, Madeira, Ilhas Canárias e Cabo Verde). Cabo Verde é constituído por dez ilhas e oito ilhéus de origem vulcânica. Atualmente não há registo de erupções, exceto na ilha do Fogo, onde o seu vulcão, altamente ativo, teve a sua última erupção em 1955 (Ramalho, 2011).



Figura 1 – Mapa com a localização das ilhas pertencentes ao grupo da Macaronésia
Adaptado de <http://www3.uma.pt/cem/aegro.ecpgr.symp/macaronesia.html>

As dez ilhas do arquipélago de Cabo Verde estão divididas, consoante a sua localização, em dois grupos. A norte situa-se o grupo Barlavento, constituído pelas ilhas de Santo Antão, São Vicente, Santa Luzia, São Nicolau, Sal e Boa Vista. O grupo Sotavento, constituído pelas ilhas Maio, Santiago, Fogo e Brava, situa-se a sul do arquipélago (Figura 2). A capital, Praia, localiza-se na ilha de Santiago e representa o maior aglomerado populacional do país (Willie, 2001). O arquipélago tem uma área total de 4068 km², sendo a maior ilha Santiago (área de 991 km²) e a menor ilha Santa Luzia (área de 35 km²). A distância entre as ilhas varia de 8 km (entre São Vicente e Santa Luzia) a 270 km (entre Santo Antão e Maio) (Ramalho, 2011).

O clima em Cabo Verde é tropical seco com influência oceânica, com chuvas de carácter torrencial, particularmente nas ilhas de relevo acentuado. O território é, na sua maioria, montanhoso, árido e pobre em terreno fértil. As temperaturas são mais elevadas em setembro (média de 26,7°C) e mais baixas em janeiro e fevereiro (médias de 18,4°C) (Ramalho, 2011). Foi em condições algo árduas, para os navegadores, que Cabo Verde foi descoberto no séc. XV por portugueses que exploravam a costa Africana Ocidental. A quando da sua descoberta, as ilhas não se encontravam habitadas.

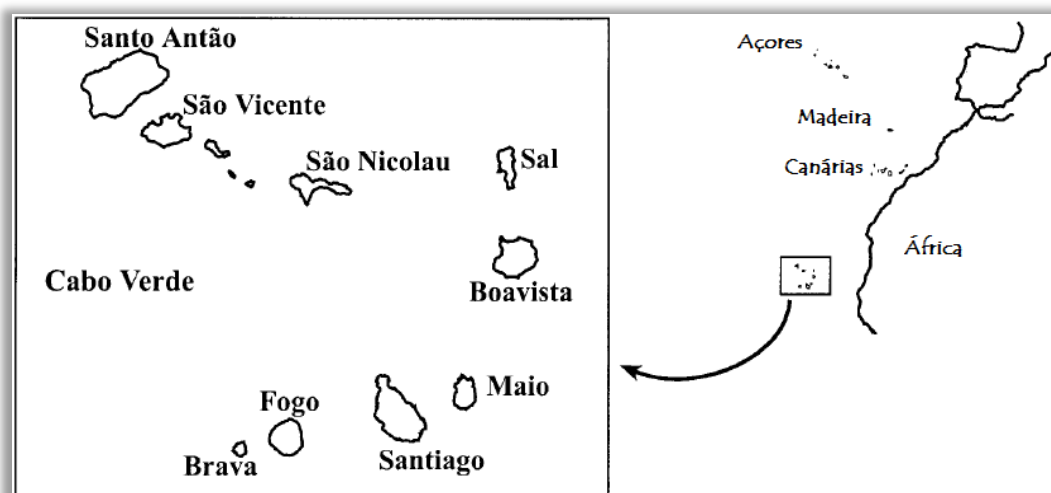


Figura 2 - Mapa de Cabo Verde;
Adaptado de Brehm *et al.* (2002)

1.2. O Império Colonial Português e a descoberta de Cabo Verde

A descoberta de Cabo Verde resultou da expansão do Império Colonial Português, que se iniciou em 1415 com a conquista de Ceuta e que se estendeu por África, Ásia e América do Sul. A descoberta de uma rota marítima através do Cabo Bojador pelo marinheiro português Gil Eanes permitiu a exploradores e mercadores Europeus partirem à descoberta de novas rotas na costa de África e, mais tarde, na Índia.

O Infante Dom Henrique, uma eminente personalidade histórica do Portugal quatrocentista, lançou navegadores lusos para além do Cabo Bojador. Entre eles estava um grupo de navegadores que contribuiu para a descoberta de Cabo Verde. Luís Cadamosto, explorador italiano, descobriu as primeiras ilhas do arquipélago, entre elas Santiago, em 1458. Diogo Gomes (Figura 3), António de Noli e Diogo Afonso (1461) descobriram as restantes ilhas. As ilhas de Cabo Verde foram descobertas, por portugueses, sem indícios de presença humana anterior. Acredita-se, contudo, que os Mouros tenham procurado sal nestas mesmas ilhas, nos séculos que se antecederam (Carreira, 1983).



Figura 3 – Fotografia da estátua de Diogo Gomes, um dos navegadores envolvidos na descoberta de Cabo Verde, em Praia

Adaptado de

http://old.encyclopedia.com.pt/articles.php?article_id=1184

1.3. Colonização e crescimento de Cabo Verde

Os primeiros colonizadores do arquipélago de Cabo Verde chegaram em 1462 à ilha de Santiago e alguns anos depois à ilha vizinha de Fogo. Foi em Santiago que surgiu a primeira povoação, Ribeira Grande (atualmente denominada Cidade Velha, de forma a evitar confusão com a cidade Ribeira Grande na ilha de Santo Antão). Ribeira Grande foi a primeira povoação Europeia permanente nos trópicos (Willie, 2001). Inicialmente, a colonização decorreu lentamente devido ao clima árduo da região. Só no séc. XVII as ilhas a norte, Santo Antão e São Nicolau, receberam os primeiros habitantes (Brehm *et al.*, 2002).

Em 1466 os colonizadores portugueses em Cabo Verde foram autorizados pelo Rei Afonso V a negociar livremente com as cidades da costa da Guiné, do Senegal e da Serra Leoa (Brehm *et al.*, 2002). Após estabelecerem postos de comércio nas ilhas já povoadas, começaram a estabelecer postos em pleno continente africano, o que fortaleceu uma relação comercial entre os colonizadores e as sociedades africanas com as quais comercializavam. Fruto desta relação teve início o comércio de escravos, que levou a um rápido aumento da população em Cabo Verde (Willie, 2001). Alguns dos primeiros escravos eram originários do Senegal e da Mauritânia e entre eles existiam também *Guanches* das Ilhas Canárias (Barry, 1998). Santiago tornou-se uma paragem obrigatória para barcos a caminho do Golfo da Guiné, Angola, São Tomé e Príncipe e Brasil e a região rapidamente se tornou numa estação para comércio transatlântico de escravos, atividade que permitiu a Cabo Verde prosperar durante o séc. XVI (Russell-Wood, 1998).

Cabo Verde não servia apenas para as atividades costeiras de Portugal no Oceano Atlântico. Serviu também como local de exílio para marginais e indivíduos acusados de crimes civís e políticos em território português. O arquipélago seria também inicialmente povoado por judeus e nobres, de Portugal e de outros países da Europa, que se mobilizaram para Cabo Verde sem as respetivas famílias.

As primeiras povoações de Cabo Verde eram, portanto, constituídas na sua maioria, por escravas africanas provenientes da costa Africana Ocidental e homens de diversas classes provenientes da Europa. Os homens acabariam por se cruzar com as mulheres indígenas, o que deu origem a uma nova classe mista de indivíduos, os *mulatos* ou *crioulos*. Indivíduos desta nova geração facilmente se tornaram intermediários de comércio, graças à sua familiarização com ambas as culturas, a Europeia e a Africana (Willie, 2001).

1.4. Destabilização do estado económico e Independência de Cabo Verde

A localização de Cabo Verde no Atlântico e a prosperidade do comércio tornaram o arquipélago vulnerável a ataques de piratas provenientes de vários países Europeus, entre eles a França e a Holanda, o que acabaria por afetar o comércio de

escravos. A influência portuguesa na região diminuiu ao longo dos sécs. XVII e XVIII e o comércio de escravos terminou em 1876.

A opressão portuguesa que mantinha Cabo Verde com o estatuto de colónia começou a sofrer resistência na segunda metade do séc. XX. Em 1956, um movimento nacionalista de libertação liderado por Amílcar Cabral, designado Partido Africano da Independência da Guiné e Cabo Verde, permitiu estabelecer uma resistência anticolonial que viria a dar origem a uma insurreição armada e sustentada e uma cerimónia para declaração da independência a 5 de julho de 1975 (Willie, 2001). Cabo Verde tornava-se independente depois de 500 anos de controlo colonial. Foi instalado um regime de partido único após a independência, substituído por um modelo de multipartidarismo em 1990 (Carvalho, 2010).

A abolição do comércio de escravos no séc. XIX perturbou o estado económico do arquipélago, já agravado por secas e escassez de alimentos. Contudo, a posição do arquipélago no Atlântico tornou-o na localização ideal para abastecimento e reparação de navios, o que permitiu atenuar a crise económica que Cabo Verde atravessava (Willie, 2001). Esta situação promoveu, contudo, um forte movimento migratório, que se viria a estabelecer como uma forte tradição no país.

1.5. Migração em Cabo Verde

A migração está presente na realidade histórica e social de Cabo Verde desde o estabelecimento da sociedade cabo-verdiana. Os primeiros sinais de expressão migratória dos cabo-verdianos deram-se no final do séc. XVII e início do séc. XVIII, sendo a América o principal destino. Desde então e até à atualidade a emigração tem levado os cabo-verdianos a mais de 25 países em todo o globo (Carling & Akesson, 2009).

Na história de Cabo Verde são conhecidos três períodos distintos de grande fluxo migratório ao longo do séc. XX. O primeiro ocorreu de 1900 a 1926 para os Estados Unidos da América. De 1927 a 1945 foram vários os destinos escolhidos pelos cabo-verdianos, entre eles a América Latina (Brasil e Argentina), África (Senegal, Guiné, São Tomé e Príncipe e Angola) e Portugal. De 1946 a 1973 os emigrantes escolheram como principal destino a Europa, nomeadamente a Holanda, França, Luxemburgo, Itália e Suíça (Carvalho, 2010).

Segundo o Censo de 2000 (INE de Cabo Verde), no período compreendido entre 1995 e 2000 emigraram 12.206 cabo-verdianos, a maioria do sexo masculino (54%). Nesse período, o principal país de destino foi Portugal com cerca de 6,590 emigrantes, seguido dos Estados Unidos da América, França, Holanda, Itália, Espanha e Suíça (Carvalho, 2010) (Figura 4). De acordo com os dados mais recentes do INE de Portugal, o nosso país tinha, em 2011, 38,895 emigrantes provenientes de Cabo Verde.

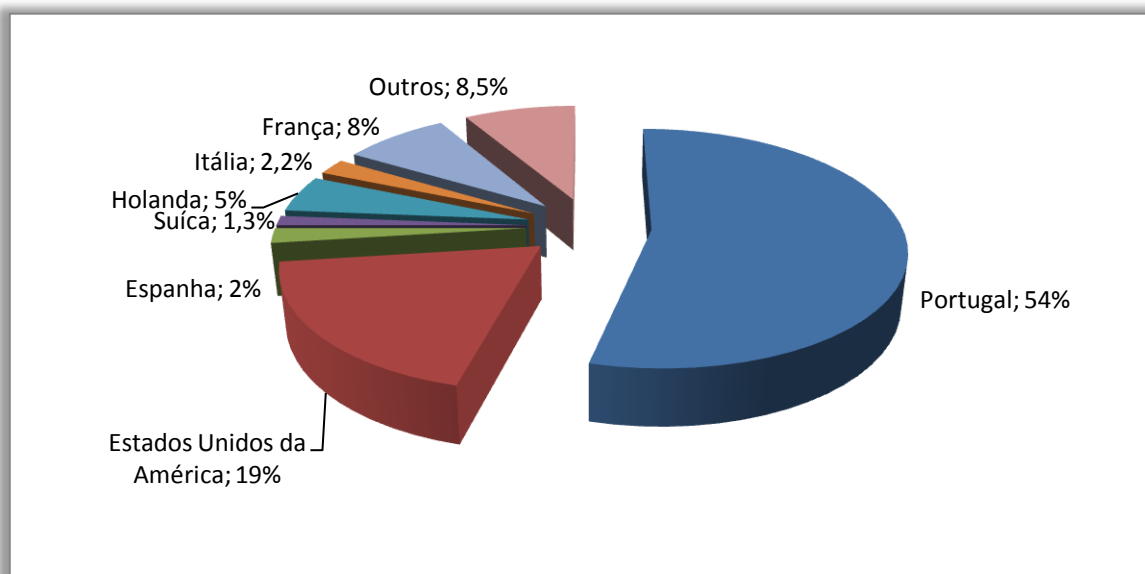


Figura 4 - Representação esquemática referente aos emigrantes Cabo-verdianos entre 1995 e 2000 por país, em %, segundo dados do Censo de 2000; Fonte: Carvalho (2010)

Este elevado fluxo migratório que se iniciou há 500 anos e que atingiu o seu auge no séc. XX permite caracterizar o caráter cultural atual do país.

1.6. Atualidade de Cabo Verde

Devido ao seu papel, como centro de comércio no Oceano Atlântico, Cabo Verde assimila, atualmente, impressões culturais não só de Portugal e oeste de África mas também de outras regiões do globo. A língua oficial é a Portuguesa mas o Crioulo é a linguagem falada mais comum. A religião dominante é a Católica, mas outras religiões africanas tradicionais também são comuns. A maioria dos Cabo-verdianos são hoje de ascendência mista, Europeia e Africana (Willie, 2001).

Atualmente, a história da população de Cabo Verde poderá ser corroborada pela análise de variações no DNA mitocondrial (mtDNA), que permitirá estimar o

impacto das populações que estiveram na origem da população cabo-verdiana e na formação da sua linhagem materna atual.

2. Mitocôndria

2.1. Estrutura e funções

O DNA mitocondrial localiza-se num dos compartimentos subcelulares das células eucariotas, a mitocôndria. As mitocôndrias são organelos pleomórficos (com 1 a 2 μm de comprimento) e altamente dinâmicos, presentes em elevado número no citoplasma, em todos os organismos que utilizem oxigénio como fonte de energia (Logan, 2006). São constituídas por duas membranas fosfolipídicas, uma externa, mais densa e uma interna, mais fina, que se dobra formando vilosidades, denominadas cristas, que lhe conferem uma área de superfície mais extensa. Na matriz mitocondrial, limitada pela membrana interna, é onde estão presentes cópias de DNA mitocondrial, RNA de transferência (tRNA), ribossomas mitocondriais e enzimas (Figura 5). A membrana externa contém poros que a tornam permeável a moléculas com massa molecular superior a 10 mil daltons (Da) e enzimas envolvidas na síntese de lípidos da mitocôndria e na sua respetiva conversão para metabolização

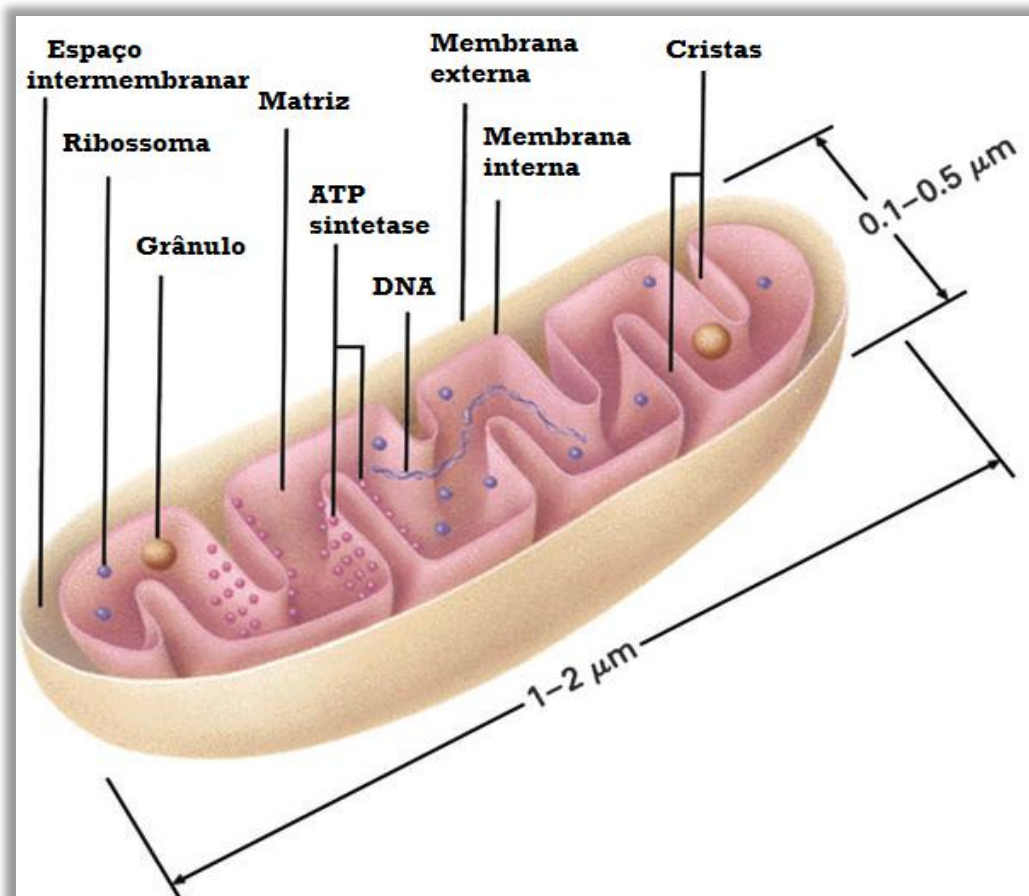


Figura 5 – Representação esquemática da estrutura da mitocôndria; Adaptado de Logan (2006)

na matriz. A membrana interna é constituída por proteínas responsáveis por reações de oxidação envolvidas no transporte de eletrões, ATP sintetases para produção de ATP na matriz e proteínas de transporte que permitem a entrada e saída de metabolitos da matriz (Alberts *et al.*, 2006).

O principal papel deste organelo consiste, enquanto centro de produção de energia da célula, na síntese de ATP durante o metabolismo aeróbio (Lodish, 2004). Está envolvida, adicionalmente, noutros processos metabólicos, entre eles a biossíntese de ácidos gordos, o Ciclo de Krebs e a biossíntese de aminoácidos (Logan, 2006). Provou-se também, ao longo do tempo, que a mitocôndria está envolvida na fotorrespiração (Douce & Neuburger, 1999), na sinalização celular para transporte de Ca^{2+} (Vandecasteele *et al.*, 2001) e na apoptose (Youle & Karbowski, 2005). O envolvimento da mitocôndria com o processo de envelhecimento está bem descrito e está associado ao aumento da produção de intermediários reativos de oxigénio (ROIs) e danos no DNA mitocondrial (Peterson *et al.*, 2012).

A mitocôndria possui o seu próprio DNA e maquinaria de síntese de proteínas sendo, no entanto, semiautónomas.

2.2. Origem da mitocôndria e semi-autonomia

A origem da mitocôndria terá ocorrido algures durante a evolução e reflete uma endossimbiose acompanhada por transferência de genes do endossimbionte ao hospedeiro. Existem, atualmente, duas teorias sobre a origem das mitocôndrias. A primeira, a tradicional teoria endossimbiótica, refere que as mitocôndrias seriam procariotas aeróbios estritos incorporados por um hospedeiro eucariota com características anaeróbias, dando origem a uma relação de simbiose. Uma teoria alternativa indica que o hospedeiro que incorporou a mitocôndria era um procariota. A mitocôndria ancestral serviria como uma fonte versátil de energia e eletrões. O hospedeiro, fruto desta relação de simbiose, viria a adquirir mais tarde características de um organismo eucariota (Gray, 2012; Martin & Mentel, 2010).

O antepassado da mitocôndria perdeu, eventualmente, a capacidade de se duplicar sozinho e, atualmente, a maioria das proteínas mitocondriais são codificadas no núcleo da célula, sintetizadas no citosol e depois importadas para a mitocôndria (Duby & Boutry, 2002). No entanto, a mitocôndria conserva uma semiautonomia

genética em relação à restante célula, já que possuiu um genoma próprio, o genoma mitocondrial.

3. DNA Mitocondrial

Designamos por genoma humano a totalidade da informação genética contida nas nossas células. É constituído pelo genoma nuclear, organizado em 23 pares de cromossomas e presente no núcleo das células (Makalowski, 2001) e pelo genoma mitocondrial, contido nas mitocôndrias presentes no citoplasma das células. São várias as diferenças existentes entre o DNA nuclear e o DNA mitocondrial (Tabela 1), adiante exploradas detalhadamente. Não são, contudo, entidades isoladas. A forte interação entre ambos os genomas é demonstrada pela necessidade de síntese de polipéptidos pelos dois genomas para posterior produção de ATP na mitocôndria (Pinheiro, 2009). O genoma mitocondrial representa cerca de 0,00055% do total do genoma humano (Ballard & Whitlock, 2004).

Tabela 1 – Características discriminantes entre o DNA nuclear e o DNA mitocondrial; Adaptado de Butler (2011)

Características	DNA Nuclear	DNA Mitocondrial
Localização	Núcleo	Mitocôndria (Citoplasma)
Número de cópias por célula	2 (1 de cada progenitor)	Podem ser superior a 1000
% de DNA	99,75%	0,25% (por célula)
Genoma	Diploide	Haploide
Estrutura	Linear, dupla hélice	Circular, cadeia dupla
Herança	Paterna e materna	Materna
Funcionamento	Autónomo	Semiautónomo, cooperação do DNA nuclear
Recombinação	Sim	Não
Reparação	Sim	Não
Taxa de mutação	Baixa	Elevada (5 a 10 vezes superior)
Poder de individualização	Único para cada indivíduo, excepto gémeos monozigóticos	Não individualiza (partilhado por todos os indivíduos da mesma linhagem materna)
Sequência de referência	2001 (projeto do genoma humano)	1981 (Anderson e colaboradores)

3.1. Genoma mitocondrial

O DNA mitocondrial é um tipo de DNA extranuclear circular de dupla cadeia, com uma cadeia pesada (H – *heavy*) rica em purinas e uma cadeia leve (L – *light*) rica em pirimidinas (Figura 6). É constituído por aproximadamente 16569 pares de bases (pb), número que varia de acordo com possíveis inserções e deleções que possam ocorrer. Apresenta-se compacto, sem intrões, com uma utilização bastante económica do genoma, já que possui poucas ou nenhuma bases entre genes no interior da sua região codificante (Butler, 2011). A estrutura circular sem histonas associadas torna-o estável e resistente à degradação já que não permite a ação de exonucleases. A sua localização (encerrada no interior das 2 membranas da mitocôndria) também favorece a sua estabilidade (Budowle *et al.*, 2003).

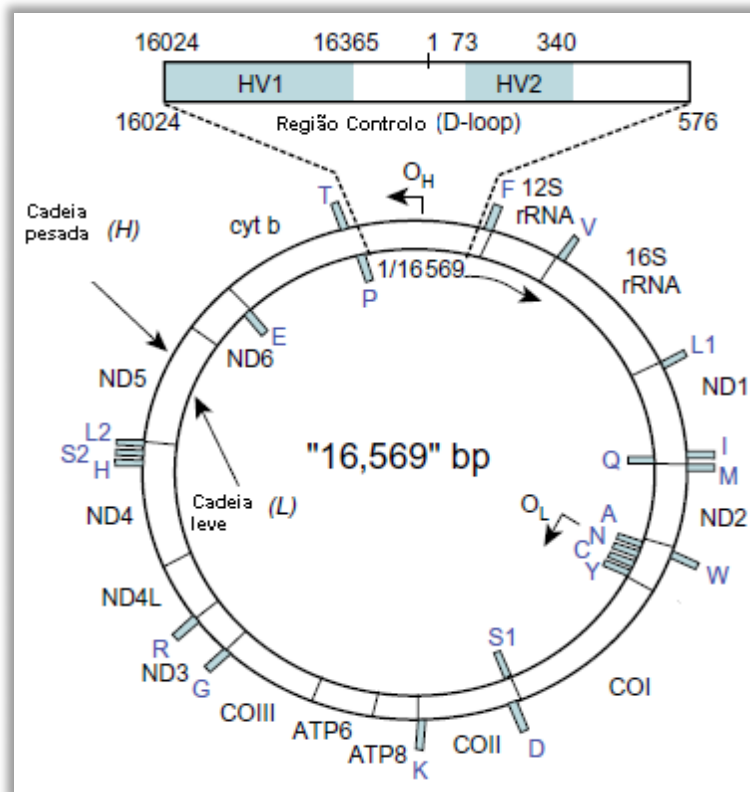


Figura 6 – Representação esquemática do genoma mitocondrial; Adaptado de Butler (2011)

A sequência completa do genoma mitocondrial foi publicada pela primeira vez em 1981 por Anderson e colaboradores (1981), viria a ser considerada a sequência de referência e é designada por *Cambridge Reference Sequence* (CRS). A CRS acabaria por ser revista por Andrews e colaboradores em 1999, que corrigiram 11 nucleótidos sendo, atualmente, denominada de *Revised Cambridge Reference Sequence* (rCRS), com o número de acesso, no *GenBank*, J01415.2. Nesta sequência

foi atribuído um número a cada uma das bases nucleotídicas. Um perfil de mtDNA é apresentado, somente, como o conjunto de alterações encontradas na cadeia leve da amostra em estudo, relativamente à rCRS (Budowle *et al.*, 2003).

A região codificante constitui cerca de 90% do genoma mitocondrial e contém 37 genes, entre eles 13 que codificam para polipéptidos envolvidos na fosforilação oxidativa, 2 que codificam para RNA ribossômico e 22 que codificam para RNA de transferência (Butler, 2011).

A região não codificante, também designada de “laço de deslocamento” (*displacement loop* ou *D-loop*), corresponde a uma porção de 1122 pb (cerca de 10% do genoma mitocondrial) e é onde se inclui a origem de replicação (O_H). A replicação do mtDNA inicia-se na cadeia pesada e é bidirecional e assincrónica, ou seja, é realizada por deslocamento de uma cadeia em relação à outra. Esta região também recebe a designação de região controlo já que está envolvida na regulação da replicação e transcrição do genoma mitocondrial, por possuir os promotores da transcrição de ambas as cadeias (P_I e P_{II}) e sequências conservadas associadas à iniciação da replicação e respetiva terminação (Coskun *et al.*, 2003).

Foi convencido por Anderson e colaboradores (1981) que a numeração das cadeias desta molécula se inicia na região controlo próxima da origem de replicação. A região controlo vai da posição 16024 à 16569 e da posição 1 à 576, e é nesta região que se encontram 3 segmentos hipervariáveis (HV1, HV2 e HV3), altamente polimórficos e por isso alvo de numerosos estudos. A região HV1 estende-se da posição 16024 à 16365, a HV2 da posição 73 à 340 e a HV3 da posição 430 à 576. No presente estudo toda a região controlo, ou seja, todas as regiões hipervariáveis foram estudadas.

3.2. Características do genoma mitocondrial

O DNA mitocondrial apresenta características particulares, designadamente um elevado número de cópias por célula, herança uniparental materna, elevada taxa de mutação e ausência de recombinação. Este conjunto de características conferem ao mtDNA a qualidade de marcador de eleição em estudos de genética populacional (Alves-Silva *et al.*, 2000; Salas *et al.*, 2002; Harich *et al.*, 2010; Afonso Costa *et al.*, 2010; Batini *et al.*, 2011; Castro *et al.*, 2012) e em diversas outras áreas como a

antropologia (Pääbo *et al.*, 1988; Krings *et al.*, 1997), a genética médica (Oikawa *et al.*, 2002; Herrnstadt & Howell, 2004; Van der Walt *et al.*, 2004) e as ciências forenses (Allen & Engström, 1998; Cerri *et al.*, 2004; Cruz *et al.*, 2004; Edson *et al.*, 2004; Seo *et al.*, 2012).

3.2.1. Código genético

À semelhança do que acontece no genoma nuclear, a informação contida no mtDNA está organizada em sequências de 3 bases cada (codões), que codificam para determinadas estruturas e funções. No entanto, no mtDNA determinados codões codificam estruturas e funções diferentes das que esses mesmos codões codificam no DNA nuclear (Scheffler, 1999). O codão para transcrição do aminoácido triptofano é o UGA no genoma mitocondrial, enquanto no genoma nuclear esse codão funciona como um codão STOP. Entre outros exemplos, no mtDNA o codão AUA codifica para uma metionina em vez de uma isoleucina e os codões AGA e AGG funcionam como codões STOP, enquanto no genoma nuclear codificam para uma arginina.

3.2.2. Número de cópias

O mtDNA está representado por 2 a 10 cópias na mitocôndria e cerca de 10^3 a 10^4 cópias por célula, ao contrário do DNA nuclear que está representado por apenas 2 cópias por célula humana (Iborra *et al.*, 2004; Torrini *et al.*, 2006). O elevado número de cópias de mtDNA é vantajoso em aplicações forenses para estudo de amostras altamente degradadas ou com pouca quantidade de DNA nuclear como ossos, dentes ou cabelos onde não seja possível obter um perfil genético com marcadores presentes no DNA nuclear (Budowle *et al.*, 2003).

3.2.3. Taxa de mutação

O mtDNA é uma molécula altamente polimórfica. Apresenta uma elevada taxa de mutação em comparação ao DNA nuclear, associada a uma baixa fidelidade da polimerase do mtDNA (pol γ), que permite um maior número de mutações durante o processo de replicação. Estudos sobre a fidelidade de polimerização pela pol γ indicam que produz um erro em cada 280 mil pares de bases (Johnson & Johnson, 2001). Por outro lado, o mtDNA é particularmente suscetível a danos oxidativos já que está próximo do principal local de produção de ROI, não está

associado a histonas e possui mecanismos de reparação pouco eficazes. A diminuição do número de cópias de mtDNA à medida que se envelhece é acompanhada por um aumento de danos no mtDNA e, conseqüentemente, mutações (Peterson *et al.*, 2012).

O mtDNA apresenta uma maior frequência de substituições pontuais do que de inserções ou deleções (Carvalho *et al.*, 2003). A maioria das substituições na região controle são substituições por transições, mais frequentemente, entre pirimidinas (Malyarchuk *et al.*, 2002).

A taxa de mutação varia entre diferentes regiões do mtDNA e vai desde $0,075 \times 10^{-6}$ a $0,0165 \times 10^{-6}$ substituições/posição/ano (Pakendorf & Stoneking, 2005). Esta variação pode ser explicada pela existência de 3 regiões específicas, os segmentos hipervariáveis (HV1, HV2 e HV3) presentes na região não codificante do mtDNA, que apresentam elevada heterogenidade e parecem possuir posições com maior incidência de mutações que evoluem 5 a 10 vezes mais rápido que os genes nucleares de cópia única e do que o restante genoma mitocondrial (Malyarchuk *et al.*, 2002; Galtier *et al.*, 2006; Howell *et al.*, 2007; Van Oven & Kayser, 2009). Verificou-se que mutações no mtDNA de amostras antigas, estudadas por Gilbert e colaboradores em 2003, ocorrem preferencialmente em determinadas posições definidas filogeneticamente na região controle, sugerindo que tais posições são hipermutáveis. A taxa de mutação da região HV1 é superior à das restantes regiões hipervariáveis (Meyer *et al.*, 1999). Contudo, também já foram descritas posições com maior incidência de mutações ao longo da região codificante, o que revela que existem zonas para além da região controle com taxa de mutação significativa (Freitas & Pereira, 2008). As mutações podem traduzir-se em situações associadas a patologias quando ocorrem em seqüências de genes altamente conservadas, alterando a sua função (Coskun *et al.*, 2003).

A taxa de fixação de uma mutação na população pode variar devido ao efeito de “afunilamento” (*bottleneck*), onde o número efetivo de moléculas de mtDNA transmitidas à descendência é significativamente reduzida em relação ao seu número presente em oócitos maduros (mais de 100 mil). Estão envolvidas 1 a 27 moléculas de mtDNA por fenómeno de *bottleneck* (Bendall *et al.*, 1997). Isto favorece uma rápida deriva genética na proporção de mtDNA que contenha a mutação dentro de

uma determinada população, podendo um indivíduo herdar um conjunto homogêneo de moléculas ou um conjunto heterogêneo de moléculas de mtDNA, com algumas dessas moléculas a possuir a mutação (Jenuth *et al.*, 1996; Howell & Smejkal, 2000).

3.2.4. Herança materna

Foi demonstrado pela primeira vez por Giles e colaboradores (1980) que o genoma mitocondrial, um genoma haploide, é herdado estritamente da mãe (Figura 7).

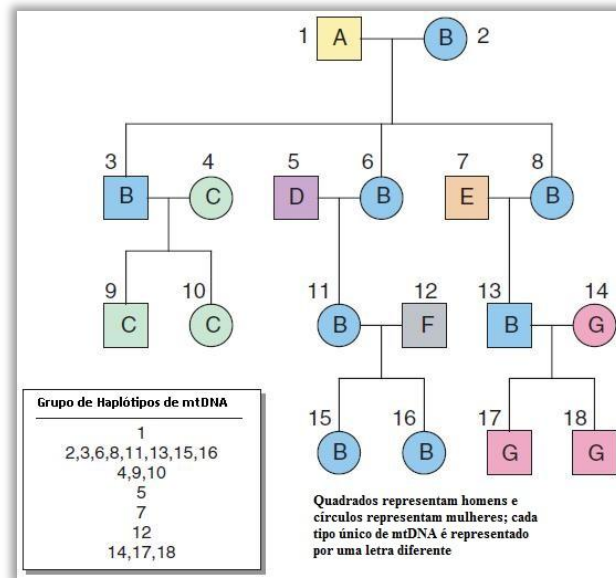


Figura 7 – Demonstração esquemática de um exemplo de transmissão materna de mtDNA para 18 indivíduos.
(Adaptado de Butler, 2011)

Quando o zigoto se divide e o blastócito se desenvolve, o citoplasma e restantes componentes celulares são consistentes com o oócito original da mãe. O oócito possui cerca de 100 mil moléculas de mtDNA, criando um cenário de diluição extrema para qualquer molécula de mtDNA do espermatozóide que possa existir no zigoto (Chen *et al.*, 1995; Holland & Parsons, 1999). Para reforçar a teoria da herança materna, qualquer mitocôndria do espermatozóide que entre num ovo em fertilização é seletivamente destruída pela maquinaria celular do embrião recém formado (Sutovsky *et al.*, 2004). Thompson e colaboradores (2003) verificaram que as mitocôndrias dos espermatozoides sofrem ubiquitinação durante a espermatogênese, promovendo a sua degradação no ovo.

Schwartz e Vissing (2002) descreveram um caso de herança paterna de mtDNA num homem de 28 anos com miopatia mitocondrial resultante de uma deleção de 2 pb no gene ND2. O mtDNA com a mutação tinha origem paterna e constituía 90% de mtDNA do músculo do paciente. Isto suscitou o interesse de toda a comunidade científica em determinar com exatidão se ocorre contribuição ou não de mtDNA paterno para a respetiva descendência. No entanto, não foram detetados indícios de herança paterna em estudos posteriores (Marchington *et al.*, 2002; Sutovsky *et al.*, 2004). Oikawa e colaboradores (2002) afirmaram que ocorrem distúrbios na herança materna em casos de síndrome de Klinefelter, contudo, múltiplos artefactos foram encontrados nas sequências dos 8 indivíduos em estudo e o caso anteriormente descrito de miopatia mitocondrial com herança paterna acabou por ser considerado uma exceção à regra (Bandelt *et al.*, 2005), regressando-se à teoria de herança estritamente materna do genoma mitocondrial.

3.2.5. Ausência de recombinação

A não recombinação é uma das características essenciais do mtDNA. As variações encontradas no genoma mitocondrial resultam exclusivamente de mutações adquiridas ao longo das gerações. A ausência de recombinação biparental facilita o estabelecimento de relações filogenéticas com aplicação em genética populacional. Paralelamente, no DNA nuclear apenas uma porção do cromossoma Y recombina (Underhill & Kivisild, 2007).

A observação de um caso de herança paterna levantou a hipótese de que a recombinação no mtDNA seria possível (Schwartz & Vissing, 2002). A mitocôndria possui, de facto, um sistema de recombinação funcional (Thyagarajan *et al.*, 1996). Foi também colocada a hipótese de que a ocorrência de recombinação no genoma mitocondrial poderia estar na causa das variações encontradas em determinadas posições entre indivíduos (Hagelberg *et al.*, 1999; Eyre-Walker & Awadalla, 2001). Este tema suscitou o interesse de diversos investigadores, que não detetaram qualquer evidência de recombinação no mtDNA (Inman & Nordborg, 2002; Herrnstadt *et al.*, 2002; Piganeau & Evre-Walker, 2004). A heterogenidade no genoma mitocondrial pode ser explicada pela existência de regiões com maior incidência de mutações, e constitui a teoria mais aceite atualmente.

A não recombinação característica do mtDNA, em conjunto com a sua herança materna, permite a utilização de familiares de uma mesma linhagem materna para resolução de casos de investigação da identidade, em indivíduos desaparecidos ou vítimas de desastres em massa, já que, excluindo mutações pontuais que possam ocorrer e ser adquiridas, a mãe passa para os filhos o seu mtDNA na íntegra, ou seja, o seu haplótipo. O haplótipo é o conjunto de variações na sequência de mtDNA de um indivíduo em comparação com a rCRS. Não é único, é partilhado por todos os indivíduos da mesma linhagem materna (Budowle *et al.*, 2003), pelo que o resultado obtido não é um perfil genético individualizante. No entanto, apesar de não ser individualizante o mtDNA permite, desde logo, a exclusão, caso o haplótipo não seja idêntico (exceto heteroplasmias) ao dos alegados parentes maternos.

3.2.6. Heteroplasmia

Considerado um genoma monoclonal, é expectável que cada indivíduo possua uma única sequência de mtDNA em todas as suas células, condição denominada de homoplasmia. Ocorre por vezes, contudo, heteroplasmia, que corresponde à coexistência de diferentes sequências de mtDNA no mesmo indivíduo com pequenas variações na sua composição nucleotídica (Howell & Smejkal, 2000; Alonso *et al.*, 2002; Brandstätter & Parson, 2003).

Existem dois tipos de heteroplasmia, a de comprimento e a de posição. A heteroplasmia de comprimento está associada a mutações envolvendo inserções ou deleções, e é observada com maior frequência nas regiões de poli-citosinas (poli-C) em HV1 (Bendall & Skyes, 1995; Tully *et al.*, 2000) e em HV2 (Prieto *et al.*, 2004). Na região HV1 a região poli-C estende-se da posição 16184 à posição 16193. Ocorrem variações no comprimento entre sequências quando se observa uma transição na posição 16189 (T-C). Na região HV2 a região poli-C vai da posição 303 à posição 315, com uma timina na posição 310. Essa timina pode ser substituída por uma citosina, dando origem à heteroplasmia. Heteroplasmia de comprimento é causada durante o processo de replicação por erros da polimerase do mtDNA, que sintetiza uma cadeia complementar com variação no tamanho em relação à sequência principal, num processo denominado de “deslize de replicação” (*replication slippage*). Este fenómeno dificulta a análise do genoma mitocondrial (Butler, 2011). As heteroplasmias de posição são mutações pontuais no genoma mitocondrial, com

substituição de uma base nucleotídica por outra numa determinada posição, apenas em algumas moléculas de mtDNA. Numa mesma posição do mtDNA umas cadeias apresentam uma base nucleotídica e outras cadeias apresentam uma outra base nucleotídica diferente.

A heteroplasma é observada, geralmente, numa única posição. Ocorre, contudo e com muita raridade, em duas ou até três posições - triplasma (Tully *et al.*, 2000). Estudos associados a um maior número de amostras e a metodologias de sequenciação com maior sensibilidade, confirmaram que a heteroplasma é um fenómeno comum, mesmo que em alguns casos não seja detetável (Tully *et al.*, 2000; Lutz *et al.*, 2004; Roberts & Calloway, 2011).

Calloway e colaboradores (2000) defendem que a frequência de heteroplasma aumenta com a idade, podendo ser uma condição hereditária ou um fenómeno de natureza somática. Contudo, Lagerström-Fermér e colaboradores (2001) realizaram um estudo retrospectivo em 4 mulheres com heteroplasma durante uma década e demonstraram que a heteroplasma na região controlo do mtDNA é herdada e mantém-se em níveis estáveis ao longo do tempo, não havendo relação com o aumento da idade. Estes resultados foram confirmados por Roberts e Calloway (2011) que não encontraram correlação significativa entre heteroplasma, idade e/ou género.

Pode ocorrer a coexistência de duas ou mais populações de mtDNA numa simples mitocôndria, célula ou tecido. É frequente encontrar diferentes níveis de heteroplasma em diferentes partes de uma mesma amostra de cabelo, e os níveis são mais acentuados nesse tipo de amostra quando comparados com amostras de sangue ou saliva (Alonso *et al.*, 2002; Roberts & Calloway, 2011). Diferentes níveis de heteroplasma entre tecidos podem ser explicados por um processo de replicação segregativa. No surgimento de uma nova mutação e após divisão de uma célula heteroplásmica, as mitocôndrias são distribuídas aleatoriamente pelas células filhas (Dimauro & Davidzon, 2005).

Em investigações forenses é necessário um cuidado especial na análise de heteroplasmas. Não é possível excluir ligação entre dois indivíduos que não possuam heteroplasma na mesma região se a restante sequência for idêntica. A presença de

heteroplasmia permitirá apenas reforçar a relação já identificada (Bär *et al.*, 2000). As heteroplasmias não são consideradas para a determinação de haplogrupos em estudos de genética populacional (Prieto *et al.*, 2004).

4. Estudo do genoma mitocondrial

O mtDNA constitui, em conjunto com a porção não recombinante do cromossoma Y, que é herdado exclusivamente pela linhagem paterna, um sistema genético haploide com importantes aplicações em estudos de evolução e também em genética médica, antropologia e ciências forenses. A evolução das metodologias de análise e respetiva progressão da qualidade e confiança nos resultados obtidos, além da capacidade de discriminação entre amostras possivelmente relacionadas, foi bastante evidente ao longo dos últimos 30 anos. A sua aplicação à genética populacional iniciou-se quando Brown e colaboradores (1980) analisaram o mtDNA de 21 indivíduos recorrendo à técnica *Restriction Fragment Length Polymorphism* (RFLP) de baixa resolução. Estudos posteriores desenvolveram a mesma técnica (Giles *et al.*, 1980; Scozzari *et al.*, 1988) até que, perto do final da década de 80, surgem RFLP de alta resolução (Cann *et al.*, 1987; Chen *et al.*, 1995; Torroni *et al.*, 1996). A partir da análise de RFLP na região codificante do mtDNA ficou demonstrado que existem locais polimórficos estáveis entre as populações que definem grupos de mtDNA designados por haplogrupos. Haplogrupo é definido por um conjunto de indivíduos que partilham variações e que normalmente são restrictos a determinada área geográfica, possuindo um ancestral comum. Na década de 90 surge o recurso a métodos de sequenciação com análise da sequência da região controlo, não codificante, que é, ainda hoje, o método mais utilizado.

Até muito recentemente apenas as primeiras duas regiões hipervariáveis, HV1 e HV2, foram alvo de análise (Oota *et al.*, 1999; Crespillo *et al.*, 2000; Alonso *et al.*, 2002; Carvalho *et al.*, 2003; Afonso Costa *et al.*, 2010), já que a maioria das alterações polimórficas se encontra nestes segmentos (Galtier *et al.*, 2006). Em alguns estudos tem-se, inclusivé, analisado apenas a região HV1 (Brandstätter & Parson, 2003; Nagai *et al.*, 2003; Tavares, 2007). Também é bastante frequente o recurso a um estudo misto, com análise das regiões hipervariáveis da região controlo e RFLP de determinados locais na região codificante (Alves-Silva *et al.*, 2000; Chen *et al.*, 2000; Salas *et al.*, 2002). A análise de apenas 2 regiões hipervariáveis (HV1 e

HV2) pode resultar na não análise de posições nucleotídicas características de vários haplogrupos, como por exemplo a posição 461, característica do haplogrupo M6, ou da posição 462, característica do haplogrupo J1 (Parson & Bandelt, 2007). A introdução do estudo do segmento HV3 permitiu aumentar a discriminação entre amostras analisadas para o genoma mitocondrial. É, atualmente, consensual na comunidade científica que é mais eficiente efetuar a amplificação da totalidade da região controlo, adotada em estudos recentes (Brandstätter *et al.*, 2008; Fendt *et al.*, 2012a; Gómez-Carballa *et al.*, 2012) e aplicada no presente estudo. O estudo da totalidade da região controlo garante um aumento do poder de discriminação entre amostras e uma maior confiança na determinação dos haplogrupos a que pertencem.

A sequência da região controlo pode não ser suficiente para determinar o seu haplogrupo. Determinadas posições na região codificante podem contribuir para um aumento da discriminação individual no estudo de mtDNA (Ingman *et al.*, 2006). Têm sido desenvolvidos métodos baseados na amplificação múltipla de *single nucleotide polymorphisms* (SNP) por todo o genoma mitocondrial (Coble *et al.*, 2006; Alvarez *et al.*, 2007; Dário *et al.*, 2009). Esta técnica apresenta-se como uma técnica de baixo custo e com elevada eficácia. É bastante utilizada atualmente em rastreio para identificação humana e em estudos antropológicos, procedendo-se depois, se necessário, à sequenciação do genoma mitocondrial. Na genética populacional esta técnica tem também sido aplicada como reforço ao estudo das regiões hipervariáveis da região controlo do mtDNA (Hill *et al.*, 2006; Brandstätter *et al.*, 2008; Álvarez-Iglesias *et al.*, 2009; Afonso Costa *et al.*, 2011).

Herrnstadt e colaboradores (2002) desenvolveram uma metodologia para sequenciar grandes extensões da região codificante. Vários investigadores têm, nos últimos anos, recorrido à análise de todo o genoma mitocondrial (Behar *et al.*, 2006; Ingman *et al.*, 2006; Malyarchuk *et al.*, 2008; Soares *et al.*, 2008; Gómez-Carballa *et al.*, 2012), garantindo uma eficácia total na análise das variações características de cada haplogrupo. A análise de múltiplos SNP da região codificante e/ou a sequenciação da totalidade do genoma mitocondrial poderão ser abordagens a aplicar em projetos futuros para estudo da origem da população imigrante de Cabo Verde residente em Lisboa.

4.1. DNA mitocondrial e Genética Populacional

Características como a herança estritamente materna e a ausência de recombinação do mtDNA permitem reconstruir a origem e evolução das populações. Uma vez que refletem a partilha de um ancestral comum, a distribuição geográfica dos haplogrupos e as suas respetivas ramificações permitem determinar a composição genética da linhagem materna de uma população e os eventos migratórios do passado, acompanhando a evolução do ser humano desde a sua origem, em África, e à medida que se dispersava por todos os continentes (Figura 8). A acumulação de mutações na sequência de mtDNA de indivíduos de uma mesma linhagem materna ao longo de gerações contínuas reflete a sua evolução. As variações na sequência de mtDNA e a idade do respetivo haplogrupo são importantes na determinação do tempo e processo a partir do qual os diversos eventos migratórios tenham ocorrido. A idade do haplogrupo reflete o momento em que ocorreu a mutação pontual que define o haplogrupo e não quando ocorreu a migração (Pakendorf & Stoneking, 2005). É também importante calcular as distâncias genéticas entre as populações e criar árvores filogenéticas que as relacionem para se perceber melhor a forma e via de migração.

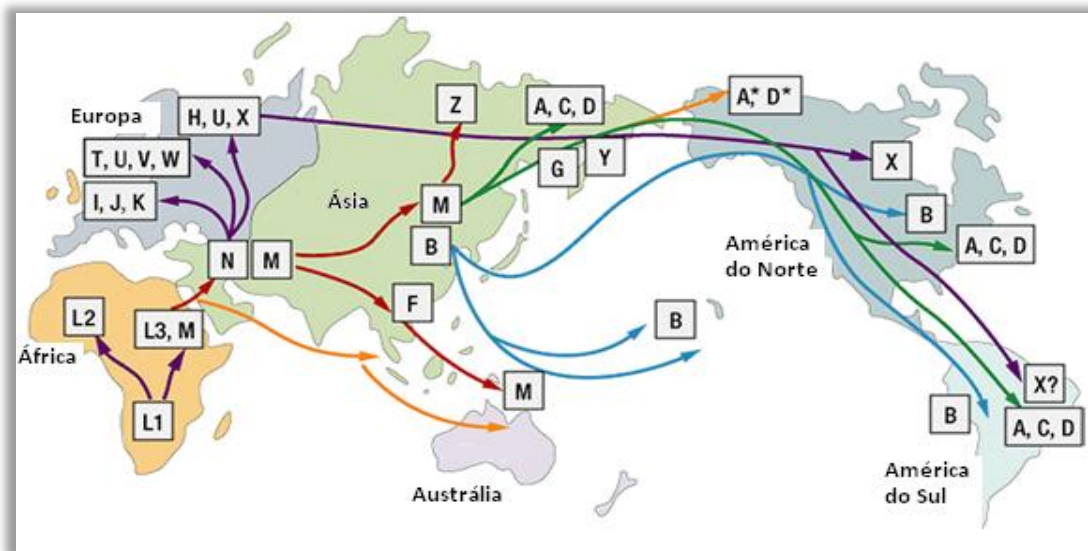


Figura 8 – Mapa com uma representação esquemática da expansão do ser humano pelas diversas regiões do globo e respectiva distribuição geográfica de haplogrupos; Adaptado de <http://www.familytreedna.com/>

A classificação atual dos principais haplogrupos é efetuada com letras maiúsculas, de A a Z. O mtDNA das populações africanas está posicionado na base da árvore filogenética, correspondendo ao macrohaplogrupo L. Este ramifica-se nos

haplogrupos L0, L1, L2, L3, L4, L5 e L6, característicos de populações africanas (Chen *et al.*, 2000; Salas *et al.*, 2002; Allard *et al.*, 2005; Behar *et al.*, 2008; Rosa & Brehem, 2011). L0 está próximo da “Eva Mitocondrial”, ancestral materno comum a todos os seres humanos. O haplogrupo L3 ramificou-se após mobilização de populações para fora de África há cerca de 60 a 70 mil anos em 2 macrohaplogrupos, o M e o N, presentes na Ásia e na Europa (Quintana-Murci *et al.*, 1999; Mishmar *et al.*, 2003). O macrohaplogrupo N ramificou-se nos haplogrupos H, I, J, K, T, U, V, W e X (Torroni *et al.*, 1996; Finnilä *et al.*, 2001; Behar *et al.*, 2012), que constituem 98% da população europeia. Os haplogrupos A, B, F e Y, que ramificaram do macrohaplogrupo N e os haplogrupos C, D, E, G e Z, que ramificaram do macrohaplogrupo M, são característicos da Ásia e Oceânia (Fucharoen *et al.*, 2001; Kivisild *et al.*, 2002; Kong *et al.*, 2011). O continente americano é constituído por, na sua grande maioria, haplogrupos asiáticos A, B, C, D e X, o que sugere que populações asiáticas entraram no continente americano antes da colonização Europeia (Bolnick & Smith, 2003; Kumar *et al.*, 2011).

Estudos de evolução têm identificado os *loci* polimórficos característicos de cada haplogrupo, quer da região codificante, quer da região controlo do mtDNA e estão hoje em dia, devidamente organizados numa árvore filogenética concebida por Van Oven e Kayser (2009). Como exemplo e recorrendo apenas a polimorfismos da região controlo, o haplogrupo T é caracterizado pela presença da substituição por transição C-T na posição 16294 do mtDNA. Sofre depois ramificação em 2 subhaplogrupos: T1, com uma substituição por transição G-A na posição 16163 e uma substituição por transição T-C na posição 16189; e T2, com uma substituição por transição C-T na posição 16296 (Kivisild *et al.*, 2006).

4.1.1. Origem do homem moderno

Os antropólogos têm sido incansáveis na procura da resposta à questão de qual a origem do homem moderno. A antropologia e a genética atuam em conjunto numa relação de complementaridade a favor da genética populacional. Quanto à origem do homem moderno, atualmente, existem duas teorias, a hipótese multirregional e a hipótese *Out of Africa*. A hipótese multirregional indica que não há uma origem geográfica única para a espécie *Homo sapiens* subespécie *Homo sapiens sapiens* e que esta evoluiu de forma simultânea a partir do *Homo erectus* em

diferentes regiões do Globo, há cerca de um milhão de anos. No entanto, a hipótese que prevalece atualmente é a hipótese *Out of Africa*, onde se afirma que o homem moderno tem origem recente em África (Ingman *et al.*, 2000; Forster, 2004), tendo posteriormente migrado para a Euro-Ásia e foi substituindo gradualmente a população de *Homo erectus* (Stringer, 2003).

As características únicas do mtDNA permitem afirmar que todas as linhagens matrilineares possuem uma raiz comum, situada no ancestral comum mais recente, denominado de “Eva Mitocondrial” (Cann *et al.*, 1987). Esta seria, teoricamente, a primeira mulher. Nascida no sul ou este de África há mais de 130 mil anos, esteve na origem da transição para o homem moderno e na origem de todas as linhagens maternas atuais (Forster, 2004; Torroni *et al.*, 2006; Behar *et al.*, 2008). Estudos sobre variabilidade genética corroboram a hipótese de que o mtDNA ancestral humano se localiza em África, com origem há cerca de 125 a 165 mil anos e com posterior fluxo migratório para os restantes continentes, já que as populações africanas apresentam o mtDNA com maior variabilidade (Chen *et al.*, 2000). Ocorrem em média no mtDNA 8 variações nucleotídicas em Caucasianos e 15 variações nucleotídicas em indivíduos de descendência africana, em relação à rCRS (Budowle *et al.*, 1999). Estudos de polimorfismos da porção não recombinante do cromossoma Y corroboram também a hipótese de uma origem do homem moderno recente em África (Cruciani *et al.*, 2011).

4.1.2. Migrações humanas e dispersão da população africana

A análise das variações no mtDNA tem sido usada para reconstruir as migrações do homem moderno. A maioria das sequências de mtDNA de africanos subsarianos encaixa numa das ramificações do macrohaplogrupo L, de L0 a L6 (excluindo os haplogrupos M e N, que derivaram do haplogrupo L3 e que representam toda a diversidade mitocondrial não subsariana). Os sub-haplogrupos do haplogrupo L encontram-se dispersos entre diversas populações do continente africano e surgiram através de um processo evolutivo que durou 80 a 120 mil anos, de acordo com a hipótese *Out of Africa*. O haplogrupo L0 constitui um dos grupos mais antigos com origem há cerca de 140 a 160 mil anos (Soares *et al.*, 2008). O primeiro grande evento migratório registou-se com o surgir do haplogrupo L1 há cerca de 130 mil anos, encontrado-se disperso pelo oeste e centro de África, inclusive

pela costa para o Atlântico (Brehm *et al.*, 2002; Behar *et al.*, 2008). O haplogrupo L5 é considerado um haplogrupo intermédio, com 120 a 140 mil anos, entre os haplogrupos L1 e L2 a L6 e é encontrado essencialmente no este africano (Brandstätter *et al.*, 2004; Kivisild *et al.*, 2004). O haplogrupo L2 corresponde, em conjunto com o haplogrupo L3, a 70% da variação materna subsariana (Chen *et al.*, 2000). Surgiu há 90 mil anos e está presente em elevadas frequências no oeste e centro de África. Sequências de mtDNA com origem no este africano, há 60 a 75 mil anos, são geralmente atribuídas ao haplogrupo L3, que eventualmente se dispersou por todo o continente africano (Salas *et al.*, 2002; Behar *et al.*, 2008). Ramificações do haplogrupo são encontradas por todas as populações africanas. Os haplogrupos L2 e L3 dispersaram do este de África para a Euro-Ásia, e o haplogrupo L3 é considerado o precursor mais provável dos haplótipos asiáticos e europeus. Os haplogrupos L4 e L6 têm origem no este de África e remontam há cerca de 100 mil e 22 mil anos, respetivamente (Kivisild *et al.*, 2004; Behar *et al.*, 2008).

Estudos anteriores demonstraram que, na Europa, os haplogrupos L estão presentes em quantidades inferiores a 1% (Salas *et al.*, 2004). A análise de sequências da região controlo de populações europeias indica que linhagens dos haplogrupos L não terão evoluído no continente Europeu, cenário compatível com uma chegada recente, a partir do continente Africano (Malyarchuk & Czarny 2005). Um estudo de 69 genomas mitocondriais de linhagens dos haplogrupos L inseridas em populações europeias demonstrou que 65% dessas linhagens têm, sem qualquer dúvida, origem subsariana, o que confirma uma migração recente de populações africanas para a Europa. Os restantes 35% correspondem a ramificações do macrohaplogrupo L específicos do continente europeu, revelando um fluxo genético desde a África subsariana até à Europa, estimando-se que tenha ocorrido há cerca de 11 mil anos (Cerezo *et al.*, 2012). Os investigadores desta matéria têm tentado determinar a origem das linhagens L na Europa, concluindo que estas chegaram recentemente ao continente europeu, provavelmente durante o período do domínio Romano, na conquista Ibérica dos Árabes ou no comércio de escravos no Atlântico (Pereira *et al.*, 2000; Álvarez-Iglesias *et al.*, 2009).

4.2. DNA mitocondrial e Ciências Forenses

São vários os métodos de Biologia Molecular, desde a análise de repetições em *tandem* de variado número (VNTR) por RFLP (Budowle *et al.*, 1991), a análise de SNPs (Comey *et al.*, 1991) e a sequenciação direta de mtDNA (Cerri *et al.*, 2004; Cruz *et al.*, 2004; Seo *et al.*, 2012), que têm sido aplicados ao longo dos tempos nas ciências forenses.

A utilidade, aplicação e validação do mtDNA nas ciências forenses está bem documentada (Wilson *et al.*, 1995; Budowle *et al.*, 2003; Parson & Bandelt, 2007; Pinheiro, 2009). Tal utilidade é justificada pelas características únicas do genoma mitocondrial já exploradas: hereditariedade uniparental, ausência de recombinação, elevada taxa de mutação e elevado número de cópias por célula. Este último fator é especialmente vantajoso na resolução de casos onde a quantidade de DNA extraído é muito pequena ou na presença de amostras altamente degradadas. O mtDNA apresenta-se, de facto, como uma molécula bastante resistente às condições ambientais (Holland & Parsons, 1999; Cerri *et al.*, 2004).

O mtDNA é utilizado, especialmente, em amostras onde exista uma reduzida probabilidade de obter resultados aplicando o estudo de marcadores polimórficos no DNA nuclear. Como exemplos temos cabelos (onde só é possível obter resultados de DNA nuclear se existir bolbo) e em determinadas circunstâncias, ossos e dentes (Budowle *et al.*, 2003). Balogh e colaboradores (2003) demonstraram também que é possível obter perfis de mtDNA a partir de impressões digitais em simples folhas de papel.

As regiões hipervariáveis são de elevado interesse na identificação humana devido à sua elevada taxa de mutação. O pequeno tamanho de cada uma das 3 regiões permite a amplificação por *Polymerase Chain Reaction* (PCR) e a sequenciação, técnicas de utilização frequente nos laboratórios de Genética Forense. A natureza haploide e monoclonal do mtDNA simplifica a interpretação de resultados. No entanto, há que ter especial atenção na análise de heteroplasmias, já que não servem de critério de exclusão, podendo apenas reforçar a relação entre amostras conhecidas.

A utilização do mtDNA na Genética Forense tem sido recorrente. Excluindo mutações e tendo em atenção que o mtDNA é herdado por via exclusivamente materna, a sequência do mtDNA é idêntica entre todos os indivíduos aparentados pela via materna. Esta característica pode auxiliar a análise de restos cadavéricos de pessoas desaparecidas, onde o mtDNA de qualquer parente materno pode servir de amostra de referência para comparação direta com o mtDNA do desconhecido. Graças à ausência de recombinação, diversas gerações podem igualmente servir como amostras de referência. Os marcadores nucleares, à exceção da porção não recombinante do cromossoma Y, não possuem esta característica. Um dos estudos mais conhecidos utilizando a sequenciação de mtDNA resultou na identificação do Czar Nicholas II a partir da análise do mtDNA dos seus ossos. Utilizando o mtDNA obtido de familiares da mesma linhagem materna, foi efetuada uma comparação com a sequência de mtDNA extraída dos seus ossos, suportando a hipótese de que se tratavam realmente dos ossos do Czar. Os supostos ossos do Czar partilhavam com o seu irmão Georgij Romanov, inclusivé, heteroplasmia de comprimento na região HV1 na posição 16168. De salientar que a amostra do irmão estava mergulhada em águas profundas há 60 anos (Holland & Parsons, 1999). Allen e Engström (1998) conseguiram estabelecer uma relação entre 3 crimes em locais diferentes ao revelar a presença de sequências de mtDNA idênticas em todos os locais a partir de hastes de cabelo e unhas. Cerri e colaboradores (2004) conseguiram identificar os restos mortais de um alegado indivíduo presumivelmente assassinado e posteriormente queimado a partir do mtDNA, com recurso à amostra de mtDNA da sua mãe. Cruz e colaboradores (2004) e Amorim e colaboradores (2011) identificaram cadáveres em avançado estado de decomposição através da comparação do mtDNA extraído a partir de ossos e um dente, com amostras de referência de familiares maternos.

4.3. Outras aplicações do genoma mitocondrial

As variações no mtDNA têm sido extensamente estudadas e aplicadas à genética médica, já que diversas patologias estão associadas a mutações no mtDNA. Diversos estudos associaram determinados haplogrupos a fenótipos clínicos (Dimauro & Davidzon, 2005; Herrstadt & Howell, 2004). O haplogrupo J está associado a um maior risco de desenvolver cegueira em indivíduos com neuropatia optica hereditária de Leber (LHON) (Brown *et al.*, 2002). Ruiz-Pesini e

colaboradores (2000) verificaram que a motilidade do esperma se encontra reduzida em homens pertencentes ao haplogrupo T e aumentada em homens incluídos no haplogrupo H. No entanto, alguns haplogrupos conferem resistência à doença de Alzheimer (Van der Walt *et al.*, 2004) ou à doença de Parkinson (Van der Walt *et al.*, 2003) e outros haplogrupos estão associadas a maior longevidade (Niemi *et al.*, 2004).

5. Bases de dados genéticos de DNA mitocondrial

Existem, mundialmente, bases de dados que possuem haplótipos de mtDNA, onde é possível comparar os haplótipos e estudar a frequência dos mesmos na população. Entre outras bases de dados existem o MITOMAP (Ruiz-Pesini *et al.*, 2007), o mtDB (Ingman & Gyllensten, 2006) e a EMPOP (Parson *et al.*, 2004).

A base de dados com maior relevância na área a nível mundial é a EMPOP e está estabelecida no Instituto de Medicina Legal de Innsbruck. A EMPOP tem como objetivo criar uma base de dados com elevada qualidade e diretamente acessível a toda a comunidade científica. O principal fundamento da sua criação esteve no facto de os laboratórios forenses não terem normalmente capacidade de gerar quantidade significativa de dados relativos a sequências mitocondriais para as diferentes populações. A reunião de dados populacionais obtidos por cada grupo de trabalho torna-se, portanto, um conjunto relevante. A EMPOP submete as sequências estudadas a um rigoroso controlo de qualidade, antes de as integrar na sua base de dados (<http://empop.org>). A EMPOP assegura a infraestrutura bioinformática, certificando-se da análise, transcrição, armazenamento de dados e difusão dos mesmos (Parson *et al.*, 2004).

6. Cabo Verde – composição genética atual

Recorrendo ao estudo do mtDNA é-nos permitido estimar o impacto das populações que estiveram na origem da população imigrante cabo-verdiana residente em Lisboa e na formação da sua atual linhagem materna. A elevada diversidade genética da população de Cabo Verde, característica de populações africanas, já foi demonstrada com recurso à análise do genoma mitocondrial, com prevalência de haplogrupos L, superior a 90% (Brehm *et al.*, 2002; Tavares, 2007). No entanto, estes estudos focaram-se apenas na análise da região HV1. Estudo de marcadores

polimórficos do DNA nuclear foram igualmente efetuados para esta população (Amorim *et al.*, 2012; Afonso Costa *et al.*, 2012). Gonçalves e colaboradores (2003) obtiveram o perfil do cromossoma Y de 201 indivíduos de Cabo Verde, confirmando a influência paterna europeia e do Médio Oriente na origem da população cabo-verdiana.

Estudos de mtDNA para populações de Cabo Verde realizados até hoje basearam-se apenas na sequenciação da região HV1 (Tavares, 2007) ou na combinação entre HV1 e análise da região codificante por RFLP (Brehm *et al.*, 2002). No presente estudo é reportada pela primeira vez a região controlo total desta população. A introdução desta metodologia, recomendada atualmente como já referido, permite aumentar o poder de discriminação entre amostras e reforça as determinações prévias dos respetivos haplogrupos. Indivíduos que partilhem as mesmas variações na região HV1 e/ou o mesmo perfil de RFLPs poderão apresentar diferenças quando a restante região controlo é analisada, permitindo diferenciar sequências que antes se julgava serem iguais e que seriam, de forma errada, agrupadas no mesmo haplótipo e, eventualmente, no mesmo haplogrupo.

Objetivos

O objetivo genérico deste projeto é caracterizar a diversidade genética da população de imigrantes oriundos de Cabo Verde, residentes em Lisboa, e respetiva ascendência estudando, pela primeira vez nesta população, a totalidade da região controlo do seu DNA mitocondrial. É esperado que os resultados se assemelhem à composição genética de mtDNA característica do oeste de África, na região entre o Senegal e a Serra Leoa. Este estudo pretende acrescentar nova informação sobre a origem da população imigrante cabo-verdiana residente em Lisboa, sobre a sua ascendência e património genético do arquipélago de Cabo Verde.

Outro dos objetivos é, após a definição dos haplótipos de mtDNA dos indivíduos pertencentes à população em estudo, incluir os mesmos na base de dados de mtDNA com maior relevância na área das Ciências Forenses, a *EDNAP Forensic mtDNA Population Database* (EMPOP). É pretendido diferenciar os haplogrupos em que os haplótipos se incluem e, posteriormente, realizar uma comparação interpopulacional, em particular, com populações africanas, a fim de determinar qual a sua proximidade. Desta forma, é pretendido acrescentar entendimento sobre o

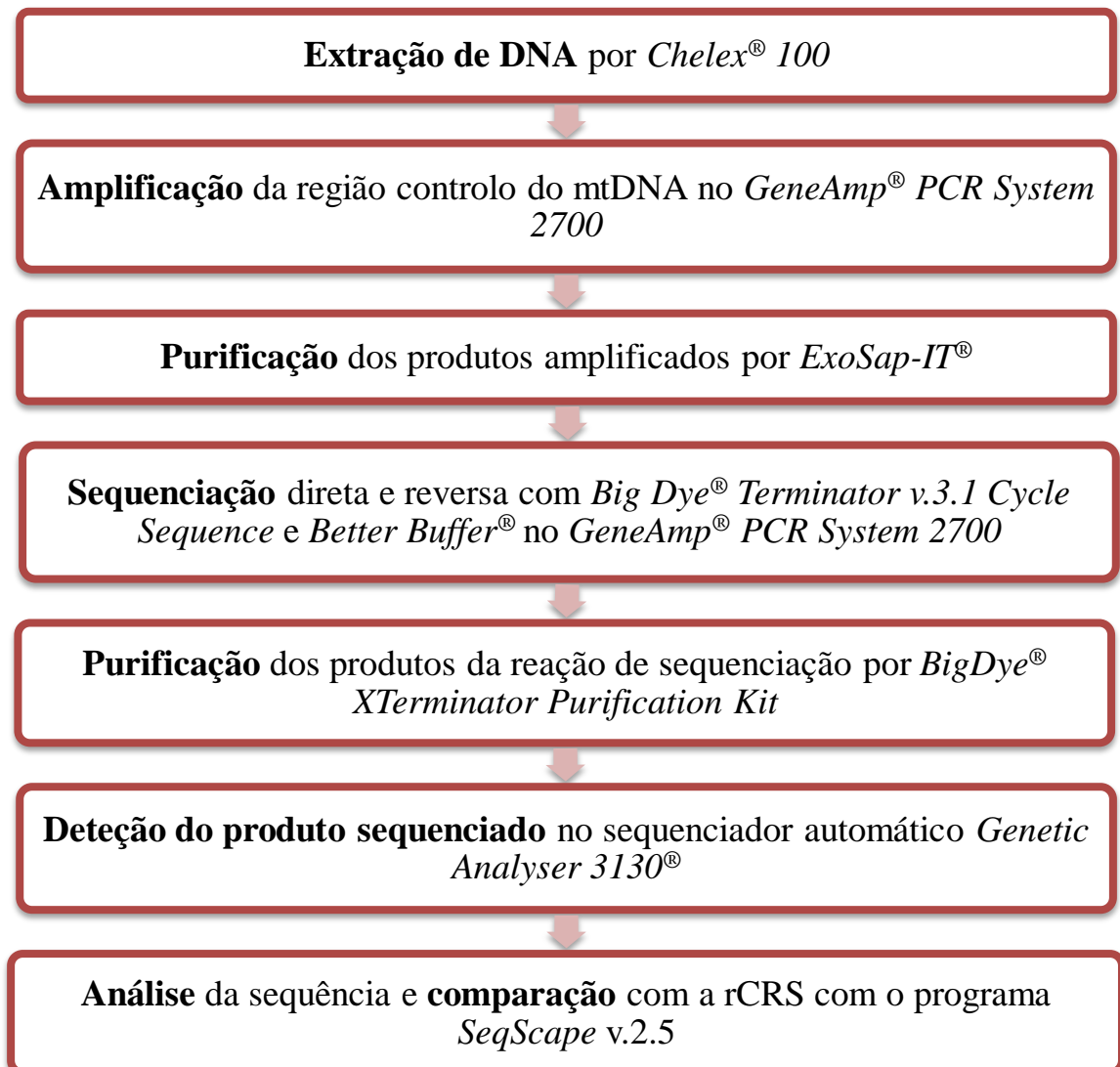
impacto das migrações das populações do passado que terão estado na origem do arquipélago de Cabo Verde, acrescentado-se portanto mais conhecimento acerca da origem e evolução do ser humano ao longo dos tempos.

Materiais e Métodos

7. Procedimento experimental

7.1. Técnicas utilizadas

As várias fases laboratoriais do procedimento experimental realizado estão representadas no seguinte fluxograma.



7.2. Seleção de amostras

O estudo foi realizado numa amostra de 103 indivíduos aparentemente saudáveis e não aparentados, com ascendência materna e paterna de Cabo Verde, atualmente a viver na zona da Grande Lisboa. As amostras estudadas foram manchas de sangue, colhidas em *Bloodstain Card* (Whatman®), a intervenientes em ações judiciais de investigação de parentesco biológico a decorrer no Instituto Nacional de Medicina Legal e Ciências Forenses, portanto, uma amostra de conveniência.

7.3. Extração de DNA

A metodologia aplicada neste procedimento consistiu na utilização de uma resina quelante adicionada diretamente à amostra em suspensão, técnica introduzida em 1991 e denominada de *Chelex*® 100 (Bio-Rad Laboratories, Hercules, CA).

O *Chelex* atua como uma substância quelante que se liga a iões metálicos polivalentes como o magnésio (Walsh *et al.*, 1991). Aplicando altas temperaturas, dá-se a lise das células e a libertação do DNA, e o *Chelex* agrega os restantes materiais (Rudin & Inman, 2002). Removendo o magnésio da reação, as enzimas responsáveis pela destruição do DNA, as nucleases, ficam inativas e as moléculas de DNA são, portanto, protegidas. A extração por *Chelex* envolve poucos passos e, conseqüentemente, a probabilidade de contaminação entre amostras é reduzida.

Foi utilizado, de cada indivíduo, um fragmento de papel com amostras de sangue, com a forma de um círculo com cerca de 3 mm², recortado com um *punch*, que foi colocado num tubo de *ependorf* de 1,5 mL devidamente rotulado com o número de extração da amostra. A cada tubo juntou-se 1 mL de água desionizada e autoclavada. Após incubação durante 30 minutos à temperatura ambiente, as amostras foram agitadas num vórtex em rotação baixa durante 5 a 10 segundos e posteriormente centrifugadas a 15000 x g durante 3 minutos. Desta primeira fase resultou a separação das células, contendo o DNA, do suporte de papel e deposição das mesmas no fundo do tubo. Foi efetuada, de seguida, a remoção do sobrenadante, ficando retidos no tubo cerca de 30 µL da solução. Adicionou-se 180 µL de solução de *Chelex*® 100 a 5% a cada amostra. Os tubos foram colocados em banho de água a 56°C durante 15 minutos. Após uma ligeira agitação foram colocadas em banho de

água em ebulição durante 8 minutos. Por fim, foi efetuada nova agitação no vórtex durante alguns segundos e centrifugação a 15000 x g durante 5 minutos. O DNA extraído foi então armazenado, até ser utilizado para amplificação, a 20°C negativos.

7.4. Amplificação

Para amplificar a região controlo total do mtDNA, foram utilizados 2 pares de primers (L15997 e H016; L16555 e H599) (*Thermo Scientific, Waltham, MA*) (Tabela 2). A nomenclatura dos primers utiliza a cadeia correspondente ao primer, L para leve e H para pesada, e a posição nucleotídica na extremidade 3'. O par de primers L15997 (direto) e H016 (reverso) permitiu a amplificação dos fragmentos entre as posições 16024 e 16599, enquanto que o par de primers L16555 (direto) e H599 (reverso) permitiu a amplificação dos fragmentos entre as posições 1 e 576.

Tabela 2 – Primers utilizados na amplificação da região controlo total do DNA mitocondrial humano

Primers	Sequência	Tamanho do produto amplificado
L15997	5' CAC CAT TAG CAC CCA AAG CT 3'	588 pb
H016	5' CCC GTG AGT GGT TAA TAG GGT 3'	
L16555	5' CCC ACA CGT TCC CCT TAA AT 3'	590 pb
H599	5' TTG AGG AGG TAA GCT ACA TA 3'	

Para a amplificação foi preparada uma solução com volume final de 10 µL (Tabela 3), contendo água ultra pura, Multiplex PCR *Master Mix*® (*Qiagen, Düsseldorf, Germany*) - mistura que contém tampão de reação, DNA polimerase, dNTPs e MgCl₂ - primers (utilizados numa concentração de 2 µM) e DNA de cada uma das amostras. Cada tubo de PCR foi devidamente rotulado com o número da amostra e fragmento a amplificar. Em cada reação de amplificação foi realizado um controlo negativo ao qual não foi adicionado DNA de forma a controlar a qualidade dos reagentes. A reação de amplificação processou-se num termociclador *GeneAmp PCR system 9700* (*Applied Biosystems, Foster City, CA*), de acordo com as condições apresentadas na tabela 4. Após a amplificação as amostras foram purificadas.

Tabela 3 – Reagentes da reação de amplificação da região controlo total do DNA mitocondrial humano e respetivo volume por amostra

Reagentes	Volume
Água ultra pura	3 μ L
Multiplex PCR Master Mix®	5 μ L
Primers a 2 μ M	1 μ L
Volume final da Mix	9 μL
+ DNA	1 μL

Tabela 4 – Condições da reação em cadeia da polimerase para a região controlo total do DNA mitocondrial humano, efetuada no termociclador *GeneAmp PCR system 9700*

1 ciclo	95°C	15 min
35 ciclos	94°C	30 seg
	60°C	90 seg
	72°C	60 seg
1 ciclo	72°C	10 min
Hold	4°C	∞

7.5. Purificação dos produtos da reação de amplificação

A purificação do produto amplificado permite remover todos os elementos que possam interferir com a sequenciação e posterior análise da sequência. Esta fase foi realizada recorrendo ao método *ExoSap-IT*[®] (*Affymetrix, Santa Clara, CA*), que utiliza duas enzimas hidrolíticas. A Exonuclease I remove primers residuais e produtos de PCR parcialmente amplificados, enquanto que a rSAP (Fosfatase Alcalina Recombinante de Camarão) remove dNTPs não incorporados.

A técnica de purificação por *ExoSap-IT*[®] consistiu na adição de 2 μ L de *ExoSap-IT*[®] a 5 μ L do produto amplificado. Esta solução passou por uma incubação a 37°C durante 15 minutos e posterior incubação a 80°C durante 15 minutos para inativação do *ExoSap-IT*[®]. Ambas as incubações foram programadas e efetuadas no termociclador *GeneAmp PCR system 9700* (*Applied Biosystems, Foster City, CA*). Concluído o processo de purificação as amostras ficaram aptas a serem sequenciadas.

7.6. Sequenciação

O objetivo desta fase foi determinar para todas as posições da região controle do mtDNA as suas bases nucleotídicas. Para tal foi utilizado o método de Sanger, descrito há mais de 30 anos e bastante utilizado atualmente para a sequenciação de DNA (Sanger *et al.*, 1977). Este processo envolve a incorporação de didesoxinucleótidos trifosfatados (ddNTPs) como terminadores do processo de amplificação. Na extremidade 3' dos ddNTPs não existe um grupo hidróxilo, o que torna impossível, a partir deles, estabelecer uma ligação fosfodiéster com um novo nucleótido na cadeia em desenvolvimento. Ou seja, a extensão termina quando a polimerase incorpora um ddNTP na cadeia sintetizada. No final da reação de sequenciação são obtidas sequências de mtDNA que diferem umas das outras em apenas uma base, partilhando a extremidade 5' mas possuindo um ddNTP diferente na extremidade 3' (Figura 9).

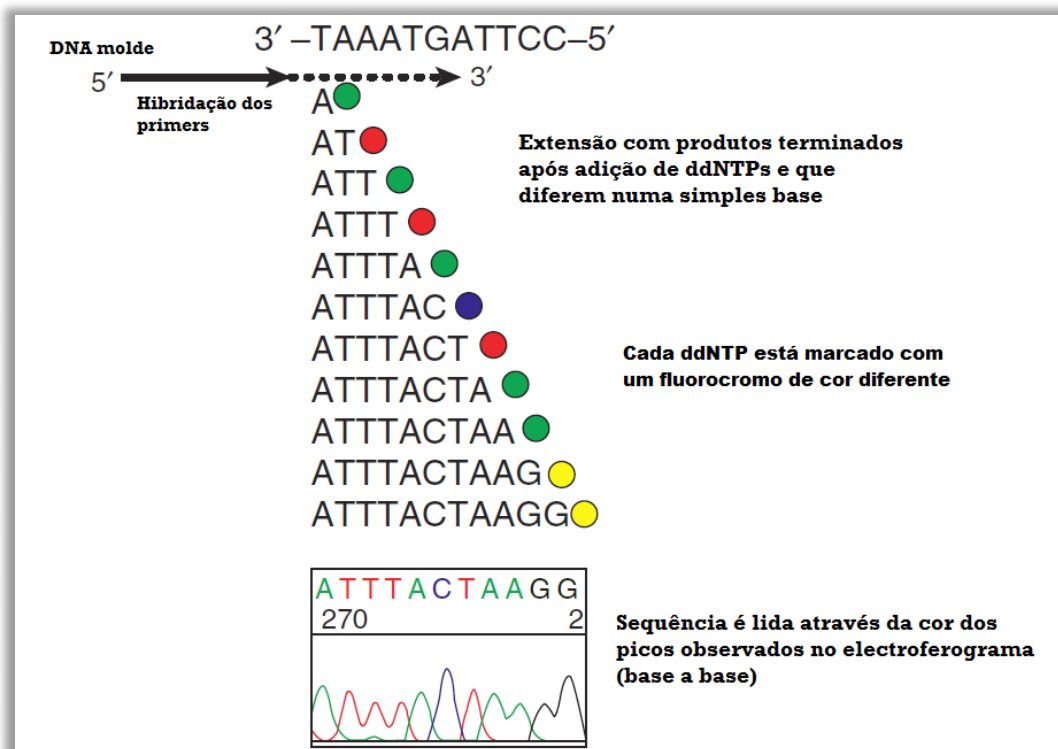


Figura 9 – Técnica de sequenciação com ddNTPs marcados com fluorocromos de diferentes cores; (Adaptado de Butler, 2011)

Os 4 tipos diferentes de ddNTPs (ddCTPs, ddATPs, ddGTPs e ddTTPs) foram adicionados em conjunto a cada amostra. Quatro marcadores fluorescentes com diferentes cores estão ligados a cada um dos diferentes ddNTPs, permitindo a distinção de cada tipo de nucleótido no momento de análise da sequência. O ddTTP

está marcado com um fluorocromo vermelho (dROX), o ddCTP está marcado com um fluorocromo azul (dR110), o ddATP está marcado com um fluorocromo verde (dR6G) e o ddGTP está marcado com um fluorocromo amarelo (dTAMRA), embora seja visualizado a preto.

A preparação da reação de sequenciação foi efetuada numa placa de 96 poços. Foi preparada uma solução com um volume final de 9 µL (Tabela 5), contendo tampão *Better Buffer*[®] (*Microzone Ltd., West Sussex, UK*), *Big Dye*[®] *Terminator v.3.1 Cycle Sequence* (*Applied Biosystems, Foster City, CA*) - mistura que contém DNA polimerase, dNTPs, ddNTPs (em menor quantidade que os dNTPs), MgCl₂, primers da cadeia a amplificar, a 2,5 µM, e o produto amplificado e purificado. A cadeia leve e a cadeia pesada de cada um dos dois fragmentos amplificados foram sequenciadas no termociclador *GeneAmp*[®] *PCR system 9700* (*Applied Biosystems, Foster City, CA*) segundo as condições apresentadas na tabela 6. O controlo negativo foi efetuado com as mesmas porções de reagentes e para cada um dos 4 primers utilizados na reação, substituindo-se o volume de DNA por 2 µL de água ultra pura.

As amostras sequenciadas foram submetidas a uma nova purificação antes da eletroforese.

Tabela 5 – Reagentes da reação de sequenciação e volume por amostra

Reagentes	Volume
<i>Better Buffer</i> [®]	4 µL
<i>Big Dye</i> [®] <i>Terminator v.3.1 Cycle Sequence</i>	2 µL
Primers a 2,5 µM	1 µL
Volume final da Mix	7 µL
DNA	2 µL

Tabela 6 – Condições da reação de sequenciação efetuada no termociclador *GeneAmp PCR system 9700*

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

7.7. Purificação dos produtos da reação de sequenciação

De forma a remover o excesso de ddNTPs e dNTPs não incorporados na sequenciação, bem como sais livres, o produto sequenciado foi purificado utilizando o kit *BigDye[®] X Terminator Purification Kit* (Applied Biosystems, Foster City, CA).

Foi preparada uma mistura de reação contendo, por amostra, 3,6 μL da solução *X Terminator* (constituído por esferas responsáveis pela captura dos ddNTPs e dNTPs não incorporados e também de sais livres nas amostras sequenciadas), 16,4 μL da solução SAM[™] (que aumenta a eficácia da solução *X Terminator* e estabiliza a amostra após a purificação) e água pura, num volume final de 50 μL . Esta solução foi adicionada a todas as amostras sequenciadas e colocada em constante agitação durante 30 minutos. Durante a agitação, os reagentes do kit capturam e imobilizam os componentes da solução não desejados. Por último, foi efetuada uma centrifugação a 1000 x g durante 2 minutos. Este passo permitiu que a fração insolúvel e os componentes capturados ficassem retidos no fundo do tubo permitindo que apenas os produtos sequenciados em suspensão fossem aspirados pelo capilar durante a eletroforese capilar (Figura 10). As amostras foram imediatamente submetidas a eletroforese capilar para deteção e posterior análise dos produtos sequenciados.

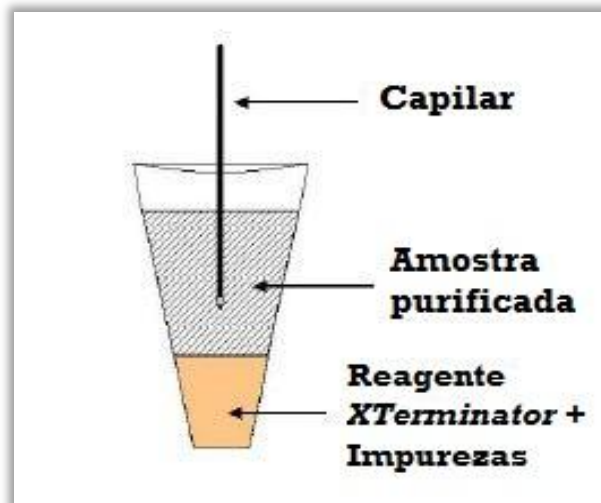


Figura 10 – Esquema da purificação por *X Terminator* e aspiração automática pelo capilar;
Adaptado de: http://www.appliedbiosystems.com/etc/medialib/appliedbio-media-library/images/application-and-technology/dna-sequencing-and-fragment-analysis/overview-of-dna-sequencing/sample-cleanup0/data-images.Par.78648.Image.235.229.1.gif.samplecleanup_data_fig2_gif.gif

7.8. Detecção dos produtos sequenciados

A sequência nucleotídica da região controlo das amostras em estudo foi obtida por eletroforese capilar no sequenciador automático *3130 Genetic Analyzer* (*Applied Biosystems, Foster City, CA*). Por cada eletroforese capilar foi preparada uma solução tampão contendo 3,5 mL de tampão com *EDTA* (10x, *Applied Biosystems*) e 31,5 mL de água ultra pura. O polímero utilizado foi o POP6, o mais adequado para a separação e detecção das sequências de mtDNA. As condições em que a eletroforese capilar ocorreu estão descritas na tabela 7.

Tabela 7 – Condições da eletroforese capilar realizada no *3130 Genetic Analyzer*

Sequenciador	<i>3130 Genetic Analyzer (Applied Biosystems)</i>
Kit de Reação	<i>Big Dye® Terminator v.3.1 Cycle Sequence (Applied Biosystems)</i>
Polímero utilizado	POP6
Temperatura	40°C
Estabilidade da corrente	5 µAmps
Voltagem da pré-corrída	15 kVolts
Tempo da pré-corrída	180 seg
Voltagem de injeção	1,2 kVolts
Tempo de injeção	12 seg
Voltagem da corrida	15 kVolts
Tempo de corrida	3500 seg

O sequenciador foi previamente preparado com a identificação das amostras através do programa *Genetic Analyzer Data Collection Software v3.0* (*Applied Biosystems, Foster City, CA*). A eletroforese capilar consistiu na migração das cadeias de mtDNA ao longo dos capilares (4 capilares, 1 capilar por amostra) do sequenciador do polo negativo (cátodo) para o polo positivo (ânodo). Durante a eletroforese os fragmentos de menor tamanho migram mais rapidamente que os fragmentos de maior tamanho. Os fragmentos passam por um feixe de laser que incide no capilar e excita os fluorocromos dos ddNTPs, o que provoca uma emissão de fluorescência de acordo com os diferentes fluorocromos. Esta fluorescência é detetada individualmente sob a forma de picos. Quanto maior for o sinal de fluorescência, maior a correspondência à respetiva base nucleotídica na sequência. No final é obtido um electroferograma que representa a sequência de mtDNA da região sequenciada, consoante o primer utilizado, o que permite a sua análise (Figura 11).

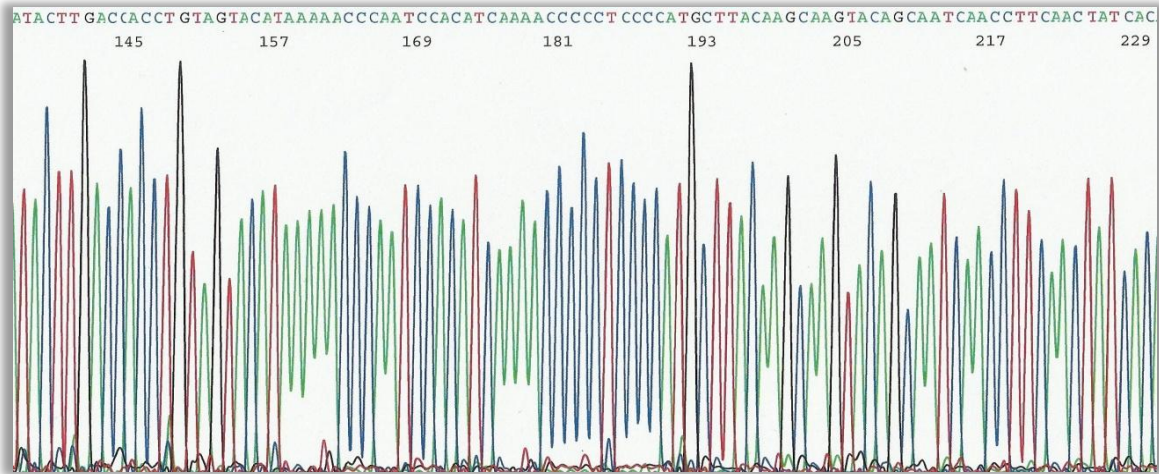


Figura 11 – Imagem de eletroferograma obtido após a eletroforese capilar no 3130 Genetic Analyzer

7.9. Análise dos produtos sequenciados

A sequenciação do mtDNA foi efetuada para a totalidade da região controle (entre a posição 16024 e a posição 576) e em ambas as cadeias, direta e reversa, de forma a comparar as sequências das cadeias complementares, com o objetivo de confirmar, com 2 resultados de ensaios diferentes, cada nucleótido das sequências analisadas (Parson & Bandelt, 2007).

A análise dos resultados da sequenciação da totalidade da região controle de mtDNA foi realizada recorrendo ao programa *Sequencing Analysis* v.5.2, que permitiu verificar a qualidade das sequências obtidas, e recorrendo ao programa *SeqScape* v.2.5 (Applied Biosystems, Foster City, CA), que permitiu comparar as sequências obtidas com a rCRS (Andrews *et al.*, 1999). A análise das sequências foi efetuada segundo as considerações da EDNAP (European DNA Profiling Group) e as recomendações de nomenclatura da IUPAC (International Union of Pure & Applied Chemistry) (Tabela 8) (Bär *et al.*, 2000). Na análise das sequências das amostras em estudo foram apenas identificadas as diferenças observadas após alinhamento com a rCRS, indicando a posição nucleotídica e o nucleótido presente nessa mesma posição. O haplótipo de cada amostra correspondeu às diferenças encontradas em relação à rCRS e respectivas posições nucleotídicas.

Tabela 8 – Nomenclatura de bases nucleotídicas de acordo com os códigos da IUPAC

Código da IUPAC	Base nucleotídica	Código da IUPAC	Base nucleotídica
A	Adenina	K	G ou T
C	Citosina	M	A ou C
G	Guanina	B	C ou G ou T
T (or U)	Timina (ou Uracilo)	D	A ou G ou T
R	A ou G	H	A ou C ou T
Y	C ou T	V	A ou C ou G
S	G ou C	N	Qualquer base
W	A ou T		

Na detecção de substituições na sequência de mtDNA relativamente à rCRS, ou seja, alteração de uma base nucleotídica por outra base nucleotídica numa determinada posição, apenas se apontou a posição onde se observou a alteração e o nucleótido que diferia da rCRS. Mutações por substituição podem ser transições (substituição de uma purina por outra purina ou de uma pirimidina por outra pirimidina) ou transversões (substituição de uma purina por uma pirimidina ou de uma pirimidina por uma purina). Citando um exemplo, a rCRS apresenta um “T” (timina) na posição 16519, mas alguns indivíduos apresentaram um “C” (citosina) nessa mesma posição, constituindo uma transição. Tal polimorfismo é assinalado como 16519C.

As inserções nas sequências de mtDNA analisadas, relativamente à rCRS, foram descritas indicando a posição imediatamente anterior à inserção, seguido de um ponto (.) e um (1) (para a primeira inserção), dois (2) (para uma segunda inserção) e assim sucessivamente, consoante o caso, e o nucleótido inserido. Como exemplo, a rCRS apresenta apenas 5 citosinas entre as posições 311 e 315. A existência de uma citosina extra antes da posição 316 é descrita como 315.1C.

As deleções nas sequências de mtDNA analisadas relativamente à rCRS foram identificadas assinalando a posição correspondente à deleção seguida de um d (ou DEL). No presente estudo foi frequente observar a ocorrência de uma dupla deleção nas posições 523 e 524, polimorfismo descrito como 523DEL e 524DEL.

Nas sequências de mtDNA com heteroplasmia de comprimento, devido a uma sobreposição de diversos padrões de leitura, não é possível analisar a sequência após a ocorrência de heteroplasmia de comprimento (Figura 12). Nestes casos, para

confirmar e seguir a regra estabelecida de 2 resultados de ensaios diferentes, foi necessário efetuar a sequenciação de cadeia leve e da cadeia pesada em duplicado.

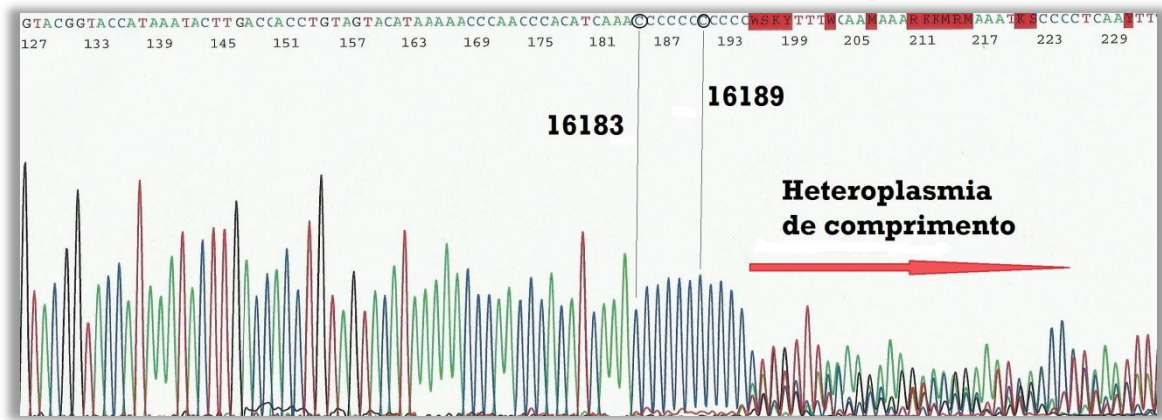


Figura 12 – Imagem de um eletroferograma de sequência da região controlo do mtDNA com heteroplasmia de comprimento

A determinação dos haplogrupos a que cada haplótipo pertence foi realizada de acordo com os polimorfismos detetados em relação à rCRS, encontrados na região controlo e com recurso à árvore filogenética desenhada por van Oven e Kayser (2009).

Os resultados obtidos foram organizados conforme os critérios da EMPOP (Parson *et al.*, 2004) e submetidos para análise e armazenamento.

7.10. Análise estatística e filogenética

A análise estatística foi efetuada calculando a diversidade nucleotídica e a diversidade de sequência (Nei & Tajima, 1981) da totalidade da região controlo do mtDNA das 103 amostras estudadas.

Para efetuar a análise molecular da variância (AMOVA) e obter medidas de distância genética interpopulacional, a fim de estabelecer as relações filogenéticas entre a população estudada e populações selecionadas da bibliografia, foi utilizado o programa *Arlequin* v.3.5.1.2 (Excoffier *et al.*, 2005). As distâncias genéticas foram avaliadas com base em valores de índice de fixação (F_{ST}). Para calcular as distâncias moleculares entre as sequências de mtDNA analisadas nas populações em estudo foi utilizado o programa *Phylip* v3.69. A obtenção da representação filogenética, em forma de árvore filogenética, foi alcançada recorrendo ao programa *Treeview* 1.5.2.

8. Resultados

8.1. Estudo dos haplótipos representados

Através do estudo efetuado foi possível analisar a totalidade da região controlo do mtDNA dos 103 indivíduos pertencentes à amostra e construir uma base de dados contendo os haplótipos de cada um dos indivíduos. O objetivo principal do presente trabalho consistiu na caracterização da diversidade genética da população imigrante de Cabo Verde residente em Lisboa através da comparação das suas sequências de mtDNA com a rCRS.

Na tabela 9 estão descritas as alterações polimórficas e respetivas posições nucleotídicas em relação à rCRS, resultantes da análise da totalidade da região controlo do mtDNA, que definem os haplótipos das amostras. Amostras com o mesmo haplótipo foram agrupadas com o mesmo código. Foi calculada a frequência de cada haplótipo na população estudada e os haplótipos foram agrupados no haplogrupo a que pertencem recorrendo à última versão da árvore desenhada por van Oven e Kayser (2009).

Os haplótipos obtidos neste estudo, após serem submetidos a um rigoroso controlo de qualidade, foram aceites para integrar a mais conceituada base de dados a nível internacional – EMPOP - e estarão disponíveis para consulta com o número de acesso EMP00616 (Parson & Dür, 2007) após a publicação do artigo científico que resultará do presente estudo.

No presente estudo foram identificados 75 haplótipos, 56 dos quais únicos (Tabela 10). A sequência de mtDNA mais frequente, foi observada em 6 das 103 sequências de mtDNA analisadas, sendo o seu haplótipo: **16111T, 16223T, 16278T, 16294T, 16309G, 16390A, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 523DEL, 524DEL**

Tabela 10 – Número de haplótipos únicos, repetidos e diferentes, considerando a região controlo total do mtDNA dos 103 indivíduos estudados

	Região controlo
Número de haplótipos únicos	56
Número de haplótipos repetidos	19
Número total de haplótipos diferentes	75

De acordo com os resultados apresentados na tabela 9, as sequências determinadas no estudo da totalidade da região controlo do mtDNA diferem da rCRS em 112 posições nucleotídicas, num total de 1651 polimorfismos.

Da análise da tabela 11, verifica-se que o tipo de polimorfismo mais frequente no presente estudo, na região controlo total, são os polimorfismos por substituição (1334 substituições) com uma razão entre transição/transversão de 24,65 (1282/52). O tipo de transição predominante foi entre pirimidinas (854).

Tabela 11 – Número e tipo de alterações polimórficas da região controlo total do mtDNA detetadas nas 103 sequências de mtDNA analisadas

Tamanho da região estudada	1178 pb		
Número de posições polimórficas	112		
Transições entre pirimidinas	854	T → C	392
		C → T	462
Transições entre purinas	428	A → G	307
		G → A	121
Transversões	52	C → G	2
		C → A	16
		A → T	1
		A → C	18
		G → T	13
		G → C	2
Inserções	155 (C)		
	1 (A)		
Delecções	83 (A)		
	78 (C)		
Número total de polimorfismos	1651		

Em todas as sequências verificou-se que existe uma transição de uma adenina (A) para uma guanina (G) na posição 263 e a inserção de uma citosina na posição 315. Apesar da presença, na rCRS, de uma adenina (A) na posição 263 e de cinco citosinas entre a posição 311 e 315, foram consideradas por Andrews e colaboradores (1999) como um polimorfismo raro detetado na CRS obtida por Anderson e colaboradores em 1981, não estando presentes na maioria dos seres humanos.

A posição 73 da região controlo do mtDNA apresenta na rCRS uma adenina (A). Em todas as sequências analisadas foi detetada uma substituição da adenina (A) por uma guanina (G), sendo este o polimorfismo por substituição mais frequente. Outras posições polimórficas apresentam elevada taxa de substituição, entre elas a posição 16223, que apresenta uma citosina (C) na rCRS e apresenta uma timina (T) em 98 das sequências analisadas (95,14%). A posição 152 apresenta uma timina (T) na rCRS e apresenta uma citosina (C) em 75 das sequências analisadas (72,82%), a posição 16278 apresenta uma citosina (C) na rCRS e apresenta uma timina (T) em 69 das sequências analisadas (66,99%). A posição 16519 apresenta uma timina (T) na rCRS e apresenta uma citosina (C) em 59 das sequências analisadas (57,28%).

Na região HV3 da região controlo do mtDNA é possível observar uma repetição de dinucleótidos AC (com 5 repetições no total) entre a posição 515 e a posição 524. Em 78 sequências das 103 analisadas (75,73%) verificou-se uma deleção de dinucleótidos AC, sendo este polimorfismo, segundo a nomenclatura (Bär *et al.*, 2000), assinalado como 523DEL e 524DEL. Destas 78 sequências, em 3, esta dupla deleção está associada a uma substituição por transição na posição 513 de uma guanina (G) para uma adenina (A). Numa sequência verificou-se a inserção de um dinucleótido AC na posição 524 (sendo esta sequência constituída por 6 repetições desse mesmo dinucleótido). Em 5 das amostras analisadas verificou-se a deleção de uma adenina (A) na posição 249, associada numa das 5 amostras a uma substituição por transição de uma guanina (G) para uma adenina (A) na posição 247.

As inserções foram detetadas, na sua maioria, no interior da região HV2 na zona dos poli-C (sequência de citosinas entre as posições 303 e 315, interrompida por uma timina na posição 310) com inserções de uma ou mais citosinas na posição 309. Foi também observada numa sequência a inserção de uma citosina na posição 356 e em 2 amostras a inserção de uma citosina na posição 16193.

Verificou-se a existência de grande variabilidade na zona de poli-C entre as posições 16183 e 16193. Esta zona é caracterizada por uma sequência de citosinas interrompida por uma timina na posição 16189. É frequente existirem substituições nesta zona e detetar a timina (T) noutra posição entre as posições 16183 e 16193. Observou-se a presença de uma timina (T) nas posições 16187, 16188 e 16192 (Tabela 12). O polimorfismo mais comum nesta região foi uma dupla transição, nas posições 16187 de citosina (C) para timina (T), e na posição 16189, de timina (T) para citosina (C).

Tabela 12 – Diferentes polimorfismos observados na zona de poli-C da região HV1 do mtDNA, entre as posições 16183 e 16193 e número de sequências onde ocorreu heteroplasmia de comprimento

	16183	16185	16187	16188	16189	16192	16193.1	
rCRS	A	C	C	C	T	C	-	Amostras
Polimorfismos			T		C			17
	C				C			10*
	C				C		C	1*
	C				C	T		1
					C	T		3
				T	C			2
					C		C	1*
					C			2
							T	1

* - Amostras cujas sequências apresentam, sempre, heteroplasmia de comprimento na região HV1

Na região controlo do mtDNA existem zonas onde o número de alterações polimórficas em relação à rCRS é menor. Entre as posições 16400 e 16500 apenas foram detetadas alterações polimórficas em 5 sequências, 2 transições de uma citosina (C) para uma timina (T) na posição 16400 e 3 transições de uma citosina (C) para uma timina (T) na posição 16465. Entre as posições 400 e 500 foram detetados apenas 14 polimorfismos, distribuídos pelas posições 418, 466, 482, 489, 494 e 499.

Nenhuma das sequências analisadas apresenta total concordância de bases nucleotídicas com a rCRS. O menor número de alterações polimórficas detetadas numa sequência, em relação à rCRS, foi de 9, em 9 das 103 sequências analisadas, enquanto que o maior número de alterações polimórficas encontrado numa sequência, em relação à rCRS, foi de 27, ocorrendo em 3 das 103 sequências analisadas. A frequência absoluta de cada polimorfismo encontrado nas 103 sequências analisadas encontra-se sumariada na tabela 13.

Tabela 13 – Frequência absoluta de cada nucleótido na região controle do mtDNA, considerando as alterações polimórficas relativamente à rCRS

Posição polimórfica	rCRS	A	G	C	T	Posição polimórfica	rCRS	A	G	C	T
16048	G	1				16259	C				1
16051	A		6			16261	C				2
16086	T			1		16264	C				20
16093	T			8		16269	A		2		
16111	C				13	16270	C				13
16114	C	9	2			16274	G	3			
16124	T			20		16278	C				69
16126	T			15		16290	C				1
16129	G	16				16292	C				1
16145	G	4				16293	A		17		
16148	C				1	16294	C				20
16163	A		3			16295	C				2
16169	C				2	16299	A		3		
16172	T			6		16304	T			4	
16177	A		1			16309	A		14		1
16178	T			1		16311	T			28	
16179	C				2	16320	C				10
16183	A			10		16355	C				8
16185	C				1	16360	C				5
16187	C				17	16362	T			20	
16188	C				2	15368	T			1	
16189	T			35		16380	C				1
16192	C				4	16390	G	36			
16193.1				2i		16399	A		4		
16209	T			2		16400	C				2
16212	A		1			16465	C				3
16213	G	7				16519	T			59	
16215	A		1			16527	C				1
16218	C				1	73	A		103		
16219	A		3			93	A		10		
16221	C				1	95	A			1	
16223	C				98	143	G	11			
16234	C				3	146	T			37	
16239	C				4	150	C				46
16249	T			1		151	C				12
16256	C	2				152	T			75	
16257	C				2	182	C				38

(C ontinua na próxima página.)

Na tabela: i = Inserção; d = Deleção

Tabela 13 – (continuação)

Posição polimórfica	rCRS	A	G	C	T	Posição polimórfica	rCRS	A	G	C	T
183	A		2			309.2					1i
185	G	2		2	13	315.1					103i
186	C	5				316	G	6			
189	A		11	7		325	C				13
195	T			63		356.1					1i
198	C				28	357	A		12		
200	A		2			373	A		1		
204	T			10		418	C				6
207	G	13				466	T				1
234	A		1			482	T				2
235	A		3			489	T				1
247	G	18				494	C				3
249	A	5d				499	G	1			
257	A		3			513	G	3			
263	A		103			523	A	78d			
264	C				3	524	C				78d
294	T			2		524.1		1i			
297	A		1			524.2					1i
309.1				47i		534	C				3

Na tabela: i = Inserção; d = Deleção

8.2. Estudo de heteroplasmias

Analisando a totalidade da região controlo do mtDNA dos 103 indivíduos com ascendência cabo-verdiana, foi observada uma taxa de heteroplasmia de comprimento de 13,59% (ocorrendo em 14 das 103 sequências obtidas). Todos os casos de heteroplasmia de comprimento encontrados foram confirmados pela sequenciação da cadeia leve e da cadeia pesada do mtDNA em duplicado.

Na tabela 14 é apresentada a frequência de haplótipos homoplásmicos e de haplótipos heteroplásmicos encontrados, com indicação da região hipervariável onde ocorreu a heteroplasmia. Entre as 103 sequências de mtDNA obtidas, 86,41% corresponderam a haplótipos homoplásmicos. Em 11 das 103 sequências (10,68%) observou-se a presença de heteroplasmia de comprimento associada à zona de poli-C da região HV1, entre a posição 16183 e a posição 16193. Em 2 das 103 sequências

(1,94%) observou-se a presença de heteroplasmia de comprimento associada à zona de poli-C da região HV2, entre a posição 303 e a posição 315. Em apenas 1 das 103 sequências foi observada heteroplasmia de comprimento em ambas as regiões hipervariáveis (HV1 e HV2) em simultâneo.

Tabela 14 – Frequência de haplótipos homoplásmicos e heteroplásmicos nas 103 sequências analisadas e região hipervariável de ocorrência

Haplótipos	Região hipervariável	Frequência
Homoplásmicos		89
Heteroplásmicos	HV1	11
	HV2	2
	HV1+HV2	1
	Total	14



Em nenhuma das 103 sequências de mtDNA analisadas foi observada heteroplasmia de posição.

8.3. Cálculo dos parâmetros de diversidade

O valor da diversidade nucleotídica (π) foi calculado para as sequências obtidas segundo a equação proposta por Nei e Tajima (1981):

$$\pi = (n/n-1) \times (1/l) \times \sum (1 - x_i^2)$$

Na fórmula apresentada n representa o número de indivíduos estudados, l o tamanho em pares de bases da sequência em estudo e x_i^2 representa a frequência relativa de cada nucleótido na posição i . Para o cálculo proposto são, portanto, necessários os seguintes dados:

-  **Tamanho da sequência em estudo – 1178 pb;**
-  **Número de indivíduos estudados – 103 indivíduos;**

O cálculo de x_i^2 foi efetuado calculando a frequência relativa para cada nucleótido (frequência absoluta sobre o tamanho da amostra). Para tal, foi determinado o número de vezes que cada nucleótido ocorreu como mutação na amostra populacional (Tabela 13). Por exemplo, a frequência absoluta da base nucleotídica adenina (A) é igual ao número de vezes que ela ocorreu como inserção,

delecção ou substituição, sobre o número de haplótipos diferentes encontradas para a região controlo total.

 **Número de sequências diferentes – 75 haplótipos diferentes.**

$$F(A) = 221/75 = 2,95$$

A frequência relativa é:

$$f(A) = 2,95/103 = 0,02835$$

Foi necessário, portanto, calcular a frequência relativa (a partir da tabela 13) para as quatro bases nucleotídicas, calculando-se depois o somatório de $1 - xi^2$.

Tabela 15 – Frequência absoluta e frequência relativa de cada nucleótido nas posições polimórficas da região controlo do mtDNA, considerando as alterações relativamente à rCRS, apresentadas na tabela 9

Nucleótido	A	G	C	T
Frequência absoluta	221/75	309/75	645/75	476/75
Frequência relativa (xi)	2,95/103	4,12/103	8,6/103	6,35/103
xi^2	0,00082	0,0016	0,00697	0,0038

$$\Sigma (1 - xi^2) = (1 - 0,00082) + (1 - 0,0016) + (1 - 0,00697) + (1 - 0,0038) = 0,99918 + 0,9984 + 0,99303 + 0,9962 = 3,98618$$

Por aplicação da fórmula apresentada, a diversidade nucleotídica (π) para a região controlo total do mtDNA no presente estudo, foi de:

$$\pi = (103/102) \times (1/1178) \times 3,98618 = 0,003417$$

Foi também calculado o valor da diversidade de sequência (h) de acordo com a equação proposta por Nei e Tajima (1981):

$$h = (n/n-1) \times (1 - \Sigma pi^2)$$

Na fórmula apresentada n representa o número de indivíduos estudados e pi^2 representa a frequência das diferentes sequências na população estudada. Como é possível verificar na tabela 9, todas as sequências analisadas apresentam alterações polimórficas em relação à rCRS. Considerando a região controlo total do mtDNA, 56 sequências surgem apenas uma vez, 14 sequências surgem, cada uma, 2 vezes, 3

sequências surgem, cada uma, 3 vezes, 1 sequência surge 4 vezes e 1 sequência surge 6 vezes.

O cálculo de π^2 foi realizado calculando a frequência de cada sequência na amostragem. Por exemplo, a frequência de sequências únicas é igual ao número de vezes que ela ocorre, neste caso uma vez, sobre o número total de indivíduos, 103, vezes o total de sequências únicas, 56.

$\pi^2 = (\text{frequência da sequência na amostra}/\text{n}^\circ \text{ total de amostras})^2 \times \text{número de sequências com a referida frequência}$

$$\begin{aligned} \pi^2 &= (1/103)^2 \times 56 = \mathbf{0,0052785} \\ \pi^2 &= (2/103)^2 \times 14 = \mathbf{0,0052785} \\ \pi^2 &= (3/103)^2 \times 3 = \mathbf{0,002545} \\ \pi^2 &= (4/103)^2 \times 1 = \mathbf{0,001508} \\ \pi^2 &= (6/103)^2 \times 1 = \mathbf{0,003393} \end{aligned}$$

Após o cálculo de π^2 , calculou-se o seu somatório:

$$\Sigma \pi^2 = 0,0052785 + 0,0052785 + 0,002525 + 0,001508 + 0,003393 = \mathbf{0,017983}$$

Por aplicação da fórmula apresentada, a diversidade de sequência (h) para a região controlo total, no presente estudo, é de:

$$h = (103/102) \times (1 - 0,017983) = 1,0098 \times 0,982017 = \mathbf{0,99164}$$

Estudos anteriores da população de Cabo Verde, Portugal e de populações de diversos países africanos, analisaram, principalmente, a região HV1 do mtDNA. De forma a ser possível efetuar uma comparação entre tais estudos e o presente trabalho, foram calculados os valores de diversidade nucleotídica e de diversidade de sequência apenas para essa região, das 103 amostras estudadas. Os valores correspondem à região HV1 do mtDNA da posição 16024 à posição 16365 e, exceto para os valores do presente estudo, não têm em conta inserções (Tabela 16).

De forma a compreender melhor os parâmetros calculados para a totalidade da região controlo, foi efetuada a comparação entre os valores de diversidade nucleotídica e diversidade de sequência, obtidos no presente estudo, com os valores

de diversidade encontrados em estudos de outras populações de origem africana nos quais toda a região controlo (16024-576) tenha, também, sido analisada (Tabela 17).

Tabela 16 – Parâmetros de diversidade da região HV1 do mtDNA de populações africanas e portuguesas

Populações	N	Diversidade de sequência	Diversidade nucleotídica	Referência bibliográfica
Presente estudo	103	0,9839	0,011	---
Cabo Verde	78	0,98335	0,0078	Tavares (2007)
Cabo Verde (SE e NW)	184; 108	0,984; 0,900	-	Brehm <i>et al.</i> (2002)
S. Tomé e Príncipe	16	0,912	-	Trovoada <i>et al.</i> (2004)
Angola	62	0,985	-	Trovoada <i>et al.</i> (2004)
Mauritânia	64	0,96	-	González <i>et al.</i> (2006)
Senegal	268	0,99	-	Rando <i>et al.</i> (1998)
Serra Leoa	277	0,99	-	Jackson <i>et al.</i> (2005)
Gana	191	0,98	-	Fendt <i>et al.</i> (2012b)
Marrocos	56	0,95	-	Turchi <i>et al.</i> (2009)
Moçambique	308	-	0,0247	Salas <i>et al.</i> (2003)
Portugal (Norte)	100	-	0,13	Pereira <i>et al.</i> (2000)
Portugal (Centro)	82	-	0,14	Pereira <i>et al.</i> (2000)
Portugal (Sul)	59	-	0,13	Pereira <i>et al.</i> (2000)
Madeira	155	0,9647	-	Brehm <i>et al.</i> (2003)
Açores	179	0,9415	-	Brehm <i>et al.</i> (2003)

Na tabela: N = Tamanho da amostra.

Tabela 17 – Parâmetros de diversidade da região controlo do mtDNA de populações africanas

Populações (etnia)	N	Diversidade de sequência	Diversidade nucleotídica	Referência bibliográfica
Presente estudo	103	0,992	0,003	---
Camarões (Baka)	27	0,855	0,003	Batini <i>et al.</i> (2011)
Camarões (Bakola)	11	0,909	0,001	Batini <i>et al.</i> (2011)
Gana	191	0,987	-	Fendt <i>et al.</i> (2012b)
Gabão (Baka)	20	0,889	0,001	Batini <i>et al.</i> (2011)
Gabão (Fang)	17	1	0,004	Batini <i>et al.</i> (2011)
Gabão (Nzebi)	19	0,977	0,004	Batini <i>et al.</i> (2011)
República Central Africana (Biaka)	23	0,968	0,003	Batini <i>et al.</i> (2011)
República Central Africana (Mbenzele)	30	0,954	0,001	Batini <i>et al.</i> (2011)
República Democrática do Congo (Babinga)	23	0,889	0,002	Batini <i>et al.</i> (2011)
República Democrática do Congo (Mbuti)	21	0,981	0,003	Batini <i>et al.</i> (2011)

Na tabela: N = Tamanho da amostra.

8.4. Estudo dos haplogrupos representados

No estudo dos 103 indivíduos, foram obtidos 75 haplótipos diferentes. Os haplogrupos foram definidos para cada haplótipo a partir das alterações polimórficas que cada sequência apresenta em relação à rCRS e da coincidência dessas mesmas variações relativamente à classificação existente. A classificação das 103 amostras em haplogrupos foi efetuada com recurso aos dados presentes na árvore filogenética desenhada por van Oven e Kayser (2009). Para uma correta classificação, é necessário que os haplótipos possuam todos os polimorfismos específicos do haplogrupo em questão, apesar de poderem apresentar outros para além dos específicos. Como exemplo, os haplótipos das amostras com os códigos CBV016 e CBV017, apesar de diferentes, possuem ambos os polimorfismos característicos do haplogrupo L3b (16124C, 16287T e 16362C), daí serem ambos incluídos no mesmo haplogrupo. O elevado número de polimorfismos obtidos em cada sequência, graças à análise da totalidade da região controlo do mtDNA, permitiu classificar todos os haplótipos nos haplogrupos. Como exemplo, o haplótipo da amostra com o código CBV004 apresenta o polimorfismo típico do macrohaplogrupo L (16223T). Apresenta ainda um polimorfismo partilhado pelos haplogrupos L3c e L3d (152C). A presença do polimorfismo 16124C, presente apenas no haplogrupo L3d, e o polimorfismo 16111T, característico do haplogrupo L3d2, permitem então efetuar a classificação final do haplótipo em questão no haplogrupo L3d2.

Os haplogrupos representados na tabela 8 apresentam diferente frequência na população em estudo (Figura 13). Apenas 8 das sequências não foram classificadas no macrohaplogrupo L característico de linhagens africanas (Behar *et al.*, 2008). Verificou-se que 95 indivíduos pertencem ao macrohaplogrupo L, o mais comum, com uma frequência de 92,24%. Entre os diversos haplogrupos do macrohaplogrupo L, o que teve maior representação (42 das 103 sequências em estudo) foi o haplogrupo L3, seguido do haplogrupo L2 (35 das 103 sequências em estudo) e do haplogrupo L1 (18 das 103 sequências em estudo). Não foi encontrada nenhuma amostra pertencente aos restantes haplogrupos do macrohaplogrupo L (L0, L4, L5 ou L6).

Dos haplogrupos representados na população em estudo, o que apresentou maior expressão foi o haplogrupo L2a1a, atribuído a 13 das 103 amostras em estudo (12,62%). Os restantes haplogrupos do haplogrupo L2 apresentaram uma expressão mais reduzida, entre eles o haplogrupo L2c (5,83%) e o haplogrupo L2b1a (3,88%). No que diz respeito ao haplogrupo L1, encontram-se representados em maior frequência nas amostras o haplogrupo L1b1 (4,85%) e o haplogrupo L1b1a2 (3,88%). No que diz respeito ao haplogrupo L3, encontram-se representados em maior frequência na população em estudo o haplogrupo L3d (8,74%) e o haplogrupo L3e4a (5,83%).

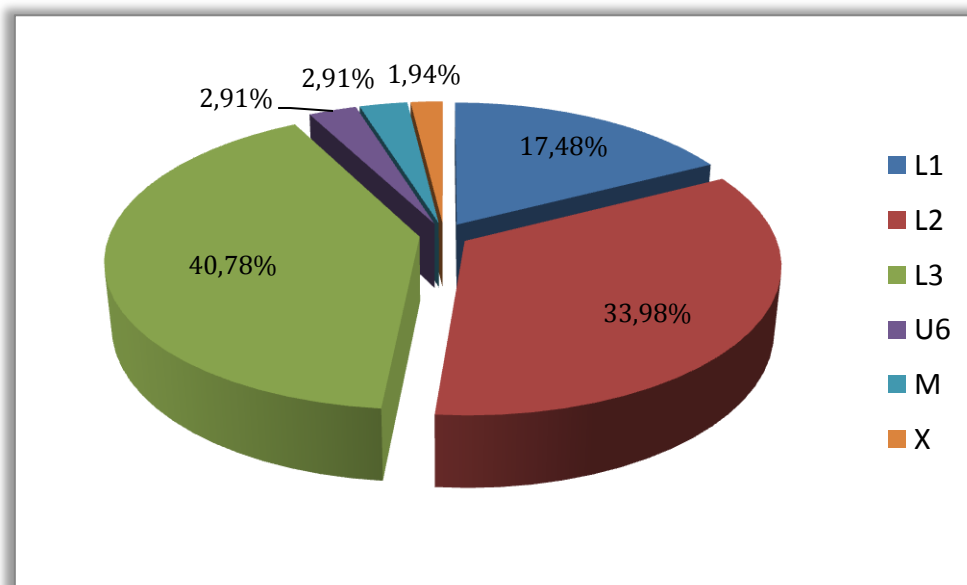


Figura 13 – Representação esquemática das percentagens observadas por cada haplogrupo para os 103 indivíduos com ascendência cabo-verdiana.

Entre os haplótipos não pertencentes ao macrohaplogrupo L, o haplogrupo U6, característico de populações do Norte de África (Pereira *et al.*, 2003) e definido pela presença dos polimorfismos 16172C, 16219G e 16278T, está representado em 3 das amostras estudadas (2,91%). Para o haplótipo da sequência com o código CBV029, a presença dos polimorfismos 16189C e 16278T permitiu a sua classificação no haplogrupo X. Contudo, não foi possível, com base nas restantes alterações polimórficas, classificar este haplótipo num dos sub-haplogrupos do haplogrupo X.

O haplótipo da sequência com o código CBV034 foi classificado no haplogrupo M e respetivo sub-haplogrupo M1 graças à presença dos polimorfismos

489C, 195C, 16129A, 16189C, 16249C e 16311C. Para o haplótipo codificado como CBV028, apesar de possuir os polimorfismos característicos do haplogrupo M62 (16295T e 150T), não possui o principal polimorfismo característico do macrohaplogrupo M (489C). O mesmo se passa com o haplótipo codificado como CBV047 por não possuir os polimorfismos característicos do haplogrupo M30 e respetivo sub-haplogrupo M30c (195A e 146C).

8.5. Análise filogenética

A análise molecular de variância (AMOVA) permite estimar a variabilidade genética entre grupos populacionais. As populações selecionadas da literatura, utilizadas para o estudo filogenético, são as constantes da tabela 18.

Tabela 18 – População do presente estudo e populações selecionadas da literatura, utilizadas no estudo filogenético

Populações (etnia)	N	Referência bibliográfica
Angola (Khoe-San)	129	Fendt <i>et al.</i> (2012a)
Presente estudo	103	---
Gana	192	Fendt <i>et al.</i> (2012b)
Somália	190	Mikkelsen <i>et al.</i> (2012)

Na tabela: N = Tamanho da amostra

A AMOVA foi aplicada a cada par de populações, permitindo obter a distância genética entre as populações (Tabela 19) e os respetivos valores de p (Tabela 20). A árvore obtida a partir das distâncias moleculares está ilustrada na Figura 14.

Tabela 19 – Valores de F_{st} entre as populações do estudo de distâncias genéticas

	Angola	Presente estudo	Gana	Somália
Angola	0,00000			
Presente estudo	0,30102	0,00000		
Gana	0,27968	0,02067	0,00000	
Somália	0,28884	0,06419	0,05866	0,00000

Tabela 20 – Valores de p relativos às populações comparadas no estudo

	Angola	Presente estudo	Gana	Somália
Angola	0.00000+- 0.0000			
Presente estudo	0.00000+- 0.0000	0.00000+- 0.0000		
Gana	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000	
Somália	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000

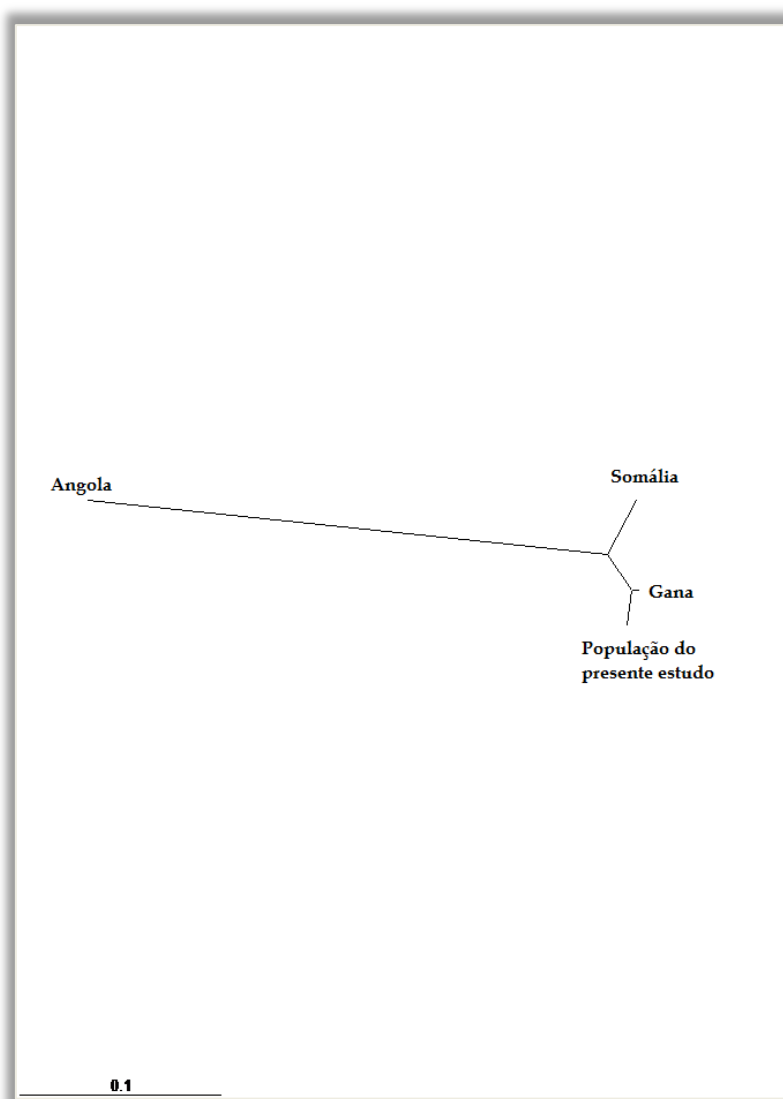


Figura 14 – Árvore resultante da análise das distâncias genéticas obtidas entre as populações referidas na tabela 18

9. Discussão

Este estudo traduz a, indiscutível, aplicabilidade da análise do mtDNA para a determinação da diversidade genética das populações humanas, bem como para o fornecimento das informações acerca da sua origem e evolução ao longo dos anos. A análise do mtDNA de 103 indivíduos com ascendência, quer paterna quer materna, de Cabo Verde permitiu recolher dados importantes sobre a origem do arquipélago, complementar outros estudos efetuados anteriormente com o mesmo objetivo e actualizar o conhecimento da composição genética da população de Lisboa. Cerca de 48% da população selecionada corresponde a indivíduos do sexo feminino, enquanto que cerca de 52% da população selecionada corresponde a indivíduos do sexo masculino.

9.1. Estudo da totalidade da região controlo do genoma mitocondrial

A utilidade do mtDNA em diversas áreas, entre elas a genética forense e a genética populacional, está bem documentada. Estudos anteriores baseados na análise da região controlo do mtDNA utilizavam primers destinados à análise de regiões mais reduzidas, entre elas as regiões HV1 e HV2, que por si só não permitiam obter o nível de discriminação que é possível alcançar nos dias de hoje com o estudo da totalidade da região controlo. Diversos autores demonstraram a possibilidade de utilizar apenas um par de *primers* mais afastados na amplificação do mtDNA, tornando evidente que se torna mais eficiente amplificar toda a região controlo numa única PCR. O presente estudo adotou um procedimento recente que se baseia na utilização de 2 pares de *primers* para amplificar toda a região controlo do mtDNA na mesma PCR em 2 fragmentos em separado, o que permite reduzir os custos, rentabilizar o tempo e, com maior eficácia e solidez, obter a informação de toda a região controlo do DNA mitocondrial (Parson & Bandelt, 2007).

Brehm e colaboradores (2003) estudaram 292 amostras de indivíduos de 7 ilhas do arquipélago de Cabo Verde e, para além da análise de 13 posições por RFLP da região codificante, sequenciaram apenas a região HV1 do mtDNA, entre as posições 16024 e 16365. Tavares (2007) analisou a região HV1 do mtDNA de 78 indivíduos com ascendência cabo-verdiana, da posição 16365 até à posição 16569, o que permitiu diferenciar alguns haplótipos idênticos no intervalo 16024-16365.

O presente estudo confirma que é possível diferenciar alguns haplótipos idênticos no intervalo 16024-16365 ao analisar a região HV1 do mtDNA até à posição 16569, dando como exemplo os haplótipos com os códigos CBV061 e CBV066 (diferentes polimorfismos assinalados a negrito):

CBV061	16114A	16129A	16213A	16223T	16278T	16355T	16362C	16390A	
CBV066	16114A	16129A	16213A	16223T	16278T	16355T	16362C	16390A	16465T

O conjunto de alterações polimórficas para os haplótipos codificados como CBV061 e CBV066 é idêntico considerando apenas o intervalo entre as posições 16024-16365. Contudo, o presente estudo permitiu detetar variações para além desse intervalo sequenciando até à posição 16569. O haplótipo com o código CBV066 apresenta uma alteração polimórfica, 16465T, não presente no haplótipo com o código CBV061, ou seja, e tendo em conta apenas a região HV1, os dois haplótipos são diferentes.

A análise da região HV1, sem considerar a restante região controlo, pode apresentar uma grande desvantagem na determinação dos haplótipos e classificação nos respetivos haplogrupos. Indivíduos que partilhem as mesmas alterações polimórficas na região HV1 poderão apresentar diferenças quando a restante região controlo é analisada, permitindo diferenciar sequências que antes se julgava serem iguais e que seriam, de forma errada, agrupadas no mesmo haplótipo e, conseqüentemente, no mesmo haplogrupo (Alvarez-Iglesias *et al.*, 2007). Como exemplo, temos os haplótipos CBV005 e CBV006 e os haplótipos CBV036 e CBV042 (diferentes polimorfismos assinalados a negrito):

CBV005	16223T	16320T	16519C	73G	150T	152C	195C	263G	315.1C
CBV006	16223T	16320T	16519C	73G	150T	195C	200G	263G	315.1C
CBV036	16223T	16278T	16390A	73G	146C	150T	152C	182T	195C
	198T	263G	315.1C	325T	523d	524d			
CBV042	16223T	16278T	16390A	73G	93G	95C	146C	150T	152C
	182T	195C	198T	263G	315.1C	325T	513A	523d	524d

Considerando apenas as alterações polimórficas detetadas na região HV1, os haplótipos com os códigos CBV005 e CBV006 e os haplótipos com os códigos CBV036 e CBV042 são coincidentes. Contudo, a análise da restante região controlo

do mtDNA permitiu identificar diferenças em ambos os pares de haplótipos, tendo em consideração as alterações polimórficas identificadas em toda a região estudada.

As diferenças apresentadas poderão ser cruciais no agrupamento dos haplótipos nos respetivos haplogrupos, podendo dois haplótipos coincidentes analisando apenas a região HV1 e, conseqüentemente, agrupados no mesmo haplogrupo, apresentar um conjunto diferente de alterações polimórficas e serem agrupados em haplogrupos diferentes ao analisar a totalidade da região controlo do mtDNA.

Como primeiro exemplo, os haplótipos codificados como CBV017 e CBV024 são idênticos considerando apenas as alterações polimórficas da região HV1. Ambos são agrupados no haplogrupo L3b por apresentarem os três polimorfismos característicos desse haplogrupo na região controlo do mtDNA (16124C, 16278T e 16362C). Contudo, os haplótipos são diferentes ao analisar a totalidade da região controlo. O haplótipo com o código CBV024 é agrupado no haplogrupo L3b1b por apresentar o polimorfismo 152C, característico desse mesmo haplogrupo, enquanto que o haplótipo com o código CBV017 se mantém agrupado no haplogrupo L3b.

CBV017	16124C	16223T	16234T	16278T	16362C	16519C	73G	263G	309.1C
	315.1C	523DEL	524DEL						
CBV024	16124C	16223T	16234T	16278T	16362C	16519C	73G	151T	152C
	263G	315.1C	523DEL	524DEL					

Como segundo exemplo, os haplótipos com os códigos CBV014 e CBV026 são coincidentes considerando apenas as alterações polimórficas na região HV1 do mtDNA. Ambos os haplótipos são agrupados no haplogrupo L3e2b por apresentarem os dois polimorfismos característicos desse haplogrupo na região HV1 (16172C e 16189C). Contudo, os haplótipos são diferentes ao analisar a totalidade da região controlo. O haplótipo com o código CBV026 é agrupado no haplogrupo L3e2b2 por apresentar o polimorfismo 152C, característico desse mesmo haplogrupo, enquanto o haplótipo com o código CBV014 se mantém agrupado no haplogrupo L3e2b.

CBV014	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C
	263G	315.1C							
CBV026	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	152C
	195C	263G	309.1C	315.1C					

As regiões HV1 e HV2 foram, na década anterior, o principal alvo de diversos estudos da região controlo do mtDNA (Alonso *et al.*, 2002; Brandstätter & Parson, 2003; Carvalho *et al.*, 2003; Nagai *et al.*, 2003). A atual análise da totalidade da região controlo permite introduzir o estudo do segmento HV3 e, conseqüentemente, a análise de posições nucleotídicas características de vários haplogrupos nessa mesma região. A importância do estudo dessas mesmas posições foi mais uma vez verificada no presente estudo.

O haplótipo codificado como CBV010 está agrupado no haplogrupo L3k, caracterizado por um conjunto de três alterações polimórficas na região controlo do mtDNA (152C, 235G e 494T). Sem a análise da totalidade da região controlo, incluindo a região HV3 à qual pertence o polimorfismo 494T foi identificado, não seria possível agrupar com o mesmo grau de confiança o haplótipo codificado como CBV010 no haplogrupo L3k. A introdução do estudo do segmento HV3 permite aumentar a discriminação da análise da região controlo do mtDNA.

CBV010	16223T	16390A	73G	150T	152C	235G	263G	309.1C
	315.1C	494T						

Os exemplos apresentados evidenciam a vantagem e maior poder de discriminação da análise da totalidade da região controlo efetuada no presente estudo, em amostras de indivíduos com ascendência cabo-verdiana, em comparação com os estudos anteriormente efetuados, garantindo maior confiança na determinação dos respetivos haplogrupos. Este método tem sido utilizado em estudos recentes de outras populações (Brandstätter *et al.*, 2008; Fendt *et al.*, 2012a; Gómez-Carballa *et al.*, 2012).

Apesar de a análise da totalidade da região controlo ser ainda considerada por toda a comunidade científica como um dos procedimentos mais eficazes na análise do DNA mitocondrial, a sequência da região controlo pode não ser suficiente para determinar o haplogrupo. A árvore filogenética desenhada por Van Oven e Kayser (2009) tem em consideração alterações polimórficas quer da região controlo quer de toda a região codificante do mtDNA, ou seja, determinadas posições na região

codificante podem ser importantes na análise dos haplótipos obtidos, garantindo uma total eficácia na determinação do haplogrupo. Como exemplo:

O haplótipo com o código CBV003 é agrupado no haplogrupo L3d devido à presença dos polimorfismos 152C e 16124C. Contudo, o haplogrupo L3d é também caracterizado pela presença dos polimorfismos 5147A, 7424G, 8618C, 13886C e 14284T, todos eles da região codificante. A análise da região codificante e respetiva deteção de todas as alterações polimórficas apresentadas permitiria agrupar o haplótipo no haplogrupo L3d com um maior grau de confiança.

CBV003 16086C 16124C 16223T 73G 152C 263G 315.1C 523DEL 524DEL

As alterações polimórficas detetadas na região controlo permitem o agrupamento do haplótipo com o código CBV070 no haplogrupo L2b1a, devido à presença dos polimorfismos 16114A, 16129A, 16213A, 16355T, 16362C, 198T, 204C e 418T. Contudo, a análise da região codificante poderia permitir que este haplótipo fosse agrupado num dos sub-haplogrupos de L2b1a. O haplogrupo L2b1a2 é caracterizado pela presença do polimorfismo 7569G, o haplogrupo L2b1a3 é caracterizado pela presença do polimorfismo 8856A e o haplogrupo L2b1a4 é caracterizado pela presença do polimorfismo 12406A.

CBV070 16114A 16129A 16213A 16223T 16261T 16278T 16355T 16362C 16390A
16465T 16519C 73G 150T 152C 182T 195C 198T 204C
249DEL 263G 309.1C 315.1C 418T 523DEL 524DEL

Segundo os exemplos apresentados, a aplicação da análise da região codificante na amostra estudada, para além dos resultados obtidos para a região controlo, permitiria maior qualidade e rigor no momento de analisar e agrupar o conjunto de haplótipos obtidos nos seus respetivos haplogrupos. Apresenta-se como uma abordagem a aplicar em futuros projetos de estudo do mtDNA da população de Cabo Verde e da população cabo-verdiana imigrante residente em Lisboa, seja através do estudo de SNPs por todo o genoma mitocondrial ou da sequenciação da totalidade do genoma mitocondrial, métodos já utilizados em estudos muito recentes em outras populações (Malyarchuk *et al.*, 2008; Soares *et al.*, 2008; Alvarez-Iglesias *et al.*, 2009; Gómez-Carballa *et al.*, 2012). Não obstante os impedimentos das normativas legais nacionais e internacionais à utilização de regiões codificantes no

âmbito de identificação humana, no âmbito exclusivo de estudos de investigação, quer forenses quer populacionais, nada impede o estudo de todo o genoma mitocondrial.

9.2. Análise dos resultados obtidos

Os resultados obtidos no presente estudo foram publicados com sucesso, após cuidada deliberação, na base de dados com maior importância a nível mundial, a EMPOP, o que confirma a viabilidade dos resultados e nos permitir tirar ilações a partir dos mesmos.

O estudo da região controlo do mtDNA de 103 indivíduos com ascendência cabo-verdiana permitiu a obtenção de 75 haplótipos diferentes, entre os quais 56 (cerca de 75%) são únicos. Esta elevada frequência de haplótipos únicos ocorre em populações isoladas, no local onde atualmente habitam há muito tempo (Batini *et al.*, 2011; Cardoso *et al.*, 2012), o que se verifica no caso de Cabo Verde ao estarmos na presença de um arquipélago que se estabeleceu como colónia do ex-Império Colonial Português há mais de 500 anos.

O haplótipo mais comum na amostra em estudo (CBV035), constituído pelas seguintes posições polimórficas - **16111T, 16223T, 16278T, 16294T, 16309G, 16390A**, 73G, 143A, 146C, 152C, 195C, 263G, 315.1C, 523DEL, 524DEL - foi também o mais comum (considerando apenas HV1, correspondendo às posições descritas a negrito) nos estudos anteriormente efetuados para a população cabo-verdiana, quer por Brehm e seus colaboradores (2003) quer por Tavares (2007). Apenas a análise da totalidade da região controlo permitiu agrupar o haplótipo com o código CBV035 no haplogrupo L2a1a, devido à presença do polimorfismo 143A, não detetável nos estudos anteriores onde apenas a região HV1 foi analisada pelo que, nesses estudos haplótipo foi agrupado no haplogrupo L2a.

A existência de haplótipos coincidentes entre indivíduos não aparentados demonstra que o mtDNA não é um marcador genético indicado para individualização. A utilidade do mtDNA para casos forenses dependerá da frequência com que os haplótipos obtidos se observam na população em geral. Alguns haplótipos poderão ser bastante raros e serem encontrados apenas, por exemplo, numa determinada família ou numa determinada região, o que permite

restringir a identidade de um indivíduo de acordo com o haplótipo obtido. Bases de dados, como a EMPOP, são de crucial importância. O haplótipo com o código CBV035, apesar de encontrado em 6 das 103 sequências analisadas no presente estudo, apenas coincide com 1 sequência entre 25701 dos haplótipos presentes na base de dados da EMPOP, o que indica que não é um haplótipo frequente.

Quanto às mutações observadas, existe uma maior frequência de substituições do que de inserções e deleções na região controle do mtDNA das 103 amostras em estudo, o que vai de encontro a estudos semelhantes como já tinha sido demonstrado por Carvalho e colaboradores (2003). O tipo de polimorfismo predominante na região controle do mtDNA foi a substituição por transição pirimidínica. O menor número de alterações polimórficas encontradas, numa sequência em relação à rCRS foi de 9 alterações polimórficas, enquanto o maior número de alterações polimórficas, em relação à rCRS, foi de 27 alterações polimórficas. A média de variações nucleotídicas foi de 17, o que está de acordo com os números de Budowle e colaboradores (1999) que determinaram a média característica de variações em populações de descendência africana em 15.

No estudo efetuado foram detetadas 112 posições polimórficas em relação à rCRS, com diferentes frequências de ocorrência na amostra (Tabela 13). Foram encontradas algumas alterações polimórficas com uma frequência bastante reduzida na população em estudo. O polimorfismo (A-G) na posição 16177 foi encontrado em apenas uma das sequências analisadas no presente estudo mas já foi descrito anteriormente por Tavares (2007) em cinco sequências e em uma das sequências analisadas por Brehm e colaboradores (2002). O polimorfismo 16114G foi observado em apenas duas das sequências analisadas no presente estudo mas foi descrito por Tavares (2007) e por Fendt e colaboradores (2012b). O polimorfismo 294C, característico do haplogrupo L3h1b2, apenas foi encontrado, para além de em 2 sequências analisadas no presente estudo, na sequência de um indivíduo de Guiné-Bissau (Behar *et al.*, 2008) e na sequência de um indivíduo de Marrocos (Harich *et al.*, 2010). O polimorfismo 494T foi encontrado em 2 das 103 sequências analisadas, tendo sido registado em algumas populações africanas, como exemplo na sequência de um indivíduo do Gana (Fendt *et al.*, 2012b), na sequência de um indivíduo do

Burkina Faso (Pereira *et al.*, 2010) e na sequência de um indivíduo da Líbia (Behar *et al.*, 2008).

A inserção de uma citosina na posição 315 ocorreu em todas as sequências analisadas. A inserção de citosinas (C) na posição 309 e na posição 16193 ocorrem com frequência em populações africanas (Fendt *et al.*, 2012b; Mikkelsen *et al.*, 2012), o que também se registou no presente estudo. A inserção de uma citosina (C) na posição 356 foi detetada em apenas uma das amostras analisadas no presente estudo. Esta inserção foi detetada em apenas 3 das 4265 sequências analisadas num estudo recente por Behar e colaboradores (2012). A inserção de uma repetição do dinucleótido AC na posição 524, detetada em 1 das 103 sequências analisadas, também já foi referenciada em estudos anteriores (Brandstätter *et al.*, 2004; Fendt *et al.*, 2012b). A dupla deleção (AC) nas posições 523 e 524 ocorreu em 78 das 103 sequências analisadas. Esta dupla deleção tem elevada frequência em populações africanas tal como demonstrado em estudos recentes (Fendt *et al.*, 2012b; Mikkelsen *et al.*, 2012).

A taxa de heteroplasmia de comprimento para a região HV1 foi de 11,65% e para a região HV2 foi de 2,91%. Analisando a totalidade da região controlo do mtDNA, a taxa de heteroplasmia de comprimento foi de 13,59%. No estudo efetuado por Tavares (2007), foi observada uma taxa de heteroplasmia de comprimento na região HV1 de 11,54%, pelo que os valores encontrados no presente estudo se enquadram nos anteriormente referenciados. Valores similares foram também encontradas em estudos realizados para outras populações africanas, entre eles por Fendt e colaboradores (2012b), na população do Gana, onde foi verificada uma taxa de heteroplasmia de comprimento para a totalidade da região controlo de cerca de 20%, e também por Mikkelsen e colaboradores (2012), na população da Somália, onde foi verificada uma taxa de heteroplasmia de comprimento para a totalidade da região controlo de cerca de 26%. Note-se que em ambos os estudos referenciados anteriormente a taxa de heteroplasmia de comprimento em HV1 foi superior à taxa de heteroplasmia de comprimento em HV2, o que está de acordo com os resultados alcançados no presente trabalho de investigação.

A não presença de qualquer heteroplasmia de posição na população de Cabo Verde analisada no presente estudo vai de encontro aos estudos anteriores realizados

para a população cabo-verdiana, onde apenas um caso foi detetado (16266Y) (Tavares, 2007).

O cálculo dos parâmetros de diversidade foi efectuado tendo em consideração a totalidade da região controlo, tendo sido alcançado um valor de diversidade nucleotídica elevado, de 0,00341 e um valor de diversidade de sequência também elevado, de 0,99164. Um estudo comparativo entre os valores de diversidade encontrados no presente trabalho e os valores de diversidade encontrados em estudos de populações de outros países africanos foi efetuado tendo em conta a totalidade da região controlo. A diversidade de sequência encontrada na população estudada foi a mais elevada de entre as populações africanas incluídas na análise (Tabela 17), encontrado-se próxima dos valores da população do Gana. O valor da diversidade nucleotídica encontrado na população estudada foi similar aos valores apresentados por todas as populações africanas incluídas na análise (Tabela 17). Os elevados índices de variabilidade obtidos estão de acordo com o estado de arte, de acordo com a qual são as populações africanas que apresentam maior variabilidade (Chen *et al.*, 2000).

Ambos os valores de diversidade foram calculados apenas para a região HV1, com o objetivo de realizar o estudo comparativo entre os valores de diversidade do presente trabalho e os valores de diversidade de estudos anteriores da população de Cabo Verde, nos quais se analisou apenas essa região hipervariável. Os valores encontrados no presente estudo são semelhantes aos valores de diversidade já publicados (Tabela 16). Os valores de diversidade obtidos podem estar relacionados com o estabelecimento da ascendência materna cabo-verdiana através de escravas provenientes de diversos territórios do continente africano e o isolamento geográfico do arquipélago que se localiza a mais de 500 km da costa ocidental africana, permitindo a retenção da elevada diversidade.

Todos os haplótipos dos 103 indivíduos, não aparentados e com ascendência cabo-verdiana, foram analisados e agrupados nos seus respetivos haplogrupos, sendo que 92,24% dos haplótipos pertencentes a haplogrupos característicos do continente africano (macrohaplogrupo L). Estudos anteriormente realizados à população cabo-verdiana alcançaram resultados semelhantes aos encontrados no presente estudo. Brehm e colaboradores (2002) obtiveram uma percentagem de haplótipos

pertencentes a haplogrupos característicos da África subsariana superior a 93%, enquanto Tavares (2007) obteve também uma percentagem próxima dos 90%. Estes valores vão de encontro aos dados históricos que atestam que a origem da composição genética de Cabo Verde está, na sua maioria, associada ao contributo de escravas africanas durante a colonização do arquipélago (Willie, 2001).

Comparando os haplogrupos obtidos com maior frequência no presente estudo com os estudos anteriores do mtDNA de Cabo Verde, Brehm e colaboradores (2002) obtiveram resultados semelhantes, com predominância dos haplogrupos L2a (29,1%), L2c (16,1%), L3e4 (10,9%), L1b (7,8%), L3b (7,2%), L1c (6,8%), L3d (6,8%), L2b (4,1%) e L3e2 (4,1%). Tavares (2007) por sua vez obteve uma predominância dos haplogrupos L2a (19,2%), L2c (14,1%), L1b (11,5%), L3e4 (9%), L3b (6,4%), L3d (3,8%) e L1c1 (3,8%). Aparentemente, a composição genética da população imigrante de Cabo Verde a residir atualmente em Lisboa é representativa da composição genética do arquipélago de Cabo Verde.

O haplogrupo com maior representação na amostra foi o haplogrupo L2a1a. O haplogrupo L2a (e seus respetivos sub-haplogrupos) é o mais frequente e com uma representação em todo o continente africano, o que torna a sua origem geográfica difícil de identificar. Tem sido encontrado com frequência em diversos países de África, entre eles Senegal, Moçambique, Gana e Somália (Rando *et al.*, 1998; Pereira *et al.*, 2001; Fendt *et al.*, 2012b; Mikkelsen *et al.*, 2012). Estudos anteriores têm sugerido, contudo, que tenha tido origem no centro de África, expandindo-se depois, quase em simultâneo, para o oeste e para o este africano (Kivisild *et al.*, 2004; Behar *et al.*, 2008). O haplogrupo L2a1a tem sido associado à expansão de grupos populacionais ao longo da costa ocidental africana até ao sudeste africano, tal como o comércio transatlântico de escravos entre o séc. XV e o séc. XIX. Esta teoria é suportada por estudos que revelaram a presença do haplogrupo L2a1a em sequências de mtDNA do Mali e da Mauritânia (González *et al.*, 2006) e em populações do sudeste de África, especialmente em Moçambique (Pereira *et al.*, 2001), uma das colónias do ex-Império Colonial Português.

Segundo os registos históricos, o comércio de escravos estabelecido em Cabo Verde foi efetuado entrando em contacto com postos de comércio espalhados por países da costa ocidental do continente africano, especialmente no Senegal, no Gana

e outras colônias portuguesas como São Tomé e Príncipe e Angola (Willie, 2001). Os dados obtidos no presente estudo vão de encontro aos registos históricos, com uma elevada frequência de haplogrupos com origem na região oeste de África. O haplogrupo L1b observa-se com elevada frequência em populações do oeste de África (Rosa *et al.*, 2004; González *et al.*, 2006; Fendt *et al.*, 2012b), especialmente nos Mandenka e nos Wolof do Senegal (Rando *et al.*, 1998; Jackson *et al.*, 2005). Os haplogrupos L2b e L2c têm também como origem o oeste de África, um pouco por toda a costa ocidental, com uma expansão semelhante à dos haplogrupos L1b, encontrando-se em elevada frequência em países como na Mauritânia, Senegal, Serra Leoa e Gana (Rando *et al.*, 1998; Jackson *et al.*, 2005; González *et al.*, 2006; Behar *et al.*, 2008; Fendt *et al.*, 2012b). Entre o haplogrupo L3, que se encontra disperso por todo o continente africano, quer os haplogrupos L3b quer os haplogrupos L3d ocorrem com maior frequência no quadrante oeste de África subsariana (Rando *et al.*, 1998; Rosa *et al.*, 2004; Jackson *et al.*, 2005; González *et al.*, 2006), com uma prevalência média de 10%, de acordo com os resultados obtidos no presente estudo (8,74% para ambos os haplogrupos). O haplogrupo L3e2 encontra-se com moderada frequência no oeste e centro de África (Salas *et al.*, 2002; Batini *et al.*, 2011; Fendt *et al.*, 2012b). O haplogrupo L3e4 está essencialmente restrito à costa ocidental africana (Rosa *et al.*, 2004).

Duas das 103 sequências de mtDNA foram agrupadas no haplogrupo L3h1b (uma delas no sub-haplogrupo L3h1b2). Foi descrito pela primeira vez em 2004 numa população de Guiné-Bissau (Rosa *et al.*, 2004). O haplogrupo L3k, verificado em três das 103 sequências de mtDNA (duas delas no sub-haplogrupo L3k1), foi descrito por Behar e colaboradores (2008) em sequências de indivíduos da Tunísia e Líbia. O haplogrupo L3f1b, verificado em uma das 103 sequências de mtDNA, foi detetado em vários países africanos, entre eles Guiné-Bissau, Egito e Etiópia (Behar *et al.*, 2008).

Entre os haplótipos que não pertencem a haplogrupos L, o haplogrupo U6 apresentou uma expressão de 2,91%, com 2 sequências agrupadas no haplogrupo U6a1 (1,94% da amostra estudada) e 1 sequência agrupada no haplogrupo U6a1a (0,98% da amostra estudada). O haplogrupo U6 é característico de populações do norte de África (Pereira *et al.*, 2003), sendo o sub-haplogrupo U6a o que apresenta

maior frequência e diversidade no noroeste africano (Rosa & Brehm, 2011). Os resultados obtidos estão de acordo com os indicados em estudos anteriores da população de Cabo Verde, com Tavares (2007) a obter uma percentagem de haplótipos pertencentes ao haplogrupo U6 de 5,12% e Brehm e colaboradores (2002) a obter uma percentagem de haplótipos pertencentes ao haplogrupo U6 de 3,96%. A introdução de linhagens mitocondriais do norte de África poderá dever-se ao facto de alguns dos primeiros escravos levados por portugueses para Cabo Verde serem originários da costa da Mauritânia, teoria avançada por Brehm e colaboradores (2002) e reforçada no presente estudo. Não foram obtidos dados sobre um possível contributo de *Guanches* das Ilhas Canárias na colonização do arquipélago, devido à ausência do haplogrupo U6b, característico desse grupo populacional (Rando *et al.*, 1998), na amostra em estudo.

Duas das 103 sequências de mtDNA analisadas foram agrupadas no haplogrupo X. Brehm e colaboradores (2003) também agruparam um dos 122 haplótipos da sua amostra de Cabo Verde nesse haplogrupo. O haplogrupo X é tipicamente encontrado em determinadas populações da América do Norte e do continente europeu (Brown *et al.*, 1998). Este haplogrupo foi também encontrado em pequena percentagem em grupos populacionais da Península Ibérica (Lima *et al.*, 2006; Álvarez *et al.*, 2007; Cardoso *et al.*, 2012) e do norte e nordeste de África (Reidla *et al.*, 2003).

Um dos 75 haplótipos obtidos no presente estudo foi agrupado no haplogrupo M1. É o único haplogrupo M presente em África, encontrando-se um pouco por todo o continente africano, principalmente no nordeste e este de África. Existem também, contudo, ocorrências ocasionais do haplogrupo M1 no oeste e noroeste de África (Rosa *et al.*, 2004; Ottoni *et al.*, 2009). González e colaboradores (2007) propuseram que o haplogrupo M1 se expandiu principalmente para o noroeste africano e, inclusivé, até à Península Ibérica, o que poderá explicar a ocorrência do haplótipo na população estudada.

A obtenção de uma reduzida frequência de haplótipos agrupados em haplogrupos europeus reflete a pequena contribuição europeia de linhagens femininas, na origem das atuais linhagens maternas de Cabo Verde, durante a colonização portuguesa levada a cabo no arquipélago. O mesmo cenário foi

verificado por Brehm e colaboradores (2002), que obtiveram apenas 7 sequências de mtDNA agrupadas nos haplogrupos N, X, T, pre-HV e pre-V; e por Tavares (2007) que obteve apenas 3 sequências de mtDNA classificadas nos haplogrupos J2, K e V, característicos de populações europeias. Estes resultados estão de acordo com dados históricos onde severamente se limitou a participação de mulheres europeias entre os colonizadores (Willie, 2001).

Os resultados obtidos apontam, como esperado, para uma elevada contribuição de escravas provenientes da costa ocidental africana na origem da composição genética das linhagens maternas atuais do arquipélago de Cabo Verde, fruto de uma atividade significativa de comércio transatlântico de escravos protagonizado pelo ex-Império Colonial Português do séc. XV ao séc. XIX. O haplótipo do cromossoma Y da população de Cabo Verde, obtida por Gonçalves e colaboradores (2003), permite confirmar a influência paterna europeia e do Médio Oriente na origem da atual população cabo-verdiana. Estes dados permitem reforçar os dados históricos onde se consta que a colonização e povoamento do arquipélago de Cabo Verde se efectuou inicialmente através do cruzamento de homens provenientes, essencialmente, de território europeu, e de escravas africanas.

A integração de linhagens de outros continentes em Lisboa, incluindo as linhagens L detetadas no presente estudo com origem em Cabo Verde, indicam que a composição genética da capital portuguesa está a ser alterada, como resultado de recentes movimentos de imigração de diversos países. É necessário, portanto, atualizar as bases de dados existentes para que possam acompanhar tais eventos e para que sejam corretamente representativas da população atual de Lisboa.

A obtenção de, na sua maioria, haplogrupos característicos da população africana reforça a ideia de que a origem de linhagens L em território europeu se deve a uma chegada recente a partir do continente africano e não de uma evolução de haplogrupos L no território europeu. No estudo atual, a presença de linhagens L em Lisboa terá sido provocada pelo aumento significativo do fluxo migratório de Cabo Verde para vários países europeus, durante, e especialmente após, o cessar das actividades de comércio de escravos na região no séc. XIX e XX.

9.3. Significado filogenético

Para o estudo filogenético foram selecionadas populações do continente africano para as quais foi analisada a totalidade da região controlo do mtDNA. O número reduzido de estudos até hoje para a totalidade da região controlo em populações do continente africano explica o reduzido número de populações consideradas para o estudo comparativo.

A determinação da distância genética interpopulacional foi efetuada para cada par de populações, através da obtenção da matriz de distâncias (Tabela 19) e respetivos valores de p (Tabela 20). Em relação à matriz de distâncias, verificou-se que a população de Cabo Verde se encontra geneticamente mais próxima da população do Gana (Figura 14). Em relação aos valores de p, valores maiores ou iguais a 0,05 indicam que não existe diferenças significativas entre as populações. Segundo os resultados obtidos, verificou-se que a população analisada no presente estudo apresente valores de p inferiores a 0,05 quando comparada com qualquer uma das populações do estudo, o que significa que Cabo Verde apresenta diferenças significativas quando comparada com todos os outros grupos populacionais.


Fendt e colaboradores (2012a) analisaram 129 sequências de mtDNA de indivíduos Khoe-San de Angola, uma população indígena dedicada essencialmente à agricultura e restritamente associada a territórios do sul de África, tais como Angola, Botswana e África do Sul. A população estudada por Fendt e colaboradores foi agrupada na sua maioria em haplogrupos L0d e L0k, haplogrupos praticamente restritos da população Khoe-san (Kivisild *et al.*, 2004; Behar *et al.*, 2008), o que explica a distância genética encontrada entre esta população e a do presente estudo.


As populações do Gana e da Somália consideradas para o estudo comparativo são compostas por 190 sequências de mtDNA de indivíduos da Somália (Mikkelsen *et al.*, 2012) e 192 sequências de mtDNA de indivíduos do Gana (Fendt *et al.*, 2012b). Entre elas, a amostra da população do Gana, estudada por Fendt e colaboradores, é a que se encontra geneticamente mais próxima da população do presente estudo. O Gana é um país da costa ocidental africana, próximo de Cabo Verde e com o qual o ex-Império Colonial Português estabeleceu contacto no séc. XVI, proporcionando o comércio de diversas mercadorias, incluindo escravos, com


outros territórios da costa oeste de África (Willie, 2001). Os haplogrupos mais comuns no estudo efetuado por Fendt e colaboradores são característicos do oeste africano e são semelhantes aos obtidos com maior frequência, na população de Cabo Verde em estudo, entre eles os haplogrupos L1b1, L2a1, L3b, L3d e L3e2. A Somália, por sua vez, é um país localizado na costa este do continente africano. No estudo efetuado por Mikkelsen e colaboradores verificou-se uma predominância de haplogrupos característicos do este e sudeste africano, entre eles os haplogrupos L0a1d, L2a1h e L3f, e haplogrupos característicos do nordeste e Médio Oriente, entre eles os haplogrupos M1 e N1.


Conclusões


Foi estudada a totalidade da região controlo do mtDNA de 103 indivíduos com ascendência cabo-verdiana, não aparentados e, atualmente, a viver em Lisboa.


 A metodologia de estudo do mtDNA consistiu na utilização de dois pares de primers para amplificação da totalidade da região controlo do mtDNA, sendo todas as amostras estudadas, numa extensão de 1178 pb;


 A sequenciação da totalidade da região controlo demonstrou ser a metodologia mais adequada na análise de sequências de mtDNA, ao permitir um aumento do poder de discriminação entre amostras, quando comparado com outros estudos;


 Os resultados obtidos no presente estudo foram integrados com sucesso, após cuidada deliberação, na base de dados com maior importância a nível mundial, a EMPOP, e estarão, após publicação, disponíveis com o número de acesso EMP00616;

 A população apresenta uma grande variabilidade genética interpopulacional, manifestada pela grande frequência de haplótipos únicos, com 75 haplótipos diferentes em 103 sequências de mtDNA analisadas;

 O menor número de alterações polimórficas encontradas numa sequência em relação à rCRS foi de 9, enquanto que o maior número de alterações polimórficas, em relação à rCRS, foi de 27;

 No conjunto de todas as sequências analisadas, foram detetadas 112 posições polimórficas, com diferentes frequências na amostra em estudo;

 O estudo das sequências obtidas confirmou a existência de uma zona muito polimórfica na região HV1, entre a posição 16183 e a posição 16193;

 A taxa de heteroplasmia de comprimento, tendo em conta a totalidade da região controlo, foi de 13,59%, com a maioria dos casos de heteroplasmia a ocorrer na região HV1;

Ⓐ A diversidade de sequência obtida foi de 0,99164 e a diversidade nucleotídica foi de 0,0034114, valores que demonstram uma grande diversidade genética associada à amostra em estudo e semelhantes a valores das populações africanas;

Ⓐ Entre os 75 haplótipos obtidos, 92,24% foram classificados em haplogrupos tipicamente africanos L1, L2 e L3. O haplogrupo L2a1a foi o que apresentou maior expressão na amostra, englobando 12,62% do total das sequências estudadas. Tal facto aponta para uma elevada contribuição de escravos provenientes da costa ocidental africana na origem da composição genética das linhagens maternas atuais do arquipélago de Cabo Verde. Estes resultados estão de acordo com estudos anteriores efetuados na população de Cabo Verde;

Ⓐ A obtenção de uma reduzida frequência de haplótipos agrupados em haplogrupos com ocorrência no norte de África e na Europa poderá refletir uma pequena contribuição feminina destas populações na origem das linhagens mitocondriais atuais de Cabo Verde durante a colonização portuguesa levada a cabo no arquipélago, o que está de acordo com dados históricos onde severamente se limitou a participação de mulheres europeias entre os colonizadores.

Ⓐ O estudo filogenético efetuado permitiu observar de forma sucinta as relações filogenéticas entre populações selecionadas da bibliografia e a população de Cabo Verde. Cabo Verde parece estar mais próximo geneticamente da população do Gana em comparação com a população da Somália e a população indígena Khoe-San de Angola;

Ⓐ O presente estudo contribuiu para uma atualização do conhecimento da composição genética da população de Lisboa, com a recente introdução de linhagens L, características de populações africanas, com origem em Cabo Verde e como resultado do aumento significativo do fluxo migratório do arquipélago para o território português durante, e especialmente após, o cessar das actividades de comércio de escravos na região de Cabo Verde no séc. XIX e XX.

Bibliografia

Afonso Costa, H., Carvalho, M., Lopes, V., Balsa, F., Bento, A. M., Serra, A., Andrade, L., et al. (2010) "Mitochondrial DNA sequence analysis of a native Bolivian population" *Journal of forensic and legal medicine*, 17(5), 247–53.

Afonso Costa, H., Carvalho, M., Bento, A., Balsa, F., Anjos, M. J., & Corte-Real, F. (2011) "Identification of Azores islands haplogroups by mitochondrial DNA analysis" *Forensic Science International: Genetics Supplement Series*, 3(1), e131–e132.

Afonso Costa, H., Vieira-Silva, C., Amorim, A., Morais, P., Matos, S., Marques-Santos, R., Espinheira, R., et al. (2012) "Genetic Portrait of 12 X-STR Loci in Cabo Verde Immigrant Population Living in Lisboa" *International Journal of Legal Medicine*, 126(S1), 121–384.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2007) *Molecular Biology of the Cell*, 5th ed. *Garland*, New York.

Alonso, A., Salas, A., Albarrán, C., Arroyo, E., Castro, A., Crespillo, M., Di Lonardo, A. M., et al. (2002) "Results of the 1999-2000 collaborative exercise and proficiency testing program on mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor" *Forensic science international*, 125(1), 1–7.

Allard, M. W., Polanskey, D., Miller, K., Wilson, M. R., Monson, K. L., & Budowle, B. (2005) "Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set" *Forensic science international*, 148(2-3), 169–79.

Allen, M., & Engström, A. (1998) "Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities" *Journal of Forensic Sciences*, 43(3), 453–64.

Alvarez, J. C., Johnson, D. L. E., Lorente, J. A., Martinez-Espin, E., Martinez-Gonzalez, L. J., Allard, M., Wilson, M. R., et al. (2007). Characterization of human

control region sequences for Spanish individuals in a forensic mtDNA data set. *Legal medicine (Tokyo, Japan)*, 9(6), 293–304.

Alvarez-Iglesias, V., Jaime, J. C., Carracedo, A., & Salas, A. (2007) "Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups" *Forensic science international. Genetics*, 1(1), 44–55.

Alvarez-Iglesias, V., Mosquera-Miguel, A., Cerezo, M., Quintáns, B., Zarrabeitia, M. T., Cuscó, I., Lareu, M. V., et al. (2009) "New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0" *PloS one*, 4(4), e5112.

Alves-Silva, J., da Silva Santos, M., Guimarães, P. E., Ferreira, C., Bandelt, H. J., Pena, S. D., & Prado, V. F. (2000) "The ancestry of Brazilian mtDNA lineages" *American journal of human genetics*, 67(2), 444–61.

Amorim, A., Afonso-Costa, H., Espinheira, R., Costa, R., Cunha, E., & Costa-Santos, J. (2011) "Human identification with combined anthropologic and genetic tools: two case reports in forensic medicine practice" *Folia Societatis Medicinae Legalis Slovacae*, 1(1), 11–14.

Amorim, A., Marques-Santos, R., Vieira-Silva, C., Afonso-Costa, H., Espinheira, R., Ferreira-Gomes, P., & Costa-Santos, J. (2012) "Genetic portrait of a native population of Cabo Verde living in Lisboa" *Forensic science international. Genetics*, 6(6), e166–9.

Anderson, S., Bankier, A., & Barrell, B. (1981) "Sequence and organization of the human mitochondrial genome" *Nature*, 290, 457–465.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999) "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA" *Nature genetics*, 23(2), 147.

Ballard, J. W. O., & Whitlock, M. C. (2004) "The incomplete natural history of mitochondria" *Molecular Ecology*, 13(4), 729–744.

Balogh, M. K., Burger, J., Bender, K., Schneider, P. M., & Alt, K. W. (2003) "STR genotyping and mtDNA sequencing of latent fingerprint on paper" *Forensic Science International*, 137(2-3), 188–195.

Bandelt, H. J., Kong, Q. P., Parson, W., & Salas, A. (2005) "More evidence for non-maternal inheritance of mitochondrial DNA?" *Journal of medical genetics*, 42(12), 957–60.

Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Holland, M., Lincoln, P. J., et al. (2000) "DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing" *International journal of legal medicine*, 113(4), 193–6.

Barry, B. (1998) "Senegambia and the Atlantic slave trade" *Cambridge University Press*, Cambridge, England.

Batini, C., Lopes, J., Behar, D. M., Calafell, F., Jorde, L. B., Van der Veen, L., Quintana-Murci, L., et al. (2011) "Insights into the demographic history of African Pygmies from complete mitochondrial genomes" *Molecular biology and evolution*, 28(2), 1099–110.

Behar, D. M., Metspalu, E., Kivisild, T., Achilli, A., Hadid, Y., Tzur, S., Pereira, L., et al. (2006) "The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event" *American journal of human genetics*, 78(3), 487–97.

Behar, D. M., Villems, R., Soodyall, H., Blue-smith, J., Pereira, L., Metspalu, E., Scozzari, R., et al. (2008) "The Dawn of Human Matrilineal Diversity" (May), 1130–1140.

Behar, D. M., Van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., Kivisild, T., et al. (2012) "A "Copernican" reassessment of the human mitochondrial DNA tree from its root" *American journal of human genetics*, 90(4), 675–84.

Bendall, K. E., & Sykes, B. C. (1995) "Length heteroplasmy in the first hypervariable segment of the human mtDNA control region" *American journal of human genetics*, 57(2), 248–56.

Bendall, K. E., Macaulay, V. A., & Sykes, B. C. (1997) "Variable levels of a heteroplasmic point mutation in individual hair roots" *American journal of human genetics*, 61(6), 1303–8.

Brandstätter, A., & Parson, W. (2003) "Mitochondrial DNA heteroplasmy or artefacts--a matter of the amplification strategy?" *International journal of legal medicine*, 117(3), 180–4.

Brandstätter, A., Peterson, C. T., Irwin, J., Mpoke, S., Koech, D. K., Parson, W., & Parsons, T. J. (2004) "Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database" *International journal of legal medicine*, 118(5), 294–306.

Brandstätter, A., Zimmermann, B., Wagner, J., Göbel, T., Röck, A. W., Salas, A., Carracedo, A., et al. (2008) "Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East" *BMC evolutionary biology*, 8, 191.

Brehm, A., Pereira, L., & Bandelt, H. (2002) "Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade" *Annals of human*, 49–60.

Brown, W. M. (1980) "Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis" *Proceedings of the National Academy of Sciences of the United States of America*, 77(6), 3605–9.

Brown, M. D., Hosseini, S. H., Torroni, a, Bandelt, H. J., Allen, J. C., Schurr, T. G., Scozzari, R., et al. (1998). mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *American journal of human genetics*, 63(6), 1852–61.

Brown, M. D., Starikovskaya, E., Derbeneva, O., Hosseini, S., Allen, J. C., Mikhailovskaya, I. E., Sukernik, R. I., et al. (2002) "The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J" *Human genetics*, 110(2), 130–8.

Bolnick, D. A., & Smith, D. G. (2003) "Unexpected patterns of mitochondrial DNA variation among Native Americans from the southeastern United States" *American journal of physical anthropology*, 122(4), 336–54.

Butler, J. M. (2011) "Advanced Topics in Forensic DNA Typing" pp.405-444. *Academic Press*, Virginia.

Budowle, B., Chakraborty, R., Giusti, M., Eisenberg, J., & Allen, R. C. (1991) "Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE" *American journal of human genetics*, 48(1), 137–44.

Budowle, B., Wilson, M. R., DiZinno, J., Stauffer, C., Fasano, M., Holland, M. M., & Monson, K. L. (1999) "Mitochondrial DNA regions HVI and HVII population data" *Forensic science international*, 103(1), 23–35.

Budowle, B., Allard, M. W., Wilson, M. R., & Chakraborty, R. (2003) "Forensics and mitochondrial DNA: applications, debates, and foundations" *Annual review of genomics and human genetics*, 4, 119–41.

Calloway, C. D., Reynolds, R. L., Herrin, G. L., & Anderson, W. W. (2000) "The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age" *American journal of human genetics*, 66(4), 1384–97.

Cann, R. L., Stoneking, M., & Wilson, A. C. (1987) "Mitochondrial DNA and human evolution" *Nature*, 325(1).

Cardoso, S., Villanueva-Millán, M. J., Valverde, L., Odriozola, A., Aznar, J. M., Piñeiro-Hermida, S., & De Pancorbo, M. M. (2012) "Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country" *Forensic science international. Genetics*, 6(4), e106–8.

Carling, J. & Akesson L. (2009) "Mobility at Heart of a Nation: Patterns and Meaning of CapeVerdean Migration" *International Migration*, vol. 47, 123-155.

Carreira, A. (1983) "Migrações nas Ilhas de Cabo Verde" *Instituto Caboverdeano do Livro*, 2ª ed., Lisboa.

Carvalho, F. (2010) "Migração em Cabo Verde, Perfil Nacional 2009" *Organização Internacional para as Migrações*. [consultado em 12-10-2012].

Disponível em: www.un.cv/files/Cape_Verde_Profile_2009.pdf

Carvalho, M., Mendes, C., Antunes, H., Anjos, M. J., Andrade, L., Lopes, V., Vieira, D. N., et al. (2003) "mtDNA analysis in Portuguese populations (Central Portugal and Azores Islands): polymorphic sites in control region sequences" *International Congress Series*, 1239, 535–539.

Castro de Guerra, D., Figuera Pérez, C., Bravi, C. M., Saunier, J., Scheible, M., Irwin, J., Coble, M. D., et al. (2012) "Sequence variation of mitochondrial DNA control region in North Central Venezuela" *Forensic science international. Genetics*, 6(5), e131–3.

Cerezo, M., Achilli, A., Olivieri, A., Perego, U., Gómez-Carballa, A., Brisighelli, F., Lancioni, H., et al. (2012) "Reconstructing ancient mitochondrial DNA links between Africa and Europe" *Genome research*, 22(5), 821–6.

Cerri, N., Verzeletti, A., & De Ferrari, F. (2004) "MtDNA: a successful identification case in burned remains" *International Congress Series*, 1261, 425–427.

Chen, X., Prosser, R., Simonetti, S., Sadlock, J., Jagiello, G., & Schon, E. A. (1995) "Rearranged mitochondrial genomes are present in human oocytes" *American journal of human genetics*, 57(2), 239–47.

Chen, Y. S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K., & Wallace, D. C. (2000) "mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations" *American journal of human genetics*, 66(4), 1362–83.

Coble, M. D., Vallone, P. M., Just, R. S., Diegoli, T. M., Smith, B. C., & Parsons, T. J. (2006) "Effective strategies for forensic analysis in the mitochondrial DNA coding region" *International journal of legal medicine*, 120(1), 27–32.

Comey, C. T., & Budowle, B. (1991) "Validation Studies on the Analysis of the HLA DQ α Locus Using the Polymerase Chain Reaction" (March), 1633–1648.

- Coskun, P., Ruiz-Pesini, E., & Wallace, D. C. (2003) "Control region mtDNA variants: longevity, climatic adaptation, and a forensic conundrum" *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 2174–2176.
- Crespillo, M., Luque, J. A., Paredes, M., Fernández, R., Ramírez, E., & Valverde, J. L. (2000) "Mitochondrial DNA sequences for 118 individuals from northeastern Spain" *International Journal of Legal Medicine*, 114(1-2), 130–132.
- Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., & Scozzari, R. (2011) "A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa" *American journal of human genetics*, 88(6), 814–8.
- Cruz, C., Ribeiro, T., Vieira-Silva, C., Lucas, I., Geada, H., & Espinheira, R. (2004) "Identification by mtDNA of exchanged human body remains" *International Congress Series*, 1261, 374–376.
- Dário, P., Bom, J., Ribeiro, T., & Geada, H. (2009) "MtSNP typing before mtDNA sequencing: Why do it?" *Forensic Science International: Genetics Supplement Series*, 2(1), 187–188.
- Dimauro, S., & Davidzon, G. (2005) "Mitochondrial DNA and disease" *Annals of medicine*, 37(3), 222–32.
- Douce, R., & Neuburger, M. (1999) "Biochemical dissection of photorespiration" *Current opinion in plant biology*, 2, 214–222.
- Duby, G., & Boutry, M. (2002) "Mitochondrial protein import machinery and targeting information" *Plant Science*, 162(4), 477–490.
- Edson, S. M., Ross, J. P., Coble, M. D., Parsons, T. J., & Barritt, S. M. (2004) "Naming the Dead — Confronting the Realities of Rapid Identification of Degraded Skeletal Remains" *Forensic science review*, 16(1), 63–90.
- Eyre-Walker, A., & Awadalla, P. (2001) "Does human mtDNA recombine?" *Journal of Molecular Evolution*, 53(4-5), 430–435.

Excoffier, L., Laval, G., & Schneider, S. (2005) "Arlequin (version 3.0): an integrated software package for population genetics data analysis" *Evolutionary bioinformatics online*, 1, 47–50.

Felsenstein J. "PHYLIP: phylogeny interface package (version 3.2)" *Cladistics* 1989;5:164-6.

Fendt, L., Huber, G., Röck, A. W., Zimmermann, B., Bodner, M., Delpont, R., Schmidt, K., et al. (2012a). Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages. *Forensic science international. Genetics*, 6(5), 662–3.

Fendt, L., Röck, A., Zimmermann, B., Bodner, M., Thye, T., Tschentscher, F., Owusu-Dabo, E., et al. (2012b) "MtDNA diversity of Ghana: a forensic and phylogeographic view" *Forensic science international. Genetics*, 6(2), 244–9.

Finnilä, S., Lehtonen, M. S., & Majamaa, K. (2001) "Phylogenetic network for European mtDNA" *American journal of human genetics*, 68(6), 1475–84.

Freitas, F., & Pereira, L. (2008) "Heterogeneity in coding mtDNA mutation rates: Implications in forensic genetics" *Forensic Science International: Genetics Supplement Series*, 1(1), 274–276.

Fucharoen, G., Fucharoen, S., & Horai, S. (2001) "Mitochondrial DNA polymorphisms in Thailand" *Journal of human genetics*, 46(3), 115–25.

Galtier, N., Enard, D., & Radondy, Y. (2006) "Mutation hot spots in mammalian mitochondrial DNA" *Genome Research*, (2000), 215–222.

Gilbert, M. T. P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N., & Cooper, A. (2003) "Distribution patterns of postmortem damage in human mitochondrial DNA" *American journal of human genetics*, 72(1), 32–47.

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980) "Maternal inheritance of human mitochondrial DNA" *Proceedings of the National Academy of Sciences of the United States of America*, 77(11), 6715–9.

Gómez-Carballa, A., Ignacio-Veiga, A., Alvarez-Iglesias, V., Pastoriza-Mourelle, A., Ruíz, Y., Pineda, L., Carracedo, A., et al. (2012) "A melting pot of multicontinental mtDNA lineages in admixed Venezuelans" *American journal of physical anthropology*, 147(1), 78–87.

Gonçalves, R., Rosa, A., Freitas, A., Fernandes, A., Kivisild, T., Villems, R., & Brehm, A. (2003) "Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers" *Human genetics*, 113(6), 467–72.

González, A M., Cabrera, V. M., Larruga, J. M., Tounkara, A., Noumsi, G., Thomas, B. N., & Moulds, J. M. (2006). "Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations" *Annals of human genetics*, 70(Pt 5), 631–57.

González, A. M., Larruga, J. M., Abu-Amero, K. K., Shi, Y., Pestano, J., & Cabrera, V. M. (2007) "Mitochondrial lineage M1 traces an early human backflow to Africa" *BMC genomics*, 8(2), 223.

Gray, M. W. (2012) "Mitochondrial evolution" *Cold Spring Harbor perspectives in biology*, 4(9), 1–14.

Hagelberg, E., Goldman, N., Lió, P., Whelan, S., Schiefenhövel, W., Clegg, J. B., & Bowden, D. K. (1999) "Evidence for mitochondrial DNA recombination in a human population of island Melanesia" *Proceedings. Biological sciences / The Royal Society*, 266(1418), 485–92.

Harich, N., Costa, M. D., Fernandes, V., Kandil, M., Pereira, J. B., Silva, N. M., & Pereira, L. (2010) "The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages" *BMC evolutionary biology*, 10, 138.

Herrnstadt, C., Elson, J. L., Fahy, E., Preston, G., Turnbull, D. M., Anderson, C., Ghosh, S. S., et al. (2002) "Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups" *American journal of human genetics*, 70(5), 1152–71.

Herrnstadt, C., & Howell, N. (2004) “An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders” *Mitochondrion*, 4(5-6), 791–8.

Hill, C., Soares, P., Mormina, M., Macaulay, V., Meehan, W., Blackburn, J., Clarke, D., et al. (2006) “Phylogeography and ethnogenesis of aboriginal Southeast Asians” *Molecular biology and evolution*, 23(12), 2480–91.

Holland, M., & Parsons, T. (1999) “Mitochondrial DNA sequence analysis-validation and use for forensic casework” *Forensic Science Review*.

Howell, N., & Smejkal, C. B. (2000) “Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction” *American journal of human genetics*, 66(5), 1589–98.

Howell, N., Elson, J. L., Howell, C., & Turnbull, D. M. (2007) “Relative rates of evolution in the coding and control regions of African mtDNAs” *Molecular biology and evolution*, 24(10), 2213–21.

Iborra, F., Kimura, H., & Cook, P. (2004) “The functional organization of mitochondrial genomes in human cells” *BMC biology*, 14, 1–14.

Ingman, M., Kaessmann, H., Pääbo, S., & Gyllensten, U. (2000) “Mitochondrial genome variation and the origin of modern humans” *Nature*, 408(6813), 708–13.

Ingman, M., & Gyllensten, U. (2006) “mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences” *Nucleic acids research*, 34(Database issue), D749–51.

Inman, H., & Nordbrg, M. (2002) “Recombination or mutational hot spots in human mtDNA?” *Molecular biology and evolution*, 19(7), 1122–7.

Instituto Nacional de Estatística de Cabo Verde (2010). “IVº Recenseamento geral da população e de habitação - Censo 2010” [consultado em 25-10-2012].

Disponível em:

<http://www.ine.cv/censo/files/Resumo%20indicadores%20RGPH%202010.pdf>

Instituto Nacional de Estatística de Portugal, INE (2012). “População Estrangeira em Portugal” [consultado em 23-10-2012].

Disponível em:

http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=150133806&att_display=n&att_download=y

Jackson, B. A., Wilson, J. L., Kirbah, S., Sidney, S. S., Rosenberger, J., Bassie, L., Alie, J. A. D., et al. (2005) “Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone” *American journal of physical anthropology*, 128(1), 156–63.

Jenuth, J. P., Peterson, A. C., Katherine, F., & Shoubridge, E. A. (1996) “Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA” *Nature genetics*, 14.

Johnson, A., & Johnson, K. A. (2001) “Fidelity of nucleotide incorporation by human mitochondrial DNA polymerase” *The Journal of biological chemistry*, 276(41), 38090–6.

Kivisild, T., Tolk, H.-V., Parik, J., Wang, Y., Papiha, S. S., Bandelt, H.-J., & Villems, R. (2002) “The emerging limbs and twigs of the East Asian mtDNA tree” *Molecular biology and evolution*, 19(10), 1737–51.

Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., et al. (2004) “Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears” *American journal of human genetics*, 75(5), 752–70.

Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K., Passarino, G., et al. (2006) “The role of selection in the evolution of human mitochondrial genomes” *Genetics*, 172(1), 373–87.

Kong, Q.-P., Sun, C., Wang, H.-W., Zhao, M., Wang, W.-Z., Zhong, L., Hao, X.-D., et al. (2011) “Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region” *Molecular biology and evolution*, 28(1), 513–22.

Krings, M., Stone, A., & Schmitz, R. (1997) "Neandertal DNA sequences and the origin of modern humans" *Cell*, 90, 19–30.

Kumar, S., Bellis, C., Zlojutro, M., Melton, P. E., Blangero, J., & Curran, J. E. (2011) "Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins" *BMC evolutionary biology*, 11(1), 293.

Lagerström-Fermér, M., Olsson, C., Forsgren, L., & Syvänen, C. (2001) "Heteroplasmy of the human mtDNA control region remains constant during life" *American journal of human genetics*, 68(5), 1299–301.

Lima, G., Pontes, M. L., Abrantes, D., Cainé, L., Pereira, M. J., Matos, P., & Pinheiro, M. F. (2006). HVI and HVII sequence polymorphisms of the human mtDNA in the North of Portugal: Population data and maternal lineages. *International Congress Series*, 1288, 133–135.

Lodish, H., Berk, A., Darnell, J., Kaiser C. A., Krieger M., Matsudaira P., Scott M. P. & Zipursky L. (2004) "Molecular cell biology" *W. H. Freeman*, 5^a ed., 171-172.

Logan, D. C. (2006) "The mitochondrial compartment" *Journal of experimental botany*, 57(6), 1225–43.

Lutz-Bonengel, S., Sängler, T., Pollak, S., & Szibor, R. (2004) "Different methods to determine length heteroplasmy within the mitochondrial control region" *International journal of legal medicine*, 118(5), 274–81.

Marchington, D. R., Scott Brown, M. S. G., Lamb, V. K., Van Golde, R. J. T., Kremer, J. M., Tuerlings, J. H. M., Mariman, E. C. M., et al. (2002) "No evidence for paternal mtDNA transmission to offspring or extra-embryonic tissues after ICSI" *Molecular human reproduction*, 8(11), 1046–9.

Makalowski, W. (2001) "The human genome structure and organization" *Acta Biochim. Pol*, 48(3), 587–598.

Malyarchuk, B. a, Rogozin, I. B., Berikov, V. B., & Derenko, M. V. (2002) "Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region" *Human genetics*, 111(1), 46–53.

Malyarchuk, B. A., & Czarny, J. (2005) "African DNA lineages in mitochondrial gene pool of Europeans" *Molekuliarnaia biologii*, 39(5), 806–12.

Malyarchuk, B., Derenko, M., Perkova, M., Grzybowski, T., Vanecek, T., & Lazur, J. (2008) "Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs" *European journal of human genetics*, 16(9), 1091–6.

Martin, W. & Mentel, M. (2010) "The Origin of Mitochondria" *Nature Education*, 3(9):58.

Meyer, S., Weiss, G., & Von Haeseler, A. (1999) "Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA" *Genetics*, 152(3), 1103–10.

Mikkelsen, M., Fendt, L., Röck, A. W., Zimmermann, B., Rockenbauer, E., Hansen, A. J., Parson, W., et al. (2012) "Forensic and phylogeographic characterisation of mtDNA lineages from Somalia." *International journal of legal medicine*, 126(4), 573–9.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., et al. (2003). Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), 171–6.

Nagai, A., Nakamura, I., Shiraki, F., Bunai, Y., & Ohya, I. (2003) "Sequence polymorphism of mitochondrial DNA in Japanese individuals from Gifu Prefecture" *Legal medicine*, 5, 210–213.

Nei, M., & Tajima, F. (1981) "Genetic drift and estimation of effective population size" *Genetics*, 98(3), 625–40.

Niemi, A.-K., Hervonen, A., Hurme, M., Karhunen, P. J., Jylhä, M., & Majamaa, K. (2003) "Mitochondrial DNA polymorphisms associated with longevity in a Finnish population" *Human genetics*, 112(1), 29–33.

Nomenclature Committee of the International Union of Biochemistry, NC-IUB.
<http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html>

Oikawa, H., Tun, Z., Young, D. R., Ozawa, H., Yamazaki, K., Tanaka, E., & Honda, K. (2002) "The specific mitochondrial DNA polymorphism found in Klinefelter's syndrome" *Biochemical and Biophysical Research Communications*, 297(2), 341–345.

Oota, H., Saitou, N., Matsushita, T., & Ueda, S. (1999) "Molecular genetic analysis of remains of a 2,000-year-old human population in China-and its relevance for the origin of the modern Japanese population" *American journal of human genetics*, 64(1), 250–8.

Otonari C., Martinez-Labarga C., Loogväli E.L., Pennarun E., Achilli A., De Angelis F., Trucchi E., Contini I., Biondi G. & Rickards O. (2009). "First genetic insight into Libyan Tuaregs: a maternal perspective." *Annal Human Genetics*, 73: 438-448.

Pääbo, S., Gifford, J., & Wilson, A. (1988) "Mitochondrial DNA sequences from a 7000-year old brain" *Nucleic Acids Research*, 16(20), 9775–9787.

PageRDM "Treeview: an application to display phylogenetic trees on personal computers" *Comput Appl Biosci* 1996;12:357-8.

Pakendorf, B., & Stoneking, M. (2005) "Mitochondrial DNA and human evolution" *Annual review of genomics and human genetics*, 6, 165–83.

Parson, W., Brandstätter, A., Alonso, A., Brandt, N., Brinkmann, B., Carracedo, A., Corach, D., et al. (2004). The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic science international*, 139(2-3), 215–26.

Parson, W., & Dür, A. (2007) "EMPOP--a forensic mtDNA database" *Forensic science international. Genetics*, 1(2), 88–92.

Parson, W., & Bandelt, H.-J. (2007) "Extended guidelines for mtDNA typing of population data in forensic science" *Forensic science international. Genetics*, 1(1), 13–9.

Pereira, L., Prata, M. J., & Amorim, A. (2000) "Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation" *Annals of human genetics*, 64(Pt 6), 491–506.

Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J., & Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Annals of human genetics*, 65(Pt 5), 439–58.

Pereira, L., Macaulay, V., Prata, M. ., & Amorim, A. (2003). Phylogeny of the mtDNA haplogroup U6. Analysis of the sequences observed in North Africa and Iberia. *International Congress Series*, 1239, 491–493.

Pereira, L., Cerný, V., Cerezo, M., Silva, N. M., Hájek, M., Vasíková, A., Kujanová, M., et al. (2010) "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 : EJHG*, 18(8), 915–23.

Peterson, C., Johannsen, D., & Ravussin, E. (2012) "Skeletal Muscle Mitochondria and Aging: A Review" *Journal of Aging Research*, 27(7), 647–653.

Piganeau, G., & Eyre-Walker, A. (2004) "A reanalysis of the indirect evidence for recombination in human mitochondrial DNA" *Heredity*, 92(4), 282–8.

Pinheiro, M. F. T. (2009) "CSI Criminal" pp. 90-91. *Edições Univ. Fernando Pessoa*, Porto, Portugal.

Prieto, L., Montesino, M., Rodríguez-Monge, A., García, C., García, E., Rivas, E., & Salas, A. (2004) "Measuring by fragment analysis the proportion of length variants in samples carrying length heteroplasmy at the homopolymeric C-stretch in mitochondrial HVII region" *International Congress Series*, 1261, 103–105.

Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K., & Santachiara-Benerecetti, S. (1999) "Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa" *Nature genetics*, 23(4), 437–41.

Ramalho, R. (2011) "Building the Cape Verde Islands" *Springer Theses*, 1st ed., 13-17.

Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M., & Bandelt, H. J. (1998) "Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations" *Annals of human genetics*, 62(Pt 6), 531–50.

Reidla, M., Kivisild, T., Metspalu, E., Kaldma, K., Tambets, K., Tolk, H.-V., Parik, J., et al. (2003) "Origin and diffusion of mtDNA haplogroup X" *American journal of human genetics*, 73(5), 1178–90.

Roberts, K., & Calloway, C. (2011) "Characterization of mitochondrial DNA sequence heteroplasmy in blood tissue and hair as a function of hair morphology" *Journal of forensic sciences*, 56(1), 46–60.

Rosa A., Brehm A., Kivisild T., Metspalu E. & Villems R. (2004). "MtDNA profile of West 28 African mtDNA at a glance Africa Guineans: towards a better understanding of the Senegambia region". *Annals of Human Genetics*, 68: 340-352.

Rosa, A., & Brehem, A. (2011) "African human mtDNA phylogeography at-a-glance" *Journal of anthropological sciences = Rivista di antropologia: JASS / Istituto italiano di antropologia*, 89, 25–58.

Rudin, N. and Inman, K. (2002) "An Introduction to Forensic DNA Analysis (2nd Edition)" *CRC Press*.

Ruiz-Pesini, E., Lapeña, A. C., Díez-Sánchez, C., Pérez-Martos, A., Montoya, J., Alvarez, E., Díaz, M., et al. (2000) "Human mtDNA haplogroups associated with high or reduced spermatozoa motility" *American journal of human genetics*, 67(3), 682–96.

Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J. C., Brandon, M. C., Mishmar, D., Yi, C., et al. (2007) "An enhanced MITOMAP with a global mtDNA mutational phylogeny" *Nucleic acids research*, 35(Database issue), D823–8

Russell-Wood, A. (1998) "The Portuguese empire, 1415-1808: a world on the move. *Johns Hopkins Univ. Press*, Baltimore.

Salas, A., Richards, M., De la Fe, T., Lareu, M. V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., et al. (2002) "The making of the African mtDNA landscape" *American journal of human genetics*, 71(5), 1082–111.

Salas, A., Fé, T. D. La, & Sobrino, B. (2003) "Mitochondrial DNA variability patterns in Southeast Africa and forensic implications" *International Congress*, 1239, 541–545.

Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., et al. (2004) "The African diaspora: mitochondrial DNA and the Atlantic slave trade" *American journal of human genetics*, 74(3), 454–65.

Sanger, F., Nicklen, S., Coulson, R. (1997) "DNA sequencing with chain-terminating inhibitors" *Pro. Natl Acad Sci USA.*, 74(12): 5463-5467.

Seo, Y., Uchiyama, D., Kuroki, K., & Kishida, T. (2012) "STR and mitochondrial DNA SNP typing of a bone marrow transplant recipient after death in a fire" *Legal medicine (Tokyo, Japan)*, 14(6), 331–5.

Scheffler, I.E., 1999. *Mitochondria*, 1st edn. *Wiley*, New York.

Scozzari, R., Torroni, A., & Semino, O. (1988) "Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms" *American journal of human genetics*, 43(4), 534–544.

Schwartz, M., & Vissing, J. (2002) "Paternal inheritance of mitochondrial DNA" *New England Journal of Medicine*, 347(8), 576–580.

Soares, P., Trejaut, J. A., Loo, J. H., Hill, C., Mormina, M., Lee, C. L., Chen, Y. M., et al. (2008) "Climate change and postglacial human dispersals in southeast Asia." *Molecular biology and evolution*, 25(6), 1209–18.

Stringer, C. (2003) "New perspectives on the Neanderthals" *Evolutionary Anthropology: Issues, News, and Reviews*, 11(S1), 58–59.

Sutovsky, P., Leyen, K. Van, & McCauley, T. (2004) "Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance" *Reproductive BioMedicine Online*, 8(1), 24–33.

Tavares, A. (2007). "Estudo do DNA mitochondrial numa população de Cabo Verde. Contribuição para uma base de dados" Lisboa: Faculdade de Ciências. Dissertação de Mestrado.

Thompson, W. E., Ramalho-Santos, J., & Sutovsky, P. (2003) "Ubiquitination of prohibitin in mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control" *Biology of reproduction*, 69(1), 254–60.

Thyagarajan, B., Padua, R., & Campbell, C. (1996) "Mammalian mitochondria possess homologous DNA recombination activity" *Journal of Biological Chemistry*, 271(44), 27536–27543.

Torrioni, A., Huoponen, K., & Francalacci, P. (1996) "Classification of European mtDNAs from an analysis of three European populations" *Genetics*, 144(1835-1850).

Torrioni, A., Achilli, A., Macaulay, V., Richards, M., & Bandelt, H. J. (2006) "Harvesting the fruit of the human mtDNA tree" *Trends in genetics : TIG*, 22(6), 339–45.

Trovoada, M. J., Pereira, L., Gusmão, L., Abade, a, Amorim, a, & Prata, M. J. (2004) "Pattern of mtDNA variation in three populations from São Tomé e Príncipe" *Annals of human genetics*, 68(Pt 1), 40–54.

Tully, L. A., Parsons, T. J., Steighner, R. J., Holland, M. M., Marino, M. A., & Prenger, V. L. (2000) "A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region" *American journal of human genetics*, 67(2), 432–43.

Turchi, C., Buscemi, L., Giacchino, E., Onofri, V., Fendt, L., Parson, W., & Tagliabracci, A. (2009) "Polymorphisms of mtDNA control region in Tunisian and

Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data" *Forensic science international. Genetics*, 3(3), 166–72.

Underhill, P. & Kivisild, T. (2007) "Use of y chromosome and mitochondrial DNA population structure in tracing human migrations" *Annual review of genetics*, 41, 539–64.

Vandecasteele, G., Szabadkai, G., & Rizzuto, R. (2001) "Mitochondrial calcium homeostasis: mechanisms and molecules" *IUBMB life*, 52(3-5), 213–9.

Van der Walt, J. M., Nicodemus, K. K., Martin, E. R., Scott, W. K., Nance, M. a, Watts, R. L., Hubble, J. P., et al. (2003) "Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease" *American journal of human genetics*, 72(4), 804–11.

Van der Walt, J. M., Dementieva, Y. a, Martin, E. R., Scott, W. K., Nicodemus, K. K., Kroner, C. C., Welsh-Bohmer, K. a, et al. (2004) "Analysis of European mitochondrial haplogroups with Alzheimer disease risk" *Neuroscience letters*, 365(1), 28–32.

Van Oven, M., & Kayser, M. (2009) "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation" *Human mutation*, 30(2), E386–94.

Walsh, P., Metzger, D., & Higuchi, R. (1991) "Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material" *Biotechniques*, 10(4).

Willie, F. (2001) "Encyclopedia of African History and Culture" Vol. 1-5, *Facts on File*.

Wilson, M. R., DiZinno, J., Polanskey, D., Replogle, J., & Budowle, B. (1995) "Validation of mitochondrial DNA sequencing for forensic casework analysis" *International journal of legal medicine*, 108(2), 68–74.

Youle, R. J., & Karbowski, M. (2005) "Mitochondrial fission in apoptosis" *Nature reviews. Molecular cell biology*, 6(8), 657–63.