UNIVERSIDADE DE ÉVORA

GENOMA MITOCONDRIAL

CONTRIBUTO PARA O ESTUDO DE MARCADORES GENÉTICOS COM INTERESSE MÉDICO-LEGAL, FORENSE E POPULACIONAL EM IMIGRANTES ORIUNDOS DE PAÍSES AFRICANOS INTEGRADOS NA POPULAÇÃO DE LISBOA

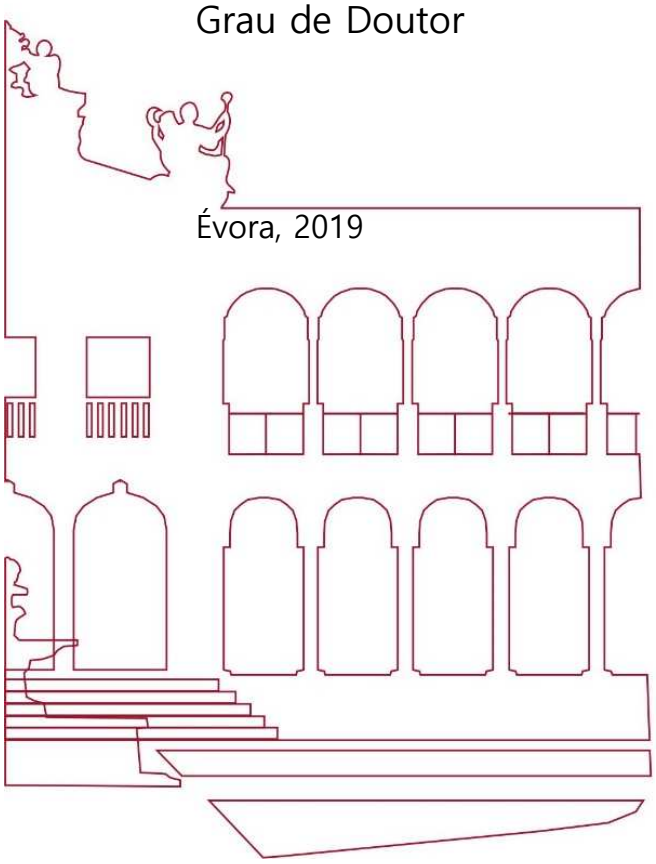
António Joaquim Amorim Santos

Orientadores

Professora Doutora Maria Teresa Ribeiro Matos Fernandes

Professor Doutor Nuno Eduardo Moura dos Santos da Costa Taveira

Tese apresentada à Universidade de Évora para obtenção do Grau de Doutor



Évora, 2019

INSTITUTO DE INVESTIGAÇÃO E FORMAÇÃO AVANÇADA

UNIVERSIDADE DE ÉVORA

GENOMA MITOCONDRIAL

CONTRIBUTO PARA O ESTUDO DE MARCADORES GENÉTICOS COM INTERESSE MÉDICO-LEGAL, FORENSE E POPULACIONAL EM IMIGRANTES ORIUNDOS DE PAÍSES AFRICANOS INTEGRADOS NA POPULAÇÃO DE LISBOA

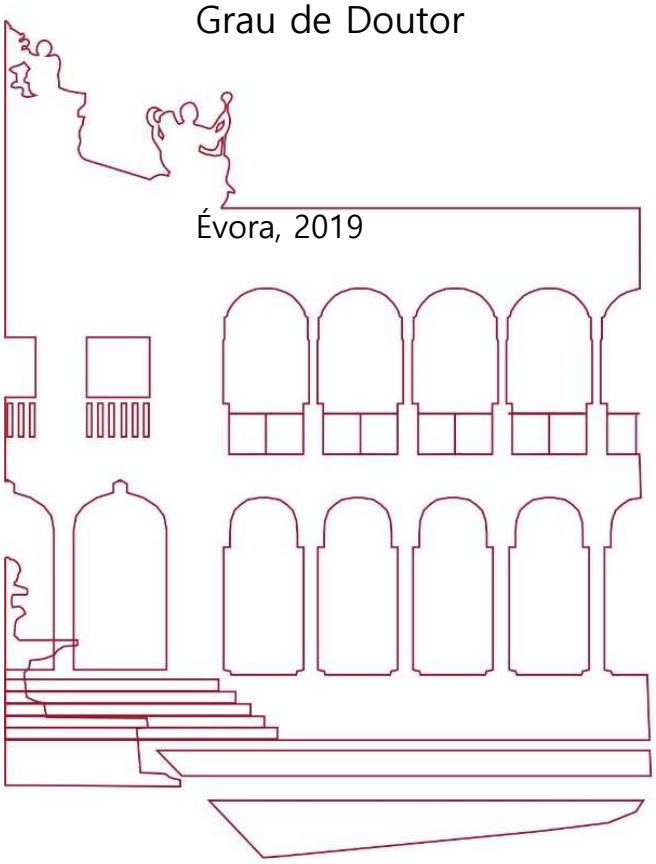
António Joaquim Amorim Santos

Orientadores

Professora Doutora Maria Teresa Ribeiro Matos Fernandes

Professor Doutor Nuno Eduardo Moura dos Santos da Costa Taveira

Tese apresentada à Universidade de Évora para obtenção do Grau de Doutor



Évora, 2019

INSTITUTO DE INVESTIGAÇÃO E FORMAÇÃO AVANÇADA

GENOMA MITOCONDRIAL

CONTRIBUTO PARA O ESTUDO DE MARCADORES GENÉTICOS COM INTERESSE MÉDICO-LEGAL, FORENSE E POPULACIONAL EM IMIGRANTES ORIUNDOS DE PAÍSES AFRICANOS INTEGRADOS NA POPULAÇÃO DE LISBOA

MITOCHONDRIAL GENOME

AN APPROACH TO THE STUDY OF GENETIC MARKERS WITH MEDICO-LEGAL, FORENSIC AND POPULATION INTEREST IN IMMIGRANTS FROM AFRICAN COUNTRIES LIVING IN LISBOA

o júri

Presidente **Doutor DIOGO FRANCISCO CAEIRO FIGUEIREDO**
Professor Catedrático
Universidade de Évora

vogais
Doutora EUGÉNIA MARIA GUEDES PINTO ANTUNES DA CUNHA
Professora Catedrática
Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Doutor NUNO EDUARDO MOURA DOS SANTOS DA COSTA TAVEIRA
Professor Catedrático
Instituto Universitário Egas Moniz

Doutor AGOSTINHO JOSÉ CARVALHO DOS SANTOS
Professor Auxiliar com Agregação convidado
Faculdade de Medicina da Universidade do Porto

Doutora MARIA TERESA RIBEIRO MATOS FERNANDES
Professora Auxiliar
Universidade de Évora

Doutora LAURA SOFIA RAMOS MENDES CAINÉ
Professora Auxiliar convidada
Faculdade de Medicina da Universidade do Porto

Doutor LICÍNIO MANUEL MENDES MANCO
Técnico Superior
Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Tese em formato de compilação de artigos, conforme Regulamento da Universidade de Évora e conforme aprovação do CCP-IIFA de 17/02/2016

À minha Querida Mãe Alice, que me traz no seu regaço tão, incondicionalmente, dedicado ao meu amparo. Um regaço tão singular e sempre tão meu e tão para mim desde o dia em que nasci. Não viveria sem o teu regaço minha Querida Mãe ...

Ao meu Pai Olívio. Desajeitado nas emoções e de poucas e muito rígidas e austeras palavras e atitudes, mas que não tenho dúvida tem um orgulho sem tamanho, ainda que não expresso, em ter-me como filho, e um orgulho sem tamanho no respeito que sempre, incondicionalmente, tive e terei por ele ...

À minha Querida Avó Angélica, que Deus levou tão cedo, mas que como ninguém sempre me protegeu e sempre me tratou como se eu fosse, imagine-se, o melhor neto, o melhor menino do Mundo ...

Ao meu Querido Padrinho António, marido da Avó Angélica, que quis Deus fosse o Avô que os laços de sangue não dariam, por ventura, com tamanha entrega e dedicação ...

À minha Querida Tia Idalina, que tem sido também minha Mãe ao longo de toda a minha vida e que me viu sempre como se eu fosse, imagine-se, o maior e melhor Homem do Mundo ...

Ao tio Rogério. Membro entrado na família por casamento com a tia Idalina, mas que sempre me viu como sobrinho de sangue que eu não era. Sempre firme com a tia Idalina no amor, no cuidado e na dedicação que me deu sempre como seu sobrinho, e sinto eu, como seu sobrinho preferido …

Ao tio Joaquim. Tão recentemente casado com a tia Idalina, mas que me trata como se eu fosse seu sobrinho desde que nasci. Penso que nunca conheci, nem conhecerei, outro Homem no Mundo dono de tamanha bondade, dedicação e pacatez. Obrigado Quim, por ser meu tio ...

Ao meu Querido irmão Francisco, mesmo que de poucas palavras, sempre incondicionalmente ao meu lado ...

Ao meu Querido Paulo Ferreira Gomes, que quis Deus, tive o privilégio de ter a meu lado, acima de tudo, como irmão, desde os meus 18 anos. Nunca te poderei agradecer o que altruisticamente me tens dado na vida meu Querido Paulo ...

Sem vos, muito da minha vida não teria sido possível ...

Ao mais extraordinário lugar de província português, no qual a minha família se instalou e onde todos sempre voltam há, pelo menos, 300 anos - o lugar dos Canaviais, Grijó -. Neste recato de província organizei, sedimentei, escrevi a grande parte desta tese e grande parte dos textos e documentos com importância na minha vida ...

Ao Senhor Professor Roberto Salema, ao Senhor Professor José Manuel Amorim, ao Senhor Professor Aires da Penha Gonçalves, à Senhora Professora Amália Gomes Ferreira, ao Senhor Professor Magalhães Sant'Ana, à Senhora Professora Eugénia Cunha, ao Senhor Professor Jorge Costa Santos, ao Senhor Professor Duarte Nuno Vieira, ao Senhor Professor Nuno Taveira, à Senhora Professora Teresa Fernandes, que acharam que eu poderia dar algum contributo à investigação científica e ao ensino universitário com algum sucesso ...

Ao Instituto Nacional de Medicina Legal e Ciências Forenses que acolheu o projeto de investigação que deu origem a esta tese.

Às minhas Queridas Amigas Heloísa Afonso Costa e Claudia Vieira da Silva ...

A Deus, que sempre me acompanhou e a quem peço que me acompanhe e ilumine sempre... Continuarei a tentar fazer a parte que me cabe...

GENOMA MITOCONDRIAL: CONTRIBUTO PARA O ESTUDO DE MARCADORES GENÉTICOS COM INTERESSE MÉDICO-LEGAL, FORENSE E POPULACIONAL EM IMIGRANTES ORIUNDOS DE PAÍSES AFRICANOS INTEGRADOS NA POPULAÇÃO DE LISBOA

Palavras chave Genoma mitocondrial, população de Lisboa, imigrantes africanos

Resumo

Comparativamente ao DNA nuclear, do qual possuímos somente duas representações por célula, uma das caraterísticas do DNA mitocondrial, de grande utilidade na área médico-legal e forense, consiste no facto de possuirmos cerca de 500 a 2000 cópias de DNA mitocondrial por célula. O comprimento do DNA mitocondrial, 100 000 vezes inferior ao DNA nuclear, e a sua forma circular também lhe conferem uma menor probabilidade de degradação em relação ao DNA nuclear. Outras caraterísticas como o facto de ser herdado somente por via materna e de apresentar reduzidas taxas de heteroplasmia e de mutação tornam este genoma de grande utilidade quando várias gerações separam os investigados dos seus familiares vivos. No entanto, a robustez da utilização do DNA mitocondrial no âmbito de investigações médico-legais e forenses está em muito dependente da existência de bases de dados de DNA mitocondrial ou de ferramentas equiparadas que permitam uma estimativa sobre a frequência de determinada sequência numa população e a sua possível distribuição geográfica.

Nas últimas décadas o número de imigrantes em Portugal aumentou consideravelmente. No final de 2014, o número de imigrantes a residir em Portugal ultrapassou os 400 000 indivíduos, entre os quais cerca de 80 000 oriundos de países africanos de língua oficial Portuguesa - PALOP -. Deparamo-nos, portanto, em Portugal e muito particularmente na região de Lisboa com uma realidade populacional diferente da existente anteriormente a este fluxo de imigrantes. Além desta nova realidade populacional, os estudos que versam sobre toda a região controlo do DNA mitocondrial de indivíduos, grupos ou populações de Portugal, ou aí residentes, são em número muito reduzido.

Sequenciámos, com o método de Sanger, toda a região controlo do DNA mitocondrial de amostras de sangue de 439 imigrantes com residência oficial na região de Lisboa e com naturalidade e nacionalidade em PALOP, designadamente Angola, Cabo Verde, Guiné-Bissau e Moçambique.

Obtivemos sequências de DNA mitocondrial com cerca de 1 122 pares de bases de todos os 439 imigrantes estudados. Não contrariando a generalidade dos estudos anteriores, nos 439 indivíduos que estudámos não identificamos heteroplasmia de posição em nenhum caso. Os imigrantes estudados encontram-se distribuídos pelos haplogrupos L0, L1, L2, L3, L4, L5, M, R, J, T, H, K, U e X. Globalmente, as caraterísticas observadas nos imigrantes PALOP foram amplamente coincidentes e não apresentaram contradições relativamente às caraterísticas descritas nas correspondentes populações de origem - populações da região de África subsariana -.

A qualidade das sequências obtidas no nosso estudo e o interesse populacional dos grupos estudados determinaram a sua aceitação para inserção na EMPOP sendo, atualmente, possível a toda a comunidade científica a nível internacional aceder às sequências de DNA mitocondrial dos imigrantes de Cabo Verde (EMP00616), Angola (EMP00662), Moçambique (EMP00681) e Guiné-Bissau (EMP00704) em www.empop.org. Esta base de dados genéticos internacional não tinha, até então, nenhum indivíduo de Cabo Verde nem de Moçambique com a caraterização de toda a região controlo do seu DNA mitocondrial.

MITOCHONDRIAL GENOME: AN APPROACH TO THE STUDY OF GENETIC MARKERS WITH MEDICO-LEGAL, FORENSIC AND POPULATION INTEREST IN IMMIGRANTS FROM AFRICAN COUNTRIES LIVING IN LISBOA

Keywords mitochondrial genome, Lisbon population, african imigrants

Abstract

Compared to nuclear DNA, of which we have only two copies per cell, one of the characteristics of mitochondrial DNA, which is very useful in medical-legal and forensic area, is that we have about 500 to 2000 copies of mitochondrial DNA per cell. The length of mitochondrial DNA, 100 000 times less than nuclear DNA, and its circular shape also give it a lower probability of degradation in comparison to nuclear DNA. Other characteristics such as maternal inheritance, low rates of heteroplasmy and mutation, make this genome very useful when several generations separate the victims from their live offspring. However, the robustness of mitochondrial DNA utilization in medical-legal and forensic investigations is highly dependent on the existence of mitochondrial DNA databases or similar tools that allow an estimation of the frequency of a given sequence in a population and his possible geographical distribution.

In the last decades the number of immigrants in Portugal has increased considerably. By the end of 2014, the number of immigrants living in Portugal exceeded 400 000, including about 80 000 immigrants from Portuguese-speaking African countries (PALOP). We are therefore in Portugal and particularly in Lisbon region with a different population reality. In addition to this new population reality, studies that deal with the entire mitochondrial DNA control region of individuals, groups or populations of Portugal, or resident there, are very few.

We sequenced the entire mitochondrial DNA control region of blood samples of 439 immigrants officially living Lisbon region and with naturalness and nationality in PALOP, namely Angola, Cape Verde, Guinea-Bissau and Mozambique.

We obtained mitochondrial DNA sequences with about 1 122 base pairs from all 439 studied immigrants. In line with the generality of previous studies, in the 439 studied individuals we did not identify point heteroplasmy in any case. The studied immigrants are distributed by the haplogroups L0, L1, L3, L4, L5, M, R, J, T, H, K, U and X. Globally, the characteristics observed in PALOP immigrants were largely coincident and not presented contradictions in relation to the characteristics described in the corresponding populations of origin - populations of the sub - Saharan Africa region.

The quality of the sequences obtained in our study and the population interest of the studied groups determined their acceptance for EMPOP insertion, and it is now possible for the entire international scientific community to access the mitochondrial DNA sequences of Cape Verdean immigrants (EMP00616), Angola (EMP00662), Mozambique (EMP00681) and Guinea-Bissau (EMP00704) at www.empop.org. This international genetic database did not, until then, have any individuals from Cape Verde or Mozambique with the characterization of the whole mitochondrial DNA control region.

Abreviaturas, Acrónimos e Siglas

AMOVA - análise de variância molecular

DNA - ácido desoxiribonucleico

DNAmt - DNA mitocondrial

CRS - sequência de referência de Cambridge

dNTP - desoxinucleótidos trifosfatados

dNTP - didesoxinucleótidos trifosfatados

EDNAP - european DNA profiling group

EMPOP - EDNAP mtDNA population database

HVI - região hipervariável um, do DNAmt

HVII - região hipervariável dois, do DNAmt

InDel - insertion/deletion polimorphisms

INMLCF - instituto nacional de medicina legal e ciências forenses

PALOP - países africanos de língua oficial portuguesa

pb - pares de bases

rCRS - sequência de referência de Cambridge revista

STR - short tandem repeats

Lista de Tabelas

Lista de Figuras

Lista de Gráficos

Índice

Capítulo I – Introdução

A teoria mais aceite atualmente relativamente à origem do homem moderno é a que corresponde à hipótese *Out of Africa*, na qual se preconiza que o homem teve uma origem africana, a partir da qual se terá dispersado para o resto do mundo (Cann, Stoneking, & Wilson, 1987; Ingman, Kaessmann, Paabo, & Gyllensten, 2000; Stringer, 2003). Estima-se que o ancestral feminino comum terá surgido há aproximadamente 200 000 anos, em África, tendo divergido para os restantes continentes, conforme representado na figura 1, e transmitido a sequência de DNAmt à população atual (Behar et al., 2008; Cann et al., 1987). Este antepassado feminino é designado por Eva mitocondrial (Harrub, Ph, & Thompson, 2010).

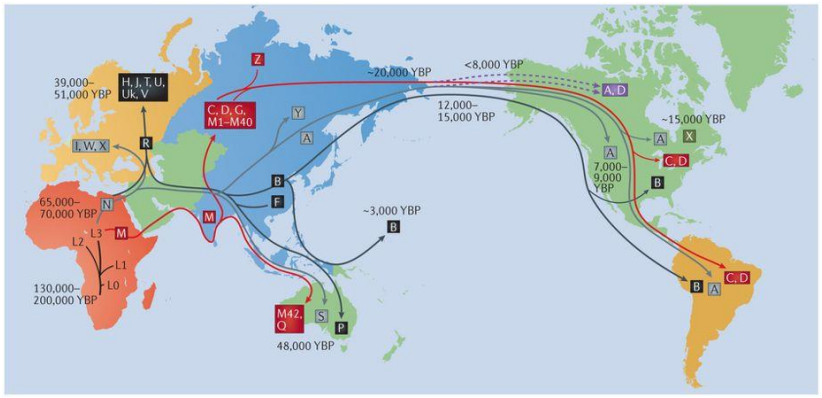


Figura 1 - Mapa representativo da expansão do Homem pelo mundo e respetiva distribuição geográfica dos haplogrupos. Adaptado de Kivisild et al., 2015.

A evolução de uma dada população é refletida nas suas sequências do DNAmt, pela acumulação de mutações, sendo que o conjunto destas mutações definem haplótipos que, por sua vez, se agrupam em haplogrupos (Pakendorf & Stoneking, 2005). Os haplogrupos de DNAmt podem, assim, estar associados a determinadas regiões geográficas e/ou a grupos étnicos em particular.

O conjunto das alterações observadas nos indivíduos é organizado e utilizado para construir árvores filogenéticas (Howell, Kubacka, & Mackey, 1996), sendo a Phylotree a árvore filogenética mais utilizada atualmente (van Oven & Kayser, 2008). Esta árvore contém polimorfismos quer da região controlo quer da região codificante do DNAmt, sendo permanentemente atualizada. A base da árvore filogenética engloba todos os haplogrupos caraterísticos de populações africanas, nomeadamente de regiões de África subsariana, no macrohaplogrupo L (Behar et al., 2008; Gonder, Mortensen, Reed, De Sousa, & Tishkoff, 2007; Pakendorf & Stoneking, 2005), que se ramifica nos haplogrupos L0, L1, L2, L3, L4, L5 e L6 (van Oven & Kayser, 2008). Foi a partir deste macrohaplogrupo L, principalmente do haplogrupo L3, que os restantes haplogrupos divergiram. Este haplogrupo L3, dispersou para regiões europeias e asiáticas, dando origem aos macrohaplogrupos M e N, que caraterizam as linhagens presentes nestas regiões. O macrohaplogrupo N, típico de populações europeias, divergiu nos haplogrupos H, I, J, K, T, U, V, W e X, que constituem aproximadamente 99% dos haplogrupos desta região (Behar et al., 2008; Budowle, Allard, Wilson, & Chakraborty, 2003; Pakendorf & Stoneking, 2005). As sequências de DNAmt, caraterísticas de regiões asiáticas, englobam haplogrupos que pertencem tanto ao macrohaplogrupo M como ao N (Toomas Kivisild, 2015).

O haplogrupo L0 é considerado como a primeira ramificação do DNAmt, que terá surgido há cerca de 140/160 000 anos, sendo a subestrutura mais primitiva observada em comunidades das regiões sul e este de África (Rosa & Brehm, 2011). A primeira classe a derivar do haplogrupo L0, há cerca de 100 000 anos, terá sido a L0d que se restringe às

populações Khoe-San, do sul de África, e às populações da Tanzânia e Angola (Antonio Salas et al., 2002).

O haplogrupo L1 terá surgido há aproximadamente 140/150 000 anos com origem no centro e ocidente do continente africano (Atkinson, Gray, & Drummond, 2009; Gonder et al., 2007; Toomas Kivisild et al., 2004; Schlebusch, Lombard, & Soodyall, 2013), sendo observado em populações do oeste de África (Fendt et al., 2012; González et al., 2006; Rosa, Brehm, Kivisild, Metspalu, & Villems, 2004) e muito especialmente em populações Mandenka e Wolof do Senegal (Rando et al., 1998). Encontra-se um pouco por toda a costa ocidental de África, em países como Mauritânia, Senegal, Serra Leoa e Gana (Behar et al., 2008; Fendt et al., 2012; González et al., 2006; Rando et al., 1998).

O haplogrupo L2 terá surgido há cerca de 90/105 000 anos na região oeste de África, apresentando ampla dispersão por todo o continente, e designadamente a sudeste e centro-oeste, como consequência das expansões Bantu e a noroeste como consequência do comércio transaariano de escravos (Harich et al., 2010; Silva et al., 2015; Tishkoff et al., 2007).

O haplogrupo L3 terá surgido há 60/75 000 anos na região este de África (Rosa & Brehm, 2011; Antonio Salas et al., 2002). Encontra-se amplamente disperso por todo o continente africano e muito especialmente em populações Bantu a este, sudoeste e oeste (Atkinson et al., 2009; Rosa & Brehm, 2011; van Oven & Kayser, 2008).

O haplogrupo L4 terá surgido antes do haplogrupo L3, também na região este de África, estando representado em países como a Etiópia e populações Barbers (Behar et al., 2008; Toomas Kivisild et al., 2004).

O haplogrupo L5 terá surgido há cerca de 110 000 anos na região este de África tendo especial representação em populações de países como Egipto, Sudão, Etiópia, Quénia, Ruanda e Tanzânia (Gonder et al., 2007; L. Pereira et al., 2001; Plaza et al., 2004; Rosa & Brehm, 2011) e populações Bantu (Rosa & Brehm, 2011).

O haplogrupo U, apesar de associado a regiões europeias, tem, também, sido identificado em populações indianas e populações africanas, designadamente do norte de África (Bermisheva et al., 2004; Finnilä, Lehtonen, & Majamaa, 2001; Rando et al., 1998). Terá surgido na Europa há cerca de 50 000 anos e ter-se-á dispersado para a Índia, próximo oriente e norte e noroeste de África (Bermisheva et al., 2004; Finnilä et al., 2001; T. Kivisild et al., 1999; Rando et al., 1998; Richards, 2000).

O haplogrupo K é considerado um subgrupo da haplogrupo U, sendo comum por toda a Europa (Coia et al., 2016; Torroni et al., 2000).

O haplogrupo H, terá surgido há 30/35 000 anos no médio oriente e ter-se-á dispersado para a Europa (Achilli et al., 2004). É o haplogrupo europeu mais comum, apresentando uma distribuição ampla no continente Europeu e em outras regiões, designadamente no norte de África, médio e próximo oriente, Índia e Ásia central (Achilli et al., 2004; Brotherton et al., 2014; Ottoni et al., 2010; Luísa Pereira et al., 2005; Roostalu et al., 2007; Turchi et al., 2016).

Os haplogrupos J, M, R, T e X são originários e encontram-se por toda a Europa e designadamente Europa central, Balcãs, Ucrânia, próximo oriente, caucaso e Ásia, mas também oeste europeu (Brandst??tter et al., 2008; Pala et al., 2012). O haplogrupo X surge também em populações da América do

norte, Península Ibérica e nordeste de África (Alvarez et al., 2007; Toomas Kivisild et al., 2004; Lima et al., 2006).

De todas as populações estudadas, as que apresentam maior variabilidade e heterogeneidade genética são as populações do continente africano (Y.-S. Chen et al., 2000). Em média, as alterações nucleotídicas destas populações são aproximadamente o dobro das variações observadas nos indivíduos euroasiáticos (Budowle et al., 1999).

Em 2010 foi publicado um dos primeiros estudos de genética de populações Portuguesas no qual foi analisada a totalidade da Região Controlo do DNAmt. Pereira e colaboradores apresentaram a caracterização de 285 indivíduos da região Sul de Portugal, através da análise da totalidade da Região Controlo do DNAmt (V. Pereira et al., 2010). Estudaram 160 indivíduos do município de Coruche, 50 indivíduos do município de Alcácer do Sul e 75 indivíduos da freguesia de Pias. A população foi descrita como uma população, globalmente, com haplogrupos típicos da Europa, mas, no entanto, em Alcácer do Sal 22% dos indivíduos estudados foram identificados como indivíduos com haplótipos integrados no macrohaplogrupo L, aliás, representação de indivíduos com haplogrupos tipicamente Africanos nunca registada em nehuma população Portuguesa estudada (V. Pereira et al., 2010).

Já em 2011 e 2013, Teixeira e colaboradores e Mairal e colaboradores, respetivamente, estudam indivíduos da população Portuguesa, analisando a totalidade da Região Controlo do DNAmt e identificam, maioritariamente, indivíduos com haplótipos correspondentes a haplogrupos preponderantemente Europeus, designadamente haplogrupos incluídos nos haplogrupos H, T2, U2, N1 e R0 (Mairal et al., 2013; Teixeira et al., 2011).

Mais recentemente, em 2015, foi divulgado um estudo com a caraterização da totalidade da Região Controlo do DNAmt da população portuguesa, que integrou 98 indivíduos da região Norte, 94 indivíduos da região Centro e 100 indivíduos do Sul de Portugal. Este estudo descreve Portugal como uma população típica da Europa ocidental com uma percentagem de linhagens maternas subsarianas inferior a 6% (Marques et al., 2015). No entanto, será importante sublinhar que este estudo apresenta aquilo que refere ser uma atualização da caraterização genética da população Portuguesa, que conta com cerca de 10 000 000 habitantes, através da análise de pouco mais que 300 indivíduos. A caraterização da população do Sul de Portugal, que conta com cerca de 4 000 000 habitantes (INE, 2014), é feita pelo estudo do DNAmt de 100 indivíduos. Na tabela 1 apresentamos os principais estudos populacionais com DNAmt realizados em populações ou grupos de indivíduos de Portugal.

Nas últimas décadas o número de imigrantes em Portugal, bem como aliás em grande parte dos países europeus, tem vindo a aumentar consideravelmente. De acordo com os dados divulgados na Base de Dados de Portugal Contemporâneo, PORDATA, no final de 2014, o número de imigrantes a residir em Portugal ultrapassou os 400 000 indivíduos, entre os quais cerca de 80 000 (20%) eram oriundos de países africanos de língua oficial Portuguesa, PALOP. Dos cerca de 80 000 imigrantes oriundos de PALOP, 65 000 tinham residência oficial na região de Lisboa. Destes cerca de 65 000 imigrantes, cerca de 34 000 são oriundos de Cabo Verde (51%), cerca de 15 000 são oriundos de Angola (23%), cerca de 15 000 são oriundos da Guiné-Bissau (23%) e cerca de 1 700 são oriundos de Moçambique (3%).

Deparamo-nos, portanto, atualmente, em Portugal e muito particularmente na região de Lisboa com uma realidade populacional diferente da existente anteriormente a este fluxo de imigrantes. Introduzirão estes emigrantes oriundos de PALOP linhagens maternas diferentes das linhagens maternas africanas já observadas em estudos anteriores na população Portuguesa?

Uma vez que as populações nativas de PALOP não estão estudadas para a totalidade da Região Controlo do DNAmt, ou se estão não existem dados disponíveis sobre as mesmas, que linhagens maternas vêm os imigrantes de PALOP introduzir na população da região de Lisboa?

A caraterização genética dos imigrantes de PALOP, a consequente determinação da variabilidade genética que vêm introduzir na população da região de Lisboa e a inserção deste grupo de indivíduos numa base de dados de DNA pública e internacional permitirá a utilização segura do estudo de DNAmt no âmbito de casos da rotina forenses, prática que, aliás, se tem vindo a sedimentar desde a década de 90 do século passado (Allen et al., 1998; Gill et al., 1994; Holland et al., 1993; Ivanov PL, Wadhams MJ, Roby RK, Holland MM, 1996; Lutz S, Weisser HJ, Heizmann J, 1996; Pfeiffer, Hühne, Ortmann, Waterkamp, & Brinkmann, 1999; Sullivan, Hopgood, & Gill, 1992; Wilson, DiZinno, Polanskey, Replogle, & Budowie, 1995).

Se por um lado, do ponto de vista médico-legal e forense, o DNAmt não permite a individualização de pessoas que o estudo de marcadores genéticos autossómicos, atualmente, permite e assegura, por outro lado pode permitir, no entanto, corroborar conclusões no âmbito da identificação humana médico-legal positiva de indivíduos ou de cadáveres desconhecidos e, sobretudo, faz um diagnóstico seguro nos

casos de exclusão e, mais ainda, no caso de amostras muito degradadas ou com DNA degradado, o DNAmt pode ser, em muitas situações, o único DNA que é possível recuperar para estudo.

Tabela 1 - Principais estudos populacionais com DNAmt realizados em populações ou grupos de indivíduos de Portugal

Referência Bibliográfica	Regiões DNAmt estudadas	N.º Indivíduos e Regiões	Macrohaplogrupo ou Haplogrupo com maior representação (%)	N.º ou Percentagem de indivíduos do Macrohaplogrupo L	N.º Indivíduos Hapl. L0
Côrte-Real H et al. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Am Hum Gene. 1996;60:331–50.	HVI	54 Portugal	N/D	N/D	N/D
Pereira L et al. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet. 2000;64(Pt 6):491–506.	HVI HVII	100 Norte 82 Centro 59 Sul	H (41%) Norte H (37.81%) Centro H (50%) Sul	5 Norte 8 Centro 4 Sul	0
Brehm A et al. Mitochondrial portraits of the Madeira and Açores archipelagos witness different genetic pools of its settlers. Hum Genet. 2003;114(1):77–86.	HVI	115 Madeira 179 Açores	H (36.2%) Madeira H (45.2%) Açores	12.9% Madeira 3.4% Açores	0
Carvalho M et al. mtDNA analysis in Portuguese populations: Polymorphic sites in control region sequences. Int Congr Ser.2003;1239:535–9.	HVI HVII	81 Centro 48 Açores	N/D	N/D	N/D
Gonzalez AM et al. Mitochondrial DNA affinities at the atlantic fringe of Europe. Am J Phys Anthropol. 2003;120(4):391–404.	HVI	84 Norte 78 Centro 137 Sul	H (20.1%) Norte H (31.5%) Centro H (26.5%) Sul	3.3% Norte 6.8% Centro 10.6% Sul	0
Santos C et al. Genetic structure and origin of peopling in the Azores islands (Portugal): the view from mtDNA. Ann Hum Genet. 2003;67(Pt 5):433–56.	HVI	146 Açores	H (32.19%)	4	0
Pereira V et al. Genetic characterization of uniparental lineages in populations from Southwest Iberia with past malaria endemicity. Am J Hum Biol. 2010;22(5):588–95.	RCT	160 Coruche 75 Pias 50 Alcácer do Sal	H (49.9%) Coruche H (34.7%) Pias H (40%) Alcácer do Sal	8.7% Coruche 3.9% Pias 22% Alcácer do Sal	2 Pias
Teixeira JC et al. Mitochondrial DNA-control region sequence variation in the NE Portuguese Jewish community. Forensic Sci Int Genet Suppl Ser. 2011;3(1):e51–2.	RCT	56 Bragança (Judeus)	H (35.7%)	0	0
Mairal Q et al. Linguistic isolates in Portugal: insights from the mitochondrial DNA pattern. Forensic Sci Int Genet. 2013;7(6):618–23.	RCT	121 Miranda do Douro	H (53.7%)	6	0
Marques SL et al. Portuguese mitochondrial DNA genetic diversity—An update and a phylogenetic revision. Forensic Sci Int Genet. 2015;15:27–32.	RCT	98 Norte 94 Centro 34 Sul	H (26.5%) Norte H (27.6%) Centro H (22%) Sul	3.1% Norte 6.4% Centro 6% Sul	1 Sul

Objetivos do Estudo

O principal objetivo do estudo será a caracterização genética dos imigrantes de PALOP a residir na região de Lisboa, através da análise da totalidade da região controlo do DNA mitocondrial.

Objectivos específicos:

1. Determinação de haplótipos/sequência da totalidade da região controlo do DNAmt de todos os imigrantes incluídos no estudo
2. Determinação dos haplogrupos correspondentes a todos os haplótipos determinados
3. Determinação da representação relativa de cada haplogrupo na população em estudo
4. Comparação dos resultados obtidos com informação anterior disponível quer relativamente às populações de origem quer relativamente à população de acolhimento
5. Caracterização dos indivíduos em estudo com um marcador genético nuclear autossómico com poder de informação em termos filogenéticos e de afiliação geográfica
6. Comparação os resultados obtidos com DNAmt e com marcador genético nuclear autossómico
7. Inserção das populações de imigrantes estudados na base de dados internacional EMPOP

Capítulo II - Estado da Arte: Mitochondrial DNA in Human Identification: a review

Amorim A, Fernandes T, Taveira N (2019) ***Mitochondrial DNA in Human Identification: a review***. *PeerJ Preprints*. DOI: 10.7287/peerj.preprints.27500v1

Mitochondrial DNA in human identification: a review

António Amorim [1, 2], **Teresa Fernandes** [3, 4], **Nuno Taveira** [Corresp. 5, 6]

[1] Instituto Nacional de Medicina Legal e Ciências Forenses, Lisbon, Portugal

[2] Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

[3] Escola de Ciências e Tecnologias, Universidade de Évora, Évora, Portugal

[4] Research Center for Anthropology and Health (CIAS), Universidade de Coimbra, Coimbra, Portugal

[5] Instituto Universitário Egas Moniz (IUEM), Almada, Portugal

[6] Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Lisbon, Portugal

Corresponding Author: Nuno Taveira
Email address: ntaveira@ff.ul.pt

Mitochondrial DNA (mtDNA) presents several characteristics useful for forensic studies, especially related to the lack of recombination, to a high copy number, and to matrilineal inheritance. mtDNA typing based on sequences of the control region or full genomic sequences analysis is used to analyze a variety of forensic samples such as old bones, teeth and hair, as well as other biological samples where the DNA content is low. Evaluation and reporting of the results requires careful consideration of biological issues as well as other issues such as nomenclature and reference population databases. In this work we review mitochondrial DNA profiling methods used for human identification and present their use in the main cases of humanidentification focusing on the most relevant issues for forensics.

Mitochondrial DNA in Human Identification: a review

Authors

António Amorim[1,2], Teresa Fernandes[3,4], Nuno Taveira[5,6*]

Affiliations

[1]Instituto Nacional de Medicina Legal e Ciências Forenses, Lisboa, Portugal

[2]Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

[3]Escola de Ciências e Tecnologias, Universidade de Évora, Évora, Portugal

[4]Research Center for Anthropology and Health (CIAS), Universidade de Coimbra

[5]Instituto Universitário Egas Moniz (IUEM), Almada, Portugal

[6]Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Portugal

* Corresponding author:

Nuno Taveira

Research Institute for Medicines (iMed.ULisboa), Faculty of Pharmacy, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisbon, Portugal

ntaveira@ff.ul.pt

Abstract

Mitochondrial DNA (mtDNA) presents several characteristics useful for forensic studies, especially related to the lack of recombination, to a high copy number, and to matrilineal inheritance. mtDNA typing based on sequences of the control region or full genomic sequences analysis is used to analyze a variety of forensic samples such as old bones, teeth and hair, as well as other biological samples where the DNA content is low. Evaluation and reporting of the results requires careful consideration of biological issues as well as other issues such as nomenclature and reference population databases. In this work we review mitochondrial DNA profiling methods used for human identification and present their use in the main cases of human identification focusing on the most relevant issues for forensics.

Introduction

Human genetic identification for forensic purposes is achieved through the definition of genetic profiles. A genetic profile or the genetic fingerprint of an individual is the phenotypic description of a set of genomic loci that are specific to that individual. In accordance with international recommendations, particularly with recommendations of the European DNA Profiling Group (EDNAP), currently, only genetic profiles obtained from autosomal short tandem repeats (STR) should be used for genetic fingerprinting. However, in a considerable number of situations of human identification, autosomal DNA is highly degraded or isn't available at all. In these cases the study of mitochondrial DNA (mtDNA) for human identification can be the last appeal and by that reason has become routine (Budowle, Allard, Wilson, & Chakraborty, 2003). Nevertheless, and despite the robustness of mtDNA in cases of exclusion or absence of identity between sequences, when the results are sequence identity, contrary to nuclear markers, they do not refer to an individual but to a group of individuals of the same maternal lineage.

Survey methodology

We systematically searched with PubMed Advanced Search Builder for papers Titles with the following combinations 1) mitochondrial DNA and biology, 2) mitochondrial DNA and guidelines, 3) mitochondrial DNA and nomenclature, 4) mitochondrial DNA and sequencing, 5) mitochondrial DNA and database, 6) mitochondrial DNA and data, 7) mitochondrial DNA and identification, 8) mitochondrial DNA and forensic. Papers nonrelated with human or animal mitochondrial DNA were excluded. Our search wasn't refined by publishing date, journal or impact factor of the journal, authors or authors affiliations. In addition, we used *Guideline* documents from the International Society for Forensic Genetics available at https://www.isfg.org/

Mitochondrial DNA biology and genetics

Mitochondria are cellular organelles that contain an extrachromosomal genome, which is both different and separate from the nuclear genome. The mitochondrial DNA (mtDNA) was first identified and isolated by Margit Nass and Sylvan Nass in 1963, who studied some mitochondrial fibers that according to their fixation, stabilization and staining behavior, appeared to be DNA related (Nass & Nass, 1963). However, the complete sequence of the first mtDNA was only published and established as the mtDNA Cambridge Reference Sequence (CRS) eighteen years later, in 1981 (Anderson et al., 1981).

Essentially, the mtDNA is a 5 mm histone-free circular double-stranded DNA molecule, with around 16 569 base-pairs and weighting 10^7 Daltons (Taanman, 1999). mtDNA strands have different densities due to different G+T base composition. The heavy (H) strand encodes more information, with genes for two rRNAs (12S and 16S), twelve polypeptides and fourteen tRNAs, while the light (L) strand encodes eight tRNAs and one polypeptide. All the 13 protein products are part of the enzyme complexes that constitute the oxidative phosphorylation system. Other characteristic features of the mtDNA are the intronless genes and the limited, or even absent, intergenic sequences, except in one regulatory region.

The mitochondrial D-loop is a triple-stranded region found in the major non-coding region (NCR) of many mitochondrial genomes, and is formed by stable incorporation of a third 680 bases DNA strand known as 7S DNA (Kefi-Ben Atig, Hsouna, Beraud-Colomb, & Abdelhak, 2009). The origin of replication is located at the non-coding or D-loop region, a 1,121 base pairs segment that is located between positions 16,024 and 516, according to the CRS numeration (Anderson et al., 1981) (Figure 1). The D-loop region, also comprehends two transcription promotors, one for each

strand. Nucleotide positions in the mtDNA genome are numbered according to the convention presented by Anderson et al. (Anderson et al., 1981), which was slightly modified by Andrews et al. (Andrews et al., 1999), determining the replacement of CRS for rCRS (revised Cambridge Reference Sequence). More precisely, the numerical designation of each base pair is initiated at an arbitrary position on the H strand, which continues thereafter and around the molecule for approximately 16,569 base pairs.

The apparent lack of mtDNA repair mechanisms and the low fidelity of the mtDNA polymerase lead to a significant higher mutation rate in the mitochondrial genome, when compared to the nuclear genome. For example, Sigurğardóttir and collaborators, estimated the mutation rate in the human mtDNA control region to be $0.32x10^{-6}$/site/year (Sigurðardóttir, Helgason, Gulcher, Stefansson, & Donnelly, 2000) which compares to 0.5×10^{-9}/site/year in the nuclear genome (Scally, 2016). Most of the sequence variation between individuals is found in two specific segments of the control region, namely in the hypervariable region 1 (HV1, positions 16,024 to 16,365) and in the hypervariable region 2 (HV2, positions 73 to 340) (Greenberg, Newbold, & Sugino, 1983). A third hypervariable region (HV3, positions 438 to 574), with additional polymorphic positions can be useful in the resolution of indistinguishable HV1/HV2 samples (Lutz et al., 2000). The small size and relatively high inter-person variability of the HV regions are very useful features for forensic testing purposes.

The mtDNA sequence defines the individual haplotype which is reported by the different base pairs relative to the rCRS mtDNA sequence. The collection of similar haplotypes defined by the combination of single nucleotide polymorphisms (SNPs) in mtDNA inherited from a common ancestor defines an haplogroup which was formed as a result of the sequential accumulation of mutations through maternal lineage (Mitchell et al., 2015).

A mitochondrion contains 2 to 10 copies of mtDNA and each somatic cell can have up to 1,000 mitochondria (Elson, Samuels, Turnbull, & Chinnery, 2001; Wei et al., 2017). Hence, when the amount of the extracted DNA is quite small or degraded, it is more likely that a DNA typing result can be obtained by typing the mtDNA than by typing polymorphic regions that are found in nuclear DNA.

Contrarily to the nuclear DNA, the mtDNA is exclusively maternally inherited, which justifies the fact that, apart from mutation, mtDNA sequence of siblings and all maternal relatives is identical (Case & Wallace, 1981; Giles, Blanc, Cann, & Wallace, 1980; Hutchison, Newbold, Potter, & Edgell, 1974). This specific characteristic can be very helpful in forensic cases, such as in the analysis of the remains of a missing person, where the known maternal relatives can provide some reference samples for a direct comparison to the mtDNA type. Due to the lack of recombination, maternal relatives from several generations apart from the source of evidence (or biological material) can be used for reference samples (Case & Wallace, 1981; Giles et al., 1980; Hutchison et al., 1974).

The haploid and monoclonal nature of the mtDNA in most individuals simplifies the process of interpretation of the DNA sequencing results. Still, it is possible to find heteroplasmy at occasional cases (Bendall, Macaulay, & Sykes, 1997; Bendall & Sykes, 1995; Comas, Paabo, & Bertranpetit, 1995; Gill et al., 1994; Ivanov et al., 1996; Wilson, Polanskey, Replogle, DiZinno, & Budowle, 1997). A person is considered as heteroplasmic if she/he carries more than one detectable mtDNA type. There are two classes of heteroplasmy, related to length polymorphisms and to point substitutions. Only the latter is important for forensic human identification. Most forensic laboratories worldwide do not report length polymorphisms and the guidelines on human identification with mtDNA do not point them as mandatory information (W. Parson et al., 2014;

Prinz et al., 2007). Furthermore, the information of length polymorphisms has no impact in haplogroups' definition.

Heteroplasmy manifests itself in diverse ways (Stewart et al., 2001). An individual may show more than one mtDNA type in a single tissue. An individual may be heteroplasmic in one tissue sample and homoplasmic in another one. Finally, an individual may exhibit one mtDNA type in one tissue and a different type in another tissue. Of the three possible scenarios, the last one is the least likely to occur. When heteroplasmy is found in the mtDNA of an individual, it usually differs at a single base, in HV1 or HV2.

Heteroplasmy was observed at position 16,169 of the mtDNA control region in the putative remains of Tsar Nicholas II of Russia and his brother, the Grand Duke of Russia Georgij Romanov (Gill et al., 1994; Ivanov et al., 1996). Comas et al. (Comas et al., 1995), in turn, detected heteroplasmy at two distinct positions, 16, 293 and 16,311, in the mtDNA of an anonymous donor's plucked hair. Wilson et al. (Wilson et al., 1997) found a family constituted by a mother and two children carrying a heteroplasmic mtDNA at position 16,355 both in blood and buccal swab samples.

The existence of heteroplasmic individuals and the limited knowledge about both the mechanism and the rate of heteroplasmy can be issues raised in an attempt to exclude mtDNA evidence from forensic investigations. Heteroplasmy at one nucleotide position is more frequently observed in hair samples, mainly due to genetic drift and to bottlenecks which occur due to the hair follicle's semiclonal nature (Budowle et al., 2003; Buffoli et al., 2013; Paus, 1998; Rogers, 2004). Hence, if an evidentiary hair sample contains one of the two heteroplasmic lineages that are observed in a reference sample, or vice versa, then the interpretation of exclusion may be incorrect. In this case, typing additional hairs may be required to solve the problem (Budowle et al., 2003).

As it was previously pointed out, it is accepted that the mitochondrial genome is maternally inherited. Even though the sperm contains a few mitochondria in the neck and in the tail region, the male mitochondrial genome is destroyed either during or shortly after the fertilization. More precisely, sperm mitochondria disappear in the early embryogenesis, either by selective destruction, inactivation or dilution (Cummins, Wakayama, & Yanagimachi, 1997; H. Shitara et al., 2000; Hiroshi Shitara, Hayashi, Takahama, Kaneda, & Yonekawa, 1998; D. C. Wallace, 2007). Nonethless, in the last years some cases of biparental mtDNA inheritance have been reported. In the most recent case, Luo and collaborators describe biparental mtDNA inheritance, either directly or indirectly, in 17 members of three multigenerational families, with results confirmed by 2 independent laboratories (Luo et al., 2018). Besides this report in humans, there are also a few examples of paternal inheritance of the mitochondrial genome in animals (Gyllensten, Wharton, Josefsson, & Wilson, 1991), and by that reason and despite the limited evidence for paternal inheritance of the mitochondrial genome in humans, the courtroom can be tempted to use such possibility to depreciate the use mtDNA evidences.

Mitochondrial DNA Nomenclature

Considering that listing more than 600 bases in order to describe the results from a new HV1 and HV2 sequence would be unpractical, an alternative approach was developed which essentially identifies and reports the differences relative to the reference sequence rCRS (Anderson et al., 1981).

Even though the process of naming mtDNA sequences seems simple and obvious, it is crucial to properly consider the nomenclatures, since complications might arise. Most ambiguities in the alignment/nomenclature arise due to insertions and/or deletions (indels). To avoid ambiguities,

168 facilitate haplotype identification and their assignment to existent haplogroups or to new
169 haplogroups, phylogenetic-based nomenclature guidelines have been proposed. The phylogenetic
170 approach provides an evolutionary based view of global mtDNA diversity that is scientifically
171 sound because all mtDNA lineages derive from a common maternal ancestor. The phylogenetic
172 notation of mtDNA haplogroups can be based on maximum parsimony or maximum likelihood
173 analysis of mutations present in sequences from the control region or in mitogenomes (for
174 comprehensive reviews in phylogenetic reconstruction methods see (Bianchi & Liò, 2007; De
175 Bruyn, Martin, & Lefeuvre, 2014). The most comprehensive repository of mtDNA genomes is
176 Phylotree (www.phylotree.org) a website that also provides the reference phylogenetic tree
177 describing the worldwide human mitochondrial DNA variation (van Oven & Kayser, 2008). The
178 phylogenetic tree shown in Phylotree is regularly updated with new haplogroups as found by
179 maximum parsimony analysis of new mtDNA haplotypes using the mtPhyl software
180 (https://sites.google.com/site/mtphyl/home) (van Oven, 2015). A maximum likelihood approach
181 for mtDNA haplogroup classification named EMMA has been recently described by Röck and
182 collaborators (Röck, Dür, Van Oven, & Parson, 2013).
183 Variants flanking long C tracts are subject to sequence-specific conventions. The long C tracts of
184 HVS-I and HVS-II should always be scored with 16,189C and 310C, respectively. Length
185 variation of the short A tract preceding 16,184 should be notated preferring transversions unless
186 the phylogeny suggests otherwise. Regarding deletions, these are recorded by the number of the
187 base(s) that is missing, with respect to the rCRS, followed by DEL or (-) (for example 249 DEL
188 or 249-). The bases that cannot be unambiguously determined are coded as N. Indels should be
189 placed 3' with respect to the light strand unless the phylogeny suggests otherwise. For example, if
190 the bases beyond the position 309 were out of the register by one base due to the insertion of a C

the mutation is designated as 309.1C. Two C insertions are designated as 309.1C and 309.2C. Important tools to assist with the notation of mtDNA sequences are available at http://empop.org/. These notations are used for storing haplotypes in the EMPOP database and have also been adopted by the Scientific Working Group on DNA Methods (SWGDAM) in the United States (W. Parson et al., 2014).

Overall, the large majority of individuals from African populations, and specially from sub-Saharan African populations, are categorized into one of the main haplogroup lineages that diverged from macro-haplogroup L - L0, L1, L2, L3, L4, L5 and L6 - (Allard et al., 2005; Bandelt et al., 2001; Behar et al., 2008; Chen et al., 2000; Gonder, Mortensen, Reed, De Sousa, & Tishkoff, 2007; Pakendorf & Stoneking, 2005; Rosa, Brehm, Kivisild, Metspalu, & Villems, 2004; van Oven & Kayser, 2008). On the other hand, more than 90% of the individuals of the European and USA Caucasian populations are categorized into 10 main haplogroup lineages - H, I, J, K, M, T, U, V, W and X - (Allard, Miller, Wilson, Monson, & Budowle, 2002; Behar et al., 2008; Budowle et al., 2003; Pakendorf & Stoneking, 2005; Torroni et al., 1996). Concerning to African-American populations, the most commonly observed haplogroups are L2a, L1c, L1b and L3b (Allard et al., 2005). The main haplogroups found in individuals from Asian populations are haplogroups M and N (Allard, Wilson, Monson, & Budowle, 2004; Kivisild, 2015).

Mitochondrial DNA Typing Guidelines

In 2014 the DNA Commission of the International Society of Forensic Genetics (ISFG) published updated guidelines and recommendations concerning mitochondrial DNA typing. These guidelines referred to good laboratory practices, targeted region, amplification and sequencing ranges, reference sequence, alignment and notation, heteroplasmy, haplogrouping of mtDNA

sequences, and databases and database searches. In Table 1 we present the 16 recommendations of ISFG. Overall, these are the main guidelines concerning the application of mtDNA polymorphisms in human identification, which are regularly revised and published by the International Society of Forensic Genetics (Bär et al., 2000; W. Parson et al., 2014; Prinz et al., 2007; Tully et al., 2001).

Mitochondrial DNA Sequencing Methodologies

In 1977 Sanger presented the first DNA sequencing technology (Sanger, Nicklen, & Coulson, 1977), also called the chain termination method and now known as first generation sequencing. The incorporation of ddNTPs in newly synthesized DNA strands results in termination of the elongation process and correspondent knowledge about the specific nucleotide present at the sequence at each position. Sanger sequencing method can produce reads from 25 up to 1200 nucleotides, allowing the read of a maximum of 96 kb nucleotides in 2 hours.

Since 2005 new sequencing methods, also known as next generation sequencing (NGS) methods, have been developed (Bruijns, Tiggelaar, & Gardeniers, 2018). Sequencing by synthesis methods such as Roches' - 454 Pyrosequencing -, and Illuminas'- HiSeq -, allow sequencing up to 80 million base pairs in 2 hours or up to 6 billion base pairs in 1- 2 weeks (Mascher, Wu, St. Amand, Stein, & Poland, 2013; Pukk et al., 2015). Sequencing by hybridization and ligation such as ABIs'- SOLiD 3plus platform -, yields 60 gigabases of usable DNA data per run. With these massive parallel sequencing (MPS) technologies, sequenced DNA fragments can range from 35-75 nucleotides as in the SOLiD technology (Bruijns et al., 2018; Ondov et al., 2010; Shendure et al., 2005), to 100-1,000 nucleotides as in 454 Pyrosequencing (Bruijns et al., 2018; Dames, Durtschi, Geiersbach, Stephens, & Voelkerding, 2010). Ion Torrent's Personal Genome Machine™ (PGM)

uses a detection methodology based in pH change upon addition of a nucleotide to a sequence (Rothberg et al., 2011). When this happens protons are released generating an electric signal that is proportional to the amount of protons released. Data collection is carried out by a complementary metal-oxide semiconductor (CMOS) sensor array chip with the sensor surface present at the bottom of the well plate, and these chips can measure millions to billions of simultaneous sequencing reactions (Liu et al., 2014). Finally, MinION (Oxford Nanopore Technologies), a portable real-time sequencing device, allows ultra-long read lengths (hundreds of kb) albeit with lower accuracy (Oikonomopoulos, Wang, Djambazian, Badescu, & Ragoussis, 2016).

The NGS technologies have been quickly applied in forensics (Bruijns et al., 2018). For example, Ion Torrent's PGM system has been used for sequencing complete mitogenomes (Walther Parson et al., 2013) and to study heteroplasmy (Magalhães et al., 2015) in the forensic context. Although PGM proved to be sensitive and accurate at detecting and quantifying mixture and heteroplasmy, there were some problems in the coverage of the mtDNA genome with some regions presenting extreme strand bias, and presenting false positives mostly generated by alignment problems in the analysis algorithms. More recently Ion S5 System (Thermo Fisher Scientific) and MiSeq FGx Desktop Sequencer (Illumina) were used to evaluate the Precision ID mtDNA Whole Genome Panel (Woerner et al., 2018). Both sequencing systems provided consistent estimation of mtDNA haplotypes. Many other studies on the use of NGS technologies for forensic genetics and mtDNA analysis have been published (Chaitanya et al., 2015; Churchill, Stoljarova, King, & Budowle, 2018; Hollard et al., 2017; Just, Irwin, & Parson, 2015; Just, Scheible, Fast, Sturk-Andreaggi, Higginbotham, et al., 2014; Just, Scheible, Fast, Sturk-Andreaggi, Röck, et al., 2014; Lopopolo, Børsting, Pereira, & Morling, 2016; Ma et al., 2018; Marshall et al., 2017; Ovchinnikov, Malek, Kjelland, & Drees, 2016; Park et al., 2017; Templeton et al., 2013; Young, King, Budowle, &

Armogida, 2017). However, further validation studies and specialized software functionality tailored to forensic practice should be produced in order to facilitate the incorporation of NGS processing into standard casework applications (Amorim & Pinto, 2018; Peck et al., 2016). In the meantime, and according to current international guidelines (W. Parson et al., 2014; Prinz et al., 2007), Sanger sequencing still continues to be an adequate method for mtDNA analysis for forensic human identification, and is used in most casework laboratories worldwide (Ballard, 2016). Some forensic laboratories perform Sanger sequencing for HVI and HVII fragments, while others have already extended the study to the HVIII fragment and, in recent years, most of the forensic laboratories are introducing the amplification of the entire control region as routine methodology (Chaitanya et al., 2016; Poletto, Malaghini, Silva, Bicalho, & Braun-Prado, 2019; Turchi et al., 2016; Yasmin, Rakha, Noreen, & Salahuddin, 2017). Attempting to improve the power of mtDNA in human identification, over the past decade some studies have been focused in the extension of the analyses to the whole mtDNA genome (Duan et al., 2018; Strobl, Eduardoff, Bus, Allen, & Parson, 2018; Woerner et al., 2018). Nevertheless, it should be stressed that while the information from the entire mtDNA genome can contribute to refine the haplogroup obtained with the study of HVI, HVII and HVII fragments or the entire control region, it is not supposed to change the previous results to a different haplogroup of a completely different geographic ancestry.

Mitochondrial DNA Population Data and Databases

When two mtDNA sequences, one from an evidence sample and another from a reference sample, cannot be excluded as being originated from the exact same source, it is necessary to convey some information concerning the rarity of the mtDNA profile. The current practice is to count how many times a specific sequence is observed within a population database(s) (Budowle et al., 1999).

Overall, the population databases that are used in forensics comprehend several convenience samples, representing the major population groups of the potential contributors in terms of evidence.

The most important mtDNA haplotypes database is the EDNAP Mitochondrial DNA Population Database (EMPOP, www.empop.org) (Walther Parson & Dür, 2007). In its early stages, EMPOP was designed and envisioned to serve as a reference population database, specifically to be used in the evaluation of the mtDNA evidence around the world, aiming to provide the highest quality mtDNA data. The architecture of this online database and its analysis tools have evolved over the last few years, even though the main emphasis of the EMPOP database remains to be mtDNA data quality. Therefore, and as a direct consequence, EMPOP not only serves as a reference population database, but also as a quality-control tool for scientists in forensic genetics, as well as in other disciplines. Finally, and even though there is a significant number of high-quality reference population databases for forensic comparisons, EMPOP is the most comprehensive resource, especially from the standpoint of the populations that are represented in such database (W. Parson et al., 2014).

EMPOP uses SAM, a string-based search algorithm that converts query and database sequences into alignment-free nucleotide strings and thus guarantees that a haplotype is found in a database query regardless of its alignment. SAM-E, an updated version of SAM that considers block InDels as phylogenetic events, is used currently. At EMPOP, the tool haplogroup browser represents all the established Phylotree haplogroups in convenient searchable format and provides the number of EMPOP sequences assigned to the respective haplogroups by estimating mitochondrial DNA haplogroups using the maximum likelihood approach EMMA (Röck et al., 2013). For multiple possible haplogroups, most recent common ancestor (MRCA) haplogroups are provided.

As mentioned before, PhyloTree provides an updated comprehensive phylogeny of global human mtDNA variation, based on both coding and control region mutations (van Oven & Kayser, 2008). The complete mtDNA phylogenetic tree includes previously published as well as newly identified haplogroups, is continuously and regularly updated, and is available online at http://www.phylotree.org. In Figure 2 we present the representation of the mtDNA phylogenetic tree, Build 17, which is divided into 25 subtrees. At EMPOP the geographical haplogroup patterns are provided via maps to visualize and better understand their geographical distribution (Figure 3). Another important human mtDNA database is *Mitomap* (Ruiz-Pesini et al., 2007). In 1996, this database developed into an online database, www.mitomap.org, containing published human mtDNA variation along with geographic and disease specific variants. Currently, *Mitomap* is manually curated, frequently updated and a functionally rich resource, presenting high-quality human mtDNA data for clinicians, investigators and geneticists (Ruiz-Pesini et al., 2007). *Mitomap* has three main categories for usage. It contains some background information regarding the human mitochondrial DNA, such as the general representation of mtDNA, haplogroups and their frequencies and illustrations of mtDNA, among others. Furthermore, users can also find information about other mtDNA-specific databases, tools and useful resources.

Mitomap stores the annotated listing of the mtDNA variants from both healthy individuals and patients. The frequencies of the variants are calculated from human mitogenomes retrieved from the *GenBank*. Therefore, users can retrieve information about the loci, the nucleotide change, the codon position and the number, among others, and download the most important data in different file formats.

Mitomap contains the *Mitomaster* analysis tool, currently providing the Application Programming Interface for it. The main function of this tool is to allow the identification of polymorphic

329 positions, the calculation of variant statistics and the assignment of haplogroups to complete or
330 partial mitogenomes. Such query might be performed by recurring to mtDNA sequences, to
331 *GenBank* identifiers or to single nucleotide variants (Brandon et al., 2009).
332 From another perspective, ethical and legal problems may arise in the implementation of mtDNA
333 databases. The informative potential which the analysis of mtDNA entails can generate privacy
334 questions (Guillen, Lareu, Pestoni, Salas, & Carracedo, 2000; H. M. Wallace, Jackson, Gruber, &
335 Thibedeau, 2014). Mitochondrial diseases affect between 1 in 4,000 and 1 in 5,000 people. In most
336 people, primary mitochondrial disease is a genetic condition that can be inherited. Information
337 about the mitochondrial genome composition may therefore enable the identification of the current
338 or future state of health of an individual. For this reason, the analysis of mtDNA must be carried
339 out only on non-coding regions, which have not been associated with any kind of disease or
340 phenotypical information.
341

342 ## Mitochondrial DNA in Forensic Human Identification

343 In the context of forensic analysis, both mtDNA sequences of a reference sample and an evidence
344 sample(s) are compared. If the mtDNA sequences are identical, the samples can't be excluded
345 since they must have the same origin or derive from the same maternal lineage. Similarly, samples
346 can't be excluded when heteroplasmy is observed at the same nucleotide positions in both samples.
347 Finally, when one sample is heteroplasmic and the other is homoplasmic but they both share at
348 least one mtDNA species, the samples can't be excluded since they may have the same origin.
349 Several authors have suggested that samples with mtDNA with one-base difference should be
350 further evaluated, mainly regarding their rate of mutation (Alonso et al., 2002; Bär et al., 2000;

Holland & Parsons, 1999; Tully et al., 2001). When two or more nucleotide differences exist between the two sequences, the overall interpretation is exclusion (W. Parson et al., 2014).

At this section we present some selected published cases of human identification with mtDNA. Table 2 summarizes the selected published cases. In 1991, Stoneking and collaborators presented the first report of successful application of the mtDNA typing to a case that involved the individual identification of skeletal remains (Stoneking, Hedgecock, Higuchi, Vigilant, & Erlich, 1991). This was the case of a 3-year-old child disappeared from her parents' house in October of 1984. In March of 1986, the skeletal remains of a human child were found in the desert, 2 miles away from the parents' residence. Using hybridization with 23 sequence-specific oligonucleotide probes (SSO) targeting nine regions of HV1 and HV2 on the control region, they found that the skeletal sample and the mother shared the same mtDNA types, corroborating that those skeletal remains were of the missing child. Moreover, they anticipated that the mtDNA typing would be valuable not only in linking biological remains to missing individuals, but also in the analysis of material in sexual assault cases.

In July of 1990, the body of a female, in a quite advanced state of decomposition, was discovered in an open field. Despite being impossible to identify the remains by analyzing the individual's clothes and fingerprints, her dentition was consistent with old dental records of a missing person from the same region. Some fragments of the heel bone and fibula, plus samples of the hair and skin, were provided for the DNA analysis, as well as a blood sample from a putative sister of the deceased. In 1992, Sullivan and collaborators attempted the identification of the highly decomposed remains of the corpse, amplifying and directly sequencing 2 hypervariable segments within HV1 and HV2 in the mtDNA (Sullivan, Hopgood, & Gill, 1992). No statistical value was given to the evidence, since no database of the British population sequences were available at that

time. Still, no differences were found between both sequences, the blood of the putative sister and the bone of the corpse, indicating they were sisters.

Perhaps the most well-known lineage study using mtDNA sequencing is related to the identification of Tsar Nicholas II's bones. Gill and collaborators, in 1994 (Gill et al., 1994), and Ivanov and collaborators, in 1996 (Ivanov et al., 1996), compared the sequences of HV1 and HV2 fragments of the mtDNA obtained from the putative bones of the Tsar with those of Tsar living maternal relatives, Countess Xenia Cheremeteff-Sfiri and the Duke of Fife. It was found that the sequences were very similar, corroborating the hypothesis that the bone remains were of Tsar Nicholas II.

In a distinct scenario, Deng et al. (Deng et al., 2005) used direct sequencing of the HV1 and HV2 fragments of the mtDNA control region to identify Tsunami victims in Thailand in 2004. This tsunami killed nearly 5,400 people in Southern Thailand, including foreign tourists and local residents. They succeeded in obtaining fully informative results for mtDNA markers (HV1 and HV2) from 258 tooth samples with a success rate of 51% (258/507).

More recently, in 2010, Ríos and collaborators (Ríos, García-Rubio, Martínez, Alonso, & Puente, 2010) used direct sequencing of the HV1 and HV2 fragments of the mtDNA control region to identify human skeletal remains that were exhumed from a mass grave from the Spanish Civil War (1936-1939). There was a match between the mtDNA profiles of the biologically youngest skeleton and the sister of the youngest person that was presumptively known to be buried in the grave, allowing the identification of that person.

Also in 2010, Piccinini and collaborators (Piccinini et al., 2010) attempted to identify the remains of a famous World War One Italian soldier that was killed in a battle along the Italian front in 1915. Like previous studies, they used the direct sequencing of the HV1 and HV2 fragments of

the mtDNA control region to define single mtDNA haplotypes. The availability of the offspring maternal lineage allowed the mtDNA analysis, which presented a clear exclusion scenario: the remains did not belong to the supposed war hero.

In 2012, a skeleton was excavated at the site of the Grey Friars friary, in Leicester, which is the last-known resting place of King Richard III (King et al., 2014). To determine if the remains belonged to King Richard III, the HV1, HV2 and HV3 regions of the mtDNA of the skeletal remains and of the living relatives of King Richard III were sequenced and compared. There was a perfect match between the sequences indicating that the remains belong to King Richard III.

The communist period in Poland during 1944-1956 resulted in the death of more than 50,000 people, who were buried in secret. One mass grave was found at the cemetery Powazki Military, in Warsaw, Poland. In 2016, Ossowski and collaborators (Ossowski et al., 2016) identified 50 victims, specifically by using autosomal, Y-STR and direct sequencing of the HV1 and HV2 fragments of the mtDNA control region.

In 2016, among the first studies on human identification with mtDNA using massive parallel sequencing, Ambers and collaborators proposed a protocol that includes the study of ten regions of mtDNA for the identification of historical human remains with forensic genetic markers (Ambers et al., 2016). They studied a 140-year-old human skeletal remains discovered at a historical site in Deadwood, South Dakota, United States. The remains were in an unmarked grave and there were no records available regarding the identity of the individual. The mtDNA profiles of the unidentified skeletal remains obtained with their method were consistent with H1 haplogroup. This haplogroup is the most common in Western Europe. The ancestry-informative nuclear SNPs also studied in this case indicated a European background. These genetic results are

419 consistent with the findings of previous anthropological report which determined that the
420 Deadwood unidentified skeletal remains belong to a male of European ancestry.
421 In 2017, the victims' remains from the World Trade Center terrorism act, which occurred in
422 September 11 of 2001, were still being identified by using the mtDNA sequencing technology,
423 among other techniques, with protocols and guidelines as recommended by the International
424 Society for Forensic Genetics (Goodwin, 2017).
425

426 ## Conclusions

427 Over the last 25 years, mtDNA typing has been widely used around the world to solve several
428 human identification related issues in violent crimes, lesser crimes, acts of terrorism, mass
429 disasters and missing persons' cases. The progress in mtDNA typing has been overwhelming,
430 going from the examination of small fragments in a matter of days to sequencing multiple entire
431 mtDNA genomes in a couple of hours. Being a lineage genetic marker, mtDNA genome can
432 provide information about ancestors, including health/disease information. Even though many
433 would readily accept that there are good reasons for researchers to obtain information about an
434 unknown suspect's potential ancestral background, many still find the potential to determine
435 genetic dispositions to certain disorders as being unacceptable. Hence, new technologies that
436 enable mitogenome sequencing must be wisely used and for the reasons that they are intended,
437 considering their specific focus and contribution within the field of forensic human identification.
438 Some concerns still remain regarding admissibility of mtDNA analysis in court especially related
439 with the issue of heteroplasmy and, more recently, with the possibility of biparental inheritage.
440 The complete elucidation of molecular mechanisms driving biparental inheritage of mtDNA, the
441 ability to determine the situations where this is likely to occur, and the ability to identify and

characterize heteroplasmy with high accuracy, are important issues that need to be addressed in order to ensure the robustness of mtDNA as an important and alternative tool in forensic human identification.

Acknowledgements

None to report.

References

Allard, M. W., Miller, K., Wilson, M., Monson, K., & Budowle, B. (2002). Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. Scientific Working Group on DNA Analysis Methods. *Journal of Forensic Sciences*, *47*(6), 1215–1223.

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), 169–179. https://doi.org/10.1016/j.forsciint.2004.06.001

Allard, M. W., Wilson, M. R., Monson, K. L., & Budowle, B. (2004). Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. *Legal Medicine*, *6*(1), 11–24. https://doi.org/10.1016/j.legalmed.2003.08.004

Alonso, A., Salas, A., Albarrán, C., Arroyo, E., Castro, A., Crespillo, M., Gómez, J. (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. https://doi.org/10.1016/S0379-0738(01)00602-8

Ambers, A. D., Churchill, J. D., King, J. L., Stoljarova, M., Gill-King, H., Assidi, M., Budowle, B. (2016). More comprehensive forensic genetic marker analyses for accurate human remains identification using massively parallel DNA sequencing. *BMC Genomics*, *17*(Suppl 9). https://doi.org/10.1186/s12864-016-3087-2

Amorim, A., & Pinto, N. (2018). Big data in forensic genetics. *Forensic Science International: Genetics*, *37*(August), 102–105. https://doi.org/10.1016/j.fsigen.2018.08.001

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457–465. https://doi.org/10.1038/290457a0

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. https://doi.org/10.1038/13779

Ballard, D. (2016). Analysis of Mitochondrial Control Region Using Sanger Sequencing (pp. 143–155). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3597-0_12

Bandelt, H. J., Alves-Silva, J., Guimarães, P. E. M., Santos, M. S., Brehm, A., Pereira, L., Pena, S. D. J. (2001). Phylogeography of the human mitochondrial haplogroup L3e: A snapshot of African prehistory and Atlantic slave trade. *Annals of Human Genetics*, *65*(6), 549–563. https://doi.org/10.1046/j.1469-1809.2001.6560549.x

Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Holland, M., Wilson, M. (2000). Guidelines for mitochondrial DNA typing. DNA Commission of the International Society for Forensic Genetics. *Vox Sang*, *79*(2), 121–125.

488 Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Rosset, S.
489 (2008). The dawn of human matrilineal diversity. *American Journal of Human Genetics*,
490 *82*(5), 1130–1140. https://doi.org/10.1016/j.ajhg.2008.04.002
491 Bendall, K. E., Macaulay, V. A., & Sykes, B. C. (1997). Variable levels of a heteroplasmic point
492 mutation in individual hair roots. *American Journal of Human Genetics*, *61*(6), 1303–1308.
493 https://doi.org/10.1086/301636
494 Bendall, K. E., & Sykes, B. C. (1995). Length heteroplasmy in the first hypervariable segment of
495 the human mtDNA control region. *American Journal of Human Genetics*, *57*(2), 248–256.
496 Bianchi, L., & Liò, P. (2007). Forensic DNA and bioinformatics. *Briefings in Bioinformatics*,
497 *8*(2), 117–128. https://doi.org/10.1093/bib/bbm006
498 Brandon, M. C., Ruiz-pesini, E., Mishmar, D., Procaccio, V., Marie, T., Nguyen, K. C.,
499 Douglas, C. (2009). MITOMASTER - A Bioinformatics Tool For the Analysis of
500 Mitochondrial DNA Sequences. *Human Mutation*, *30*(1), 1–6.
501 https://doi.org/10.1002/humu.20801.MITOMASTER
502 Bruijns, B., Tiggelaar, R. M., & Gardeniers, H. J. G. E. (2018). Massively parallel sequencing
503 techniques for forensics: A review. *Electrophoresis*, 1–13.
504 https://doi.org/10.1002/elps.201800082
505 Budowle, B., Allard, M. W., Wilson, M. R., & Chakraborty, R. (2003). FORENSICS AND
506 MITOCHONDRIAL DNA: Applications, Debates, and Foundations*. *Annual Review of*
507 *Genomics and Human Genetics*, *4*(1), 119–141.
508 https://doi.org/10.1146/annurev.genom.4.070802.110352
509 Budowle, B., Wilson, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., Holland, M. M., &
510 Monson, K. L. (1999). Mitochondrial DNA regions HVI and HVII population data.

Forensic Science International, *103*(1), 23–35. https://doi.org/10.1016/S0379-0738(99)00042-0

Buffoli, B., Rinaldi, F., Labanca, M., Sorbellini, E., Trink, A., Guanziroli, E., Rodella, L. F. (2013). The human hair: from anatomy to physiology. *International Journal of Dermatology*, (November 2015), 1–11. https://doi.org/10.1111/ijd.12362

Case, J., & Wallace, D. (1981). Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts. *Somatic Cell Genetics*, *7*(1), 103–108.

Chaitanya, L., Ralf, A., van Oven, M., Kupiec, T., Chang, J., Lagacé, R., & Kayser, M. (2015). Simultaneous Whole Mitochondrial Genome Sequencing with Short Overlapping Amplicons Suitable for Degraded DNA Using the Ion Torrent Personal Genome Machine. *Human Mutation*, *36*(12), 1236–1247. https://doi.org/10.1002/humu.22905

Chaitanya, L., van Oven, M., Brauer, S., Zimmermann, B., Huber, G., Xavier, C., Kayser, M. (2016). High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database. *Forensic Science International. Genetics*, *21*, 158–167. https://doi.org/10.1016/j.fsigen.2015.12.002

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. *The American Journal of Human Genetics*, *66*(4), 1362–1383. https://doi.org/10.1086/302848

Churchill, J. D., Stoljarova, M., King, J. L., & Budowle, B. (2018). Massively parallel sequencing-enabled mixture analysis of mitochondrial DNA samples. *International Journal of Legal Medicine*, *132*(5), 1263–1272. https://doi.org/10.1007/s00414-018-1799-3

Comas, D., Paabo, S., & Bertranpetit, J. (1995). Heteroplasmy in the control region of human

534 mitochondrial DNA. *Genome Research*, *5*(1), 89–90. https://doi.org/10.1101/gr.5.1.89
535 Cummins, J. M., Wakayama, T., & Yanagimachi, R. (1997). Fate of microinjected sperm
536 components in the mouse oocyte and embryo. *Zygote*, *5*(4), 301–308.
537 https://doi.org/10.1017/S0967199400003889
538 Dames, S., Durtschi, J., Geiersbach, K., Stephens, J., & Voelkerding, K. V. (2010). Comparison
539 of the illumina genome analyzer and roche 454 GS FLX for resequencing of hypertrophic
540 cardiomyopathy-associated genes. *Journal of Biomolecular Techniques*, *21*(2), 73–80.
541 De Bruyn, A., Martin, D. P., & Lefeuvre, P. (2014). Phylogenetic Reconstruction Methods: An
542 Overview (pp. 257–277). Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-62703-
543 767-9_13
544 Deng, Y. J., Li, Y. Z., Yu, X. G., Li, L., Wu, D. Y., Zhou, J., Yu, J. (2005). Preliminary DNA
545 identification for the tsunami victims in Thailand. *Genomics, Proteomics and*
546 *Bioinformatics*, *3*(3), 143–157. https://doi.org/10.1016/S1672-0229(05)03020-2
547 Duan, M., Chen, L., Ge, Q., Lu, N., Li, J., Pan, X., Wilson, M. R. (2018). Evaluating
548 heteroplasmic variations of the mitochondrial genome from whole genome sequencing data.
549 *Future Science OA*, *699*(9), FSO336. https://doi.org/10.4155/fsoa-2018-0059
550 Elson, J. L., Samuels, D. C., Turnbull, D. M., & Chinnery, P. F. (2001). Random Intracellular
551 Drift Explains the Clonal Expansion of Mitochondrial DNA Mutations with Age. *The*
552 *American Journal of Human Genetics*, *68*(3), 802–806. https://doi.org/10.1086/318801
553 Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human
554 mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States*
555 *of America*, *77*(11), 6715–6719. https://doi.org/10.1073/pnas.77.11.6715
556 Gill, P., Ivanov, P. L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Sullivan, K. (1994).

557 Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics*,
558 *6*(2), 130–135. https://doi.org/10.1038/ng0294-130
559 Gonder, M. K., Mortensen, H. M., Reed, F. A., De Sousa, A., & Tishkoff, S. A. (2007). Whole-
560 mtDNA genome sequence analysis of ancient african lineages. *Molecular Biology and*
561 *Evolution*, *24*(3), 757–768. https://doi.org/10.1093/molbev/msl209
562 Goodwin, W. H. (2017). The use of forensic DNA analysis in humanitarian forensic action: The
563 development of a set of international standards. *Forensic Science International*, *278*, 221–
564 227. https://doi.org/10.1016/j.forsciint.2017.07.002
565 Greenberg, B. D., Newbold, J. E., & Sugino, A. (1983). Intraspecific nucleotide sequence
566 variability surrounding the origin of replication in human mitochondrial DNA. *Gene*, *21*(1–
567 2), 33–49. https://doi.org/10.1016/0378-1119(83)90145-2
568 Guillen, M., Lareu, M. V., Pestoni, C., Salas, A., & Carracedo, A. (2000). Ethical-legal problems
569 of DNA databases in criminal investigation. *Journal of Medical Ethics*, *26*(4), 266–271.
570 https://doi.org/10.1136/jme.26.4.266
571 Gyllensten, U., Wharton, D., Josefsson, A., & Wilson, A. C. (1991). Paternal inheritance of
572 mitochondrial DNA in mice. *Nature*, *352*(6332), 255–257.
573 https://doi.org/10.1038/352255a0
574 Holland, M. M., & Parsons, T. J. (1999). Mitochondrial DNA Sequence Analysis - Validation
575 and Use for Forensic Casework. *Forensic Science Rev*, *11*(1), 21–50.
576 Hollard, C., Keyser, C., Delabarde, T., Gonzalez, A., Vilela Lamego, C., Zvénigorosky, V., &
577 Ludes, B. (2017). Case report: on the use of the HID-Ion AmpliSeq™ Ancestry Panel in a
578 real forensic case. *International Journal of Legal Medicine*, *131*(2), 351–358.
579 https://doi.org/10.1007/s00414-016-1425-1

580 Hutchison, C. A., Newbold, J. E., Potter, S. S., & Edgell, M. H. (1974). Maternal inheritance of
581 mammalian mitochondrial DNA. *Nature*, *251*(October 1974), 536–538.
582 Ivanov, P. L., Wadhams, M. J., Roby, R. K., Holland, M. M., Weedn, V. W., Parsons, T. J., &
583 Ivanov PL, Wadhams MJ, Roby RK, Holland MM, W. V. & P. T. (1996). Mitochondrial
584 DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the
585 authenticity of the remains of Tsar Nicholas II. *Nat Genet.*, *12*(12), 417–420.
586 https://doi.org/10.1038/ng0496-417
587 Just, R. S., Irwin, J. A., & Parson, W. (2015). Mitochondrial DNA heteroplasmy in the emerging
588 field of massively parallel sequencing. *Forensic Science International: Genetics*, *18*, 131–
589 139. https://doi.org/10.1016/j.fsigen.2015.05.003
590 Just, R. S., Scheible, M. K., Fast, S. A., Sturk-Andreaggi, K., Higginbotham, J. L., Lyons, E. A.,
591 Irwin, J. A. (2014). Development of forensic-quality full mtGenome haplotypes: Success
592 rates with low template specimens. *Forensic Science International: Genetics*, *10*(1), 73–79.
593 https://doi.org/10.1016/j.fsigen.2014.01.010
594 Just, R. S., Scheible, M. K., Fast, S. A., Sturk-Andreaggi, K., Röck, A. W., Bush, J. M., Irwin, J.
595 A. (2014). Full mtGenome reference data: Development and characterization of 588
596 forensic-quality haplotypes representing three U.S. populations. *Forensic Science*
597 *International: Genetics*, *14*, 141–155. https://doi.org/10.1016/j.fsigen.2014.09.021
598 Kefi-Ben Atig, R., Hsouna, S., Beraud-Colomb, E., & Abdelhak, S. (2009). ADN mitochondrial:
599 propriétés et applications. *Archives de l'Institut Pasteur de Tunis*, *86*(1–4), 3–14.
600 King, T. E., Fortes, G. G., Balaresque, P., Thomas, M. G., Balding, D., Delser, P. M., Schürer,
601 K. (2014). Identification of the remains of King Richard III. *Nature Communications*, *5*, 1–
602 8. https://doi.org/10.1038/ncomms6631

603 Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial
604 genomes. *Investigative Genetics*, *6*(1), 1–10. https://doi.org/10.1186/s13323-015-0022-2
605 Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Law, M. (2014). Comparison of next-generation
606 sequencing systems. *The Role of Bioinformatics in Agriculture*, *2012*, 1–25.
607 https://doi.org/10.1201/b16568
608 Lopopolo, M., Børsting, C., Pereira, V., & Morling, N. (2016). A study of the peopling of
609 Greenland using next generation sequencing of complete mitochondrial genomes. *American*
610 *Journal of Physical Anthropology*, *161*(4), 698–704. https://doi.org/10.1002/ajpa.23074
611 Luo, S., Valencia, C. A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Huang, T. (2018). Biparental
612 Inheritance of Mitochondrial DNA in Humans. *Proceedings of the National Academy of*
613 *Sciences*, *115*(51), 13039–13044. https://doi.org/10.1073/PNAS.1810946115
614 Lutz, S., Wittig, H., Weisser, H. J., Heizmann, J., Junge, A., Dimo-Simonin, N., Augustin, C.
615 (2000). Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be
616 distinguished by sequencing the HVI and HVII regions? *Forensic Science International*,
617 *113*(1–3), 97–101. https://doi.org/10.1016/S0379-0738(00)00222-X
618 Ma, K., Zhao, X., Li, H., Cao, Y., Li, W., Ouyang, J., Liu, W. (2018). Massive parallel
619 sequencing of mitochondrial DNA genomes from mother-child pairs using the ion torrent
620 personal genome machine (PGM). *Forensic Science International: Genetics*, *32*(November
621 2017), 88–93. https://doi.org/10.1016/j.fsigen.2017.11.001
622 Magalhães, S., Marques, S. L., Alves, C., Amorim, A., Alvarez, L., & Goios, A. (2015).
623 Evaluation of heteroplasmy detection in the Ion Torrent PGM. *Forensic Science*
624 *International: Genetics Supplement Series*, *5*, e13–e15.
625 https://doi.org/10.1016/j.fsigss.2015.09.006

626 Marshall, C., Sturk-Andreaggi, K., Daniels-Higginbotham, J., Oliver, R. S., Barritt-Ross, S., &
627 McMahon, T. P. (2017). Performance evaluation of a mitogenome capture and Illumina
628 sequencing protocol using non-probative, case-type skeletal samples: Implications for the
629 use of a positive control in a next-generation sequencing procedure. *Forensic Science*
630 *International: Genetics*, *31*(May), 198–206. https://doi.org/10.1016/j.fsigen.2017.09.001
631 Mascher, M., Wu, S., St. Amand, P., Stein, N., & Poland, J. (2013). Application of Genotyping-
632 by-Sequencing on Semiconductor Sequencing Platforms: A Comparison of Genetic and
633 Reference-Based Marker Ordering in Barley. *PLoS ONE*, *8*(10), 1–11.
634 https://doi.org/10.1371/journal.pone.0076925
635 Mitchell, S. L., Goodloe, R., Brown-gentry, K., Pendergrass, S. A., Murdock, D. G., &
636 Crawford, D. C. (2015). Characterization of mitochondrial haplogroups in a large
637 population-based sample from the United States. *Human Genetics*, *133*(7), 861–868.
638 https://doi.org/10.1007/s00439-014-1421-9.Characterization
639 Nass, M. M. K., & Nass, S. (1963). Intramitochondrial fibers with DNA characteristics. I.
640 Fixation and electron staining reactions. *The Journal of Cell Biology*, *19*, 593–611.
641 Oikonomopoulos, S., Wang, Y. C., Djambazian, H., Badescu, D., & Ragoussis, J. (2016).
642 Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative
643 assessment of cDNA populations. *Scientific Reports*, *6*(August), 1–13.
644 https://doi.org/10.1038/srep31602
645 Ondov, B. D., Cochran, C., Landers, M., Meredith, G. D., Dudas, M., & Bergman, N. H. (2010).
646 An alignment algorithm for bisulfite sequencing using the Applied Biosystems SOLiD
647 System. *Bioinformatics*, *26*(15), 1901–1902. https://doi.org/10.1093/bioinformatics/btq291
648 Ossowski, A., Diepenbroek, M., Kupiec, T., Bykowska-Witowska, M., Zielińska, G.,

649 Dembińska, T., & Ciechanowicz, A. (2016). Genetic Identification of Communist Crimes'
650 Victims (1944–1956) Based on the Analysis of One of Many Mass Graves Discovered on
651 the Powazki Military Cemetery in Warsaw, Poland. *Journal of Forensic Sciences*, *61*(6),
652 1450–1455. https://doi.org/10.1111/1556-4029.13205
653 Ovchinnikov, I. V., Malek, M. J., Kjelland, K., & Drees, K. (2016). Whole Human
654 Mitochondrial DNA Sequencing (pp. 157–171). Humana Press, New York, NY.
655 https://doi.org/10.1007/978-1-4939-3597-0_13
656 Pakendorf, B., & Stoneking, M. (2005). Mitochondrial Dna and Human Evolution. *Annual*
657 *Review of Genomics and Human Genetics*, *6*(1), 165–183.
658 https://doi.org/10.1146/annurev.genom.6.080604.162249
659 Park, S., Cho, S., Seo, H. J., Lee, J. H., Kim, M. Y., & Lee, S. D. (2017). Entire mitochondrial
660 DNA sequencing on massively parallel sequencing for the Korean population. *Journal of*
661 *Korean Medical Science*, *32*(4), 587–592. https://doi.org/10.3346/jkms.2017.32.4.587
662 Parson, W., & Dür, A. (2007). EMPOP—A forensic mtDNA database. *Forensic Science*
663 *International: Genetics*, *1*(2), 88–92. https://doi.org/10.1016/j.fsigen.2007.01.018
664 Parson, W., Gusmão, L., Hares, D. R. R., Irwin, J. A. A., Mayr, W. R. R., Morling, N., Parsons,
665 T. J. J. (2014). DNA Commission of the International Society for Forensic Genetics:
666 Revised and extended guidelines for mitochondrial DNA typing. *Forensic Science*
667 *International: Genetics*, *13*, 134–142. https://doi.org/10.1016/j.fsigen.2014.07.010
668 Parson, W., Strobl, C., Huber, G., Zimmermann, B., Gomes, S. M., Souto, L., Lagace, R.
669 (2013). Forensic Science International : Genetics Evaluation of next generation mtGenome
670 sequencing using the Ion Torrent Personal Genome Machine (PGM), *7*, 543–549.
671 https://doi.org/10.1016/j.fsigen.2013.06.003

672 Paus, R. (1998). Principles of hair cycle control. *Journal of Dermatology*, *25*(12), 793–802.
673 https://doi.org/10.1111/j.1346-8138.1998.tb02507.x
674 Peck, M. A., Brandhagen, M. D., Marshall, C., Diegoli, T. M., Irwin, J. A., & Sturk-Andreaggi,
675 K. (2016). Concordance and reproducibility of a next generation mtGenome sequencing
676 method for high-quality samples using the Illumina MiSeq. *Forensic Science International:*
677 *Genetics*, *24*, 103–111. https://doi.org/10.1016/j.fsigen.2016.06.003
678 Piccinini, A., Coco, S., Parson, W., Cattaneo, C., Gaudio, D., Barbazza, R., & Galassi, A. (2010).
679 World war one Italian and Austrian soldier identification project: DNA results of the first
680 case. *Forensic Science International: Genetics*, *4*(5), 329–333.
681 https://doi.org/10.1016/j.fsigen.2009.10.005
682 Poletto, M. M., Malaghini, M., Silva, J. S., Bicalho, M. G., & Braun-Prado, K. (2019).
683 Mitochondrial DNA control region diversity in a population from Parana state—increasing
684 the Brazilian forensic database. *International Journal of Legal Medicine*, *133*(2), 347–351.
685 https://doi.org/10.1007/s00414-018-1886-5
686 Prinz, M., Carracedo, A., Mayr, W. R., Morling, N., Parsons, T. J., Sajantila, A., Schneider, P.
687 M. (2007). DNA Commission of the International Society for Forensic Genetics (ISFG):
688 Recommendations regarding the role of forensic genetics for disaster victim identification
689 (DVI). *Forensic Science International: Genetics*, *1*(1), 3–12.
690 https://doi.org/10.1016/j.fsigen.2006.10.003
691 Pukk, L., Ahmad, F., Hasan, S., Kisand, V., Gross, R., & Vasemägi, A. (2015). Less is more:
692 Extreme genome complexity reduction with ddRAD using Ion Torrent semiconductor
693 technology. *Molecular Ecology Resources*, *15*(5), 1145–1152. https://doi.org/10.1111/1755-
694 0998.12392

695 Ríos, L., García-Rubio, A., Martínez, B., Alonso, A., & Puente, J. (2010). Identification process
696 in mass graves from the Spanish Civil War II. *Forensic Science International*, *219*(1–3).
697 https://doi.org/10.1016/j.forsciint.2011.11.021
698 Röck, A. W., Dür, A., Van Oven, M., & Parson, W. (2013). Concept for estimating
699 mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic*
700 *Science International: Genetics*, *7*(6), 601–609. https://doi.org/10.1016/j.fsigen.2013.07.005
701 Rogers, G. E. (2004). Hair follicle differentiation and regulation. *International Journal of*
702 *Developmental Biology*, *48*(2–3), 163–170. https://doi.org/10.1387/ijdb.15272381
703 Rosa, A., Brehm, A., Kivisild, T., Metspalu, E., & Villems, R. (2004). MtDNA profile of West
704 Africa Guineans: Towards a better understanding of the Senegambia region. *Annals of*
705 *Human Genetics*, *68*(4), 340–352. https://doi.org/10.1046/j.1529-8817.2004.00100.x
706 Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Bustillo, J.
707 (2011). An integrated semiconductor device enabling non-optical genome sequencing.
708 *Nature*, *475*(7356), 348–352. https://doi.org/10.1038/nature10242
709 Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J. C., Brandon, M. C., Mishmar, D., Wallace,
710 D. C. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny.
711 *Nucleic Acids Research*, *35*(SUPPL. 1), 823–828. https://doi.org/10.1093/nar/gkl927
712 Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating
713 inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467.
714 https://doi.org/10.1073/pnas.74.12.5463
715 Scally, A. (2016). The mutation rate in human evolution and demographic inference. *Current*
716 *Opinion in Genetics and Development*, *41*, 36–43. https://doi.org/10.1016/j.gde.2016.07.008
717 Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M.,

718 Church, G. M. (2005). Molecular biology: Accurate multiplex polony sequencing of an
719 evolved bacterial genome. *Science*, *309*(5741), 1728–1732.
720 https://doi.org/10.1126/science.1117389
721 Shitara, H., Hayashi, J. I., Takahama, S., Kaneda, H., & Yonekawa, H. (1998). Maternal
722 inheritance of mouse mtDNA in interspecific hybrids: Segregation of the leaked paternal
723 mtDNA followed by the prevention of subsequent paternal leakage. *Genetics*, *148*(2), 851–
724 857. https://doi.org/10.1093/nar/17.18.7325
725 Shitara, H., Kaneda, H., Sato, A., Inoue, K., Ogura, A., Yonekawa, H., & Hayashi, J. I. (2000).
726 Selective and continuous elimination of mitochondria microinjected into mouse eggs from
727 spermatids, but not from liver cells, occurs throughout emborogenesis. *Genetics*, *156*(3),
728 1277–1284. https://doi.org/10.1073/pnas.1303231110
729 Shokolenko, I. N. (2014). Aging: A mitochondrial DNA perspective, critical analysis and an
730 update. *World Journal of Experimental Medicine*, *4*(4), 46.
731 https://doi.org/10.5493/wjem.v4.i4.46
732 Sigurðardóttir, S., Helgason, A., Gulcher, J. R., Stefansson, K., & Donnelly, P. (2000). The
733 Mutation Rate in the Human mtDNA Control Region. *American Journal of Human*
734 *Genetics*, *66*, 1599–1609. https://doi.org/Doi 10.1086/302902
735 Stewart, J. E., Fisher, C. L., Aagaard, P. J., Wilson, M. R., Isenberg, A. R., Polanskey, D.,
736 Budowle, B. (2001). Length variation in HV2 of the human mitochondrial DNA control
737 region. *Journal of Forensic Sciences*, *46*(4), 862–870. https://doi.org/10.1520/JFS15059J
738 Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L., & Erlich, H. A. (1991). Population
739 variation of human mtDNA control region sequences detected by enzymatic amplification
740 and sequence-specific oligonucleotide probes. *American Journal of Human Genetics*, *48*(2),

370–382.

Strobl, C., Eduardoff, M., Bus, M. M., Allen, M., & Parson, W. (2018). Evaluation of the precision ID whole MtDNA genome panel for forensic analyses. *Forensic Science International: Genetics*, *35*(March), 21–25. https://doi.org/10.1016/j.fsigen.2018.03.013

Sullivan, K. M., Hopgood, R., & Gill, P. (1992). Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *International Journal of Legal Medicine*, *105*(2), 83–86. https://doi.org/10.1007/BF02340829

Taanman, J.-W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1410*(2), 103–123. https://doi.org/10.1016/S0005-2728(98)00161-3

Templeton, J. E. L., Brotherton, P. M., Llamas, B., Soubrier, J., Haak, W., Cooper, A., & Austin, J. J. (2013). DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investigative Genetics*, *4*(1), 1–13. https://doi.org/10.1186/2041-2223-4-26

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Wallace, D. C. (1996). Classification of european mtDNAs from an analysis of three European populations. *Genetics*, *144*(4), 1835–1850.

Tully, G., Bär, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., Schneider, P. (2001). Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Science International*, *124*(1), 83–91. https://doi.org/10.1016/S0379-0738(01)00573-4

Turchi, C., Stanciu, F., Paselli, G., Buscemi, L., Parson, W., & Tagliabracci, A. (2016). The mitochondrial DNA makeup of Romanians: A forensic mtDNA control region database and

764 phylogenetic characterization. *Forensic Science International: Genetics*, *24*, 136–142.
765 https://doi.org/10.1016/j.fsigen.2016.06.013
766 van Oven, M. (2015). PhyloTree Build 17: Growing the human mitochondrial DNA tree.
767 *Forensic Science International: Genetics Supplement Series*, *5*, e392–e394.
768 https://doi.org/10.1016/j.fsigss.2015.09.155
769 van Oven, M., & Kayser, M. (2008). Updated comprehensive phylogenetic tree of global human
770 mitochondrial DNA variation. *Human Mutation*, *30*(September), 386–394.
771 https://doi.org/10.1002/humu.20921
772 Wallace, D. C. (2007). Why Do We Still Have a Maternally Inherited Mitochondrial DNA?
773 Insights from Evolutionary Medicine. *Annual Review of Biochemistry*, *76*(1), 781–821.
774 https://doi.org/10.1146/annurev.biochem.76.081205.150955
775 Wallace, H. M., Jackson, A. R., Gruber, J., & Thibedeau, A. D. (2014). Forensic DNA
776 databases-Ethical and legal standards: A global review. *Egyptian Journal of Forensic*
777 *Sciences*, *4*(3), 57–63. https://doi.org/10.1016/j.ejfs.2014.04.002
778 Wei, W., Keogh, M. J., Wilson, I., Coxhead, J., Ryan, S., Rollinson, S., Chinnery, P. F. (2017).
779 Mitochondrial DNA point mutations and relative copy number in 1363 disease and control
780 human brains. *Acta Neuropathologica Communications*, *5*(1), 13.
781 https://doi.org/10.1186/s40478-016-0404-6
782 Wilson, M. R., Polanskey, D., Replogle, J., DiZinno, J. A., & Budowle, B. (1997). A family
783 exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both
784 somatic mosaicism and pronounced segregation of mitotypes. *Human Genetics*, *100*(2),
785 167–171. https://doi.org/10.1007/s004390050485
786 Woerner, A. E., Ambers, A., Wendt, F. R., King, J. L., Moura-Neto, R. S., Silva, R., & Budowle,

787 B. (2018). Evaluation of the precision ID mtDNA whole genome panel on two massively
788 parallel sequencing systems. *Forensic Science International: Genetics*, *36*(March), 213–
789 224. https://doi.org/10.1016/j.fsigen.2018.07.015
790 Yasmin, M., Rakha, A., Noreen, S., & Salahuddin, Z. (2017). Mitochondrial control region
791 diversity in Sindhi ethnic group of Pakistan. *Legal Medicine*, *26*, 11–13.
792 https://doi.org/10.1016/j.legalmed.2017.02.001
793 Young, B., King, J. L., Budowle, B., & Armogida, L. (2017). A technique for setting analytical
794 thresholds in massively parallel sequencing-based forensic DNA analysis. *PLoS ONE*,
795 *12*(5), 1–15. https://doi.org/10.1371/journal.pone.0178005
796

Legends to figures and tables

Figure 1 – The human mitochondrial DNA genome with genes and control regions labeled. Adapted from Shokolenko et al., 2014 (Shokolenko, 2014)

Figure 2 - Representation of the mtDNA tree, Build 17, 18 Feb 2016. Available at http://www.phylotree.org/tree/index.htm (van Oven & Kayser, 2008)

Figure 3 - Representation of the geographical origin of mtDNA haplogroups and main mutations that are on the origin of each haplogroup. Adapted from Kivisild et al., 2015 (Kivisild, 2015).

Table 1 - Guidelines of the DNA Commission of the International Society of Forensic Genetics, 2014.

Table 2 - Selected published cases of human identification with mtDNA.

Figure 1

Genes and control regions labeled of the human mitochondrial DNA genome

The human mitochondrial DNA genome with genes and control regions labeled. Adapted from Shokolenko et al., 2014 (11)

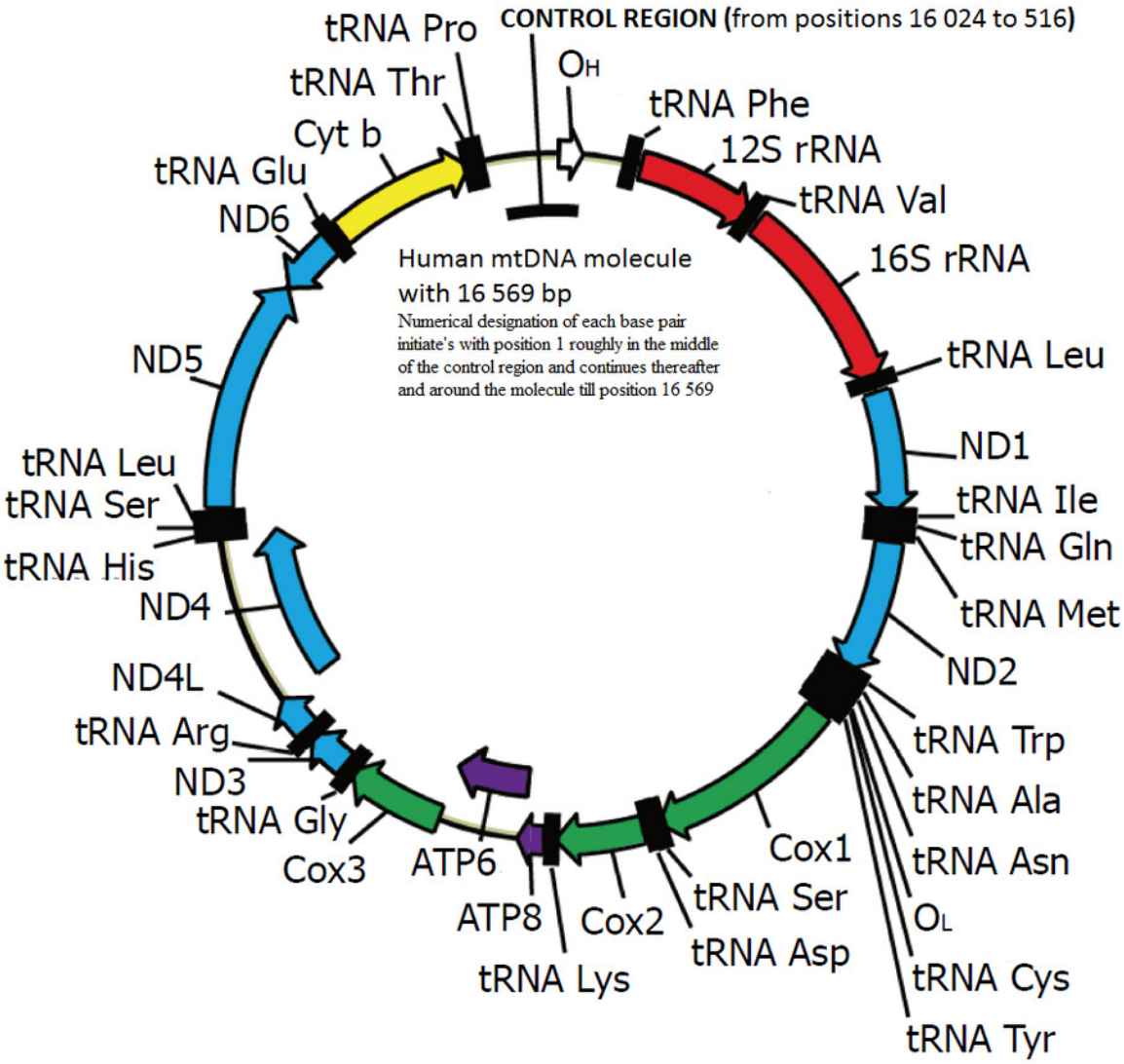


Figure 2

Phylogenetic tree of global human mitochondrial DNA variation

Phylogenetic tree of global human mitochondrial DNA variation. mtDNA tree Build 17 (18 Feb 2016). Available at http://www.phylotree.org/tree/index.htm . mtDNA-MRCA- Most recent common ancestor of mtDNA.

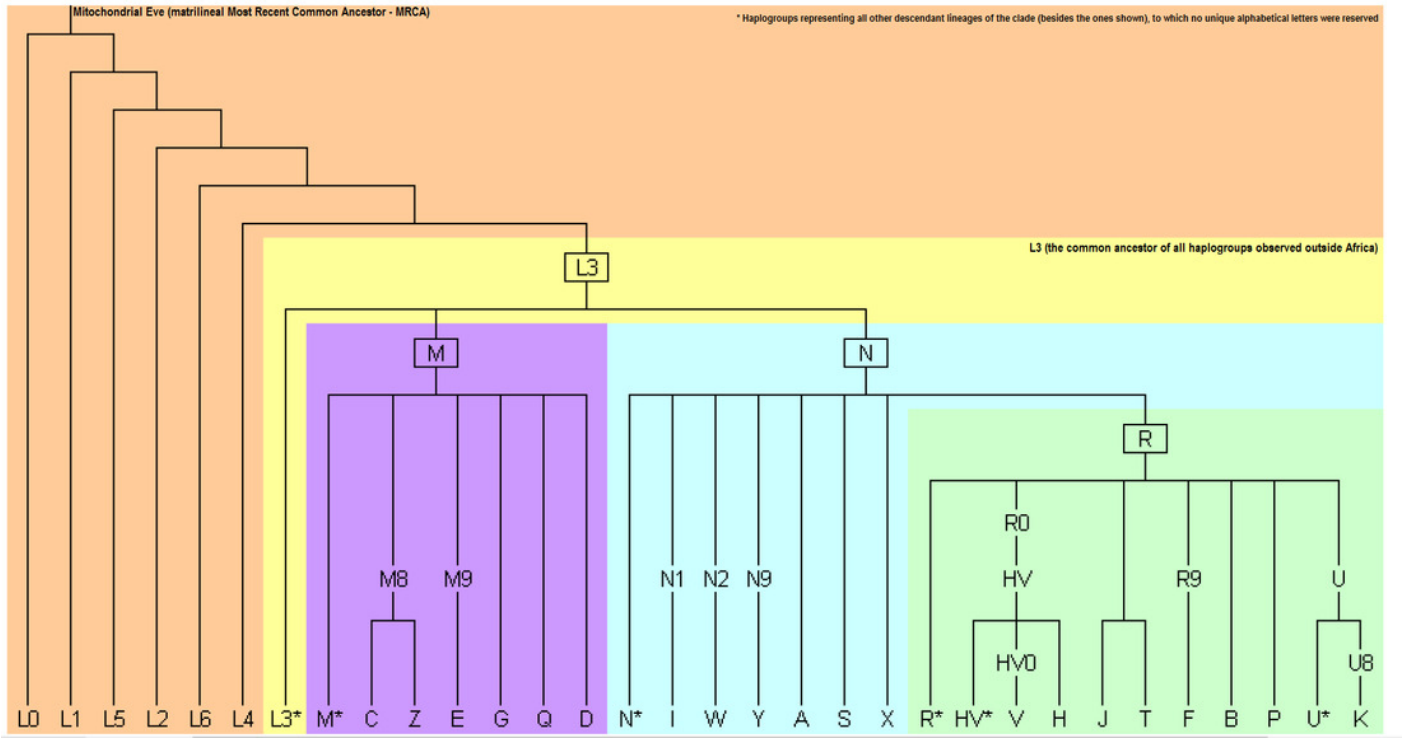


Figure 3

Geographical origin of mtDNA haplogroups and main mutations that are at the origin of each haplogroup

Representation of the geographical origin of mtDNA haplogroups and main mutations that are at the origin of each haplogroup. Adapted from Kivisild et al., 2015 (83) .

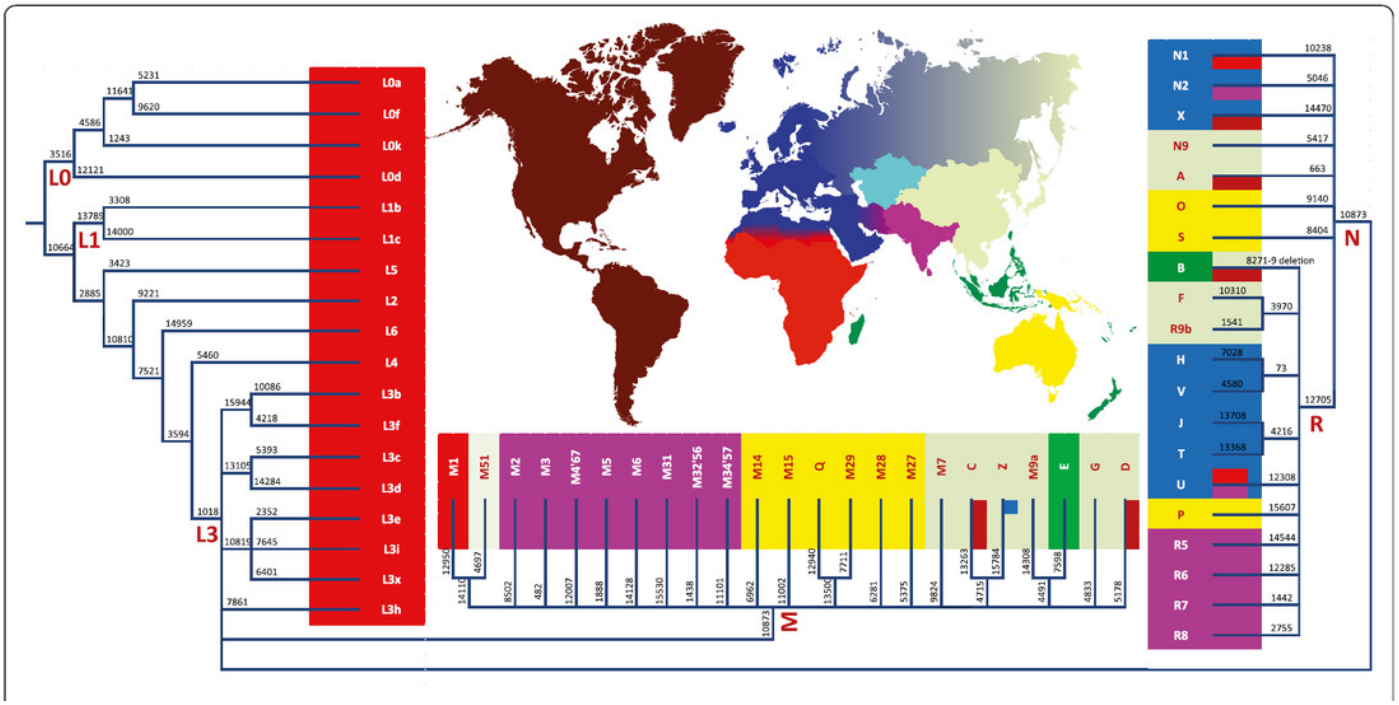


Table 1(on next page)

Guidelines of the DNA Commission of the InternationalSociety of Forensic Genetics

Guidelines of the DNA Commission of the International Society of Forensic Genetics, 2014.

1 **Table 1** - Guidelines of the DNA Commission of the International Society of Forensic Genetics, 2014

Addressement	Recommendation	Statement
General recommendations/ good laboratory practice	Recommendation #1	Good laboratory practice and specific protocols for work with mtDNA must be followed in accordance with previous guidelines
	Recommendation #2	Negative and positive controls as well as extraction reagent blanks must be carried through the entire laboratory process
	Recommendation #3	Reported consensus sequences must be based on redundant sequence information, using forward and reverse sequencing reactions whenever practical
	Recommendation #4	Manual transcription of data should be avoided and independent confirmation of consensus haplotypes by two scientists must be performed
	Recommendation #5	Laboratories using mtDNA typing in forensic casework shall participate regularly in suitable proficiency testing programs
Targeted region, amplification and sequencing ranges	Recommendation #6	In population genetic studies for forensic databasing purposes, the entire mitochondrial DNA control region should be sequenced.
Reference sequence	Recommendation #7	MtDNA sequences should be aligned and reported relative to the revised Cambridge Reference Sequence (rCRS, NC001807), and should include the interpretation range (excluding primer sequence information)
Alignment and notation	Recommendation #8	IUPAC conventions using capital letters shall be used to describe differences to the rCRS and (point heteroplasmic) mixtures. Lower case letters should be used to indicate mixtures between deleted and non-deleted (inserted and non-inserted) bases. N-designations should only be used when all four bases are observed at a single position (or if no base call can be made at a given position). For the representation of deletions, ‘‘DEL’’, ‘‘del’’ or ‘‘S’’ shall be used
	Recommendation #9	The alignment and notation of mtDNA sequences should be performed in agreement with the mitochondrial phylogeny (established patterns of mutations). Tools to assist with the notation of mtDNA sequences are available at http:// empop.org/
Heteroplasmy	Recommendation #10	In forensic casework, laboratories must establish their own interpretation and reporting guidelines for observed length and point heteroplasmy. The evaluation of heteroplasmy depends on the limitations of the technology and the quality of the sequencing reactions as well as the experience of the laboratory. Differences in both PHP and LHP do not constitute evidence for excluding two otherwise identical haplotypes as deriving from the same source or same maternal lineage
	Recommendation #11	For population database samples, length heteroplasmy in homopolymeric sequence stretches should be interpreted by calling the dominant variant, which can be determined by identifying the position with the highest representation of a non-repetitive peak downstream of the affected stretch
Haplogrouping of mtDNA sequences	Recommendation #12	MtDNA population data should be subjected to analytical software tools that facilitate phylogenetic checks for data quality control. A comprehensive suite of QC tools is provided by EMPOP
Databases and database searches	Recommendation #13	The entire database of available sequences should be searched with respect to the sequencing (interpretation) range to avoid biased query results
	Recommendation #14	Laboratories must be able to justify the choice of database(s) and statistical approach used in reporting
	Recommendation #15	Laboratories must establish statistical guidelines for use in reporting an mtDNA match between two samples
	Recommendation #16	Highly variable positions such as length variants in homopolymeric stretches should be disregarded from searches for determining frequency estimates. Heteroplasmic calls should be queried in a manner that

does not exclude any of the heteroplasmic variants

Table 2(on next page)

Selectedpublished cases of human identification with mtDNA.

Selected published cases of human identification with mtDNA.

1 **Table 2** - Selected published cases of human identification with mtDNA

Reference/Year	Studied samples	mtDNA studied regions	Used methodologies	Reference samples	Results
Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet. **1991**;48(2):370–82.	Skeletal remains of a human child, found in 1986	HVI, HVII	PCR for amplification Hybridization with oligonucleotide probes for sequence determination	Parents of a 3-year-old child disappeared from home in 1984	Identical mtDNA sequence in skeletal remains and sample of the 3-year-old child mother Positive ID
Sullivan KM, Hopgood R, Gill P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Legal Med. **1992**;105(2):83–6.	Body of a female, in an advanced state of decomposition discovered in 1990	HVI, HVII	PCR for amplification Sanger sequencing	Blood sample from a sister of a deceased female at the same region	No differences were observed between the corpse and blood from the putative sister Positive ID
Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, et al. Identification of the remains of the Romanov family by DNA analysis. Nat Genet. **1994**;6(2):130–5.	Nine skeletons found in a grave in Ekaterinburg, Russia, 1991	HVI, HVII	PCR for amplification Sanger sequencing	Blood sample from Gt. Gt. Grandson of Louise of Hesse-Cassel and from Gt. Gt. Gt. Granddaughter of Louise of Hesse-Cassel	Exact sequence between putative Tsarina Alexandra and putative three children. Exact mtDNA results between putative Tsar Nicholas II and two living maternal relatives of the Tsar
Ivanov PL, Wadhams MJ, Roby RK, Holland MM WV& PT. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat Genet. **1996**;(12):417–20.	Skeleton of putative Tsar Nicholas II	HVI, HVII	PCR for amplification Sanger sequencing	Skeleton of Grand Duke of Russia Georgij Romanov (Tsar's brother) Blood sample from Countess Xenia Cheremeteff-Sfiri (maternal Tsar's relative)	Establishment of the authenticity of the remains of Tsar Nicholas II
Deng YJ, Li YZ, Yu XG, Li L, Wu DY, Zhou J, et al. Preliminary DNA identification for the tsunami victims in Thailand. Genomics, Proteomics Bioinforma. **2005**;3(3):143–57.	258 tooth samples from killed people at the 2004 Southeast Asia Thailand	HVI, HVII	PCR for amplification Sanger sequencing	200 relatives of the tsunami victims	200 tsunami victims have been identified, including both Thai nationals and foreign tourists from several nations

	Tsunami				
Ríos L, García-Rubio A, Martínez B, Alonso A, Puente J. Identification process in mass graves from the Spanish Civil War II. Forensic Sci Int. **2010**;219(1–3).	Skeletal remains exhumed from a mass grave from the Spanish Civil War (1936–1939)	HVI, HVII	PCR for amplification Sanger sequencing	Sister of the youngest person presumptively known to be buried in the grave	Match between mtDNA profiles of the biologically youngest skeleton and the sister of the youngest person presumptively known to be buried in the grave
Piccinini A, Coco S, Parson W, Cattaneo C, Gaudio D, Barbazza R, et al. World war one Italian and Austrian soldier identification project: DNA results of the first case. Forensic Sci Int Genet. **2010**;4(5):329–33.	Remains of missing soldiers occasionally found during excavations	HVI, HVII	PCR for amplification Sanger sequencing	Offspring of the italian soldier Libero Zugni Tauro	Both mtDNA and Y-STR data showed clear exclusion scenarios between the human remains and the reference samples
King TE, Fortes GG, Balaresque P, Thomas MG, Balding D, Delser PM, et al. Identification of the remains of King Richard III. Nat Commun. **2014**;5:1–8.	Skeleton excavated at the presumed site of the Grey Friars friary in Leicester, 2012	Whole mitochondrial genome	PCR for amplification Massive parallel sequencing	Saliva samples of the modern relatives of Richard III	Positive mtDNA match between the only known female-line of Richard III and studied modern relatives of Richard III
Ossowski A, Diepenbroek M, Kupiec T, Bykowska-Witowska M, Zielińska G, Dembińska T, et al. Genetic Identification of Communist Crimes' Victims (1944–1956) Based on the Analysis of One of Many Mass Graves Discovered on the Powazki Military Cemetery in Warsaw, Poland. J Forensic Sci. **2016**;61(6):1450–5.	Remains of eight people buried in one of many mass graves, which were found at the cemetery Powazzki Military in Warsaw, Poland	HVI, HVII	PCR for amplification Sanger sequencing	Reference material was collected from the closest living relatives of Communist Crimes' Victims (1944–1956)	Positive mtDNA match between 6 putative victims and 6 living relatives

Capítulo III - Material e Métodos

1 - Amostra Estudada

Estudamos 439 imigrantes com residência oficial na região de Lisboa com naturalidade e nacionalidade em PALOP e com ascendência declarada, também, como natural e nacional de PALOP. O número total de imigrantes pretendeu levar em conta, para cada subgrupo ou país de origem, um número de imigrantes proporcionalmente aproximado à percentagem de imigrantes que cada um desses subgrupos têm, atualmente, a residir em Portugal. No entanto, estávamos limitados ao número máximo de indivíduos que nos foi possível integrar no estudo e que correspondia ao número máximo de imigrantes presentes ao INMLCF por ordem dos Tribunais para perícias de investigação de parentesco. Assim, estudamos 103 imigrantes oriundos de Cabo Verde (percentagem correspondente a cerca de 23% da amostra estudada. Não nos foi possível estudar um número de imigrantes correspondente aos cerca de 51% de imigrantes de Cabo Verde residentes em Portugal), 173 imigrantes oriundos de Angola (percentagem correspondente a cerca de 39% da amostra estudada. Foi-nos possível estudar um número superior aos cerca de 24% de imigrantes de Angola residentes em Portugal), 80 imigrantes oriundos da Guiné-Bissau (percentagem correspondente a cerca de 18% da amostra estudada, valor muito aproximado dos cerca de 22% de imigrantes da Guiné-Bissau residentes em Portugal), e 83 imigrantes oriundos de Moçambique (percentagem correspondente a cerca de 19% da amostra estudada e a mais que os cerca de 3% de imigrantes de Moçambique residentes em Portugal). Na figura 2 apresentamos um mapa mundi no qual se encontra assinalada, com cores, a localização de Portugal e dos países

de origem dos imigrantes incluídos no estudo - Cabo Verde, Guiné-Bissau, Angola e Moçambique -.

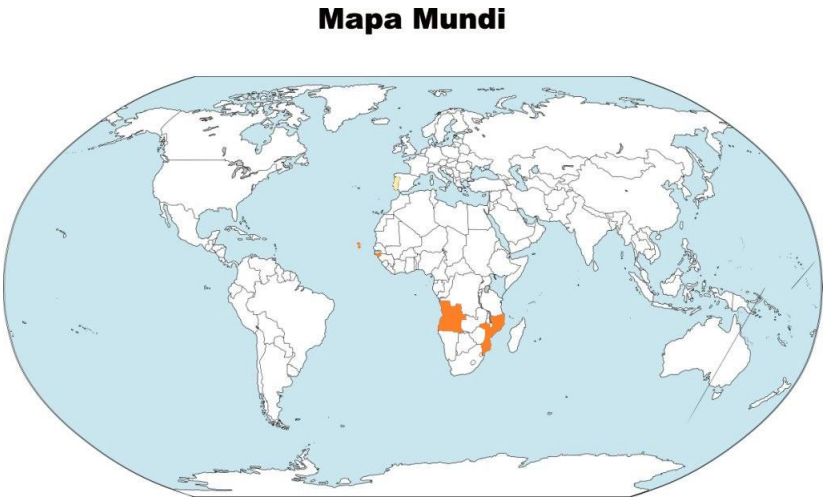


Figura 2 - Mapa mundi no qual se encontra assinalado, a cor amarela, a localização de Portugal e a cor laranja, da esquerda para a direita, a localização das ilhas de Cabo Verde, Guiné-Bissau, Angola e Moçambique.

No estudo foram utilizadas amostras de sangue colhidas no âmbito de perícias de investigação de parentesco biológico a decorrer no INMLCF.

2 - Métodos laboratoriais

2.1 - Extração de DNA

Para extração de DNA das amostras de sangue colhidas foi utilizada uma resina quelante - Chelex® 100 (BioRad) a 5% -, de acordo com a técnica desenvolvida por Walsh e colaboradores em 1991 (Walsh, Metzger, & Higuchi, 2013).

Numa fase prévia à extração foi necessária a preparação da solução de Chelex® 100 a 5% através da diluição 2,5 g da resina em 50 mL de água desionizada. Esta preparação pode ser armazenada e utilizada em extrações posteriores, sendo que o seu acondicionamento nunca ultrapassassou uma semana.

O processo iniciou-se com o corte de um círculo da mancha de sangue armazenada e conservada em papel Whatman®. O círculo de papel, efetuado com um paper-punch de diâmetro 3mm (de forma a garantir a uniformização do tamanho das amostras extraídas), foi transferido para um tubo de microcentrífuga de 1,5mL devidamente identificado com um código alfa numérico atribuído à amostra.

Adicionamos 1mL de água desionizada. Seguiu-se um período de incubação de 15minutos à temperatura ambiente para que ocorresse a transferência do conteúdo do suporte para o solvente. Após uma ligeira agitação e centrifugação da mistura à velocidade de 14 000 rpm, durante 3 minutos, o material genético ficou concentrado no fundo do tubo de microcentrífuga, de onde se retiraram 970µL de sobrenadante. Ao volume restante foram acrescentados 180 µL da solução previamente preparada com a resina, após a qual ocorreu um novo período de incubação a 56 °C durante 15 minutos, garantindo a aderência dos resíduos ao Chelex® 100. Foi efetuada uma nova agitação. De seguida, a amostra

foi exposts à temperatura de 100 ºC durante 8minutos. No último passo, uma centrifugação de 5minutos a 14 000 rpm agregou os resíduos celulares na parte inferior do tubo, tendo sido o sobrenadante utilizado na fase seguinte de amplificação. Em cada fase de extração foi efetuado um controlo negativo para o qual foi realizado todo o procedimento experimental anteriormente descrito, à exceção da adição da mancha de sangue.

2.2 - Amplificação de DNA

Para a amplificação da região controlo total do DNAmt, constituída pelas regiões HVI, HVII e HVIII, utilizaram-se dois pares de primers. O par L15971 (direto) e H016 (reverso) amplificou a região entre as posições 16024 e a 16569 enquanto que o par L16555 (direto) e H639 (reverso) amplificou a região entre as posições 1 a 576. A nomenclatura dos primers é definida de acordo com a sua afinidade para a cadeia leve (L) ou pesada (H) bem como pela posição nucleotídica na região 3' com a qual interagem. A sequência específica de cada primer encontra-se representada na tabela 2.

Tabela 2- Sequências dos primers utilizados na amplificação da região controlo total do DNAmt.

L15971	5' - TTA ACT CCA CCA TTA GCA CC - 3'
H016	5' - TGA TAG ACC TGT GAT CCA TCG TGA - 3'
L16555	5' - CCC ACA CGT TCC CCT TAA AT - 3'
H639	5' - GGG TGA TGT GAG CCC GTC TA - 3'

Para a realização desta técnica foram utilizados vários reagentes: Multiplex PCR Master Mix® (Quiagen), que contém tampão de reação com Cloreto de Magnésio

($MgCl_2$), HotStarTaq Plus DNA polymerase e desoxinucleotídeos trifosfatados (dNTPs), o conjunto de primers anteriormente mencionados e o DNA extraído de cada uma das amostras, num volume total de reação de 10 µL. Os volumes necessários de cada reagente encontram-se descritos na tabela 3.

Tabela 3 - Reagentes e respetivos volumes utilizados, por amostra, na preparação da reação de amplificação da região controlo total do DNA mitocondrial.

Reagentes	Volume
Multiplex PCR Master Mix	5 µL
Água mili-Q	3 µL
Primers a 2 µM	1 µL
DNA	1 µL
Total	10 µL

Para cada reação de amplificação prepararam-se dois controlos negativos, um para cada par de primers, aos quais se adicionou água mili-Q em detrimento de DNA. A reação de amplificação da região controlo total foi efetuada num termociclador GeneAmp PCR system 9700 (Applied Biosystems - AB) programado de acordo com a tabela 4, num total de 35 ciclos.

Tabela 4 - Condições de temperaturas e tempos requeridos para a amplificação da região controlo total do DNA mitocondrial.

1 ciclo	35 ciclos			1 ciclo	
Incubação Inicial	**Desnaturação**	***Annealing***	**Extensão**	**Extensão Final**	***Hold***
94°C	94°C	60°C	72°C	72°C	4°C
5 min	30 seg	90 seg	60 seg	10 min	∞

2.3 - Purificação dos Produtos Amplificados

Este passo de purificação foi realizado com recurso ao método do ExoSAP-IT® (Affymetrix) e teve como objetivo a remoção de elementos em excesso provenientes da reação

de amplificação. Este reagente contém duas enzimas hidrolíticas – exonuclease I e fosfatase alcalina recombinante de camarão (rSAP) - que são responsáveis pela remoção, respetivamente, de primers residuais e de dNTPs não incorporados no decorrer do processo de amplificação.

A técnica de purificação consistiu na adição de 4 µL de ExoSAP-TI® por cada 10 µL de produto amplificado, num volume final de 14 µL. Esta etapa foi programada e realizada num termociclador GeneAmp PCR system 9700 (AB). Inicialmente incubaram-se as amostras a 37°C, temperatura definida como ótima para as enzimas que constituem o reagente, durante 15 minutos. Seguiu-se um período de aquecimento a 80°C, também durante 15 minutos, com o intuito de inativar as enzimas do reagente.

2.4 - Sequenciação

A sequenciação da região controlo do DNAmt tem por objetivo determinar a base nucleotídica caraterística de cada posição dentro desta região. O método utilizado nesta etapa foi o método de sequenciação de Sanger (Sanger, Nicklen, & Coulson, 1977), que consiste na incorporação de didesoxinucleotídeos trifosfatados (ddNTPs) na extremidade 3' do nucleótido anterior, levando ao término da síntese da cadeia na posição onde ocorre a ligação dos ddNTPs. Estes ddNTPs apresentam na sua constituição marcadores fluorescentes, com diferentes cores, que permitem, durante a análise das sequências, a sua deteção e identificação. O ddATP encontra-se marcado a verde, o ddCTP a azul, o ddGTP marcado a amarelo, mas visualizado a preto, e o ddTTP marcado a vermelho.

Nesta fase do processo foi necessária a preparação de quatro reações, uma por cada primer (L15971, L16555, H016 e H639). Para além dos primers da cadeia a amplificar, a mistura era constituída por tampão Better Buffer® (Microzone Ltd.), BigDye® Terminator v.3.1 Cycle Sequence (AB), uma solução que contém DNA polimerase, $MgCl_2$, dNTPs e ddNTPs marcados com Dye Terminator, num volume total de 4 µL. As soluções foram distribuídas pelos poços de uma placa de 96 poços, aos quais se adicionou 1 µL de DNA das amostras previamente amplificadas e purificadas. Os volumes necessários de cada reagente encontram-se descritos na tabela 5.

Tabela 5 - Reagentes e respetivos volumes utilizados, por amostra, na preparação da reação de sequenciação da região controlo total do DNA mitocondrial.

Reagentes	Volume
Better Buffer	2 µL
BigDye Terminator	1 µL
Primers a 2.5 µM	1 µL
DNA	1 µL
Total	5 µL

Prepararam-se simultaneamente quatro controlos negativos, um por cada primer, aos quais se adicionou água mili-Q em detrimento de DNA amplificado e purificado. O processo de sequenciação da região controlo total do DNAmt foi realizado num termociclador GeneAmp PCR system 9700 (AB) programado para as temperaturas e os tempos descritos na tabela 6, num total de 35 ciclos.

2.5 - Purificação dos Produtos Sequenciados

A etapa de purificação foi realizada com a ajuda do kit BigDye® XTerminator Purification Kit (AB), que possui dois

reagentes na sua composição, o XTerminator Solution e a SAM TM Solution. Este passo é importante na remoção do excesso de dNTPs e ddNTPs não incorporados durante a reação de sequenciação, bem como na remoção do excesso de primers e sais livres.

Tabela 6 - Condições de temperaturas e tempos requeridos para a sequenciação da região controlo total do DNA mitocondrial.

1 ciclo	35 ciclos			1 ciclo	
Incubação Inicial	**Desnaturação**	***Annealing***	**Extensão**	**Extensão Final**	***Hold***
96°C	96°C	50°C	60°C	60°C	4°C
2/5 min	15 seg	9 seg	2 min	10 min	∞

Neste processo preparou-se uma solução de reação constituída por 30 µL de água, 16,4 µL de SAM Solution e 3,7 µL de XTerminator Solution, num volume final de 50,1 µL por amostra. Distribuíram-se 50 µL da mistura por cada poço utilizado da placa de 96 poços. Colocou-se a placa num agitador, durante aproximadamente 30 minutos. Por fim, efetuou-se uma centrifugação a 2800 rpm, durante 2 minutos, passo que permitiu a deposição dos excessos da reação de sequenciação, impedindo a sua interferência na eletroforese. Desta forma, durante a eletroforese capilar, os capilares aspiram apenas os produtos sequenciados retidos no sobrenadante.

2.6 - Deteção dos Produtos sequenciados

Após purificação dos produtos sequenciados, as amostras foram colocadas num sequenciador automático, o Genetic Analyser 3130 (AB) de 4 capilares, para serem submetidas a uma eletroforese capilar. A realização desta eletroforese possibilitou a deteção dos fragmentos obtidos e a aquisição da sequência nucleotídica da região controlo do DNAmt em

estudo. Durante este processo, os fragmentos migram consoante o seu peso molecular, desde o polo negativo (cátodo) até ao polo positivo (ânodo) dos capilares. Isto significa que fragmentos de tamanho menor irão migrar a uma velocidade maior do que fragmentos de tamanho maior. Aquando da migração, os fragmentos vão ser atravessados por um feixe de laser que, ao excitar os eletrões presentes nos fluorocrómos dos ddNTPs, promovem a emissão de fluorescência, facilitando assim a deteção das diferentes bases nucleotídicas. Os resultados obtidos de cada amostra por cada primer organizam-se num eletroferograma (figura 3), que constitui, através de picos, o conjunto de todos os nucleótidos que compõem a região controlo.

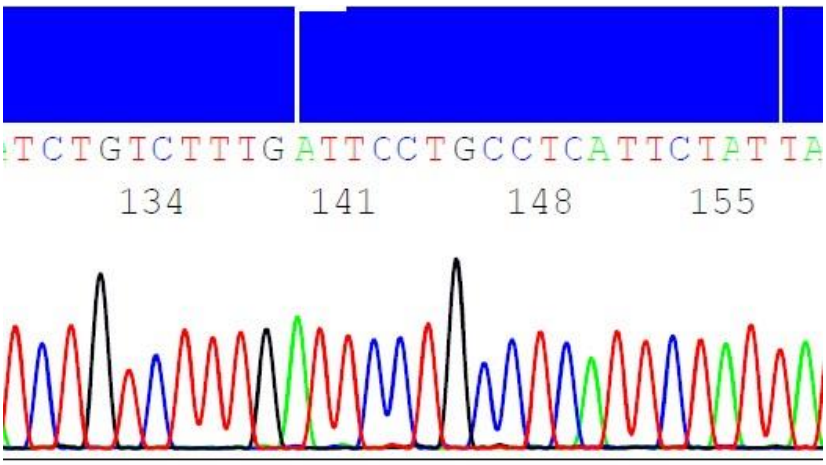


Figura 3 - Exemplo de um segmento de um eletroferograma que obtivemos com o programa Sequencing Analysis v.5.2, após sequenciação com o primer L16555.

Previamente ao início da corrida e com auxílio do programa Genetic Analyser Data Collection Software v.4.0, as amostras e os controlos presentes em cada um dos poços da placa de 96 poços foram identificados com a codificação da amostra e o primer correspondente.

2.7 - Análise dos produtos Sequenciados

Para cada amostra, a região controlo total do DNAmt, compreendida entre as posições 16024 a 16569 e 1 a 576 foi sequenciada. Ambos os intervalos mencionados foram sequenciados utilizando um par de primers (um direto e outro reverso), em que um tem afinidade para a cadeia leve e o outro para a cadeia pesada. O objetivo deste procedimento é comparar os segmentos sequenciados por cada par de primers, de forma a verificar se há concordância, em cada posição nucleotídica, entre a cadeia sequenciada pelo primer direto e reverso.

A análise dos resultados da sequenciação da região controlo do DNAmt foi efetuada com auxílio a dois programas, o Sequencing Analysis v.5.2, que avalia a qualidade das sequências por identificação dos nucleótidos em cada posição, e o SeqScape v.3 (AB), que permitiu o alinhamento e a posterior comparação das sequências obtidas com a rCRS. Este último programa identifica apenas as diferenças observadas entre a sequência em estudo e a sequência padrão, discriminando, assim, para cada amostra, todas as bases nucleotídicas e respetivas posições que se encontrem alteradas. Para cada indivíduo, o conjunto dos polimorfismos observados relativamente à rCRS dá origem ao haplótipo. A caraterização das sequências teve por base as considerações da European DNA Profiling Group (EDNAP) e as regras de nomenclatura da International Union of Pure & Applied Chemistry (IUPAC) (W. Parson et al., 2014): A (adenina), C (citosina), G (guanina), T (timina), R (adenina e guanina) e Y (citosina e timina).

A região controlo total do DNAmt é uma região hipervariável que pode acumular alterações nucleotídas. As alterações por substituição podem ser classificadas em transições, que

podem ocorrer entre purinas (A e G) ou entre pirimidinas (C e T), e transversões, que ocorrem entre uma purina e uma pirimidina ou vice-versa. Para detetar substituições é necessário comparar as sequências em estudo com a rCRS, após a qual se identifica a posição em que ocorreu a alteração e o nucleótido que originou essa substituição. Outro tipo de alterações que podem ser encontradas durante a análise do DNAmt são as inserções. Estas são identificadas pela posição nucleotídica imediatamente anterior à alteração, seguida de um ponto e do número um (.1) e do nucleótido inserido, para a primeira inserção, de um ponto e do número dois (.2) e do nucleótido inserido, na segunda inserção, e assim sucessivamente, consoante o número de inserções. As deleções são outro tipo de alterações que podem ser determinados na análise do DNAmt. Estas são representadas pela posição do nucleótido deletado relativamente à rCRS, seguida do sufixo "d", "DEL" ou "-".

Nas amostras que apresentaram heteroplasmia de comprimento, a leitura da sequência só foi possível até à posição heteroplásmica, uma vez que, a partir desta região, o eletroferograma era uma mistura de bases nucleotídica sobrepostas e ilegíveis. Para estes casos, com o intuito de confirmar a sequência obtida, efetuou-se a repetição da sequenciação do segmento heteroplásmico. Outra caraterística que se observou em algumas das sequências de DNAmt analisadas foi a presença de heteroplasmia de posição, que consiste na sobreposição de nucleótidos numa determinada posição da sequência. A sua identificação segue as diretrizes descritas no código da IUPAC e representam-se pela posição heteroplásmica e a letra correspondente.

Para se proceder à classificação dos haplótipos em haplogrupos recorreu-se a uma árvore filogenética de DNAmt, a Phylotree, desenhada em 2008, por van Oven e Kayser (van Oven & Kayser, 2008). Esta árvore exibe os polimorfismos existentes relativamente à rCRS (figura 4).

2.8 - Análise filogenética

Para determinar as relações filogenéticas entre a população em estudo e as selecionadas da literatura, efetuou-se a análise da variância molecular (AMOVA). Este parâmetro permite calcular, através do estudo dos haplótipos, a variabilidade genética entre diferentes populações. Para realizar a AMOVA e estimar as distâncias genéticas interpopulacionais, utilizou-se o programa Arlequin v. 3.5.2.2 (Excoffier & Lischer, 2010). Através deste programa foram obtidos os valores do índice de fixação (Fst), que avalia as distâncias genéticas entre as populações, e o valor de p, que indica o nível de significância (0.05).

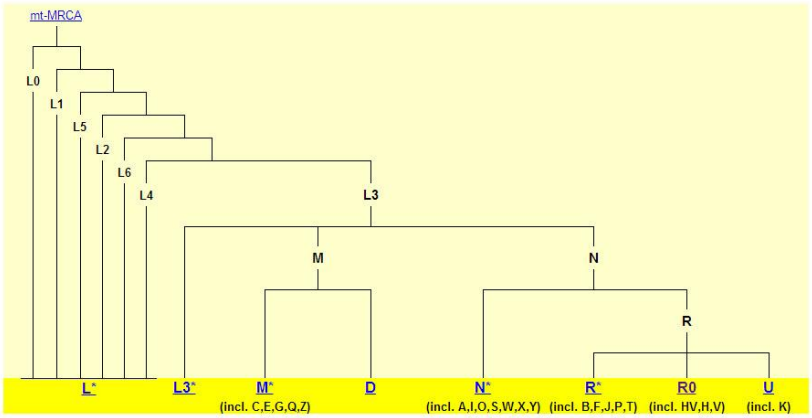


Figura 4 - Representação da árvore filogenética de DNAmt (van Oven & Kayser 2009)

Tendo por base os valores de Fst determinados e recorrendo à ferramenta Neighbor do software Phylip v.3.695 foi

possível obter as distâncias moleculares entre as sequências de DNAmt das populações analisadas. A sua representação gráfica, em forma de árvore filogenética, foi elaborada através do programa Treeview v. 1.6.6 (Page, 1996).

Após a determinação das sequências e haplótipos de DNA mitocondrial, as sequências foram submetidas à base de dados EMPOP - EDNAP Forensic mtDNA Population Database -, a mais importante base de dados de DNA mitocondrial a nível internacional para fins forenses e populacionais, de forma a disponibilizar os resultados obtidos a toda a comunidade científica (Walther Parson & Dür, 2007).

Capítulo IV - Estudo de imigrantes de Cabo Verde:

Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis

Morais P, Amorim A, Vieira da Silva C, Ribeiro T, Costa Santos J, Afonso Costa H. ***Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis.*** Journal of Genetics 94 (2015), 3:509-512.

RESEARCH NOTE

Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis

PAULO MORAIS[1,2], ANTÓNIO AMORIM[1,3,4], CLÁUDIA VIEIRA DA SILVA[1], TERESA RIBEIRO[1], JORGE COSTA SANTOS[1,5] and HELOÍSA AFONSO COSTA[1]*

[1]*Instituto Nacional de Medicina Legal e Ciências Forenses—Delegação do Sul, 1169-201, Lisboa, Portugal*
[2]*Escola Superior de Saúde Egas Moniz, and* [4]*Instituto Superior de Ciências da Saúde Egas Moniz, 2829-511, Almada, Portugal*
[3]*Faculdade de Ciências da Universidade de Lisboa, 1749-016, Lisboa, Portugal*
[5]*Faculdade de Medicina da Universidade de Lisboa, 1600-190, Lisboa, Portugal*

[Morais P., Amorim A., Vieira da Silva C., Ribeiro T., Costa Santos J. and Afonso Costa H. 2015 Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis. *J. Genet*. **94**, 509–512]

Introduction

Cabo Verde is a group of 10 volcanic islands and several uninhabited islets located on the west African coast and belongs to a group of four archipelagos located in the Atlantic Ocean (Açores, Madeira, Islas Canarias and Cabo Verde), named Macaronesia. The Portuguese colonization began soon after the discovery of the archipelago in 1460 with Santiago and Fogo being the first islands to be populated. The first settlers arrived in 1462 on the island of Santiago and were an assortment of Portuguese nobles, Jews, exiles and convicts (Willie 2001). Cabo Verde remained a colony of the Portuguese Colonial Empire until 1975 when the independence was proclaimed and the country became formally an independent nation. In the 19th century, drought and famine promoted strong migration movements between the isles of the archipelago and other regions. Migration is present in the historic and social reality of Cabo Verde archipelago since the establishment of its society.

According to the 2008 survey released by the Cabo Verde National Institute of Statistics, the country had about 500,000 inhabitants. Conjointly with Europe, the immigrant population of Portugal and particularly Lisboa, is clearly increasing. This migration contributes not only increase in the number of inhabitants, but also to increase the social, cultural, religious, linguistic, anthropological and genetic heterogeneity.According to the Portuguese Foreign Affair Services and the Portuguese National Institute of Statistics, immigrants from Cabo Verde with fixed residence in Portugal increased from 28,796 in 1990 to 43,510 individuals in 2010, and 34,234 live in Lisboa or nearby villages.

Mitochondrial DNA (mtDNA) studies allow us to estimate as which populations had matrilineal impact in the formation of the actual Cabo Verde archipelago population. While studying genetic markers on the Y chromosome of 201 individuals born in Cabo Verde, Gonçalves and coworkers confirmed in 2003 the paternal influence from Europe and Middle East on the origin of the population of Cabo Verde. Until now, mtDNA studies in the population of Cabo Verde have been restricted to the mtDNA sequencing of the HV1 region (Tavares 2007) or to the combination between sequencing of the HV1 region and the analysis of restriction fragment length polymorphism (RFLP) sites in the coding region (Brehm *et al.* 2002). The introduction of the analysis of the entire mtDNA control region for the first time when studying native individuals from Cabo Verde will increase the discriminatory power between samples and overall efficiency in determination of haplogroups.

The aims of this study were (i) to enrich mtDNA global database, (ii) obtainment of the mtDNA variability of the Cabo Verde population living in Lisboa to complement previous studies by our group using STR genetic markers (Amorim *et al*. 2012; Afonso Costa *et al*. 2014), (iii) assign haplotypes to designated haplogroups, (iv) infer whether there are genetic proximity between the studied population and previous

*For correspondence. E-mail: hcosta@dlinml.mj.pt.
Antonio Amorim, Claudia Vieira da Silva, Teresa Ribeiro, Jorge Costa Santos, and Heloisa Afonso Costa are the authors belong to CENCIFOR research group.

Keywords. Cabo Verde; Lisboa; mitochondrial DNA; population genetics.

studies according to the mtDNA profile of the Cabo Verde population, and (v) compare the studied population with other African populations, with the aim to bring more light to our understanding on the subject of the impact of migrations involving Cabo Verde archipelago's origin.

Materials and methods

This study was carried with a sample of 103 native individuals of Cabo Verde currently living in Lisboa, age ranging from 15 to 70 years. All individuals were from the biological kinship investigations cohort conducted at the Instituto Nacional de Medicina Legal e Ciências Forenses—Delegação do Sul (INMLCF-DS). According to Portuguese law that regulates INMLCF activity, samples from routine forensic cases can be used for investigation purpose including genetic population studies. All samples were blind by coded so that no personal or judicial connections could be made to personal or judicial data associated to the participants. Blood samples were extracted using the Chelex® 100 method (Walsh *et al*. 1991). Total mtDNA control region was amplified in two fragments using two sets of primers, L15997/H016 and L16555/H599. PCR was done in a GeneAmp® PCR system 9700 (ABI, Foster City, USA) thermo cycler. Reaction was done in a final volume of 10 μL according to the instructions of the QIAGEN® Multiplex PCR Kit (Qiagen, Hilden, Germany) with primers at 0.2 μM. Amplification conditions were as follows: 95°C for 5 min, followed by 35 cycles, each cycle consisting of 94°C for 30 s, 60°C for 90 s and 72°C for 60 s. The amplification ended with a final extension of 10 min at 72°C. The amplified products were purified with ExoSAP-IT® (USB Corporation, Ohio, USA) method. The light and heavy chains from each of the two fragments amplified from the mtDNA control region were sequenced in a final volume of 8 μL containing Better Buffer® (Microzone, Sussex, UK), Big Dye® Terminator ver. 3.1 Cycle Sequence (ABI, Foster City, USA) and primers at 2.5 μM, in a GeneAmp® PCR system 9700 (ABI, Foster City, USA) thermo cycler. Sequencing conditions were as follows: 96°C for 2 min, 35 cycles with 96°C for 15 s, 50°C for 9 s and 60°C for 2 min, ending with 60°C for 10 min. Sequenced products were purified with the BigDye® XTerminator Purification Kit (ABI, Foster City, USA). The nucleotide sequences of the mtDNA control region, from position 16,024 to position 576, were detected in an Applied Biosystems® 3130 Genetic Analyzer (ABI, Foster City, USA), and analysed with the ABI DNA Sequencing Analysis® ver. 5.2 software. Obtained sequences were compared with the revised Cambridge Reference Sequence (rCRS) (Andrews *et al*. 1999) using the SeqScape® ver. 2.5 software and typed following IUPAC (International Union of Pure and Applied Chemistry) recommendations (Bär *et al*. 2000; Carracedo *et al*. 2000) to define haplotypes. Haplogroup designation was carried out according to the polymorphism detected in relation to Phylotree (van Oven and Kayser 2009). Nucleotide and sequence diversity of the studied population and interpopulation F_{st} genetic distances were determined using Arlequin ver. 3.5.1.2 software (Excoffier *et al.* 2005). The phylogenetic representation was obtained using the Neighbour method of Phylip ver. 3.69 software (Felsenstein 1989) and with Treeview 1.5.2 software.

Results

All mtDNA sequences analysed in this study were submitted and accepted by the EMPOP database (http://www.empop.org) (Parson and Dür 2007) with accession number EMP00616. We identified 75 haplotypes, 56 were seen only once. Hundred and twelve polymorphic nucleotide positions were determined.

Sequence diversity for this population was 0.9914 and the nucleotide diversity was 0.0132. Haplogroup frequencies seen among the 103 native individuals of Cabo Verde currently living in Lisboa are provided in table 1. The majority of the mtDNA haplotypes (92.24%) are included in macrohaplogroup L. Haplogroup L2a1a is the most frequent haplogroup found in the present study (12.621%). Phylogram representing F_{st} distances between the Cabo Verde population of Lisbon and selected populations from the literature (Fendt *et al*. 2012a, b; Mikkelsen *et al*. 2012) are shown in figure 1.

The population of Cabo Verde shows lower genetic distance to Ghana's population ($F_{st} = 0.02067$, $P = 0.00000$) and Somalia's population ($F_{st} = 0.06419$, $P = 0.00000$). It also reveals much higher genetic distance with the Angola's Khoe-San population ($F_{st} = 0.30102$, $P = 0.00000$). Nonetheless, we note a significant differentiation between the population of Cabo Verde and any other population of the comparative study.

Discussion

High frequency of unique haplotypes found in the present study was expected and is characteristic of isolated populations with early local settlement. Previous studies that focussed on the study of mtDNA of the population of Cabo Verde (Brehm *et al.* 2002; Tavares 2007) support the results of the present study as they share the most frequent haplotype (CBV035, EMPOP00616).

The majority of the analysed sequences belong to macrohaplogroup L, characteristic of African populations (Rosa and Brehm 2011). Such results are in accordance with historical data in which Cabo Verde's mtDNA pool is linked primarily to the African slave's contribution (Willie 2001). Interestingly, the most frequent haplogroup, L2a1a in the present study has also been described in Mali, Mauritania populations (González *et al.* 2006) and in southeast African populations, namely Mozambique (Pereira *et al.* 2001), one of the colonies of the excolonial Portuguese empire. Further study may be needed as L2a1a may be linked to the expansion of population groups along the west coast of Africa to

Table 1. Haplogroup distribution among 103 native individuals of Cabo Verde, currently living in Lisboa.

Haplogroup	Number of individual	Frequency	Proportion
L1b1	5	0.0485	4.854
L1b1a12	2	0.0194	1.942
L1b1a2	4	0.0388	3.883
L1b1a8	1	0.0097	0.971
L1b1a9	1	0.0097	0.971
L1c1	3	0.0291	2.913
L1c3a1b	1	0.0097	0.971
L1c3b	1	0.0097	0.971
L2a	1	0.0097	0.971
L2a1a	13	0.1262	12.621
L2a1k	1	0.0097	0.971
L2b	1	0.0097	0.971
L2b1	2	0.0194	1.942
L2b1a	4	0.0388	3.883
L2b'c	3	0.0291	2.913
L2c	6	0.0583	5.825
L2c2	3	0.0291	2.913
L2c3	1	0.0097	0.971
L3b	5	0.0485	4.854
L3b1a5	1	0.0097	0.971
L3b1b	3	0.0291	2.913
L3b2a	1	0.0097	0.971
L3d	9	0.0874	8.738
L3d2	1	0.0097	0.971
L3d4	1	0.0097	0.971
L3e2	2	0.0194	1.942
L3e2a1	4	0.0388	3.883
L3e2b	2	0.0194	1.942
L3e2b2	1	0.0097	0.971
L3e4a	6	0.0583	5.825
L3f1b	1	0.0097	0.971
L3h1b	1	0.0097	0.971
L3h1b2	1	0.0097	0.971
L3k	1	0.0097	0.971
L3k2	2	0.0194	1.942
M1	1	0.0097	0.971
M30c*	1	0.0097	0.971
M62*	1	0.0097	0.971
U6a1a	1	0.0097	0.971
U6a1	2	0.0194	1.942
X*	2	0.0194	1.942

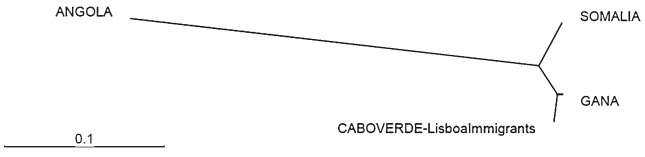


Figure 1. Phylogram derived from the analysis of genetic distances between selected populations from the literature and the studied population.

southeast Africa, like transatlantic slave trade between the 15th and 19th century.

Haplogroup L1b occurs frequently in West African populations (Rosa *et al*. 2004; González *et al*. 2006; Fendt *et al*. 2012a), especially among Senegal's Mandenka and Wolof populations (Rando *et al.* 1998; Jackson *et al.* 2005). Haplogroups L2b and L2c occur frequently on Mauritania, Senegal, Sierra Leone and Ghana (Rando *et al*. 1998; Jackson *et al*. 2005; González *et al*. 2006; Behar *et al*. 2008; Fendt *et al*. 2012a). Haplogroups L3b and L3d also occur frequently on West African populations, especially in sub-Saharan African populations, with an average of 10% (Rando *et al.* 1998; Rosa *et al.* 2004; Jackson *et al.* 2005), precisely what was verified in the present study (8.7% for each of the haplogroups). Haplogroups L3e2 and L3e4 also occur mostly on West African populations (Salas *et al.* 2002; Behar *et al.* 2008; Fendt *et al.* 2012a).

Haplogroup L3h1b, found in two of the 103 analysed sequences was first described in a study focussed on Guinea population (Rosa *et al.* 2004). Haplogroup L3k, found in three of the 103 analysed sequences has been found in Tunisia and Libya populations (Behar *et al.* 2008). Haplogroup U6a, characteristic of North African populations, especially northwest populations (Rosa and Brehm 2011) occurred in 2.91% of the analysed sequences of the present study. Comparable results were obtained by Brehm *et al.* (2002) and Tavares (2007) studies.

Haplogroup X was determined in 1.94% of the analysed sequences. This haplogroup, diverging originally from haplogroup N, is often found in European and North American populations (Reidla *et al.* 2003), was also found with low frequency in Iberian Peninsula (Pereira *et al.* 2003; Cardoso *et al.* 2012) and in North and northeast African populations (Reidla *et al.* 2003).

Haplogroup M1 occurred in one of the 75 obtained haplotypes in the present study. It was primarily found in the northeast and east Africa, although recent studies have proposed that it expanded into northwest Africa and even into Iberian Peninsula (González *et al.* 2006). Angola's Khoe-San, an indigenous population almost entirely dedicated to farming and present in some countries in South Africa, are primarily associated with haplogroups L0d and L0k (Fendt *et al.* 2012b), which may explain the high genetic distance obtained between such populations and the population from our study (figure 1). Ghana, a country in the west African coast, geographically close to Cabo Verde archipelago and involved in commercial activities along the coast, including slave trade, was the population genetically less distant to the present studied population. These populations also share the most common haplogroups. Somalia, a country in the east African coast, has as most common haplogroups L0a1d, L2a1h and L3f, characteristic of east and southeast Africa, and M1 and N1, characteristic of northeast Africa and Middle East (Mikkelsen *et al.* 2012). Analysis of the entire mtDNA control region for the first time in Cabo Verde native individuals increased the discriminatory power between samples and overall efficiency in the determination of haplogroups.

The results obtained in this study were as expected and point once again to the high contribution of slaves from the west African coast, as a result of the transatlantic salve trade executed by the excolonial Portuguese empire, in the origin

of the actual maternal lineage of Cabo Verde. Along with the Y chromosome profile of Cabo Verde (Gonçalves *et al.* 2003), these results are in accordance with the historical data where the colonization of the archipelago was initially performed through contact between male European settlers and female African slaves. The presence of haplogroups L, characteristic of African populations, in Lisboa and consequently in Portugal, as found in the present study, is probably caused by major migration movements of Cabo Verde's inhabitants to Europe during and especially after the end of slave trade in the region in the 19th and 20th centuries. This African haplogroups of the studied population introduce considerable genetic diversity in Lisboa population.

References

Afonso Costa H., Morais P., Vieira da Silva C., Matos S., Marques Santos R., Espinheira R. *et al.* 2014 X-chromosome STR markers data in a Cabo Verde immigrant population of Lisboa. *Mol. Biol. Rep.* **41**, 2559–2569.

Amorim A., Marques-Santos R., Vieira-Silva C., Afonso-Costa H., Espinheira R., Ferreira-Gomes P. *et al.* 2012 Genetic portrait of a native population of Cabo Verde living in Lisboa. *Forensic Sci. Int. Genet.* **6**, e166–e169.

Andrews R. M., Kubacka I., Chinnery P. F., Lightowlers R. N., Turnbull D. M. and Howell N. 1999 Re-analysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147.

Bär W., Brinkmann B., Budowle B., Carracedo A., Gill P., Holland M. *et al.* 2000 DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Int. J. Legal. Med.* **113**, 193–196.

Behar D. M., Villems R., Soodyall H., Blue-Smith J., Pereira L., Metspalu E. *et al.* 2008 The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1130–1140.

Brehm A., Pereira L., Bandelt H. J., Prata M. J. and Amorim A. 2002 Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann. Hum. Genet.* **66**, 49–60.

Cardoso S., Villanueva-Millán M. J., Valverde L., Odriozola A., Aznar J. M., Piñeiro-Hermida S. *et al.* 2012 Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country. *Forensic Sci. Int. Genet.* **6**, e106–e108.

Carracedo A., Bär W., Lincoln P., Mayr W., Morling N., Olaisen B. *et al.* 2000 DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Sci. Int.* **110**, 79–85.

Excoffier L., Laval G. and Schneider S. 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50.

Felsenstein J. 1989 PHYLIP: phylogeny interface package (version 3.2). *Cladistics* **5**, 164–166.

Fendt L., Röck A., Zimmermann B., Bodner M., Thye T., Tschentscher F. *et al.* 2012a MtDNA diversity of Ghana: a forensic and phylogeographic view. *Forensic Sci. Int. Genet.* **6**, 244–249.

Fendt L., Huber G., Röck A. W., Zimmermann B., Bodner M., Delport R. *et al.* 2012b Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages. *Forensic Sci. Int. Genet.* **6**, 662–663.

Gonçalves R., Rosa A., Freitas A., Fernandes A., Kivisild T., Villems R. *et al.* 2003 Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. *Hum. Genet.* **113**, 467–472.

González A. M., Cabrera V. M., Larruga J. M., Tounkara A., Noumsi G., Thomas B. N. *et al.* 2006 Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations. *Ann. Hum. Genet.* **70**, 631–657.

Jackson B. A., Wilson J. L., Kirbah S., Sidney S. S., Rosenberger J., Bassie L. *et al.* 2005 Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am. J. Phys. Anthropol.* **128**, 156–163.

Mikkelsen M., Fendt L., Röck A. W., Zimmermann B., Rockenbauer E., Hansen A. J. *et al.* 2012 Forensic and phylogeographic characterisation of mtDNA lineages from Somalia. *Int. J. Legal. Med.* **126**, 573–579.

Parson W. and Dür A. 2007 EMPOP—a forensic mtDNA database. *Forensic Sci. Int. Genet.* **1**, 88–92.

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

Pereira L., Macaulay V., Torroni A., Scozzari R., Prata M. J. and Amorim A. 2001 Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann. Hum. Genet.* **65**, 439–458.

Rando J. C., Pinto F., González A. M., Hernández M., Larruga J. M., Cabrera V. M. *et al.* 1998 Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann. Hum. Genet.* **62**, 531–550.

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. *Am. J. Hum. Genet.* **73**, 1178–1190.

Rosa A. and Brehm A. 2011 African human mtDNA phylogeography at-a-glance. *J. Anthropol. Sci.* **89**, 25–58.

Rosa A., Brehm A., Kivisild T., Metspalu E. and Villems R. 2004 MtDNA profile of West 28 African mtDNA at a glance Africa Guineans: towards a better understanding of the Senegambia region. *Ann. Hum. Genet.* **68**, 340–352.

Salas A., Richards M., De La Fe T., Lareu M.-V., Sobrino B., Sánchez-Diz P. *et al.* 2002 The making of the African mtDNA landscape. *Am. J. Hum. Genet.* **71**, 1082–1111.

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

van Oven M. and Kayser M. 2009 Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394.

Walsh P., Metzger D. and Higuchi R. 1991 Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506–513.

Willie P. 2001 *Encyclopedia of African history and culture.* Facts on File. New York, USA.

Received 1 October 2014, in revised form 27 January 2015; accepted 18 March 2015
Unedited version published online: 6 April 2015
Final version published online: 22 September 2015

Estudamos a totalidade da região controlo do DNAmt de 103 indivíduos nascidos em Cabo Verde, não aparentados, e, actualmente, integrados na população de Lisboa.

A metodologia de estudo do DNAmt consistiu na utilização de dois pares de primers para sequenciação da totalidade da região controlo do DNAmt, sendo todas as amostras estudadas numa extensão de 1122 pb. Obtivemos haplótipos individuais completos para todos os indivíduos estudados que apresentamos na tabela 7.

A população apresenta uma grande variabilidade genética intrapopulacional, manifestada pela grande frequência de haplótipos únicos, com 75 haplótipos diferentes em 103 sequências de DNAmt analisadas. O menor número de alterações polimórficas encontradas numa sequência em relação à CRS revista foi de 9, enquanto o maior número de alterações polimórficas, em relação à CRS revista, foi de 27. No conjunto de todas as sequências analisadas, foram detectadas 112 posições polimórficas, com diferentes frequências na amostra em estudo. O estudo das sequências obtidas mostrou a existência de uma zona especialmente polimórfica na região HV1, entre a posição 16183 e a posição 16193.

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. No gráfico 1 apresentamos a distribuição dos imigrantes de Cabo Verde pelos diferentes haplogrupos do macrohaplogrupo L. Tal facto aponta para uma elevada contribuição de escravas provenientes da costa ocidental africana na origem da composição genética das linhagens maternas actuais do arquipélago de Cabo Verde. Estes

resultados estão de acordo com estudos anteriores efectuados na população de Cabo Verde.

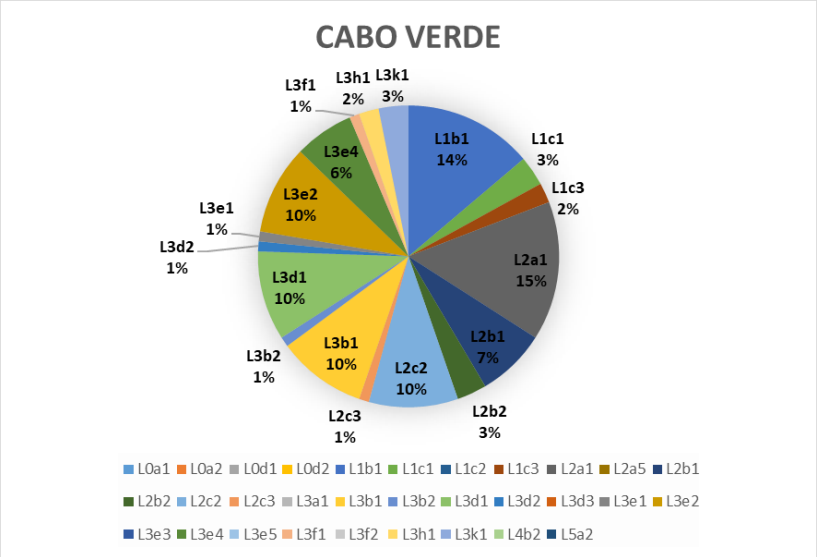


Gráfico 1 - Distribuição dos imigrantes de Cabo Verde pelos diferentes haplogrupos do macrohaplogrupo L

Entre os haplótipos que não pertencem a haplogrupos do tipo L, 2,91% foram agrupados no haplogrupo U6, típico de populações do norte de África. 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.

Os restantes haplótipos (4,85%) foram agrupados no haplogrupo M e no haplogrupo X. A obtenção de uma reduzida frequência de haplótipos agrupados em haplogrupos com ocorrência na Europa poderá reflectir uma pequena contribuição feminina de populações europeias na origem das linhagens mitocondriais actuais 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 efectuado 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.

A população de imigrantes de Cabo Verde atualmente integrados na população de Lisboa estudada no âmbito do nosso projeto está alojada na EMPOP com *accession number* EMP00616 (https://empop.online/populations).

Tabela 7 - Haplótipos e correspondentes haplogrupos dos imigrantes de Cabo Verde

CBV001	L3b1a5	16223T	16278T	16362C	16519C	73G	263G	315.1C	523DEL	524DEL																				
CBV002	L3d	16124C	16399G	73G	150T	152C	263G	315.1C	523DEL	524DEL																				
CBV003	L3d	16086C	16124C	16223T	73G	152C	263G	315.1C	523DEL	524DEL																				
CBV004	L3d2	16111T	16124C	16223T	73G	152C	263G	315.1C	523DEL	524DEL																				
CBV005	L3e2	16223T	16320T	16519C	73G	150T	152C	195C	263G	315.1C																				
CBV006	L3e2	16223T	16320T	16519C	73G	150T	195C	200G	263G	315.1C																				
CBV007/1	L3e2a1	16223T	16320T	16519C	73G	150T	195C	198T	263G	315.1C																				
CBV007/2	L3e2a1	16223T	16320T	16519C	73G	150T	195C	198T	263G	315.1C																				
CBV007/3	L3e2a1	16223T	16320T	16519C	73G	150T	195C	198T	263G	315.1C																				
CBV008/1	L3d	16093C	16124C	16223T	73G	146C	152C	263G	315.1C	523DEL	524DEL																			
CBV008/2	L3d	16093C	16124C	16223T	73G	146C	152C	263G	315.1C	523DEL	524DEL																			
CBV008/3	L3d	16093C	16124C	16223T	73G	146C	152C	263G	315.1C	523DEL	524DEL																			
CBV008/4	L3d	16093C	16124C	16223T	73G	146C	152C	263G	315.1C	523DEL	524DEL																			
CBV009	L3e2a1	16212G	16223T	16320T	16519C	73G	150T	195C	198T	263G	315.1C																			
CBV010	L3k	16223T	16390A	73G	150T	152C	235G	263G	309.1C	315.1C	494T																			
CBV011/1	L3k1	16223T	16355T	73G	150T	152C	235G	263G	309.1C	315.1C	494T																			
CBV011/2	L3k1	16223T	16355T	73G	150T	152C	235G	263G	309.1C	315.1C	494T																			
CBV012/1	L3b	16124C	16223T	16278T	16362C	73G	195C	263G	315.1C	482C	523DEL	524DEL																		
CBV012/2	L3b	16124C	16223T	16278T	16362C	73G	195C	263G	315.1C	482C	523DEL	524DEL																		
CBV013/1	L3d	16124C	16223T	16311C	16399G	73G	150T	152C	263G	315.1C	523DEL	524DEL																		
CBV013/2	L3d	16124C	16223T	16311C	16399G	73G	150T	152C	263G	315.1C	523DEL	524DEL																		
CBV014/1	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																		
CBV014/2	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																		
CBV015/1	L3e4a	16051G	16223T	16264T	16519C	73G	150T	257G	263G	315.1C	523DEL	524DEL																		
CBV015/2	L3e4a	16051G	16223T	16264T	16519C	73G	150T	257G	263G	315.1C	523DEL	524DEL																		
CBV016**	L3b	16048A	16124C	16223T	16278T	16362C	73G	263G	309.1C	309.2C	315.1C	523DEL	524DEL																	
CBV017/1	L3b	16124C	16223T	16234T	16278T	16362C	16519C	73G	263G	309.1C	315.1C	523DEL	524DEL																	
CBV017/2	L3b	16124C	16223T	16234T	16278T	16362C	16519C	73G	263G	309.1C	315.1C	523DEL	524DEL																	
CBV018	L3d4	16124C	16223T	16519C	73G	152C	189G	195C	207A	263G	315.1C	523DEL	524DEL																	
CBV019/1	L3e4a	16051G	16223T	16264T	16299G	16519C	73G	150T	263G	309.1C	315.1C	523DEL	524DEL																	
CBV019/2	L3e4a	16051G	16223T	16264T	16299G	16519C	73G	150T	263G	309.1C	315.1C	523DEL	524DEL																	
CBV019/3	L3e4a	16051G	16223T	16264T	16299G	16519C	73G	150T	263G	309.1C	315.1C	523DEL	524DEL																	
CBV020	L3e4a	16051G	16223T	16257T	16264T	16519C	73G	150T	263G	309.1C	315.1C	523DEL	524DEL																	
CBV021	U6a1a	16172C	16179T	16183C	16189C	16219G	16278T	16519C	73G	150T	263G	309.1C	315.1C																	
CBV022	L2a1a	16189C	16223T	16278T	16294T	16309G	16390*	73G	146C	152C	195C	263G	309.1C	315.1C																
CBV023	L3b1b	16124C	16223T	16278T	16362C	16519C	73G	151T	152C	263G	309.1C	315.1C	523DEL	524DEL																
CBV024	L3b1b	16124C	16223T	16234T	16278T	16362C	16519C	73G	151T	152C	263G	315.1C	523DEL	524DEL																
CBV025	L3d	16124C	16223T	16257T	16519C	73G	151T	152C	183G	263G	309.1C	315.1C	523DEL	524DEL																
CBV026***	L3e2b2	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	152C	195C	263G	309.1C	315.1C																
CBV027	L3f1b	16209C	16223T	16292T	16295T	16311C	16519C	73G	150T	189G	200G	234G	263G	315.1C																
CBV028	M62*	16129A	16223T	16295T	73G	150T	151T	152C	207A	263G	309.1C	315.1C	523DEL	524DEL																
CBV029/1	X*	16114A	16188T	16189C	16223T	16278T	16304C	16311C	73G	152C	263G	309.1C	315.1C	523DEL	524DEL															
CBV029/2	X*	16114A	16188T	16189C	16223T	16278T	16304C	16311C	73G	152C	263G	309.1C	315.1C	523DEL	524DEL															
CBV030	L2a	16093C	16148T	16223T	16278T	16294T	16355T	16390A	16519C	73G	146C	195C	263G	309.1C	315.1C															
CBV031	L2a1a	16093C	16223T	16278T	16294T	16309G	16390*	16519C	73G	143A	146C	152C	195C	263G	315.1C															
CBV032	L3b1b	16124C	16223T	16278T	16311C	16362C	16519C	73G	152C	189G	263G	309.1C	315.1C	523DEL	524DEL															
CBV033*	L3b2a	16124C	16183C	16189C	16193.1C	16223T	16278T	16362C	16527T	73G	263G	309.1C	315.1C	523DEL	524DEL															
CBV034	M1	16129A	16185T	16189C	16223T	16249C	16311C	16519C	73G	195C	263G	309.1C	315.1C	466C	489C															
CBV035/1	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
CBV035/2	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
CBV035/3	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
CBV035/4	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
CBV035/5	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
CBV035/6	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
CBV036/1	L2b'c	16223T	16278T	16390A	73G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL														
CBV036/2	L2b'c	16223T	16278T	16390A	73G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL														
CBV037	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	73G	143A	146C	152C	195C	263G	309.1C	315.1C	523DEL	524DEL													
CBV038	L2a1a	16111T	16223T	16278T	16294T	16309G	16390*	16519C	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL													
CBV039	L2c	16223T	16278T	73G	93G	146C	150T	152C	182T	185C	189G	263G	315.1C	325T	356.1C	523DEL	524DEL													
CBV040*/1	U6a1	16111T	16172C	16183C	16189C	16219G	16278T	16519C	73G	150T	185A	189G	204C	207A	263G	309.1C	315.1C													
CBV040*/2	U6a1	16111T	16172C	16183C	16189C	16219G	16278T	16519C	73G	150T	185A	189G	204C	207A	263G	309.1C	315.1C													
CBV041	L2a1a	16179T	16223T	16278T	16290T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	198T	263G	309.1C	315.1C												
CBV042	L2c3	16223T	16278T	16390A	73G	93G	95C	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	513A	523DEL	524DEL											
CBV043	L2c	16178C	16223T	16278T	16380T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL											
CBV044	L2c	16223T	16278T	16320T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL												
CBV045	L2b'c	16177G	16223T	16278T	16311C	16390A	16519C	73G	146C	150T	152C	182T	195C	263G	315.1C	325T	523DEL	524DEL												
CBV046	L3h1b2	16129A	16223T	16256A	16311C	16362C	16519C	73G	151T	152C	189C	195C	263G	294C	315.1C	499A	523DEL	524DEL												
CBV047*	M30c*	16189C	16223T	16259T	16274A	16278T	16390A	16519C	73G	146C	182T	195A	207A	263G	315.1C	316A	523DEL	524DEL												
CBV048	L2a1a	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	207A	263G	264T	309.1C	315.1C	534T											
CBV049	L2c	16223T	16278T	16390A	16519C	73G	93G	146C	150T	152C	182T	195C	198T	263G	309.1C	315.1C	325T	523DEL	524DEL											
CBV050**	L2a1a	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	207A	263G	264T	309.1C	315.1C	534T											
CBV051	L2b	16114A	16129A	16213A	16223T	16278T	16390A	73G	146C	150T	152C	182T	185C	195C	198T	204C	207A	263G	309.1C	315.1C										
CBV052/1	L2c2	16223T	16264T	16278T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	207A	263G	309.1C	315.1C	325T	523DEL	524DEL										
CBV052/2	L2c2	16223T	16264T	16278T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	207A	263G	309.1C	315.1C	325T	523DEL	524DEL										
CBV052/3	L2c2	16223T	16264T	16278T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	207A	263G	309.1C	315.1C	325T	523DEL	524DEL										
CBV053	L3h1b	16129A	16192T	16209C	16218T	16223T	16256A	16311C	16362C	16519C	73G	151T	189C	195C	263G	294C	309.1C	315.1C	523DEL	524DEL										
CBV054	L1b1	16114G	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	523DEL	524DEL								
CBV055	L1b1a2	16111T	16126C	16187T	16189C	16223T	16239T	16270T	16278T	16293G	16311C	16519C	73G	146C	152C	182T	185T	189G	207A	247A	263G	315.1C	357G	523DEL	524DEL					
CBV056	L2a1k	16183C	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	263G	264T	315.1C	524.1A	524.2C	534T									
CBV057	L1b1	16126C	16145A	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL								
CBV058/1	L2b1	16114A	16129A	16169T	16213A	16223T	16278T	16362C	16390A	73G	150T	152C	182T	195C	198T	204C	263G	309.1C	315.1C	418T	523DEL	524DEL								
CBV058/2	L2b1	16114A	16129A	16169T	16213A	16223T	16278T	16362C	16390A	73G	150T	152C	182T	195C	198T	204C	263G	309.1C	315.1C	418T	523DEL	524DEL								
CBV059/1	L1b1	16126C	16145A	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL							
CBV059/2	L1b1	16126C	16145A	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL							
CBV060/1	L1b1a2	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	189G	195C	247A	263G	315.1C	357G	523DEL	524DEL							
CBV060/2	L1b1a2	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	189G	195C	247A	263G	315.1C	357G	523DEL	524DEL							
CBV061	L2b1a	16114A	16129A	16213A	16223T	16278T	16355T	16362C	16390A	73G	150T	152C	182T	195C	198T	204C	249DEL	263G	309.1C	315.1C	418T	523DEL	524DEL							
CBV062**	L2c	16126C	16183C	16189C	16223T	16269G	16278T	16293G	16390A	73G	93G	146C	150T	152C	182T	195C	198T	263G	309.1C	315.1C	325T	523DEL	524DEL							
CBV063*	L2c	16126C	16189C	16193.1C	16223T	16269G	16278T	16293G	16390A	73G	93G	146C	150T	152C	182T	195C	198T	263G	309.1C	315.1C	325T	523DEL	524DEL							
CBV064	L1b1a8	16111T	16126C	16187T	16189C	16221T	16223T	16239T	16270T	16278T	16311C	16519C	73G	146C	152C	182T	185T	247A	263G	309.1C	315.1C	357G	523DEL	524DEL						
CBV065	L1b1a9	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16368C	16519C	73G	152C	182T	185T	189G	195C	247A	263G	309.1C	315.1C	357G	523DEL	524DEL						
CBV066	L2b1a	16114A	16129A	16213A	16223T	16278T	16355T	16362C	16390A	16465T	73G	150T	152C	182T	195C	198T	204C	249DEL	263G	309.1C	315.1C	418T	523DEL	524DEL						
CBV067	L1b1a12	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16362C	16400T	16519C	73G	152C	182T	185T	195C	198T	247A	263G	309.1C	315.1C	357G	523DEL	524DEL				
CBV068	L1b1	16114G	16126C	16187T	16189C	16223T	16264T	16270T	16274A	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	309.1C	315.1C	357G	373G	523DEL	524DEL				
CBV069	L1c3b	16129A	16145A	16163G	16187T	16189C	16223T	16239T	16278T	16293G	16294T	16309T	16311C	16360T	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL				
CBV070/1	L2b1a	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				
CBV070/2	L2b1a	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				
CBV071	L1b1a12	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16362C	16400T	16519C	73G	151T	152C	182T	185T	195C	198T	247A	263G	309.1C	315.1C	357G	523DEL	524DEL			
CBV072	L1b1a2	16093C	16126C	16187T	16189C	16223T	16239T	16264T	16270T	16278T	16293G	16311C	16399G	16519C	73G	146C	152C	182T	185T	189G	207A	247A	263G	315.1C	357G	523DEL	524DEL			
CBV073	L1c1	16093C	16129A	16187T	16189C	16274A	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	249DEL	263G	297G	315.1C	316A	523DEL	524DEL		
CBV074/1	L1c1	16129A	16163G	16187T	16189C	16223T	16278T	16293G	16294T	16304C	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	315.1C	316A	513A	523DEL	524DEL		
CBV074/2	L1c1	16129A	16163G	16187T	16189C	16223T	16278T	16293G	16294T	16304C	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	315.1C	316A	513A	523DEL	524DEL		
CBV076*	L1c3a1b	16129A	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16355T	16360T	16519C	73G	151T	152C	182T	183G	186A	189C	204C	247A	257G	263G	315.1C	316A	523DEL	524DEL			

Capítulo V - Estudo de imigrantes de Angola: Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA

Simão F, Afonso Costa H, Vieira da Silva C, Ribeiro T, Porto MJ, Costa Santos J, Amorim A. ***Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA.*** Forensic Science International: Genetics 15 (2015):33-38.

Forensic Science International: Genetics 15 (2015) 33–38

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA

Filipa Simão [a,b], Heloísa Afonso Costa [a,f], Claúdia Vieira da Silva [a,f], Teresa Ribeiro [a,f], Maria João Porto [a,f], Jorge Costa Santos [a,c,d,f], António Amorim [a,d,e,f,*]

[a] Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal
[b] Faculdade de Ciências e Tecnologias da Universidade Nova de Lisboa, Portugal
[c] Faculdade de Medicina da Universidade de Lisboa, Portugal
[d] Instituto Superior de Ciências de Saúde Egas Moniz, Portugal
[e] Faculdade de Ciências da Universidade de Lisboa, Portugal
[f] CENCIFOR – Centro de Ciências Forenses, Portugal

ARTICLE INFO

Keywords:
Mitochondrial DNA
Control region
Angola
Lisboa

ABSTRACT

Portugal has been considered a country of emigrants, nevertheless in the past decades the number of immigrants has grown throughout all the country. This migratory flux has contributed to a raise of heterogeneity at multiple levels. According to statistical data, at the end of 2012 the total number of Angolan immigrants in Portugal equalled about 20,000 individuals. A territorial predominance has been found for the metropolitan region of Lisboa. Angola is a country located in the Atlantic coast of Africa. The presence of Bantu people and the colonisation by Portuguese people on Angolan territory are considered to be the major modulators of the genetic patterns in Angola.

Mitochondrial DNA is known for its features that enable an approach to the study of human origin and evolution, as well to the different migration pathways of populations. This genetic marker can also contribute to ascertaining the identity of individuals in forensic cases. The main aim of this study was to determine the genetic structure of the Angolan immigrant population living in Lisboa. Therefore, a total of 173 individuals, inhabitants in Lisboa, nonrelated and with Angolan ancestry were studied. Total control region of mitochondrial DNA was amplified from position 16,024 to position 576 using two pairs of primers – L15997/H016 and L16555/H639.

The majority of the identified haplotypes belong to mtDNA lineages known to be specific of the sub-Saharan region. Our results show that this immigrant population inhabitant in Lisboa presents a genetic profile that is characteristic of African populations. This study also demonstrates the genetic diversity that this immigrant population introduces in Lisboa. This does not contradict the historical data concerning colonization of Angola, since this was made mainly by male European individuals, who did not contribute with their maternal information of mtDNA.

Lisboa immigrant population from Angola can be accessed via EMPOP dataset with accession number EMPOP662.

1. Introduction

Angola is a country located in the south Atlantic coast of Africa. The Khoe-san people were the first known population living in this territory. Around 3000–4000 years ago, the Bantu people, arising from the East and Central regions of Africa, began two main migratory fluxes to the South and West of the continent. Part of this movement culminated with their fixation in Angola territory. In the 15th century, with the ascension of King D. João II to the Portuguese throne and his concern on the discovery of the sea route to Índia, the recognition of the African coast progressed rapidly. In 1482, the Portuguese navigator Diogo Cão, arrived at to the Angolan territory. Later, in 1641, Angola was occupied by Netherlands, invasion that ended in 1648, when Portugal freed the territory. Angola officially became a Portuguese colony in 1886 and it became independent in 1975 [1–3]. Thus, the arrival of Bantu

* Corresponding author at: Instituto Nacional de Medicina Legal e Ciências Forenses, Serviço de Genética e Biologia Forenses – Sul, Rua Manuel Bento de Sousa, 3, 1169-201 Lisboa, Portugal. Tel.: +351 218 811 865; fax: +351 21 8864 493.
E-mail address: aamorim@dlinml.mj.pt (A. Amorim).

http://dx.doi.org/10.1016/j.fsigen.2014.09.013
1872-4973/

people and the miscegenation between different sociocultural groups that occurred during the colonial time are considered the main modulators of the genetic portrait of Angolan population.

Since the beginning of the 20th century, Portugal has been known by its number of emigrant people. Nevertheless, this scenery changed in the 70s with the increasing flow of immigrants, largely related with the post-colonial and post-independence period of African countries of Portuguese language. According to statistical data from *Serviço de Estrangeiros e Fronteiras* concerning to 2012 [4], the total number of immigrant people in Portugal equalled 417,042 individuals. One of the most representative nationalities was Angolan with a total of 20,366 individuals living in Portugal; also it was been found a territorial predominance for the metropolitan region of Lisboa was also found, with 12,704 Angolan residents [4,5].

The mitochondrial DNA (mtDNA) characteristics such as maternal inheritance [6], high copy number [7], absence of recombination [8] and high mutation rate make this genetic marker useful in different fields: in medical genetics, by analysis of pathogenic mutations; in population genetics, when investigating the differences between human populations; and in forensic genetics, by comparison between one or more evidence mtDNA sequences with a reference sequence. The study of the mitochondrial genome allows ascertaining the differences between human populations in evolutionary studies; thus, it is possible to estimate the geographic origin of a population, their dynamic and the time elapsed since the most common ancestor of modern humans. The discrimination power of mtDNA is due to the polymorphic status of its hypervariable regions located in the control region – HVI, HVII and HVIII (Hypervariable region I, II and III) – with a total length of 1122 bp [9]. In the past decades, with the development of new sequencing technologies of high performance, the study of mtDNA total control region has become more common.

Although mtDNA has a lower discrimination power when compared with nuclear DNA, since the haplotype of an individual is shared within the same maternal lineage, the study of mtDNA is also commonly used in forensic genetic. In particular cases where the samples have a small amount of nuclear DNA or it is degraded into small fragments, this genetic marker reveal to be very useful [10–12].

The phenomenon of heteroplasmy – the presence of more than one type of mtDNA in one individual – has been reported in different studies. The heteroplasmy can occur in two forms: length heteroplasmy and sequence heteroplasmy. The first one is associated with insertions or deletions in homopolymeric C-streches in HVI, HVII and HVIII regions. The second type of heteroplasmy is a result of substitutions, with presence of two different nucleotides in a single position [13–18].

The haplotype of each individual is defined by a list of polymorphic sites, comparatively to revised Cambridge Reference Sequence (rCRS) [19,20], a reference sequence, which implied that the remaining positions, not described, are equal to the rCRS. The set of polymorphisms of a given sequence of mtDNA forms a group of lineages next to each other that cluster in haplogroups, defined by the letters of the alphabet. Sequences that belong to the same haplogroup or sub-haplogroup have specific mutations, representative of a common ancestor. These haplogroups can be related with a specific geographic region or, in some cases, with an ethnic group. Thus, the study of mtDNA and its haplogroups of different populations have been used to determine the time and routs of human migration and to investigate pre historical and historical times of humanity [21,22].

An African origin of modern humans explains the increasing interest in different studies of populations with African origins [23]. Studies on maternal ancestry of several lineages based on mtDNA also allow to obtain complementary information concerning the origin of man in Africa and its dispersion on and from this continent. The model most accepted in the scientific community concerning this matter, and confirmed by mtDNA analysis, is known as Out of Africa theory [24]. This postulates that, after an initial expansion of populations from Africa to Eurasia, the different populations did not maintain contact, and evolved independently. Subsequently, the modern man emerged and evolved in Africa, from a single population of archaic *Homo sapiens*; after this stage of evolution, a new expansion resulted in the replacement and global extinction of the population for modern human, as a result of their superior constitution [25].

The high frequency of haplogroups L shown in African populations, as well as the high level of heterogeneity of the mitochondrial DNA sequences in this continent have been reported several times, being higher than in other geographical regions [26–40].

Studies regarding mtDNA total control region of African populations have been performed [31–33,37]. However, sequences of mtDNA total control region of Angolan individuals are not published. The only information, to date, on maternal inheritance of this population includes sequences of HVI and HVII mtDNA control region [34,35,41,42], or HVI region [43]. The Angolan immigrant population living in Lisboa has been studied with STR genetic markers [44].

The aims of this study include (a) identification of the mtDNA haplotypes of the Angolan immigrant population living in Lisboa, (b) obtainment of mtDNA variability within this group of samples (c) determination of the haplogroups and (d) comparision of the present population with other African populations, hoping to understand the present and future impact of the Angolan immigrant population in Lisboa.

2. Materials and methods

2.1. Sampling and DNA extraction

Blood samples were collected from 173 Angolan immigrant inhabitants in Lisbon, non-related, of a single generation, which attended the National Institute of Legal Medicine and Forensic Science (Instituto Nacional de Medicina Legal e Ciências Forenses – INMLCF), southern branch, for kinship investigations. According to the Portuguese law that regulates INMLCF's activity, samples from routine forensic cases can be used for investigation purposes. All samples are used with special codification without any connection to judicial data related to the downer. DNA was extracted using Chelex®100 method [45].

2.2. mtDNA typing

Control region of the mtDNA was amplified using two pairs of primers – L15997/H016 and L16555/H639. The amplified products were purified with ExoSap-IT (Affymetrix). Total control region sequencing was performed with BigDye® Terminator v.3.1 Cycle Sequence (AB) followed by purification with BigDye® XTerminator Purification Kit (AB). Finally, the sequenced products were detected in a sequencer Genetic Analyzer 3130 (AB) and the results were analysed by Sequencing Analysis v.5.2 and SeqScape v.3 (AB) softwares.

2.3. Sequences alignment and haplogroup classification

Obtained sequences were compared with rCRS [20] applying the nomenclature guidelines used in mtDNA analysis to identify haplotypes [46]. Haplogroups were determined based on Phylotree, build 16, developed by van Oven and Kayser, last updated in February 2014 [47].

2.4. Data analysis

The statistical analysis of the mtDNA total control region, calculating the sequence diversity and nucleotide diversity, as well the population diversity among other populations [31–33,35–38,42,48], was performed using Arlerquin v 3.5.1.2 software [49]. The phylogenetic representation of genetic distances was drawn with Neighbour method of Phylip v. 3.69 [50] and with Treeview v. 1.5.2 [51] softwares.

3. Results and discussion

3.1. Haplotype composition

The haplotypes acquired in this study were submitted and accepted on EMPOP database (www.empop.org) [52] accession number EMPOP662.

Along the 173 studied individuals, 145 different haplotypes were identified, 125 of which are unique. The other 20 haplotypes were present in more than one individual. Thus, 14 sequences shared their haplotypes within groups of 2 samples, 4 sequences within groups of 3 samples and 2 sequences within groups of 4 samples. The two most frequent haplotypes are 16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C (ANG078, ANG151, ANG166 and ANG171) and 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL (ANG012, ANG077, ANG106 and ANG152) present in four samples each.

The transition of a cytosine (C) for a thymine (T) in position 16,223 of HVI region, it is present in 153 samples. In 142 sequences was observed a transition in position 73 of HVII region was observed, where an adenine (A) is replaced by a guanine (G). In HVII region, position 152, a substitution of a thymine (T) by a cytosine (C) is observed in 103 sequences. Also, in position 16,278 of HVI, a cytosine (C) is replaced by a thymine (T) in 92 sequences. The most frequent polymorphisms present in this sampling are also found in other African populations [31–33].

In the 173 studied individuals one point heteroplasmy was identified at position 16,183 of HVI region in sample ANG163. Here, in some mtDNA molecules of the individual, the nucleotide present is an Adenine, while in others is a Cytosine, being the position classified as 16183M. Regarding length heteroplasmy, associated with homopolymeric C-streches in HVI, HVII and HVIII regions, there are a total of 55 heteroplasmic samples. In 25 samples it is present only in HVI region, in 19 it is present only in HVII region and in 1 sample it is present only in HVIII region. However, in 8 samples the length heteroplasmy is present in both HVI and HVII, and in 2 samples it is present in HVI and HVIII regions. The remaining 118 samples are homoplasmic.

3.2. Haplogroup composition

Macrohaplogrup L is the most common haplogroup in our sample, since there are 151 samples that fit into L haplogroup (about 87%). Haplogroups H and T are present in 5 samples each. Haplogroup R0 is present in 4 samples, haplogroups K and U appear in 3 samples each and the haplogroups J and M are present in one sample each.

The L lineages are known to be characteristic from Sub-Saharan regions of Africa, and it is in this continent that it is found the higher genetic diversity [27,39,40]. Within haplogroup L0 there are four sub-haplogroups - L0a, L0d, L0f and L0k. In this study the samples presented haplotypes belonging to sub-haplogroups L0d (1 sequence) and L0a (23 sequences). The sub-haplogroup L0d is known for being restricted to Khoe-san people in South Africa and to Tanzania and Angolan populations [39,40,53]; on the other hand, the L0a lineages are widely spread through Africa (eastern, central and southern areas) [39,40]. The Khoe-san L0k was not found in our samples. Regarding to haplogroup L1, five samples were included in L1b, characteristic of western-central African regions [28,40]. The L1c, which is concentrated in Central and West of Africa includes 32 of our samples. The haplogroups L3 and L2 comprise about 70% of the Sub-Saharan maternal variation [40]. Three of the five subclades of haplogroup L2 (L2a, L2b, L2c) were present in 32 of our studied individuals; these are known for existing in the West-East region of Africa. The lineage L3 is spread through Africa [39,40]; a total of 56 of our studied individuals were within this haplogroup, included in L3a, L3b, L3d and L3f subclades. The L4 lineages, despite having low frequencies, are mostly found in the East-North of Africa; in this case, 2 of our samples belong to this haplogroup [39,40].

Although present in a small range, some of the haplogroups identified in our samples, do not belong to L lineages and are not characteristic from the African population, even though they can be found in some areas of this continent. The European populations derive from haplogroup N, which branches into haplogroups H, I, J, K, T, U, V, W and X. On the other hand, the Asiatic lineages belong both to haplogroups N and M.

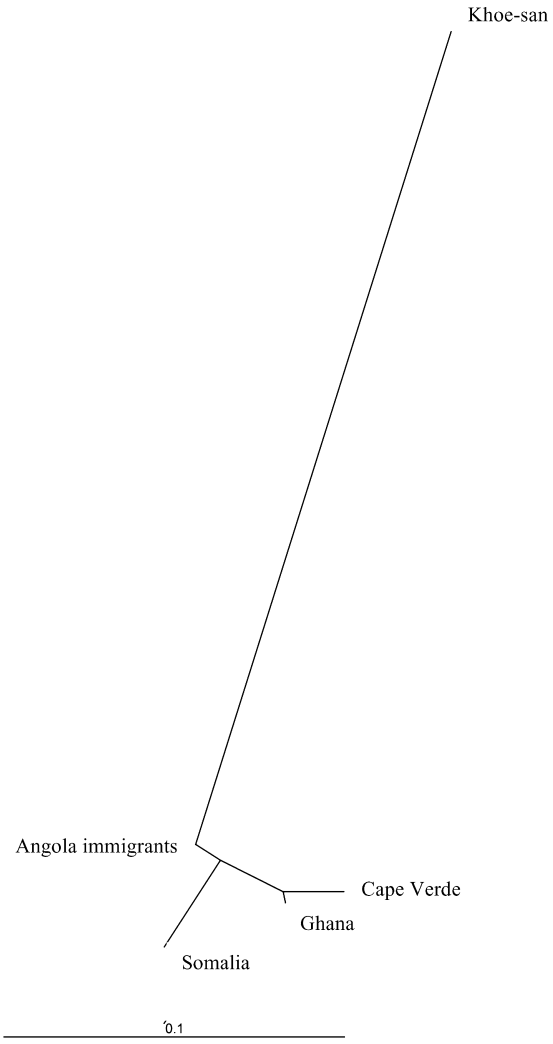


Fig. 1. Phylogenetic representation of the genetic distances for African populations regarding mtDNA total control region.

3.3. Genetic diversity

Statistical analyses were made taking into consideration the total control region of sequences. According to calculations obtained from Arlequin software v. 3.5.2.1 [49], for this population the value of the sequence diversity is 0.9964 ± 0.0012 and the nucleotide diversity is 0.016028 ± 0.007910. These values were also calculated for HVI and HVII regions, in order to be able to compare our results with the published studies concerning only to HVI and HVII mtDNA control region. Therefore the sequence diversity for HVI and HVII regions of this population is 0.9959 ± 0.0012 and the nucleotide diversity is 0.025058 ± 0.012429. The values of the sequence diversity are higher when the total control region is studied, in comparision with the study of the hypervariable regions HVI and HVII.

3.4. Inter population analysis

In Fig. 1 we present the phylogenetic tree based on molecular distances between populations, calculated with Arlequin v. 3.5.2.1 software [49], for the mtDNA total control region. Our Lisboa immigrant population from Angola was compared with

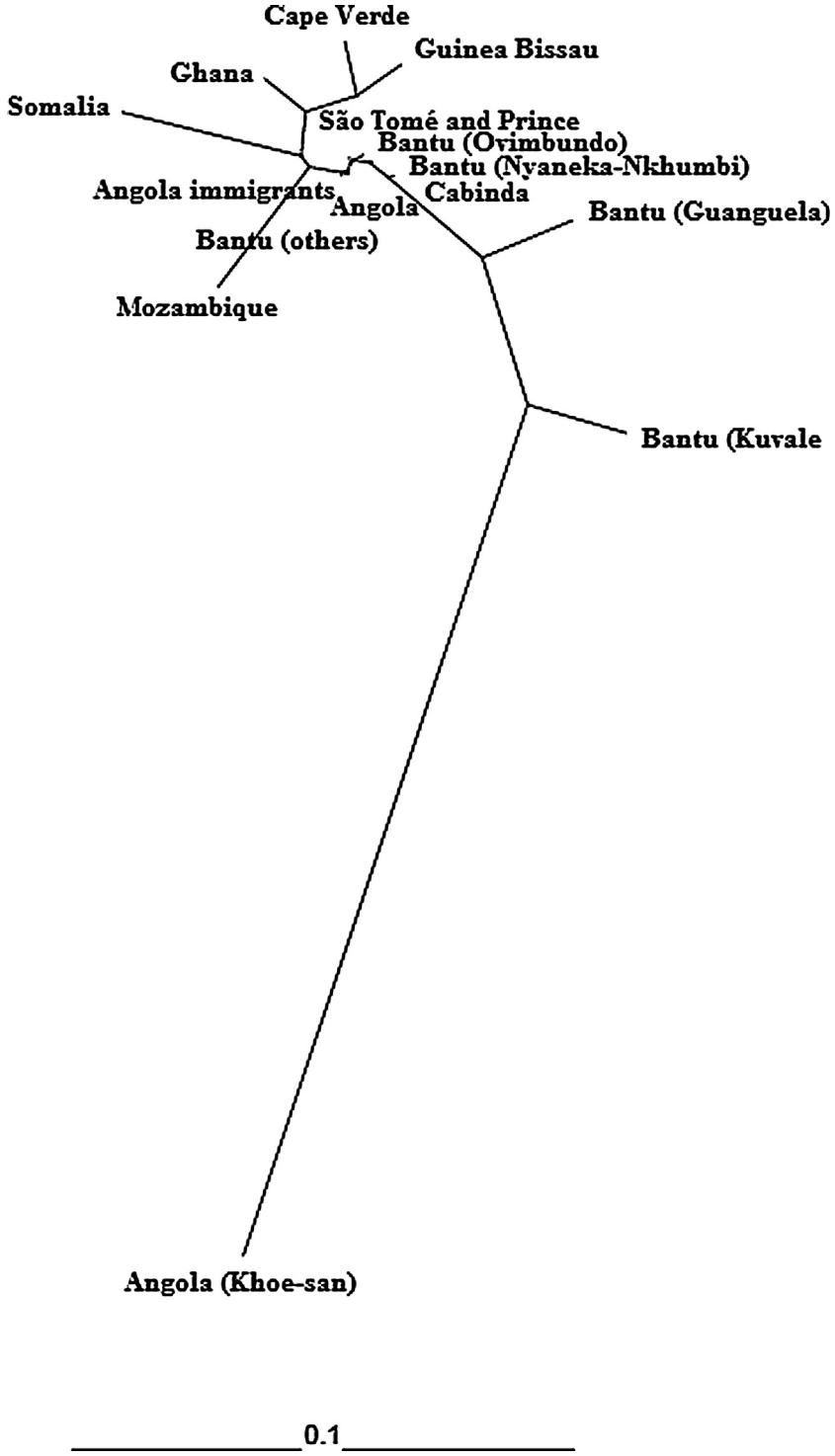


Fig. 2. Phylogenetic representation of the genetic distances for African populations regarding HVI and HVII mtDNA control region.

Table 1
Populations used in the phylogenetic study of mtDNA total control region.

Population	Sample size	References
Present study	173	–
Khoe-san	129	Frendt (2012)
Ghana	192	Frendt (2012)
Somalia	190	Mikkelsen (2012)
Cape Verde	103	Morais (2013)

Table 2
Analysis of molecular variance (AMOVA): pairwise difference genetic distance (FST/FST ρ values) between the studied population and other populations from the literature, regarding mtDNA total control region.

Populations	Sample size	Fst	ρ
Present study	173	0.00000	–
Khoe-san	129	0.24677	0.00000 ± 0.0000
Cape Verde	103	0.04408	0.00000 ± 0.0000
Somalia	190	0.03588	0.00000 ± 0.0000
Ghana	192	0.02746	0.00000 ± 0.0000

Table 3
Populations used in the phylogenetic study of mtDNA HVI and HVII regions.

Population	Sample size	References
Present study	173	–
Angola (Khoe-san)	129	Frendt (2012)
Ghana	192	Frendt (2012)
Somalia	190	Mikkelsen (2012)
Cape Verde	103	Morais (2013)
Guinea Bissau	88	Carvalho (2011)
Mozambique	109	Pereira (2001)
São Tomé and Prince	100	Trovoada (2004)
Angola	59	Carvalho (2006)
Bantu (others)	45	Coelho (2009)
Bantu (Kuvale)	54	Coelho (2009)
Bantu (Guanguela)	21	Coelho (2009)
Bantu (Nyaneka-Nkhumbi)	153	Coelho (2009)
Bantu (Ovimbundu)	92	Coelho (2009)
Cabinda	110	Beleza (2005)

populations from Ghana [32], Somalia [31], Cape Verde [33] and from indigenous Angolan Khoe-san [37] (Table 1). Although the immigrant Angolans living in Lisboa show a great distance from Khoe-san population (Fst = 0.24677; ρ = 0.00000), it is smaller than the distance between the latter one and the other populations analysed in the study. Still, the Angolan population living in Lisboa is closer to Ghana population (Fst = 0.02746; ρ = 0.00000), followed by Somalia (Fst = 0.03588; ρ = 0.00000) and Cape Verde (Fst = 0.04408; ρ = 0.00000) (Table 2). The lack of Khoe-san haplogroups contribution into southwest regions of Africa has already been described by other authors [26,36,54].

For comparison purposes, and as there are few studies regarding African population for mtDNA total control region, the sequences of the studied population were reduced to HVI and HVII regions, to allow comparison with other populations studied for HVI and HVII regions. In this case, the population of Angolan immigrants living in Lisboa was compared with populations from Guinea Bissau [29], Cape Verde [33], Somalia [31], Ghana [32], Mozambique [38], S. Tome and Principe [48], Angola [42], indigenous Khoe-san [37], Bantu ethnic groups from Angola [35,36] and Cabinda [34], all of them with sequenced HVI and HVII regions (Table 3) (Fig. 2). These results culminate in a short distance between the population of Angolan immigrants living in Lisboa and Angolan, people living in Angola as well Bantu ethnic groups from Angola, giving information that the first one maintains the genetic information of its origin, although, nowadays, inhabiting a different geographic region.

4. Conclusions

The study of the entire control region for the Angolan immigrant population living in Lisboa, obtained for the first time, increases the discrimination power of the sequences and allows a better determination of haplogroups. Due to the high frequency of unique haplotypes present in our samples, the population has a great genetic variability.

It is possible to verify the genetic similarity with African populations. From the analysis restricted to HVI and HVII regions it was determined that the Angolan immigrants living in Lisboa have a small distance to the native Angolan populations. These results were expected since, according to historical data, the colonisation of this country by the European population it was made mainly by male individuals, which would lead to a low contribution to European mtDNA sequences of possible miscegenation during colonial times. On the other hand, the presence of these immigrants in Portugal is relatively recent, which eventually results in a lower contribution of the haplotypes in our new generations; but in a closer future, these African haplotypes will, eventually, be a part of the overall Lisboa population haplotypes. On the other hand, the great distance of our studied population from Khoe-san population demonstrates the high replacement rate of Khoe-san people by Bantu populations in Angolan territory.

Acknowledgements

The authors would like to thank Walther Parson and Alexander Röck (EMPOP, University of Innsbruck, Austria).

References

[1] J. Redinha, Etnias e culturas de Angola, 2009.
[2] J. De Raeymaeker, À descoberta de Angola, 1st ed., Oficina do Livro – Sociedade Editorial Lda, Lisboa, 2012.
[3] Iniciação à história da Angola, República Popular de Agola, Ministério da Educação, 1992.
[4] P. Dias, R. Machado, J. Estrela, A.R. Bento, Relatório de imigração fronteiras e asilo 2012, Oeiras, 2012.
[5] PORDATA: Base de Datos Portugal Contemporâneo, 2014.
[6] R.E. Giles, H. Blanc, H.M. Cann, D.C. Wallace, Maternal inheritance of human mitochondrial DNA, Proc. Natl. Acad. Sci. U. S. A. 77 (1980) 6715–6719.
[7] H.H. Cells, D. Bogenhagen, D.A. Clayton, The number of mitochondrial deoxyribonucleic acid genomes in mouse the number of mitochondrial deoxyribonucleic genomes in mouse L and human HeLa cells, J. Biol. Chem. 249 (1974) 7991–7995.
[8] J. William, O. Ballard, M.C. Whitlock, The incomplete natural history of mitochondria, Mol. Ecol. 13 (2004) 729–744.
[9] J.M. Butler, Forensic DNA Typing, 2005.
[10] S. Lutz, H.J. Weisser, J. Heizmann, S. Pollak, mtDNA as a tool for identification of human remains. Identification using mtDNA, Int. J. Legal Med. 109 (1996) 205–209.
[11] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, Int. J. Legal Med. 108 (1995) 68–74.
[12] W. Parson, T.J. Parsons, R. Scheithauer, M.M. Holland, Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case, Int. J. Legal Med. 111 (1998) 124–132.
[13] D. Comas, S. Paabo, J. Bertranpetit, Heteroplasmy in the control region of human mitochondrial DNA, Genome Res. 5 (1995) 89–90.
[14] K.A. Roberts, C. Calloway, Characterization of mitochondrial DNA sequence heteroplasmy in blood tissue and hair as a function of hair morphology, J. Forensic Sci. 56 (2011) 46–60.
[15] A. Alonso, A. Salas, C. Albarrán, E. Arroyo, A. Castro, M. Cresillo, et al., 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 Sci. Int. 125 (2002) 1–7.
[16] A. Brandstätter, W. Parson, Mitochondrial DNA heteroplasmy or artefacts – a matter of the amplification strategy? Int. J. Legal Med. 117 (2003) 180–184.
[17] N. Howell, C.B. Smejkal, Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction, Am. J. Hum. Genet. 66 (2000) 1589–1598.
[18] M.R. Wilson, D. Polanskey, J. Replogle, J.A. DiZinno, B. Budowle, A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes, Hum. Genet. 100 (1997) 167–171.

[19] S. Anderson, A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, et al., Sequence and organization of the human mitochondrial genome, Nature 290 (1981) 457–465.
[20] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.
[21] H.J. Bandelt, M. van Oven, A. Salas, Haplogrouping mitochondrial DNA sequences in legal medicine/forensic genetics, Int. J. Legal Med. 126 (2012) 901–916.
[22] B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution, Annu. Rev. Genomics Hum. Genet. 6 (2005) 165–183.
[23] C.M. Schlebusch, P. Skoglund, P. Sjödin, L.M. Gattepaille, D. Hernandez, F. Jay, et al., Genomic variation in seven Khoe-San groups reveals adaptation and complex African history, Science 338 (2012) 374–379.
[24] N. Maca-Meyer, A.M. González, J.M. Larruga, C. Flores, V.M. Cabrera, Major genomic mitochondrial lineages delineate early human expansions, BMC Genet. 2 (2001) 13.
[25] I. Tattersall, Human origins: out of Africa, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 16018–16021.
[26] A. Salas, M. Richards, M.-V. Lareu, R. Scozzari, A. Coppa, A. Torroni, et al., The African diaspora: mitochondrial DNA and the Atlantic slave trade, Am. J. Hum. Genet. 74 (2004) 454–465.
[27] Y.S. Chen, A. Torroni, L. Excoffier, A.S. Santachiara-Benerecetti, D.C. Wallace, Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups, Am. J. Hum. Genet. 57 (1995) 133–149.
[28] E. Watson, K. Bauer, R. Aman, G. Weiss, A. von Haeseler, S. Pääbo, mtDNA sequence diversity in Africa, Am. J. Hum. Genet. 59 (1996) 437–444.
[29] M. Carvalho, P. Brito, A.M. Bento, V. Gomes, H. Antunes, H.A. Costa, et al., Paternal and maternal lineages in Guinea-Bissau population, Forensic Sci. Int. Genet. 5 (2011) 114–116.
[30] M. Carvalho, V. Bogas, F. Balsa, H.A. Costa, V. Lopes, M.M. Melo, et al., Mitochondrial control region data of 3 ethnic groups from angola, Forensic Sci. Int. Genet. 3 (2011) e125–e126.
[31] M. Mikkelsen, L. Fendt, A.W. Röck, B. Zimmermann, E. Rockenbauer, A.J. Hansen, et al., Forensic and phylogeographic characterisation of mtDNA lineages from Somalia, Int. J. Legal Med. 126 (2012) 573–579.
[32] L. Fendt, A. Röck, B. Zimmermann, M. Bodner, T. Thye, F. Tschentscher, et al., MtDNA diversity of Ghana: a forensic and phylogeographic view, Forensic Sci. Int. Genet. 6 (2012) 244–249.
[33] P. Morais, Estudo do DNA mitocondrial de um grupo de imigrantes oriundos de Cabo Verde residentes em Lisboa Mestrado em Biologia Molecular em Saúde, 2013.
[34] S. Beleza, L. Gusmão, A. Amorim, A. Carracedo, A. Salas, The genetic legacy of western Bantu migrations, Hum. Genet. 117 (2005) 366–375.
[35] M. Coelho, F. Sequeira, D. Luiselli, S. Beleza, J. Rocha, On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola, BMC Evol. Biol. 9 (2009) 80, http://dx.doi.org/10.1186/1471-2148-9-80.
[36] S. Plaza, A. Salas, F. Calafell, F. Corte-Real, J. Bertranpetit, A. Carracedo, et al., Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, Hum. Genet. 115 (2004) 439–447.
[37] L. Fendt, G. Huber, A.W. Röck, B. Zimmermann, M. Bodner, R. Delport, et al., Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages, Forensic Sci. Int. Genet. 6 (2012) 662–663.
[38] L. Pereira, V. Macaulay, A. Torroni, R. Scozzari, Z. Pavia, V. Ferrata, Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansion and the slave trade, Ann. Hum. Genet. 65 (2001) 439–458.
[39] A. Salas, M. Richards, T. De la Fe, M.-V. Lareu, B. Sobrino, P. Sánchez-Diz, et al., The making of the African mtDNA landscape, Am. J. Hum. Genet. 71 (2002) 1082–1111.
[40] A. Rosa, A. Brehem, African human mtDNA phylogeography at-a-glance, J. Anthropol. Sci. 89 (2011) 25–58.
[41] S. Plaza, A. Salas, F. Calafell, F. Corte-Real, J. Bertranpetit, A. Carracedo, et al., Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, Hum. Genet. (2004) 439–447.
[42] M. Carvalho, F. Balsa, V. Lopes, M.J. Anjos, L. Andrade, A. Serra, et al., Mitochondrial DNA analysis in a population from Angola, in: XXTH Congress of International Academy of Legal Medicine, 23–26 August, Budapest, Hungary, 2006.
[43] S. Tofanelli, P. Garagnani, C. Bini, X. Fosella, S. Pelotti, G. Paoli, et al., mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in Sub-Saharan Africa, Am. J. Phys. Anthropol. 311 (2009) 302–311.
[44] R.M. dos Santos, C.V. da Silva, H.A. Costa, P.F. Gomes, S. Sanches, R. Espinheira, et al., Genetic portrait of an immigrant population from Angola living in Lisboa, Forensic Sci. Int. Genet. 6 (6) (2012), e170-3.
[45] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques 10 (1991) 506–513.
[46] A. Carracedo, W. Bar, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, et al., DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 17–85.
[47] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2009) E386–E394.
[48] M.J. Trovoada, L. Pereira, L. Gusmão, A. Abade, A. Amorim, M.J. Prata, Insights from pattern of mtDNA variation into the genetic history of São Tomé e Príncipe, Int. Congr. Ser. 1261 (2004) 377–379.
[49] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0); an integrated software package for population genetics data analysis, Evol. Bionform. Online 1 (2005) 47–50.
[50] J. Felsenstein, PHYLIP: phylogeny interface package (version 3.2), Cladistics 5 (1989) 164–166.
[51] R.D.M. Page, Treeview: an application to display phylogenetic trees on personal computers, Comput. Appl. Biosci. 12 (1996) 357–358.
[52] W. Parson, A. Dür, EMPOP – a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88–92.
[53] D.M. Behar, R. Villems, H. Soodyall, J. Blue-smith, L. Pereira, E. Metspalu, et al., The dawn of human matrilineal diversity, Am. J. Hum. Genet. 82 (2008) 1130–1140.
[54] J. Alves-Silva, M. da Silva Santos, P.E. Guimarães, A.C. Ferreira, H.J. Bandelt, S.D. Pena, et al., The ancestry of Brazilian mtDNA lineages, Am. J. Hum. Genet. 67 (2000) 444–461.

Avaliamos a contribuição genética da população imigrante Angolana a residir em Lisboa, pela análise da região controlo do DNAmt de 173 indivíduos nascidos em Angola e atualmente integrados na população de Lisboa.

Amplificamos a totalidade da região controlo do DNAmt, numa extensão de 1 122pb, e a metodologia utilizada revelou ser eficiente para a realização do objetivo proposto, permitindo a obtenção de sequências referentes à totalidade da região controlo do DNAmt, com elevada qualidade. A utilização da totalidade da região controlo do DNAmt revelou ser de eficiência superior, comparativamente a estudos anteriores da molécula que apenas incluíam a sequenciação dos segmentos HVI e HVII ou do segmento HVI isolado da região controlo do DNAmt. Os polimorfimos presentes na região controlo total do DNAmt ajudam a clarificar os haplótipos e a discriminar os haplogrupos a que as sequências se inserem.

Obtivemos haplótipos individuais completos para todos os indivíduos estudados que apresentamos na tabela 8.

Os haplótipos determinados no conjunto de amostras estudadas pertencem, em grande parte, a haplogrupos característicos da região subsariana de África. No gráfico 2 apresentamos a distribuição dos imigrantes de Angola pelos diferentes haplogrupos do macrohaplogrupo L. Apenas uma pequena percentagem (≈13%) de haplogrupos é característica de populações euroasiáticas. É possível confirmar a baixa influência europeia na constituição do padrão genético de DNAmt do indivíduo Angolano e, consequentemente, no padrão genético dos imigrantes Angolanos a residir em Lisboa.

O elevado número de posições polimórficas e respetivos polimorfismos, confirma dados anteriores relativos à

elevada heterogeneidade genética presente em populações do continente Africano. A grande percentagem de haplótipos únicos presente do estudo realizado é também característica de populações Africanas.

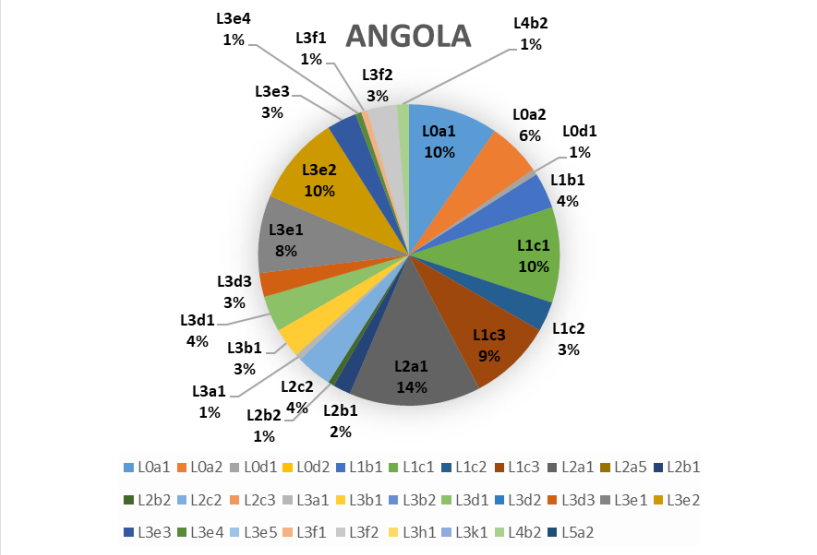


Gráfico 2 - Distribuição dos imigrantes de Angola pelos diferentes haplogrupos do macrohaplogrupo L

Através da obtenção das distâncias genéticas entre a população dos imigrantes Angolanos a residir em Lisboa com outras populações Africanas, confirmou-se a proximidade deste povo imigrante com os indivíduos Angolanos a residir em Angola e com diversas populações Bantu.

A análise do DNAmt dos indivíduos Angolanos a residir em Lisboa, reforça a história da criação do povo Angolano, estando em concordância com relatos arqueológicos e linguísticos.

Confirma-se, desta forma, que o padrão genético do DNAmt de Angola e, consequentemente de imigrantes Angolanos a residir em Lisboa resulta, em grande parte, da expansão do povo Bantu, que com a sua chegada a este território,

expulsou os primeiros habitantes da região, o povo Khoe-san. Por outro lado, colonização de Angola terá ocorrido maioritariamente por indivíduos europeus do sexo masculino, que durante possíveis miscigenações com indígenas do território, não terão contribuído com o seu material genético mitocondrial. Ainda, a presença de imigrantes Angolanos em Portugal é recente, sendo que ainda não decorreu tempo suficiente para que as miscigenações aí desenvolvidas tenham considerável contribuição europeia.

Tendo em consideração o número crescente de imigrantes Angolanos em Portugal, a informação retirada a partir do estudo do DNAmt do grupo de indivíduos Angolanos a residir em Lisboa realça a importância destes estudos na caracterização genética dos indivíduos a residir em Lisboa. Num futuro próximo, indivíduos naturais e nacionais de Lisboa, irão apresentar haplótipos, até então considerados como tipicamente Africanos.

A população de imigrantes de Angola atualmente integrados na população de Lisboa estudada no âmbito do nosso projeto está alojada na EMPOP com *accession number* EMP00662 (https://empop.online/populations).

Tabela 8 - Haplótipos e haplogrupos correspondentes dos imigrantes de Angola

ANG002	K2a1a	16066G	16182C	16183C	16189C	16223T	16224C	16278T	16390A	16519C	73G	146C	152C	182T	263G	315.1C	511T													
ANG003	L2a1g	16131C	16189C	16223T	16225T	16234T	16278T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	309.2C	315.1C												
ANG004	L3e2b	16172C	16182C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																	
ANG005	L2a1	16092C	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C															
ANG006	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	16390A	89C	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL			
ANG007	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																		
ANG008	L1c1	16038G	16129A	16187T	16189C	16223T	16271C	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	309.1C	315.1C	316A	521G	523DEL	524DEL
ANG009	L1c1b	16086C	16129A	16187T	16189C	16223T	16241G	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A	416A	518T	523DEL	524DEL
ANG010	L1c3b1a	16129A	16163G	16187T	16189C	16209C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL					
ANG011	H	16519C	152C	263G	309.1C	315.1C																								
ANG012	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL					
ANG013	L2a1a3c	16093C	16223T	16256T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	263G	315.1C															
ANG014	L4b2	16093G	16223T	16287A	16293T	16301T	16311C	16355T	16362C	16399G	73G	146C	150T	152C	195C	200G	244G	263G	315.1C	513A											
ANG018	L1c3a	16129A	16182C	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16360T	16519C	73G	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL							
ANG017	L0a2a2	16148T	16172C	16173T	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	309.1C	315.1C	523DEL	524DEL					
ANG019	L1c1b	16086C	16129A	16187T	16189C	16223T	16241G	16278T	16291T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A	416A			
ANG020	L3e1	16176T	16223T	16327T	73G	150T	189G	200G	263G	315.1C																				
ANG021	L1b1a10	16126C	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	151T	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL	567G								
ANG022	L3e2b	16164G	16172C	16189C	16223T	16320T	16519C	73G	146C	150T	195C	263G	315.1C	523DEL	524DEL															
ANG023	L3e1	16207G	16223T	16327T	73G	150T	183G	189G	200G	263G	315.1C																			
ANG024	L3f2	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	315.1C																			
ANG025	L3d	16124C	16223T	73G	152C	207A	263G	315.1C	523DEL	524DEL																				
ANG026	L3e1	16213A	16223T	16309G	16327T	73G	150T	189G	200G	263G	315.1C																			
ANG027	T1a9	16126C	16163G	16186T	16189C	16249C	16294T	16311C	16519C	73G	263G	309.1C	315.1C																	
ANG028	L1c1	16129A	16187T	16189C	16223T	16263C	16278T	16293G	16294T	16311C	16360T	16368C	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A	523DEL	524DEL		
ANG029	L3b1a	16124C	16223T	16278T	16362C	16519C	73G	152C	263G	309.1C	315.1C	523DEL	524DEL																	
ANG030	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL									
ANG031	T2c1	16126C	16292T	16294T	16296T	16519C	73G	146C	263G	315.1C	523DEL	524DEL																		
ANG032	L3f2	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	315.1C																			
ANG033	R0	16519C	263G	315.1C																										
ANG034	L2a1f3	16189C	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	198T	263G	315.1C															
ANG035	L2a1a	16223T	16265C	16278T	16294T	16309G	16368C	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C	573.1C	573.2C	573.3C	573.4C										
ANG036	L2b	16114A	16129A	16212G	16213T	16223T	16278T	16390A	73G	146C	150T	152C	182T	195C	198T	204C	207A	263G	309.1C	309.2C	315.1C										
ANG037	L3e4	16051G	16223T	16264T	16519C	73G	150T	257G	263G	315.1C	523DEL	524DEL																		
ANG038	L1c1b	16086C	16129A	16187T	16189C	16223T	16241G	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A	416A	518T	523DEL	524DEL
ANG039	L0a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	236C	247A	263G	315.1C	523DEL	524DEL									
ANG040	L0a1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16234T	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL				
ANG041	L1b1a10	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	151T	152C	182T	185T	195C	247A	263G	309.1C	315.1C	357G	523DEL	524DEL							
ANG042	L1c3c	16129A	16187T	16189C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	93G	151T	152C	182T	186A	189C	195C	247A	248G	263G	315.1C	316A	458T	523DEL	524DEL			
ANG043	J1c2c	16069T	16126C	16519C	73G	146C	185A	188G	228A	263G	295T	309.1C	315.1C	462T	489C															
ANG044	L3b	16124C	16223T	16278T	16362C	16519C	73G	263G	315.1C	523DEL	524DEL																			
ANG045	L2c2b1b	16223T	16264T	16278T	16311C	16390A	73G	93G	146C	150T	152C	182T	183G	195C	198T	263G	264T	315.1C	325T	523DEL	524DEL									
ANG046	L3b1a3	16124C	16223T	16278T	16311C	16362C	16519C	73G	263G	315.1C	523DEL	524DEL																		
ANG047	L2c2b1b	16223T	16264T	16278T	16311C	16390A	73G	93G	146C	150T	152C	182T	183G	195C	198T	263G	315.1C	325T	523DEL	524DEL										
ANG048	L1c1a2	16129A	16183C	16189C	16223T	16274A	16278T	16293G	16294T	16311C	16360T	16519C	73G	89C	93G	95C	152C	182T	186A	189C	236C	247A	263G	297G	315.1C	316A	523DEL	524DEL		
ANG049	H1b	16356C	16519C	263G	315.1C																									
ANG050	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16274A	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	309.1C	315.1C	357G	373G	523DEL	524DEL					
ANG051	L0a1b	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	152C	185A	189G	236C	247A	263G	315.1C	523DEL	524DEL					
ANG052	L2a1e1	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C	524.1A	524.2C	524.3A	524.4C												
ANG053	L3e1a	16185T	16223T	16311C	16327T	73G	150T	185A	189G	200G	263G	315.1C																		
ANG054	L1c1a	16093C	16129A	16187T	16189C	16223T	16263C	16278T	16293G	16294T	16311C	16360T	16368C	16519C	73G	151T	152C	182T	186A	189C	195C	247A	263G	297G	315.1C	316A	523DEL	524DEL		
ANG055	L2a1	16131C	16189C	16223T	16225T	16234T	16278T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	309.2C	315.1C												
ANG056	L2b2a	16114A	16129A	16213A	16223T	16278T	16354T	16390A	73G	146C	150T	152C	182T	195C	198T	204C	263G	315.1C												
ANG057	L1c1	16038G	16187T	16189C	16223T	16261T	16278T	16293G	16294T	16311C	16360T	16368C	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A	523DEL	524DEL		
ANG058	L2c2b1b	16223T	16264T	16278T	16311C	16390A	73G	93G	146C	150T	152C	182T	183G	195C	198T	263G	315.1C	325T	523DEL	524DEL										
ANG059	L3b1a8	16145A	16223T	16278T	16362C	16519C	73G	195C	263G	279C	315.1C	523DEL	524DEL																	
ANG060	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	93G	152C	189G	204C	207A	236C	247A	263G	309.1C	315.1C	523DEL	524DEL							
ANG061	L0a2a2	16086C	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL						
ANG062	L1c	16129A	16183C	16189C	16223T	16278T	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL								
ANG063	L1c3b1a	16129A	16163G	16187T	16189C	16209C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	309.1C	315.1C	316A	523DEL	524DEL				
ANG064	L3d3a	16124C	16183C	16189C	16223T	16278T	16304C	16311C	73G	152C	263G	315.1C	523DEL	524DEL																
ANG065	H	16519C	152C	263G	309.1C	315.1C																								
ANG066	L2a1c1	16086C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	198T	263G	315.1C															
ANG067	L0a'b	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16274A	16278T	16311C	16320T	89C	93G	95C	152C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL			
ANG068	L2b	16114A	16129A	16213A	16223T	16274A	16278T	16390A	16519C	73G	146C	150T	152C	182T	183G	195C	198T	204C	263G	309.1C	315.1C									
ANG069	L2a1a1	16223T	16278T	16294T	16309G	16368C	16390A	16519C	73G	146C	195C	263G	309.1C	315.1C	524.1A	524.2C	524.3A	524.4C												
ANG070	L3d3a	16124C	16183C	16189C	16223T	16278T	16304C	16311C	73G	152C	263G	315.1C	523DEL	524DEL																
ANG071	L3b	16124C	16223T	16278T	16362C	16399G	16519C	73G	263G	315.1C	523DEL	524DEL																		
ANG072	L3e2a1b1	16129A	16223T	16320T	16399G	16519C	73G	150T	195C	198T	263G	315.1C	499A																	
ANG073	L0d1b2	16129A	16187T	16189C	16223T	16230G	16243C	16294T	16311C	16519C	73G	146C	152C	195C	199C	247A	315.1C	316A	498DEL											
ANG074	L1c3a1b	16129A	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16355T	16360T	16390A	16519C	73G	151T	152C	182T	186A	189C	247A	263G	309.1C	309.2C	315.1C	316A	523DEL	524DEL			
ANG075	L3d3a	16124C	16183C	16189C	16223T	16278T	16304C	16311C	73G	152C	263G	315.1C	523DEL	524DEL																
ANG076	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	309.1C	315.1C	357G	523DEL	524DEL								
ANG077	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL					
ANG078	L3e1a	16185T	16223T	16327T	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																		
ANG079	L3f2	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																		
ANG080	L0a1ba	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	150T	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL				
ANG081	L3e1d	16176T	16223T	16327T	73G	150T	152C	189G	200G	263G	315.1C																			
ANG082	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16234T	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL				
ANG083	K1a1b1c	16222T	16224C	16311C	16327T	16519C	73G	263G	315.1C	385G	497T																			
ANG084	L1c1a2	16129A	16183C	16189C	16223T	16274A	16278T	16293G	16294T	16311C	16360T	16519C	73G	76T	89C	93G	95C	152C	182T	186A	189C	236C	247A	263G	297G	309.1C	315.1C	316A	523DEL	524DEL
ANG085	L1c2b2	16078G	16129A	16187T	16189C	16223T	16265C	16286A	16294T	16311C	16320T	16360T	16519C	16527T	73G	151T	152C	186A	189C	195C	198T	204C	247A	263G	297G	315.1C	316A	523DEL	524DEL	
ANG086	L3e1	16223T	16327T	73G	150T	189G	200G	263G	309.1C	315.1C																				
ANG087	L1c1	16093C	16129A	16187T	16189C	16274A	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	249DEL	263G	297G	315.1C	316A	523DEL	524DEL		
ANG088	L1c2b1b1	16093C	16129A	16187T	16189C	16223T	16265C	16278T	16286G	16294T	16311C	16360T	16519C	16527T	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	264T	297G	315.1C	316A		
ANG089	L2a1a2	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C															
ANG090	L3e3	16223T	16265T	16519C	73G	150T	195C	263G	315.1C	523DEL	524DEL																			
ANG091	U5a2a	16114A	16192T	16256T	16270T	16294T	16526A	73G	195C	263G	309.1C	315.1C																		
ANG092	L3e2b	16172C	16189C	16223T	16311C	16320T	16519C	73G	150T	152C	195C	263G	315.1C																	
ANG093	L1c1b	16086C	16129A	16187T	16189C	16223T	16241G	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A	416A	518T	523DEL	524DEL
ANG094	R0	16519C	263G	315.1C																										
ANG095	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	309.1C	315.1C																	
ANG096	L2a1	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	263G	309.1C	315.1C															
ANG097	L2a1	16189C	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C															
ANG098	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL						
ANG099	U5a1	16256T	16270T	16399G	73G	263G	309.1C	309.2C	315.1C																					
ANG100	L3e3b2	16223T	16265T	16519C	16527T	73G	150T	195C	263G	315.1C	523DEL	524DEL																		
ANG101	L2a1a3	16129A	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C																
ANG102	L3d1a1a	16124C	16223T	16319A	73G	146C	150T	263G	315.1C	523DEL	524DEL																			
ANG103	L3d1d	16124C	16223T	16256T	16368C	73G	152C	263G	315.1C	523DEL	524DEL																			
ANG104	L1c3a	16129A	16182C	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16360T	16519C	73G	152C	182T	186A	189C	247A	263G	315.1C	316A	573.1C	573.2C	573.3C	573.4C					
ANG105	L3e2b	16172C	16189C	16223T	16320T	16366T	16519C	73G	150T	195C	263G	309.1C	315.1C	573.1C	573.2C															
ANG106	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL					
ANG107	L0a1b1	16071T	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16261T	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL				
ANG108	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	95C	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL						
ANG109	L1c3b1a	16129A	16163G	16187T	16189C	16209C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL					
ANG110	L2a1a2	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C															
ANG111	L3f1b1a1	16129A	16209C	16223T	16292T	16295T	16311C	16519C	73G	189G	200G	263G	272G	309.1C	315.1C															
ANG112	L2c2	16223T	16264T	16278T	16309T	73G	93G	146C	150T	152C	182T	195C	198T	263G	309.1C	315.1C	325T	523DEL	524DEL											
ANG113	L3c3b	16129A	16163G	16187T	16189C	16209C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL					
ANG114	L2a1'2'3'4	16223T	16278T	16293G	16294T	16390A	73G	143A	146C	152C	182T	195C	263G	309.1C	315.1C	523DEL	524DEL													
ANG115	L1c3b'c	16129A	16187T	16189C	16223T	16274A	16278T	16293G	16294T	16311C	16360T	16519C	73G	93G	95C	151T	152C	154C	182T	186A	189C	236C	247A	263G	297G	315.1C	316A	523DEL	524DEL		
ANG116	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL							
ANG117	L3e2b	16172C	16182C	16183C	16189C	16223T	16320T	16519C	73G	150T	152C	195C	263G	315.1C																
ANG118	L1c2	16092C	16129A	16187T	16189C	16223T	16265C	16286G	16294T	16311C	16360T	16519C	16527T	73G	151T	152C	182T	186A	189C	195C	198T	200G	247A	263G	297G	315.1C	316A			
ANG119	L3e1	16223T	16327T	73G	150T	189G	200G	263G	315.1C																					
ANG120	U5a2a	16114A	16192T	16256T	16270T	16294T	16526A	73G	195C	263G	309.1C	315.1C																		
ANG121	L1c1a	16129A	16182C	16183C	16189C	16223T	16278T	16293G	16294T	16311C	16355T	16360T	16519C	73G	89C	93G	95C	151T	152C	182T	186A	189C	195C	236C	247A	263G	297G	315.1C	316A	523DEL
ANG122	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL									

ANG123	L0a1a2	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	64T	93G	95C	151T	185A	189G	200G	236C	247A	263G	315.1C	523DEL	524DEL					
ANG124	L3d	16124C	16223T	73G	152C	207A	263G	315.1C	523DEL	524DEL																				
ANG125	L2a	16086C	16223T	16278T	16294T	16390A	73G	146C	152C	195C	198T	263G	315.1C																	
ANG126	T2c1	16126C	16292T	16294T	16296T	16519C	73G	146C	263G	315.1C	523DEL	524DEL																		
ANG127	L2a1	16223T	16234T	16249C	16278T	16294T	16295T	16390A	16526A	73G	143A	146C	152C	195C	263G	315.1C														
ANG128	M1b2	16129A	16183C	16189C	16223T	16249C	16311C	16399G	16519C	73G	195C	263G	315.1C	489C																
ANG129	L3e3	16093C	16223T	16265T	16519C	73G	150T	195C	263G	315.1C	523DEL	524DEL																		
ANG130	L3e2b	16172C	16183C	16189C	16192T	16223T	16320T	16519C	73G	150T	152C	195C	263G	309.1C	315.1C															
ANG131	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	152C	195C	263G	309.1C	315.1C																
ANG132	L3e2b	16172C	16183C	16189C	16223T	16320T	16366T	16519C	73G	150T	195C	263G	309.1C	315.1C	573.1C	573.2C														
ANG133	H24	16293G	16362C	16519C	263G	309.1C	309.2C	315.1C																						
ANG134	L2a1	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C														
ANG135	L2a1a1	16223T	16278T	16294T	16309G	16368C	16390A	16519C	73G	146C	152C	195C	263G	309.1C	309.2C	315.1C	524.1A	524.2C	524.3A	524.4C										
ANG136	L1c1d	16038G	16086C	16129A	16187T	16189C	16223T	16278T	16284G	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	309.1C	315.1C	316A	523DEL	524DEL
ANG137	L2a1'2'3'4	16131C	16189C	16223T	16234T	16278T	16294T	16390A	73G	146C	152C	263G	309.1C	309.2C	315.1C															
ANG138	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16214T	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL				
ANG139	L2c2b1b	16069T	16223T	16234T	16264T	16278T	16311C	16390A	16519C	73G	93G	146C	150T	152C	182T	183G	195C	198T	263G	315.1C	325T	523DEL	524DEL							
ANG140	L3e1a	16182T	16223T	16327T	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																		
ANG141	L3f2/L3h	16192T	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																	
ANG142	L3d1a1a	16124C	16223T	16311C	16319*	73G	150T	152C	263G	315.1C	523DEL	524DEL																		
ANG143	L3e3	16172C	16223T	16265T	16519C	73G	150T	195C	263G	315.1C	523DEL	524DEL																		
ANG144	L1c1a	16093C	16129A	16187T	16189C	16223T	16263C	16278T	16293G	16294T	16311C	16360T	16368C	16519C	73G	151T	152C	182T	186A	189C	195C	247A	263G	297G	315.1C	316A	523DEL	524DEL		
ANG145	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL							
ANG146	L3f2	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	315.1C																			
ANG147	L3d3a	16124C	16183C	16189C	16223T	16234T	16278T	16304C	16311C	73G	152C	263G	309.1C	309.2C	315.1C	523DEL	524DEL													
ANG148	L1c3b'c	16129A	16187T	16189C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	16527T	73G	151T	152C	182T	186A	189C	247A	263G	265C	315.1C	316A	523DEL	524DEL					
ANG149	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																		
ANG150	L0a2	16093C	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL							
ANG151	L3e1a	16185T	16223T	16327T	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																		
ANG152	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL					
ANG153	L2a1a3c	16093C	16223T	16256T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	263G	315.1C															
ANG154	L3e2b	16172C	16183C	16189C	16223T	16301T	16320T	16519C	73G	150T	195C	263G	309.1C	315.1C																
ANG155	L3d1a1a	16124C	16223T	16319A	73G	150T	152C	263G	315.1C	523DEL	524DEL																			
ANG156	L1c3a1b	16129A	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16355T	16360T	16519C	73G	151T	152C	182T	183G	186A	189C	204C	247A	257G	263G	309.1C	315.1C	316A	523DEL	524DEL		
ANG157	K2a1a	16066G	16182C	16183C	16189C	16223T	16224C	16278T	16390A	16519C	73G	146C	152C	182T	263G	309.1C	309.2C	315.1C	511T											
ANG158	L2c	16223T	16278T	16390A	73G	89C	93G	146C	150T	182T	195C	198T	263G	309.1C	315.1C	325T	523DEL	524DEL												
ANG159	L3e1a	16185T	16223T	16327T	16519C	73G	150T	189G	200G	263G	315.1C																			
ANG160	L3e2b	16172C	16189C	16223T	16320T	16519C	73G	150T	152C	195C	263G	315.1C																		
ANG161	L4b2	16223T	16233G	16271C	16293T	16311C	16355T	16362C	16399G	44.1C	73G	146C	152C	244G	263G	309.1C	315.1C													
ANG162	L1c	16093C	16129A	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16327T	16355T	16360T	16390A	16519C	73G	151T	152C	182T	186A	189C	195C	247A	263G	309.1C	315.1C	316A			
ANG163	L1c2	16129A	16183M	16189C	16223T	16265C	16278T	16286G	16294T	16311C	16360T	16519C	16527T	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	315.1C	316A				
ANG164	R0	16468C	16519C	263G	309.1C	315.1C																								
ANG165	R0	16519C	185A	263G	309.1C	315.1C																								
ANG166	L3e1a	16185T	16223T	16327T	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																		
ANG167	T2c1d	16093C	16126C	16292T	16294T	16519C	73G	146C	263G	279C	309.1C	315.1C																		
ANG168	H11a	16293G	16311C	195C	263G	309.1C	315.1C																							
ANG169	L2b1a	16114A	16129A	16213A	16223T	16278T	16355T	16362C	16390A	73G	150T	152C	182T	195C	198T	204C	263G	309.1C	315.1C	418T	523DEL	524DEL								
ANG170	L3a1a3	16093C	16223T	16278T	16294T	16390T	73G	143A	146C	152C	182T	195C	263G	309.1C	315.1C	523DEL	524DEL													
ANG171	L3e1a	16185T	16223T	16327T	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																		
ANG172	L3e3	16223T	16265T	16519C	73G	150T	195C	263G	309.1C	315.1C	499A	523DEL	524DEL																	
ANG173	T2b3	16126C	16294T	16296T	16304C	16519C	73G	151T	263G	315.1C																				
ANG174	L3e1a1	16185T	16223T	16311C	16327T	73G	150T	185A	189G	263G	315.1C																			
ANG175	L1c2	16129A	16187T	16189C	16223T	16265C	16286G	16294T	16311C	16360T	16519C	16527T	73G	151T	152C	182T	186A	189C	195C	198T	200G	247A	263G	297G	315.1C	316A				
ANG176	L2a1	16129A	16223T	16234T	16249C	16278T	16294T	16295T	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C													

Capítulo VI - Estudo de imigrantes de Moçambique: The immigrant population from Mozambique in Lisbon: Updated mitochondrial DNA portrait

Proença de Campos M, Afonso Costa H, Vieira da Silva C, Bogas V, Ribeiro T, Porto MJ, Amorim A. ***The immigrant population from Mozambique in Lisbon: Updated mitochondrial DNA portrait.*** Forensic Science International: Genetics Supplement Series 6 (2017) e298-e300.

Forensic Science International: Genetics Supplement Series 6 (2017) e298–e300

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss

The immigrant population from Mozambique in Lisbon: Updated mitochondrial DNA portrait

Marta Proença de Campos[a], Heloísa Afonso Costa[b,d], Cláudia Vieira da Silva[b], Vanessa Bogas[b], Teresa Ribeiro[b], Maria João Porto[b], António Amorim[a,b,c,*]

[a] Faculdade de Ciências da Universidade de Lisboa, Portugal
[b] Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal
[c] Instituto Superior de Ciências da Saúde Egas Moniz, Portugal
[d] Faculdade de Ciências da Saúde da Universidade da Beira Interior, Portugal

ARTICLE INFO

Keywords:
mtDNA
Haplogroups
Population genetics
Mozambican immigrants

ABSTRACT

Since the end of the 1970s Portugal had a role in the migratory movements, becoming a destiny for immigrants of a wide range of nationalities, especially from the African continent. According to statistical data, until the end of 2015, there were approximately 3000 Mozambican immigrants living in Portugal and from those, more than a half living in Lisbon metropolitan region.

Mitochondrial DNA identical sequences are shared by matrilineal inheritance. Along with the lack of recombination, it enables to trace the ancestral origin of each population and its evolutionary history. However, not only in evolutionary and population studies but also in forensic genetics, mtDNA is an important tool.

The aim of our study is the genetic characterization of Mozambican immigrants living in Lisbon in order to emphasize their genetic variability contribution to Lisbon population.

We studied blood samples from 83 Mozambican immigrants living in Lisbon. A wide range of haplotypes belonging to L, H, U, K, R, J, T and M haplogroups were founded. The highest incidence was observed for the L haplogroup (81%), often pointed as characteristic Sub-Saharan, region where Mozambique is framed. The great genetic heterogeneity for Mozambican immigrant population was highlighted in our results. Phylogenetic analysis established that the studied population is the immigrant community genetically closer to the Portuguese population.

1. Introduction

Although Portugal has been traditionally considered an emigration country, in recent decades, an intense flow of immigrants has been acknowledged, mainly due to the individuals coming from Portuguese speaking countries like Mozambique. Migration plays an important role in shaping the patterns of genetic diversity throughout the country by increasing the genetic heterogeneity [1]. Until the end of 2015, Portugal had roughly 400,000 immigrants. Approximately 3000 were from Mozambique and from those about 1700 lived in Lisbon metropolitan region [2].

In the last decades, mtDNA has been widely used as a genetic marker in different fields, namely in population studies, forensic science, medical genetics, molecular evolution and anthropology [3]. The analysis of mtDNA haplogroups provides information about geographic origin and phylogenetic relationships of different populations, being a footprint of a given population [3,4]. Macrohaplogroup L is the African specific haplogroup and it is limited to the Sub-Saharan African region [3,5]. When it is found outside of Africa, it might represent a marker of gene flow [6]. The African mtDNA landscape is established by haplogroups L, M1, from East Africa, and U6, characteristic of North Africa [7].

The aim of this study was the genetic characterization of Mozambican immigrants living in Lisbon in order to emphasize their genetic variability contribution to Lisbon population. For this, we identified the mtDNA haplotypes of Mozambican immigrants, classified the haplotypes into haplogroups and compared the studied population with other African populations as well as with Portuguese population.

2. Materials and methods

Blood samples were obtained from 83 unrelated Mozambican immigrants currently living in Lisbon, acquired through kinship investigations carried out in Forensic Genetics and Biology Service at the

* Corresponding author at: Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal.
E-mail address: antonio.j.amorim@inmlcf.mj.pt (A. Amorim).

http://dx.doi.org/10.1016/j.fsigss.2017.09.134
Received 1 September 2017; Accepted 19 September 2017
Available online 20 September 2017
1875-1768/

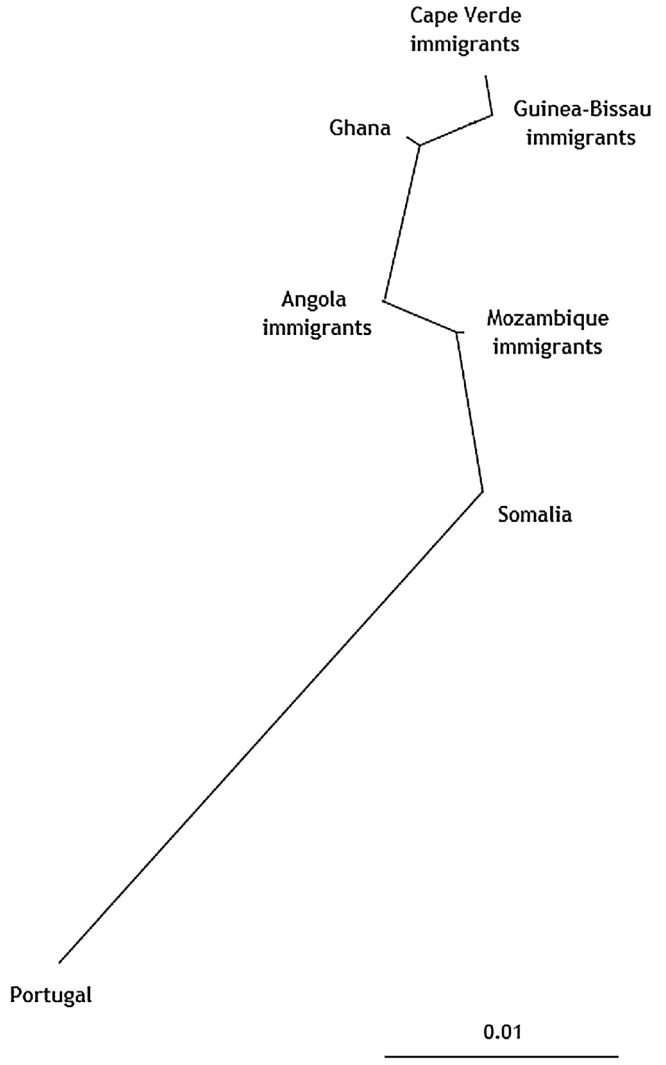


Fig. 1. Graphical representation of genetic distances between Mozambican immigrant population and other selected populations.

Instituto Nacional de Medicina Legal e Ciências Forenses – Delegação Sul (INMLCF – DS). DNA extraction was performed using Chelex® 100 method [8]. Total mtDNA control region sequences were amplified using two sets of primers – L15971/H016 and L16555/H069–comprising regions with a total of 1122 bp (16,024–16,569 and 1–576). In order to define haplotypes, the obtained sequences were compared with the revised Cambridge Reference Sequence (rCRS) [9] applying the nomenclature guidelines used in mtDNA analysis. The haplogroup classification was carried out using Phylotree build 17 tool [3]. The Arlerquin v.3.5.2.2 software [10] was used to calculate the number of polymorphic sites, haplotype diversity and nucleotide diversity, as well as to obtain F_{ST} genetic distances among studied population and other populations [11–15]. The phylogenetic representation from the genetic distances was performed using Neighbour method of Phylip v.3.695 [16] software and with Treeview v.1.6.6 software [17].

3. Results and discussion

The 83 mitochondrial haplotypes were submitted and accepted by the EMPOP database [18] with access number EMP00681. In this study, 74 different haplotypes were identified, 68 of which being unique (82%). This high genetic heterogeneity was already described in literature for different African populations including the studied population [11–14].

From the 74 haplotypes observed in this study, 50 different haplogroups were identified. The L macrohaplogroup was the most frequent haplogroup found in Mozambican immigrants, with a total of 67 sequences (81%). This is in accordance with the expected, since this macrohaplogroup is characteristic of Sub-Saharan regions. Similar results for Mozambican population have already been suggested before by Pereira [19] and Salas [20]. The remaining sequences (19%) were included in non-characteristic haplogroups from African regions. The H haplogroup, predominant in European populations, was present in 6 sequences and 8 sequences were distributed by U and K haplogroup, with 4 sequences each. Four sequences belonging to the haplogroups J1, M4, R0 and T2 were also identified. The graphical representation (Fig. 1) shows the establishment of two distinct groups: one composed by the Portuguese population and another, more distant, by the African populations, as expected. The population under study is genetically closer to the Angolan immigrants living in Lisbon (Fst = 0.01273; ρ = 0.00000) and furthest to the Portuguese population (Fst = 0.11815; ρ = 0.00000). These results are in agreement with previously published studies of other genetic markers – InDels [21].

4. Conclusion

The present study revealed a great genetic diversity for the Mozambican immigrant population, owing to the high frequency of unique haplotypes. This fact suggests that their integration among Lisbon population is going to lead to an increase in the genetic variability in this region. In this study, we described that the majority of haplotypes identified (81%) fit into the haplogroup characteristic of Sub-Saharan African regions, the L macrohaplogroup, where Mozambique is framed. After a phylogenetic analysis, we highlighted a great distance between the studied population and Portuguese population, in contrast to a smaller distance to Angolan immigrant population.

Conflict of interest

None.

Acknowledgment

None.

References

[1] C. Rego, M.F. Mendes, J. Rebelo, et al., Perfil dos imigrantes em Portugal: por países de origem e regiões de destino, Revista Portuguesa de Estudos Regionais 24 (2010).
[2] PORDATA, Base de Datos Portugal Contemporâneo, 2017.
[3] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2009) 386–394.
[4] A. Rosa, A. Brehm, African human mtDNA phylogeography at-a-glance, J. Anthropol. Sci. 89 (2011) 25–58.
[5] M.K. Gonder, H.M. Mortensen, F.A. Reed, et al., Whole-mtDNA genome sequence analysis of ancient african lineages, Mol. Biol. Evol. 24 (2007) 757–768.
[6] T. Kivisild, Maternal ancestry and population history from whole mitochondrial genomes, Investig. Genet. 6 (2015) 3.
[7] S. Plaza, A. Salas, F. Calafell, et al., Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, Hum. Genet. 115 (2004) 439–447.
[8] P.S. Walsh, D.a. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques 54 (2013) 506–513.
[9] R.M. Andrews, I. Kubacka, P.F. Chinnery, et al., Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.
[10] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resources 10 (2010) 564–567.
[11] M. Mikkelsen, L. Fendt, A.W. Röck, et al., Forensic and phylogeographic characterisation of mtDNA lineages from Somalia, Int. J. Legal Med. 126 (2012) 573–579.
[12] L. Fendt, A. Röck, B. Zimmermann, et al., MtDNA diversity of Ghana: a forensic and phylogeographic view, Forensic Sci. Int. Genet. 6 (2012) 244–249.
[13] P. Morais, A. Amorim, C. Vieira Da Silva, et al., Genetic portrait of Lisboa immigrant

population from Cabo Verde with mitochondrial DNA analysis, J. Genet. 94 (2015) 509–512.
[14] F. Simão, H.A. Costa, C.V. Da Silva, et al., Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA, Forensic Sci. Int. Genet. 15 (2014) 33–38.
[15] S.L. Marques, A. Goios, A.M. Rocha, et al., Portuguese mitochondrial DNA genetic diversity—an update and a phylogenetic revision, Forensic Sci. Int. Genet. 15 (2015) 27–32.
[16] J. Felsenstein, PHYLIP (Phylogeny Inference Package) Version 3.695. Distributed by the Author, Department of Genome Sciences, University of Washington, Seattle, 2013.
[17] R.D.M. Page, TreeView: an application to display phylogenetic trees on personal computers, Comput. Appl. Biosci. 12 (1996) 357–358.
[18] W. Parson, A. Dür, EMPOP—a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88–92.
[19] L. Pereira, V. Macaulay, A. Torroni, et al., Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade, Ann. Hum. Genet. 65 (2001) 439–458.
[20] A. Salas, M. Richards, T. De la Fe, et al., The making of the African mtDNA landscape, Am. J. Hum. Genet. 71 (2002) 1082–1111.
[21] A. Inácio, H.A. Costa, C.V. da Silva, et al., Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa, Int. J. Legal Med. 131 (2017) 657–660.

Obtivemos haplótipos individuais completos para todos os imigrantes oriundos de Moçambique estudados, que apresentamos na tabela 9.

Através da elevada percentagem de haplótipos únicos identificados na população imigrante moçambicana, num total de 68 haplótipos em 83 estudados (82%), demonstra-se a grande variabilidade genética, já descrita por outros autores, para populações africanas.

A análise das sequências de DNAmt permitiu detetar um total de 1295 alterações nucleotídicas, que correspondem a 127 posições polimórficas. Verificou-se que as substituições foram a alteração mais frequente (77%) e dentro destas, as que tiveram uma incidência maior foram as transições entre pirimidinas (50%). Este elevado número de alterações nucleotídicas também confirma dados anteriores relativos à grande heterogeneidade genética observada em populações africanas.

Foi na região HVII que se verificou a alteração mais frequente do estudo: a inserção de uma citocina na posição 315 (315.1C), com uma taxa de 100%. Este resultado vem confirmar estudos anteriormente realizados para várias populações, incluindo as africanas. Outras alterações detetadas e também comuns foram os polimorfismos 263G (93%), 16223T (81%) e 16519C (60%).

As alterações 16187T 16188G 16189C, identificadas em 17% das sequências, constituem o conjunto dos polimorfismos mais frequente que se observou no presente estudo.

A grande parte dos haplótipos determinados neste estudo, sensivelmente 81%, enquadrou-se em haplogrupos caraterísticos africanos e típicos das regiões subsarianas. Os mais abundantes foram os haplogrupos L0, L2 e L3, sendo o L2 o haplogrupo que apresentou maior expressão,

contabilizando cerca de 29% de todas as amostras. No gráfico 3 apresentamos a distribuição dos imigrantes de Moçambique pelos diferentes haplogrupos do macrohaplogrupo L. Estes resultados estão de acordo com estudos realizados anteriormente para a população de Moçambique.

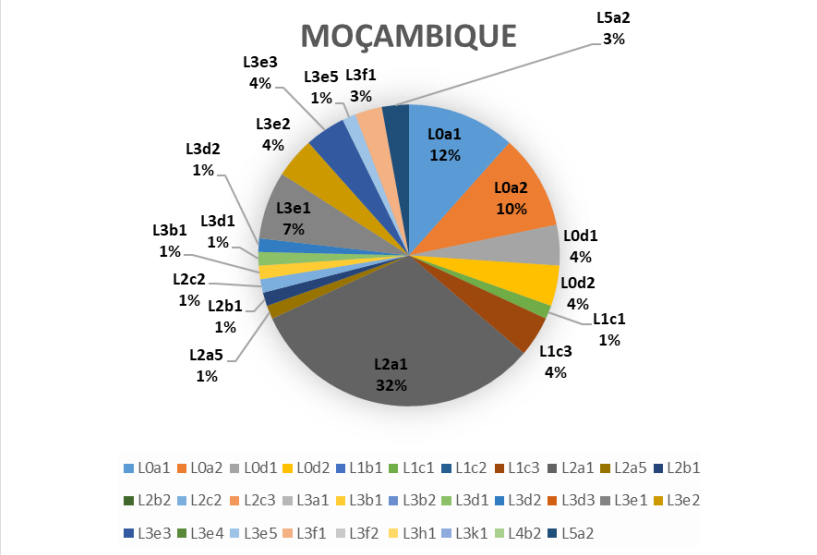


Gráfico 3 - Distribuição dos imigrantes de Moçambique pelos diferentes haplogrupos do macrohaplogrupo L

Uma pequena percentagem de haplogrupos identificados (19%) foram inseridos em haplogrupos caraterísticos de populações europeias e asiáticas, realçando a baixa contribuição destas populações no padrão genético da população imigrante moçambicana.

A análise filogenética permitiu verificar a formação de dois grandes grupos: um composto pelas populações do continente africano e outro, mais distante, composto pela população portuguesa. Através do cálculo das distâncias genéticas interpopulacionais foi possível observar que a população de imigrantes moçambicanos se encontra

geneticamente mais próxima da população de imigrantes angolanos a residir em Lisboa (Fst= 0.01273).

Contrariamente, e como seria de esperar, a população mais distante foi a população portuguesa (Fst=0.11815).

Relativamente às diferentes comunidades de imigrantes que residem atualmente em Lisboa, verificou-se que a população em estudo foi a que apresentou menor distância genética da população portuguesa e a mais distante foi a cabo-verdiana.

No que diz respeito à população portuguesa, a população geneticamente mais próxima foi a da Somália, da parte oriental de África, e a mais distante foi a de Cabo Verde, da África ocidental.

O presente estudo revelou uma recente introdução, na população de Lisboa, de linhagens caraterísticas de comunidades africanas, que poderão levar à modulação dos padrões genéticos nesta região e, consequentemente, em Portugal, se ocorrerem miscigenações entre indivíduos da população imigrante moçambicana e indivíduos de Lisboa.

A população de imigrantes de Moçambique atualmente integrados na população de Lisboa estudada no âmbito do nosso projeto está alojada na EMPOP com *accession number* EMP00681 (https://empop.online/populations).

Tabela 9 - Haplótipos e haplogrupos dos imigrantes de Moçambique

POP001	L3e1a1	16185T	16209C	16223T	16311C	16320T	16327T	73G	150T	185A	189G	263G	315.1C															
POP002	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL					
POP003	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL			
POP005	L2a1a2	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	309.2C	315.1C												
POP006	H1b	16189C	16356C	16519C	263G	309.1C	309.2C	315.1C																				
POP007	L3e1	16223T	16327T	73G	150T	189G	263G	309.1C	315.1C																			
POP008	L2a1a2	16223T	16278T	16286Y	16294T	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C														
POP009	L2a1b1a	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C												
POP010	L0a1a2	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	93G	185A	189G	200G	236C	247A	263G	315.1C	523DEL	524DEL						
POP011	U2e1	16051G	16129C	16183C	16189C	16362C	16519C	73G	152C	217C	263G	309.1C	315.1C	340T	508G													
POP012	L1c3b1a	16129A	16163G	16187T	16189C	16209C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL			
POP013	H1b1	16183C	16189C	16356C	16362C	16519C	263G	315.1C	523DEL	524DEL																		
POP014	L0a2a2	16148T	16172C	16173T	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL				
POP015	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	204Y	236C	247A	263G	315.1C	523DEL	524DEL			
POP016	L3e1b2	16223T	16239T	16325DEL	73G	150T	185A	189G	263G	309.1C	315.1C																	
POP017	H1k	16189C	16290T	16519C	263G	309.1C	315.1C																					
POP018	L0a2a2	16093C	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL				
POP019	H2a2a	16519C	263G	309.1C	309.2C	315.1C																						
POP020	L2a1b1a	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C												
POP021	L3e1b2	16223T	16239T	16325DEL	73G	150T	185A	189G	194T	263G	309.1C	315.1C																
POP022	L3b1a3	16124C	16223T	16278T	16293G	16311C	16362C	16519C	73G	263G	309.1C	315.1C	523DEL	524DEL														
POP023	U6a	16172C	16219G	16278T	16357C	16519C	73G	263G	309.1C	315.1C																		
POP024	U5a1	16256T	16270T	16311C	16399G	73G	195C	200G	263G	315.1C																		
POP025	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	150T	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL				
POP027	L3e1a1	16185T	16223T	16311C	16317G	16327T	73G	150T	185A	189G	200G	263G	309.1C	315.1C														
POP028	L1c1	16117C	16129A	16172C	16173T	16188A	16189C	16223T	16256T	16278T	16291T	16293G	16294T	16311C	16360T	16368C	16519C	73G	151T	152C	182T	186A	195C	198T	247A	263G	297G	315.1C	316A	523DEL
POP029	L3d1a1a	16124C	16223T	16293G	16319A	73G	150T	152C	263G	309.1C	315.1C	523DEL	524DEL															
POP030	L2c2	16223T	16264T	16278T	16390A	73G	93G	146C	150T	152C	182T	195C	198T	263G	309.1C	315.1C	325T	523DEL	524DEL									
POP031	L3e1	16223T	16327T	16343G	73G	150T	189G	200G	263G	315.1C																		
POP032	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	375T	523DEL	524DEL		
POP033	L2a1b1a	16182C	16183C	16189C	16192T	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C										
POP034	L2a1b1a	16182C	16183C	16189C	16192T	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C											
POP035	L3d2b	16111T	16124C	16223T	73G	152C	199C	263G	315.1C	523DEL	524DEL																	
POP036	L2a1a2	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C													
POP037	L5a1	16129A	16148T	16166G	16183DEL	16187T	16189C	16192T	16223T	16278T	16311C	16355T	16362C	73G	152C	182T	247A	263G	315.1C	444G	455.1T	459.1C	523DEL	524DEL				
POP038	L2a1a2	16223T	16278T	16286T	16291T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C	523DEL	524DEL											
POP039	L0d2c	16129A	16187T	16189C	16223T	16230G	16243C	16311C	16344T	16519C	73G	146C	152C	195C	247A	249DEL	294A	315.1C	523DEL	524DEL	498DEL							
POP041	H3c3	16519C	260A	263G	309.1C	309.2C	315.1C	524.1A	524.2C																			
POP042	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL			
POP043	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	315.1C																
POP044	L2a1a1	16129A	16223T	16278T	16294T	16309G	16368C	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C	524.1A	524.2C	524.3A	524.4C								
POP045	L0d1c	16187T	16189C	16223T	16230G	16234T	16243C	16311C	16519C	73G	146C	152C	195C	247A	315.1C	456T	498DEL											
POP046	L2b	16114A	16129A	16213A	16223T	16274A	16278T	16390A	73G	146C	150T	152C	182T	183G	195C	198T	204C	263G	309.1C	315.1C								
POP047	L0a2a2	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL					
POP048	K1a4a1a	16188T	16224C	16311C	73G	195C	263G	315.1C	497T	524.1A	524.2C																	
POP049	L3e3	16093C	16265T	16519C	73G	150T	195C	263G	315.1C	523DEL	524DEL																	
POP050	L3f1b4a	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																
POP051	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL			
POP052	L0a2	16093C	16148T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16362C	16519C	64T	93G	152C	189G	236C	247A	263G	315.1C	523DEL	524DEL					
POP053	H1e1a6	16311C	16519C	150T	263G	315.1C																						
POP054	L2a1	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C												
POP055	J1c2	16069T	16126C	16519C	73G	185A	188G	228A	263G	295T	309.1C	315.1C	462T	489C														
POP056	K2a	16224C	16278T	16311C	16519C	73G	146C	152C	263G	309.1C	315.1C	523DEL	524DEL															
POP057	L2a1b1a	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C												
POP058	L5a2	16129A	16148T	16166G	16183DEL	16187T	16189C	16192T	16223T	16278T	16311C	16355T	16362C	73G	152C	182T	247A	263G	315.1C	444G	455.1T	459.1C	523DEL	524DEL	527T			
POP059	M4a	16145A	16176T	16179T	16223T	16261T	16266T	16291T	16311C	16519C	73G	263G	315.1C	489C														
POP060	L2a1a2	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C														
POP061	L2a1a2	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C													
POP062	L2a1b1a	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C											
POP063	L0a2a2	16148T	16172C	16187T	16188A	16189C	16223T	16230G	16311C	16320T	16519C	64T	93G	152C	189G	204C	207A	236C	247A	263G	315.1C	523DEL	524DEL					
POP064	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	315.1C	523DEL	524DEL				
POP065	L3e1b2	16168T	16223T	16239T	16325DEL	73G	150T	185A	189G	263G	309.1C	315.1C																
POP068	L2a1a2	16223T	16278T	16286T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C														
POP069	L0d2a1	16129A	16187T	16189C	16212G	16223T	16230G	16243C	16311C	16390A	16519C	73G	146C	152C	195C	198T	247A	309.1C	315.1C	498DEL	523DEL	524DEL						
POP070	L0a1b1	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16278T	16293G	16311C	16320T	93G	95C	185A	189G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL			
POP071	U5b1b1	16183C	16189C	16270T	73G	150T	263G	315.1C																				
POP072	L2a1b1a	16129A	16182C	16183C	16189C	16192Y	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	315.1C										
POP073	L0d1b2b2	16129A	16187T	16189C	16223T	16239T	16243C	16294T	16311C	16519C	73G	146C	152C	195C	247A	315.1C	498DEL	523DEL	524DEL	573.1C	573.2C	573.3C	573.4C	573.5C				
POP074	L3e3	16223T	16265T	16519C	73G	150T	195C	263G	315.1C	523DEL	524DEL																	
POP075	L0d2a	16129A	16187T	16189C	16212G	16223T	16230G	16243C	16311C	16390A	16519C	73G	146C	152C	195C	247A	309.1C	315.1C	498DEL	523DEL	524DEL							
POP076	T2b3	16126C	16294T	16296T	16304C	16519C	73G	151T	263G	315.1C																		
POP077	L2a1b1a	16182C	16183C	16189C	16192T	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C										
POP078	L2a1b1a	16129R	16182C	16183C	16189C	16223T	16278T	16290T	16294T	16309G	16390A	73G	146C	152C	195C	263G	309.1C	315.1C										
POP079	L0d1b2b2	16129A	16187T	16189C	16223T	16239T	16243C	16294T	16311C	16519C	73G	146C	152C	195C	247A	315.1C	498DEL	523DEL	524DEL	573.1C	573.2C	573.3C						
POP080	L2a5	16223T	16224C	16256T	16278T	16309G	16390A	16519C	73G	146C	152C	182T	263G	315.1C	511T													
POP081	L2a1a3c	16093C	16223T	16256T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	263G	309.1C	315.1C											
POP082	L3e3a	16223T	16265T	16519C	73G	150T	195C	263G	309.1C	315.1C	523DEL	524DEL	573.1C	573.2C	573.3C	573.4C												
POP083	L3f1b4a	16209C	16223T	16311C	16519C	73G	150T	189G	200G	263G	309.1C	315.1C																
POP084	L2a1a2	16092C	16223T	16278T	16286T	16294T	16309G	16355T	16390A	16519C	73G	146C	152C	195C	263G	315.1C												
POP085	L3e5	16041G	16223T	16355T	16519C	73G	150T	263G	315.1C	398C	480C	523DEL	524DEL															
POP086	H1+152	16519C	152C	263G	309.1C	315.1C																						
POP087	L1c3a	16129A	16183C	16189C	16215G	16223T	16278T	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	309.1C	315.1C	316A	408A	494T	523DEL	524DEL		
POP088	L2a1	16189C	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C														

Capítulo VII - Estudo de imigrantes da Guiné-Bissau:

Updated mtDNA study of Guinea-Bissau immigrant population living in Lisbon

Amorim A, Ribeiro J, Vieira da Silva C, Bogas V, Ribeiro T, Porto MJ, Afonso Costa H. ***Updated mtDNA study of Guinea-Bissau immigrant population living in Lisbon.*** Forensic Science International: Genetics Supplement Series 6 (2017) e329-e331.

Forensic Science International: Genetics Supplement Series 6 (2017) e329–e331

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss

Updated mtDNA study of Guinea-Bissau immigrant population living in Lisbon

António Amorim[a,*], Joana Ribeiro[b], Cláudia Vieira da Silva[a], Vanessa Bogas[a], Teresa Ribeiro[a], Maria João Porto[a], Heloísa Afonso Costa[a]

[a] Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal
[b] Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

ARTICLE INFO

Keywords:
mtDNA haplotype
Haplogroups
Guinea-Bissau
Lisboa
Immigrants

ABSTRACT

Guinea-Bissau is a country on the west coast of African Continent.Before the arrival of Europeans and until the XVII century, the territory was part of the Gabu kingdom, Mali Empire.

In the past decades the number of immigrants in Portugal has grown and Guinean immigrant community is one of the largest immigrant communities. Most of those immigrants are based in Lisbon metropolitan region.

Over the last decades, mtDNA has become an important genetic marker in human and population evolution studies, as well as in forensic and clinical case studies. Maternal inheritance, high copy number, lack of recombination and high mutation rateare characteristics that makes this genetic marker useful instudies of human origin and evolution.The non-coding mtDNAregion presents high genetic variability and the haplotypes obtained through the study of this region fall into haplogroups with specific polymorphisms shared by all individuals with a common ancestor.

We studied 80 blood samples from Guinean immigrants living in Lisbon. mtDNA control region was amplified and sequenced between positions 16024 and 576, using two sets of primers - L15971/H016 and L16555/H639 -.

The study of the control region revealed high genetic diversity within Guinean immigrants, with a high frequency of unique haplotypes. Most of the obtained mitochondrial DNAsequences belong to African haplogroups.

As the result of the integration of African immigrants in Lisbon population we will have, in a near feature, individuals borned in Portugal with Portuguese nationality exhibiting mtDNA haplotypes typical of regions of the African Continent.

1. Introduction

Since the end of the 1970s Portugal had an important role in migratory movements, becoming a destiny for immigrants of a wide range of nationalities, mainly from African countries. According to PORDATA, until the end of 2014, there were approximately 40 000 immigrants from Cape Verde, 20 000 immigrants from Angola, 18 000 immigrants from Guinea-Bissau and 3 000 immigrants from Mozambique living in Portugal, and from those, more than 80% living in Lisbon region [1]. This reality can be one of the main contributors for genetic variation of Lisbon population at the present and in the future.

Mitochondrial DNA (mtDNA) has certain features that make it desirable for forensics, namely, high copy number, lack of recombination, and matrilineal inheritance. These mtDNA features are also important in evolutionary and population studies [2,3].

We aim to characterize mtDNA of immigrants from Guinea-Bissauliving in Lisbon and their potential contribution to genetic variation of Lisbon population.

2. Material and methods

DNA extraction of blood samples was performed with Chelex®100 method [4]. mtDNA total control region was amplified with primers L15971/H016 and primers L16555/H639. ExoSap-IT® was used to purify the amplified product. Direct and reverse sequencing was performed with BigDye® Terminator v.3.1 Cycle Sequence (AB). BigDye®XTerminator Purification Kit (AB) was used to purify the sequenced product. Sequenced products were detected with electrophoresis at genetic Analyser 3130® (ABI Prism®). Sequence analysis was performed with the software sequencing analysis v.5.2 (AB) [5,6] and sequence comparison with rCRS was performed with software SeqScape v.3 (AB) [5,6]. Haplogroup determination was

* Corresponding author.
E-mail address: antonio.j.amorim@inmlcf.mj.pt (A. Amorim).

http://dx.doi.org/10.1016/j.fsigss.2017.09.122
Received 3 September 2017; Accepted 19 September 2017
Available online 20 September 2017
1875-1768/

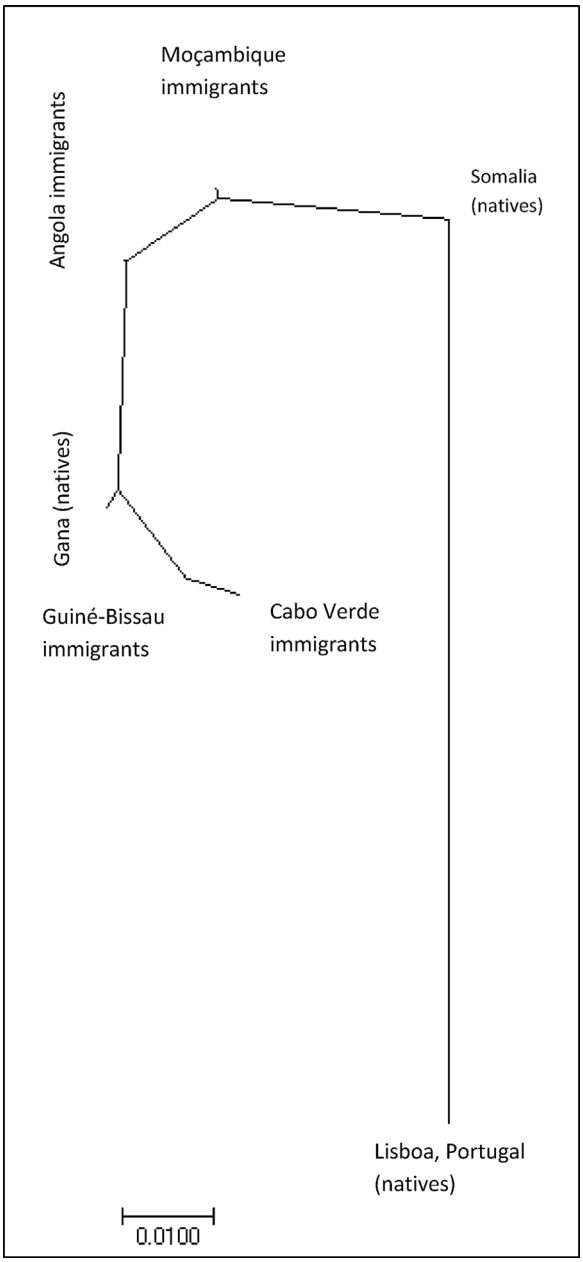


Fig. 1. Graphical representation of genetic distances between Guinean immigrant population and other populations.

achieved with software Phylotree build 17 [7]. Genetic distances were calculated with software Arlequin ver.3.5 and graphic/tree representation was performed with software MEGA7 [8].

3. Results

For each sample, the complete sequence of the total control region was obtained. The comparison of the obtained sequences with the Revised Cambridge Reference Sequence (rCRS), among the 80 analysed individuals, allowed the identification of 66 different haplotypes, 54 of which are unique. The other haplotypes were present in more than one individual. The 66determined haplotypes follow into 4 different macrohaplogroups-L, U, M, V-.The most common haplogroup is the haplogroup L with 73 haplotypes. Haplogroup U has5 haplotypes. The haplogroupsM and V appear only once.

The 80 mitochondrial haplotypes were submitted for insertion atthe EMPOP database [9].

At Fig. 1, the optimal tree with the sum of branch length = 0.16763094 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary history was inferred using the Neighbor-Joining method [10]. Evolutionary distances between all populations show the establishment of two distinct groups: one composed by the Portuguese population and another, more distant, by the African populations, as expected. Comparing the different immigrant populations living in Lisbon, the genetically closest community of Portuguese population is Mozambique and the furthest is Cape Verde, followed by Guinea-Bissau and Angola.

EMPOP has a tool that provides a summary table of all matching haplotypes that meet anyinserted sequence. Using this EMPOP matching tool we observe that 38 (47.5%) of our Guinea-Bissau immigrant haplotypes had no correspondence/match with the 26127 haplotypes present at EMPOP database.

4. Discussion and conclusions

Guinea-Bissau immigrants present high number of unique haplotypes, most of them belonging to macrohaplogroup L from Sub-Saharan regions of Africa. This haplogroup is uncommon in European and Portuguese populations.

Our results indicate that the Guinea-Bissau immigrants introduce high genetic diversity in Lisbon population not only due to high number of unique haplotypes of the studied populations but regarding to high genetic distances between all studied populations as show in the presented phylogenetic tree representative of genetic distances.Our results suggest that the integration of Guinea-Bissau immigrants among Lisbon population is going to lead to an increase in the genetic variability in this region. Therefore, it is very important to study the immigrant communities currently living in Lisbon.

The high level of determined Guinea-Bissau immigrant haplotypes with no correspondence/match at EMPOP database emphasizes the importance of characterize mtDNA of all new individuals of any population and their potential contribution to introduce genetic variation at Lisbon population.

Our results are in agreement with previously published studies with focuson mtDNA as well as in other genetic markers [11–16].

References

[1] PORDATA. Base de Dados Portugal Contemporâneo, 2017. Available in: www.pordata.pt.
[2] M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and the origin of modern humans, Nature 408 (6813) (2000) 708–713.
[3] J.C. Alvarez, D.L.E. Johnson, J.A. Lorente, E. Martinez-espin, L.J. Martinez-Gonzalez, M. Allard, M.R. Wilson, B. Budowle, Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set, Legal Med. 9 (6) (2007) 293–304.
[4] P.S. Walsh, D.A. Metzger, R. Higuchi, Biotechniques 30th anniversary gem Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques 54 (3) (2013) 506–513.
[5] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (2) (1999) 147.
[6] W. Bär, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, M. Holland, P.J. Lincoln, W. Mayr, N. Morling, B. Olaisen, P.M. Schneider, G. Tully, M. Wilson, Guidelines for mitochondrial DNA typing. DNA commission of the international society for forensic genetics, Vox Sang. 79 (2) (2000) 121–125.
[7] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2) (2009) 386–394.
[8] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.
[9] W. Parson, A. Dür, EMPOP – a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88–92.
[10] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[11] A. Inácio, H.A. Costa, C.V. da Silva, et al., Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa, Int. J. Legal Med. 131 (2017) 657–660.
[12] A. Rosa, A. Brehm, African human mtDNAphylogeography at-a-glance, J. Anthropol. Sci. 89 (2011) 25–58.

[13] M.K. Gonder, H.M. Mortensen, F.A. Reed, et al., Whole-mtDNA genome sequence analysis of ancient african lineages, Mol. Biol. Evol. 24 (2007) 757–768.
[14] L. Fendt, A. Röck, B. Zimmermann, et al., MtDNA diversity of Ghana: a forensic and phylogeographic view, Forensic Sci. Int. Genet. 6 (2012) 244–249.
[15] P. Morais, A. Amorim, C. Vieira Da Silva, et al., Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis, J. Genet. 94 (2015) 509–512.
[16] F. Simão, H.A. Costa, C.V. Da Silva, et al., Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA, Forensic Sci. Int. Genet. 15 (2014) 33–38.

Estudamos 80 amostras de sangue correspondentes a 80 imigrantes nascidos e com ascendência na Guiné-Bissau atualmente integrados na população de Lisboa. Obtivemos a sequência completa de toda a região controlo do DNAmt para todos os imigrantes incluídos no estudo e que apresentamos na tabela 10.

A comparação das sequências obtidas com a sequência de referência de Cambridge revista, permitiu identificar 66 haplótipos diferentes. Destes 66 haplótipos diferentes, 54 são haplótipos com uma única representação na população. Os restantes surgiram em mais do que um imigrante.

Os 66 haplótipos identificados incluem-se em 4 macrohaplogrupos - L, U, M, V -. O macrohaplogrupo com maior número de imigrantes foi o macrohaplogupo L com 73 indivíduos. No gráfico 4 apresentamos a distribuição dos imigrantes da Guiné-Bissau pelos diferentes haplogrupos do macrohaplogrupo L.

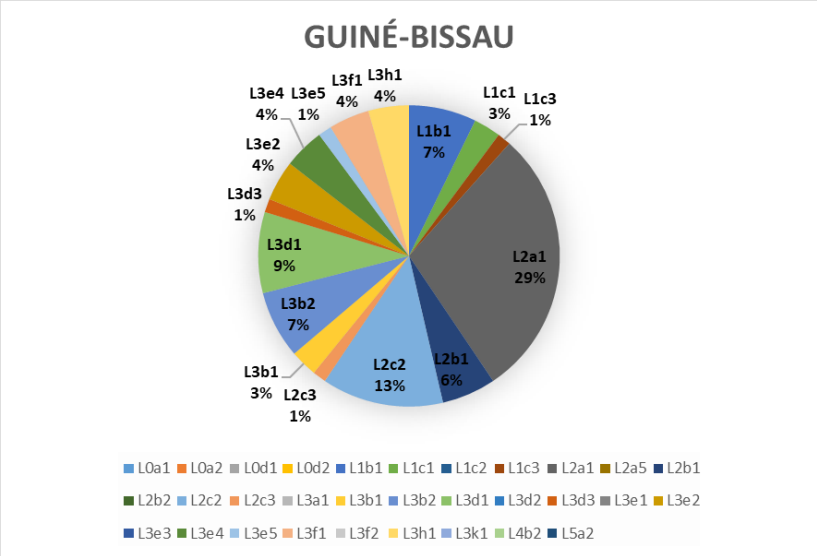


Gráfico 4 - Distribuição dos imigrantes da Guiné-Bissau pelos diferentes haplogrupos do macrohaplogrupo L

É possível confirmar a baixa influência europeia na constituição do padrão genético de DNAmt dos indivíduos da Guiné-Bissau a residir em Lisboa.

O elevado número de posições polimórficas e respetivos polimorfismos observados, confirma dados anteriores relativos à elevada heterogeneidade genética presente em populações do continente Africano. A grande percentagem de haplótipos únicos presente do estudo realizado é também característica de populações Africanas.

Através da obtenção das distâncias genéticas entre a população dos imigrantes da Guiné-Bissau a residir em Lisboa com outras populações Africanas, confirmou-se a maior proximidade deste povo imigrante com outros indivíduos Africanos.

Relativamente às diferentes comunidades de imigrantes que residem atualmente em Lisboa, verificou-se que a população em estudo foi a que apresentou menor distância genética aos imigrantes cabo-verdianos. No que diz respeito à população portuguesa, a população geneticamente mais próxima foi a da Somália, da parte oriental de África, e a mais distante foi a de Cabo Verde, da África ocidental.

A população de imigrantes da Guiné-Bissau atualmente integrados na população de Lisboa estudada no âmbito do nosso projeto está alojada na EMPOP com *accession number* EMP00704 (https://empop.online/populations).

Tabela 10 - Haplótipos e correspondentes haplogrupos dos imigrantes da Guiné-Bissau

GUI001	L3e4a	16051G	16185G	16223T	16264T	16519C	73G	150T	263G	309.1C	315.1C	523DEL	524DEL																	
GUI002	L3e2	16192T	16223T	16320T	16519C	73G	150T	152C	194T	195C	263G	315.1C																		
GUI003	U6a	16172C	16219G	16278T	16519C	73G	263G	309.1C	315.1C																					
GUI004	L2a1l	16189C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	534T															
GUI005	L3e2b	16172C	16182C	16183C	16189C	16223T	16320T	73G	150T	195C	263G	315.1C																		
GUI006	L2a1a3	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C	523DEL	524DEL														
GUI007	L3b2a	16124C	16183C	16187A	16189C	16223T	16278T	16362C	16527T	73G	263G	309.1C	315.1C	523DEL	524DEL															
GUI008	L2a1l	16189C	16223T	16274A	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	198T	263G	264T	315.1C	534T											
GUI009	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL								
GUI010	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL								
GUI011	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL								
GUI012	L2c	16223T	16278T	16291T	16390A	16519C	73G	89C	93G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL										
GUI013	L2a1e1	16223T	16230G	16278T	16294T	16390A	16519C	73G	146C	152C	195C	263G	315.1C																	
GUI014	L2c3	16223T	16278T	16390A	73G	93G	95C	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	513A	523DEL	524DEL											
GUI015	L2c	16223T	16278T	16390A	16519C	73G	93G	150T	152C	182T	195C	198T	263G	309.1C	309.2C	315.1C	325T	499A	523DEL	524DEL										
GUI016	L2a1a3	16093C	16189C	16223T	16264T	16278T	16294T	16309G	16390A	16519C	73G	143A	146C	152C	195C	263G	309.1C	315.1C	534T											
GUI017	L3d1b2	16124C	16223T	16274A	16399G	73G	150T	152C	263G	315.1C	523DEL	524DEL																		
GUI018	L2a1c	16086C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
GUI019	L3e2b	16172C	16183C	16189C	16223T	16320T	16519C	73G	150T	195C	263G	309.1C	315.1C																	
GUI020	L3d1b2	16124C	16223T	16311C	16399G	73G	150T	152C	263G	315.1C	523DEL	524DEL																		
GUI021	L2a1	16189C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL	534T													
GUI022	L2b	16114A	16129A	16213A	16223T	16278T	16390A	73G	146C	150T	152C	182T	185C	195C	198T	204C	207A	263G	309.1C	315.1C										
GUI023	L2b1a	16046C	16114A	16129A	16213†	16223T	16278T	16355T	16362C	16390A	73G	150T	152C	195C	198T	204C	263G	315.1C	418T	523DEL	524DEL									
GUI024	L1c3b	16111T	16129A	16163G	16187T	16189C	16223T	16278T	16293G	16294T	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL					
GUI025	U6a	16172C	16219G	16278T	16519C	73G	183R	263G	309.1C	315.1C																				
GUI026	M10	16129A	16185T	16189C	16193DEL	16223T	16249C	16311C	16519C	73G	195C	263G	315.1C	466C	489C															
GUI027	L3f1b	16209C	16223T	16224C	16292T	16311C	16519C	73G	189G	200G	263G	315.1C																		
GUI028	L1b	16093C	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL							
GUI029	L1b1a1´4	16114A	16126C	16189C	16223T	16264T	16270T	16278T	16293G	16311C	16399G	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523DEL	524DEL							
GUI030	U6a3f	16172C	16183C	16189C	16219G	16278T	73G	150T	185A	189G	263G	309.1C	315.1C																	
GUI032	L2c2b	16093C	16223T	16264T	16278T	16390A	73G	89C	93G	95C	146C	150T	152C	182T	183G	195C	198T	204C	263G	315.1C	325T	523DEL	524DEL							
GUI033	L2a1l	16189C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL	534T													
GUI034	L2c	16223T	16278T	16390A	16519C	73G	93G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL												
GUI035	L2a1b	16189C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C															
GUI036	L3f1b	16086C	16209C	16223T	16259T	16292T	16311C	16519C	73G	189G	200G	263G	315.1C																	
GUI037	L3h1b2	16129A	16223T	16256A	16311C	16362C	16519C	73G	151T	152C	189C	195C	263G	294C	315.1C	499A	523DEL	524DEL												
GUI038	L3h1b2	16129A	16223T	16256A	16311C	16362C	16519C	73G	151T	152C	189C	195C	263G	294C	315.1c	499A	523DEL	524DEL												
GUI039	L3f1b1	16209C	16223T	16292T	16295T	16311C	16355T	16519C	73G	189G	200G	263G	315.1C																	
GUI040	L0a1a	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16360T	64T	93G	152C	185A	189G	200G	236C	247A	263G	315.1C	523DEL	524DEL					
GUI041	U6a	16172C	16219G	16278T	16519C	73G	263G	309.1C	315.1C																					
GUI042	L2a1l	16189C	16192T	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	534T														
GUI043	L3b2a	16124C	16183C	16189C	16234T	16223T	16278T	16362C	16527T	73G	263G	315.1C	315.2C	523DEL	524DEL															
GUI044	L2a1b	16183C	16189C	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL	534T												
GUI045	L0a1a	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	64T	93G	150T	152C	185A	189G	200G	236C	247A	263G	309.1C	315.1C	523DEL	524DEL				
GUI046	L2a'b'c'd	16177G	16223T	16278T	16311C	16390A	16519C	73G	146C	150T	152C	182T	195C	263G	309.1C	315.1C	325T													
GUI047	L2a1c3b	16201T	16223T	16274A	16278T	16294T	16355T	16390A	73G	143A	146C	152C	195C	263G	315.1C															
GUI048	U5b1b	16189C	16192T	16270T	16320T	73G	150T	152C	263G	309.1C	315.1C																			
GUI049	L1c	16129A	16183C	16189C	16223T	16278T	16294T	16360T	16519C	73G	151T	152C	182T	186A	189C	247A	263G	315.1C	316A	523DEL	524DEL									
GUI050	L2c	16051G	16223T	16234T	16278T	16390A	73G	93G	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL												
GUI052	L3d1b1b	16124C	16223T	16257T	16519C	73G	151T	152C	183G	263G	309.1C	315.1C	523DEL	524DEL																
GUI053	L3d1b2	16124C	16223T	16311C	16399G	73G	150T	152C	263G	309.1C	315.1C	523DEL	524DEL																	
GUI054	HV0	16189C	16298C	16530T	72C	263G	315.1C																							
GUI055	L3b2	16124C	16182C	16183C	16189C	16214T	16223T	16278T	16362C	16527T	73G	263G	315.1C	523DEL	524DEL															
GUI056	L2c	16223T	16278T	73G	93G	146C	150T	152C	182T	185C	189G	195C	263G	309.1C	315.1C	325T	523DEL	524DEL												
GUI058	L2a'b'c'd	16177G	16223T	16278T	16311C	16390A	16519C	73G	146C	150T	152C	182T	195C	263G	309.1C	315.1C	325T													
GUI059	L2a1	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C																	
GUI060	L2c	16223T	16261T	16278T	16390A	73G	146C	150T	152C	195C	198T	263G	315.1C	325T	523DEL	524DEL														
GUI061	L2b1a	16114A	16129A	16213A	16223T	16278T	16290T	16355T	16362C	16390A	73G	150T	152C	182T	195C	198T	204C	263G	315.1C	418T	523DEL	524DEL								
GUI062	L2a1a3	16111T	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
GUI063	L3d	16124C	16223T	16241G	16519C	16549DEL	73G	152C	263G	315.1C	438T																			
GUI064	L3e4a	16051G	16223T	16264T	16519C	73G	150T	263G	309.1C	315.1C	523DEL	524DEL																		
GUI065	L3h1b2	16129A	16223T	16256A	16311C	16362C	16519C	73G	151T	152C	189C	195C	263G	294C	315.1C	499A	523DEL	524DEL												
GUI066	L1c1	16129A	16187T	16189C	16223T	16256T	16261T	16278T	16293G	16311C	16360T	16519C	73G	151T	152C	182T	186A	189C	195C	198T	247A	249DEL	263G	297G	315.1C	316A	523DEL	524DEL		
GUI067	L2b	16114A	16129A	16213A	16223T	16278T	16390A	73G	146C	150T	152C	182T	195C	198T	263G	315.1C														
GUI068	L3e2	16223T	16311C	16320T	16519C	73G	150T	152C	195C	263G	315.1C																			
GUI073	L2c	16223T	16278T	16390A	73G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL														
GUI075	L2a1	16111T	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	315.1C	523DEL	524DEL														
GUI077	L3e4	16051G	16223T	16264T	16519C	73G	150T	263G	309.1C	315.1C	523DEL	524DEL																		
GUI078	L3d3	16111T	16124C	16223T	16519C	73G	152C	199C	263G	309.1C	315.1C	523DEL	524DEL																	
GUI092	L3e5	16041G	16223T	16355T	16519C	73G	150T	263G	315.1C	398C	523DEL	524DEL																		
GUI093	L3b2	16124C	16183C	16189C	16214T	16223T	16278T	16362C	16527T	73G	263G	315.1C	315.2C	523DEL	524DEL															
GUI095	L2a1	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	263G	309.1C	315.1C	523DEL	524DEL															
GUI096	L2c*	16223T	16278T	16390A	73G	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523DEL	524DEL														
GUI098	L3e2a	16223T	16320T	16519C	73G	150T	195C	198T	263G	315.1C																				
GUI102	L3b1a	16093C	16223T	16278T	16362C	16519C	73G	263G	315.1C	523DEL	524DEL																			
GUI111	L3b1a	16093C	16223T	16278T	16362C	16519C	73G	263G	315.1C	523DEL	524DEL																			
GUI117	L2a1	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C	523DEL	524DEL														
GUI118	L3b2	16124C	16183C	16189C	16223T	16278T	16362C	16527T	73G	263G	309.1C	315.1C	523DEL	524DEL																
GUI124	L2a1*	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	315.1C	574C														
GUI125	L3d1a	16124C	16223T	16319A	73G	146C	152C	263G	315.1C	523DEL	524DEL																			

Capítulo VIII - Estudo dos imigrantes PALOP com marcadores autossómicos: Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa

Inácio A, Afonso Costa H, Vieira da Silva C, Ribeiro T, Porto MJ, Costa Santos J, Igrijas G, Amorim A. ***Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa.*** International Journal of Legal Medicine (2017) 131:657-660.

Int J Legal Med (2017) 131:657–660
DOI 10.1007/s00414-016-1484-3

POPULATION DATA

Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa

Ana Inácio[1,2] · **Heloísa Afonso Costa**[1,6] · **Cláudia Vieira da Silva**[1,6] · **Teresa Ribeiro**[1,6] · **Maria João Porto**[1,6] · **Jorge Costa Santos**[3,4,6] · **Gilberto Igrejas**[2] · **António Amorim**[1,4,5,6]

Received: 30 June 2016 / Accepted: 20 October 2016 / Published online: 29 October 2016

Abstract The migratory phenomenon in Portugal has become one of the main factors for the genetic variability. In the last few years, a new class of autosomal insertion/deletion markers—InDel—has attracted interest in forensic genetics. Since there is no data for InDel markers of Portuguese-speaking African countries (PALOP) immigrants living in Lisboa, our aim is the characterization of those groups of individuals by typing them with at least 30 InDel markers and to compare different groups of individuals/populations. We studied 454 bloodstain samples belonging to immigrant individuals from Angola, Guinea-Bissau, and Mozambique. DNA extraction was performed with the Chelex® 100 method. After extraction, all samples were typed with the Investigator® DIPplex method. Through the obtained results, allelic frequencies show that all markers are at Hardy-Weinberg equilibrium, and we can confirm that those populations show significant genetic distances between themselves, between them, and the host Lisboa population. Because of this, they introduce genetic variability in Lisboa population.

Keywords PALOP immigrants · InDel · Lisboa population

Introduction

Since the early 1970s, in Portugal, the flow of immigrants from African countries has increased. We can relate this reality with the post-colonial and post-independence period of Portuguese-speaking African countries.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-016-1484-3) contains supplementary material, which is available to authorized users.

✉ António Amorim
aamorim@dlinml.mj.pt

Ana Inácio
anainacio14@gmail.com

Heloísa Afonso Costa
hcosta@dlinml.mj.pt

Cláudia Vieira da Silva
csilva@dlinml.mj.pt

Teresa Ribeiro
tribeiro@dlinml.mj.pt

Maria João Porto
mariajoao.porto@dcinml.mj.pt

Jorge Costa Santos
costasantosjm@gmail.com

Gilberto Igrejas
gigrejas@utad.pt

[1] Instituto Nacional de Medicina Legal e Ciências Forenses, Coimbra, Portugal

[2] Departamento de Genética e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

[3] Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

[4] Instituto Superior de Ciências da Saúde Egas Moniz, Monte de Caparica, Portugal

[5] Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

[6] Centro de Ciências Forenses (CENCIFOR), Coimbra, Portugal

According to the Portugal Contemporary Base PORDATA, by the end of 2014, the total number of immigrants from Portuguese-speaking African countries (PALOP) in Portugal was about 91,000, and from those, 75,000 are part of Lisboa population.

The number of immigrants in Portugal is relevant, and the migratory phenomenon in Portugal, particularly in Lisboa, can become one of the main factors for the genetic variability [1–7].

In the last few years, a new class of autosomal insertion/deletion markers—InDel—has attracted interest in forensic genetics, mainly by their abundance in the human genome and by the simple analysis methodologies associated to their study [8]. They are characterized by the presence or absence of a specific sequence of nucleotides [9–12]. Significative differences in allele frequencies of InDel markers, between different groups or populations, can be used as ancestry and eventually evolutionary indicators [13].

InDels represent approximately 15.6 % of all the polymorphisms in human genome, suggesting that in human population, there is a minimum of 1.56 million InDels [14, 15]. Furthermore, InDels combine characteristics of short tandem repeats (STRs) and SNPs, filling some gaps of each one of these markers and introducing a new class of genetic markers for forensic purposes. Advances in forensic genetics are noticed over the past few years and show us the interest for the development of biallelic markers, such as SNPs and InDels [9, 16].

Until today, only one commercial kit is available for the study of this kind of polymorphisms—Investigator® DIPplex—developed by Qiagen. This kit allows the study of 30 InDel markers and the homologous gene of amelogenin as informative of the individual's sexual gender [12, 16, 17]. This multiplex has the advantage that amplified products can be separated and analyzed by capillary electrophoresis using an automatic sequencer, with similar technology to STR analysis.

The Investigator® DIPplex kit (Qiagen, Hilden, Germany) has already been employed in some population studies, especially in populations of European, American, and Asian origin [8–10, 12, 16–24]. Demonstrating high discrimination power and high power of exclusion has shown to be proper for forensic cases, especially in individual identification that requires analysis of degraded samples. In relation to kinship investigations, this kit can only be a complement to the STR analysis.

Since there is a reduced amount of data for InDel markers of PALOP immigrants living in Lisboa, our aim is to characterize those groups of individuals by typing them with 30 markers and compare different groups of individuals/populations.

Material and methods

Four hundred fifty-four bloodstain samples were studied, 258 from Angola, 124 from Guinea-Bissau, and 72 from Mozambique. These bloodstain samples were collected from immigrant individuals, inhabitants of Lisboa metropolitan area, undergoing forensic investigations in Instituto Nacional de Medicina Legal e Ciências Forenses (INMLCF). The number of studied individuals from each of the different African countries—Angola, Guinea-Bissau, Mozambique—represents each immigrant group within the Lisboa population.

An interview was conducted in order to register personal data of the studied individuals, particularly the name, the age, the birthplace, the individual, and the parental ethnicity. According to Portuguese legal regulations, samples from routine forensic cases ongoing at INMLCF can be used for investigation purposes, which naturally include genetic studies. In our study, all samples are used with special codification without any connection to personal or judicial data related to the donor.

DNA extraction was carried out using the Chelex®100 resin extraction method [25]. InDel typing was accomplished with Investigator® DIPplex PCR Amplification Kit (Qiagen, Hilden, Germany). DNA fragment separation, detection, and identification were achieved with capillary electrophoresis using an ABI PRISM® Genetic Analyzer 3130 xl sequencer (Applied Biosystems, Foster City, USA).

Arlequin software ver.3.5 [26] was used for the calculations of allele frequencies for each locus, as well as expected and observed heterozigosities. Beyond this, Arlequin software ver.3.5 [26] was also used to calculate Hardy-Weinberg equilibrium (HWE) and to estimate p values, which was considered to be significant at $p < 0.0001$, after Bonferroni correction [27]. Concerning forensic parameters, power of exclusion (PE), discrimination power (DP), polymorphic information content (PIC), typical paternity index (TPI), and matching probability (MP) were calculated using PowerStats software, ver.12 (Promega, UK).

Phylogenetic comparison between the three populations—Angola immigrants, Guinea-Bissau immigrants, and Mozambique immigrants—as well as others, South of Portugal [28] and Cape Verde immigrants [29], were accomplished with Arlequin software ver.3.5 and with Phylogeny Inference Package (PHYLIP) ver.3.2 [30].

Results

A sample of 454 immigrant individuals from different populations, Angola, Guinea-Bissau, and Mozambique, was studied aiming to characterize them with InDel genetic markers for future application in forensic casework samples. Supplementary Tables S1, S2, and S3 in ESM 1 show the frequencies for insertions and deletion alleles for all studied populations/groups. All studied InDel followed Hardy-Weinberg expectations ($p < 0.0001$) except the *locus* HLD97 from Guinea-Bissau immigrant population.

Expected heterozigosity (He), observed heterozigosity (Ho), PE, DP, PIC, TPI, and MP values for all populations/groups are also presented in supplementary Tables S1, S2, and S3 (ESM 1). The combined matching probability (CMP), the combined power of discrimination (CPD), and the combined power of exclusion (CPE) are presented in supplementary Table S4 (ESM 1). The achieved values for forensic parameters allow a satisfactory level of discrimination in forensic cases.

By the analysis of the obtained electropherograms, we detected microvariants in loci HLD92, HLD99, and HLD84 which influenced the interpretation of the genotypes (see ESM 2, Figs. S1, S2, and S3). The detection and confirmation of microvariants particularly in loci HLD92, HLD99, and HLD84 are an advantage of using this panel of InDel. It is important to note that microvariant of HLD92 locus is not described in the literature.

Discussion and conclusion

Through the allelic frequencies obtained for the three immigrant populations in the study, it was possible to verify the existence of genetic differences between them. Figure 1 presents the phylogenetic tree representative of those genetic distances, presented in supplementary Table S5 (ESM 1).

The populations of Angola, Guinea-Bissau, and Mozambique, before the major migration of Bantu, were inhabited by regional tribes and by Khoisan tribes [21, 22]. Thus, the great evolutionary phenomenon of the African population was due to the expansion of Bantu tribes along two main different regions—East, where Mozambique is included, and West of Africa where Angola and Guinea-Bissau are included [23, 24].

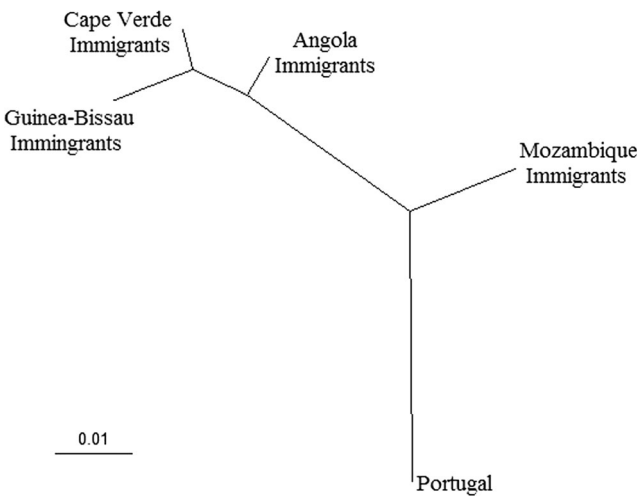


Fig. 1 Phylogenetic tree representing the genetic distances among populations under study

Concerning Cape Verde, it was discovered uninhabited and was later colonized mainly by people of the African West Coast due to the slave trade practiced by the Portuguese [25, 26]. So, it is expected that these three immigrant populations—Angola, Guinea-Bissau, and Cape Verde—are closer and Mozambique immigrants has a greater genetic distance in relation to the other three African populations [31].

The results show us that those immigrant populations of Angola, Guinea-Bissau, and Cape Verde are more approximated between them and much more separated of the Portuguese population. Mozambique immigrants are closer with Portuguese population than the other ones and can participate in an individual cluster. Angola immigrants, Guinea-Bissau immigrants, and Cape Verde immigrants are included in the other cluster and as expected taking into account the expansion of Bantu tribes along two main different regions as mentioned before.

Through the obtained results, it is possible to confirm that the studied African populations show significant genetic distances between themselves, between them, and the Lisboa population, and so, we can conclude that they introduce genetic variability in the Lisboa population.

Furthermore, the Investigator DIPplex® PCR amplification kit (Qiagen, Hilden, Germany), due to its characteristics such as short amplicon size, absence of stutters, and simplicity to implement in forensic genetic laboratories, can be very interesting as a supplement of any study of STR in forensic caseworks with the studied Portuguese and African populations.

This paper follows the guidelines for publication data request by the journal [32].

References

1. Morais P, Amorim A, Vieira da Silva C, Ribeiro T, Costa Santos J, Afonso Costa H (2015) Genetic portrait of Lisboa immigrant population from Cabo Verde with mitochondrial DNA analysis. J Genet 94(3):509–512
2. Simão F, Afonso Costa H, Vieira da Silva C, Ribeiro T, Porto MJ, Costa Santos J, Amorim A (2014) Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA. Forensic Sci Int Genet 15:33–38. doi:10.1016/j.fsigen.2014.09.013
3. Afonso Costa H, Morais P, Vieira da Silva C, Matos S, Marques Santos R, Espinheira R, Costa Santos J, Amorim A (2014) X-Chromosome STR markers data in a Cabo Verde immigrant population of Lisboa. Mol Biol Rep 41:2559–2569. doi:10.1007/s11033-014-3114-9
4. Marques dos Santos R, Vieira da Silva C, Afonso Costa H, Ferreira Gomes P, Sanches S, Espinheira R, Costa Santos J, Amorim A (2012) Genetic portrait of an immigrant population from Angola living in Lisboa. Forensic Sci Int Genet 6:170–173. doi:10.1016/j.fsigen.2012.05.010
5. Amorim A, Marques dos Santos R, Vieira da 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 Sci Int Genet 6:166–169. doi:10.1016/j.fsigen.2012.05.006

6. Amorim A, Marques dos Santos R, Vieira da Silva C, Afonso Costa H, Espinheira R, Costa Santos J (2012) Genetic portrait of Brazilian immigrant population living in Lisboa. Forensic Sci Int Genet 6: 121–124. doi:10.1016/j.fsigen.2012.01.008
7. Amorim A, Vieira da Silva C, Afonso Costa H, Sanches S, Espinheira R, Costa Santos J (2011) Allele frequencies of all CODIS and four non-CODIS STR loci of an immigrant Brazilian population living in Lisbon—preliminary results. Forensic Sci Int Genet Suppl Ser 3:127–128. doi:10.1016/j.fsigss.2011.08.063
8. Carvalho A, Pinheiro MF (2013) Population data of 30 insertion/delection polymorphisms from a sample taken in the North of Portugal. Int J Legal Med 127:65–67. doi:10.1007/s00414-012-0693-7
9. Meng H-T, Zhang Y-D, Shen C-M, Yuan G-L, Yang C-H, Jin R, Yan J-W, Wang H-D, Liu W-J, Jing H, Zhu B-F (2015) Genetic polymorphism analyses of 30 InDels in Chinese Xibe ethnic group and its population genetic differentiations with other groups. Sci Rep 5:8260. doi:10.1038/srep08260
10. LaRue BL, Lagacé R, Chang CW, Holt A, Hennessy L, Ge J, King JL, Chakraborty R, Budowle B (2014) Characterization of 114 insertion/deletion (INDEL) polymorphisms, and selection for a global INDEL panel for human identification. Leg Med 16:26–32. doi:10.1016/j.legalmed.2013.10.006
11. Zidkova A, Horinek A, Kebrdlova V, Korabecna M (2013) Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population. Int J Legal Med 127:7–10. doi:10.1007/s00414-011-0649-3
12. Pepinski W, Abreu-Glowacka M, Koralewska-Kordel M, Michalak E, Kordel K, Niemcunowicz-Janica A, Szeremeta M, Konarzewska M (2013) Population genetics of 30 INDELs in populations of Poland and Taiwan. Mol Biol Rep 40:4333–4338. doi:10.1007/s11033-013-2521-7
13. Pereira R, Phillips C, Pinto N, Santos C, dos Santos SEB, Amorim A, Carracedo Á, Gusmão L (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS One 7:1–10. doi:10.1371/journal.pone.0029684
14. Klein R, Neumann C, Roy R (2015) Detection of insertion/deletion polymorphisms from challenged samples using the Investigator DIPplex® Kit. Forensic Sci Int Genet 16:29–37. doi:10.1016/j.fsigen.2014.11.020
15. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182–1190. doi:10.1101/gr.4565806
16. Ferreira Palha TDJB, Ribeiro Rodrigues EM, Cavalcante GC, Marrero A, de Souza IR, Seki Uehara CJ, Silveira da Motta CHA, Koshikene D, da Silva DA, de Carvalho EF, Chemale G, Freitas JM, Alexandre L, Paranaiba RTF, Soler MP, Santos S (2015) Population genetic analysis of insertion–deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit. Forensic Sci Int Genet 19:10–14. doi:10.1016/j.fsigen.2015.03.015
17. Seong KM, Park JH, Hyun YS, Kang PW, Choi DH, Han MS, Park KW, Chung KW (2014) Population genetics of insertion-deletion polymorphisms in South Koreans using Investigator DIPplex kit. Forensic Sci Int Genet 8:80–83. doi:10.1016/j.fsigen.2013.06.013
18. LaRue BL, Ge J, King JL, Budowle B (2012) A validation study of the Qiagen Investigator DIPplex® kit: an INDEL-based assay for human identification. Int J Legal Med 126:533–540. doi:10.1007/s00414-012-0667-9
19. Li CT, Zhang SH, Zhao SM (2011) Genetic analysis of 30 InDel markers for forensic use in five different Chinese populations. Genet Mol Res 10:964–979. doi:10.4238/vol10-2gmr1082
20. Neuvonen AM, Palo JU, Hedman M, Sajantila A (2012) Discrimination power of Investigator DIPplex loci in Finnish and Somali populations. Forensic Sci Int Genet 6:99–102. doi:10.1016/j.fsigen.2011.09.005
21. Wei YL, Qin CJ, Dong H, Jia J, Li CX (2014) A validation study of a multiplex INDEL assay for forensic use in four Chinese populations. Forensic Sci Int Genet 9:e22–e25. doi:10.1016/j.fsigen.2013.09.002
22. Friis SL, Børsting C, Rockenbauer E, Poulsen L, Fredslund SF, Tomas C, Morling N (2012) Typing of 30 insertion/deletions in Danes using the first commercial indel kit - Mentype® DIPplex. Forensic Sci Int Genet 6:72–74. doi:10.1016/j.fsigen.2011.08.002
23. Kis Z, Zalán A, Volgyi A, Kozma Z, Domján L, Pamjav H (2012) Genome deletion and insertion polymorphisms (DIPs) in the Hungarian population. Forensic Sci Int Genet 6:125–126
24. Tomas C, Skitsa I, Steinmeier E, Poulsen L, Ampati A, Børsting C, Morling N (2015) Results for five sets of forensic genetic markers studied in a Greek population sample. Forensic Sci Int Genet 16: 132–137. doi:10.1016/j.fsigen.2015.01.001
25. Walsh P, Metzger D, Higuchi R (1991) Chelex 100 as a médium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513
26. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
27. Armstrong RA (2014) When to use the bonferroni correction. Ophthalmic Physiol Opt 34:502–508. doi:10.1111/opo.12131
28. Vieira da Silva C, Matos S, Afonso Costa H, Morais P, Dos Santos RM, Espinheira R, Costa Santos J, Amorim A (2013) Genetic portrait of south Portugal population with InDel markers. Forensic Sci Int Genet 7:10–12. doi:10.1016/j.fsigen.2013.03.009
29. Reis RA, InDel e SNP (2014) Caracterização Genética das Populações Imigrantes de Cabo Verde e Guiné-Bissau a residir na Região de Lisboa, Escola Superior de Saúde Egas Moniz
30. Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166
31. Pereira L, Gusmão L, Alves C, Amorim A, Prata MJ (2002) Bantu and European Y-lineages in Sub-Saharan Africa. Ann Hum Genet 66:369–378
32. Carracedo A, Butler JM, Gusmão L, Linacre A, Parson W, Roewer L, Schneider PM (2014) Update of guidelines for the publication of genetic population data. Forensic Sci Int Genet 10:A1–A2

ELECTRONIC SUPPLEMENTARY MATERIAL

Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa

INTERNATIONAL JOURNAL OF LEGAL MEDICINE

Ana Inácio[1,2], Heloísa Afonso Costa[1,6], Cláudia Vieira da Silva[1,6], Teresa Ribeiro[1,6], Maria João Porto[1,6], Jorge Costa Santos[3,4,6],Gilberto Igrejas[2], António Amorim[1,4,5,6]

1 - Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal

2 - Departamento de Genética e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

3 - Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

4 - Instituto Superior de Ciências da Saúde Egas Moniz, Monte de Caparica, Portugal

5 - Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

6 - CENCIFOR - Centro de Ciências Forenses, Coimbra, Portugal

Ana Inácio: anainacio14@gmail.com

Heloísa Afonso Costa: hcosta@dlinml.mj.pt

Cláudia Vieira da Silva: csilva@dlinml.mj.pt

Teresa Ribeiro: tribeiro@dlinml.mj.pt

Maria João Porto: mariajoao.porto@dcinml.mj.pt

Jorge Costa Santos: costasantosjm@gmail.com

Gilberto Igrejas: gigrejas@utad.pt

António Amorim: aamorim@dlinml.mj.pt

In this electronic supplementary material part of the manuscript gives additional information on:

- Supplementary tables S1, S2 and S3 show the frequencies for insertions and deletion alleles for all studied populations/groups also expected heterozigosity (He), observed heterozigosity (Ho), PE, DP, PIC, TPI and MP values.
- Supplementary table S4 presents the combined matching probability (CMP), the combined power of discrimination (CPD) and the combined power of exclusion (CPE).
- Supplementary table S5 show the genetic distances between the studied populations.

Supplementary Table S1: Allelic frequencies and some statistic and forensic parameters of 30 InDel genetic markers in Angola immigrant population.

InDel *Locus*	Chromosome localization	Sequence	Allelic frequencies											
			N	Deletion	Insertion	Microvariant	Ho	He	HWE	PE	DP	PIC	TPI	MP
HLD77	7q31.1	TAAG	258	0,6667	0,3333	-	0,4419	0,4453	1,0000	0,1410	0,5930	0,3500	0,9000	0,4070
HLD45	2q31.1	CACG	258	0,5756	0,4244	-	0,5310	0,4895	0,2040	0,2160	0,5970	0,3700	1,0700	0,4030
HLD131	7q36.2	TGGGCTTATT	258	0,3140	0,6860	-	0,4264	0,4316	0,8843	0,1310	0,5840	0,3400	0,8700	0,4160
HLD70	6q16.1	AGCA	258	0,1686	0,8314	-	0,2985	0,2809	0,3776	0,0630	0,4450	0,2400	0,7100	0,5550
HLD6	16q13	GCAGGACTGGGCACC	258	0,6279	0,3721	-	0,4496	0,4682	0,5937	0,1470	0,6140	0,3600	0,9100	0,3860
HLD111	17p11.2	CACA	258	0,5833	0,4167	-	0,4225	0,4871	0,0404	0,1310	0,6400	0,3700	0,8700	0,3600
HLD58	5q14.1	AGGA	258	0,7694	0,2306	-	0,3605	0,3556	1,0000	0,0920	0,5200	0,2900	0,7800	0,4800
HLD56	4q25	TAAGT	258	0,4244	0,5756	-	0,4690	0,4895	0,5264	0,1620	0,6280	0,3700	0,9400	0,3720
HLD118	20p11.1	CCCCA	258	0,6802	0,3198	-	0,4457	0,4359	0,7760	0,1440	0,5830	0,3400	0,9000	0,4170
HLD92	11q22.2	GTTT	258	0,5969	0,3973	0,0058	0,4729	0,4868	0,4028	0,1650	0,6290	0,3700	0,9500	0,3710
HLD93	12q22	ACTTT	258	0,5116	0,4884	-	0,5271	0,5007	0,4520	0,2120	0,6100	0,3700	1,0600	0,3900
HLD99	14q23.1	TGAT	258	0,3740	0,6182	0,0078	0,4806	0,4788	0,3705	0,1710	0,6180	0,3700	0,9600	0,3820
HLD88	9q22.32	CCACAAAGA	258	0,3450	0,6550	-	0,4651	0,4528	0,6829	0,1590	0,5930	0,3500	0,9300	0,4070
HLD101	15q26.1	GTAG	258	0,2636	0,7364	-	0,3566	0,3890	0,2016	0,0900	0,5540	0,3100	0,7800	0,4460
HLD67	5q33.2	CTACTGAC	258	0,3837	0,6163	-	0,4729	0,4739	1,0000	0,1650	0,6100	0,3600	0,9500	0,3900
HLD83	8p22	AAGG	258	0,4128	0,5872	-	0,4612	0,4857	0,4358	0,1560	0,6270	0,3700	0,9300	0,3730
HLD114	17p13.3	TCCTATTCTACTCTGAAT	258	0,2442	0,7558	-	0,3488	0,3698	0,4038	0,0860	0,5350	0,3000	0,7700	0,4650
HLD48	2q11.2	GACTT	258	0,3236	0,6764	-	0,5155	0,4387	0,0049	0,2010	0,5550	0,3400	1,0300	0,4450
HLD124	22q12.3	GTGGA	258	0,7461	0,2539	-	0,3450	0,3796	0,1441	0,0840	0,5450	0,3100	0,7600	0,4550
HLD122	21q22.11	GAAGTCTGAGG	258	0,6047	0,3953	-	0,4884	0,4790	0,7936	0,1780	0,6090	0,3600	0,9800	0,3910
HLD125	22q11.23	ATTGCC	258	0,6725	0,3275	-	0,4923	0,4414	0,0680	0,1810	0,5690	0,3400	0,9800	0,4310
HLD64	5q12.3	GACAAA	258	0,2810	0,7190	-	0,3992	0,4049	0,8779	0,1130	0,5640	0,3200	0,8300	0,4360
HLD81	7q21.3	GTAAGCATTGT	258	0,4574	0,5426	-	0,4961	0,4973	1,0000	0,1840	0,6230	0,3700	0,9900	0,3770
HLD136	22q13.1	TGTTT	258	0,2422	0,7578	-	0,3760	0,3678	0,8654	0,1000	0,5310	0,3000	0,8000	0,4690
HLD133	3p22.1	CAACCTGGATT	258	0,6008	0,3992	-	0,4264	0,4806	0,0903	0,1310	0,6330	0,3600	0,8700	0,3670
HLD97	13q12.3	AGAGAAAGCTGAAG	258	0,5426	0,4574	-	0,4186	0,4973	0,0124	0,1260	0,6520	0,3700	0,8600	0,3480
HLD40	1p32.3	GGGACAGGTGGCCACTAGGAGA	258	0,6318	0,3682	-	0,4574	0,4662	0,7882	0,1530	0,6090	0,3600	0,9200	0,3910
HLD128	1q31.3	ATTAAATA	258	0,3876	0,6124	-	0,5039	0,4757	0,3624	0,1910	0,5980	0,3600	1,0100	0,4020
HLD39	1p22.1	CCTAAACAAAAATGGGAT	258	0,5116	0,4884	-	0,4729	0,5007	0,3827	0,1650	0,6370	0,3700	0,9500	0,3630
HLD84	8q24.12	CTTTC	258	0,3023	0,6841	0,0136	0,4070	0,4413	0,1167	0,1180	0,6110	0,3500	0,8400	0,3890

Ho: Observed heterozigoty; He: Expected heterozigoty; HWE: Hardy-Weinberg equilibrium; PE: Power of exclusion; DP: Discrimination power; PIC: Polymorphic information content; TPI: Typical paternity index; MP: Matching probability

Supplementary Table S2: Allelic frequencies and some statistic and forensic parameters of 30 InDel genetic markers in Guinea-Bissau immigrant population.

InDel *Locus*	Chromosome localization	Sequence	Allelic frequencies											
			N	Deletion	Insertion	Microvariant	Ho	He	HWE	PE	DP	PIC	TPI	MP
HLD77	7q31.1	TAAG	123	0,6748	0,3252	-	0,4065	0,4407	0,4140	0,1180	0,5980	0,3400	0,8400	0,4020
HLD45	2q31.1	CACG	123	0,5935	0,4065	-	0,4553	0,4845	0,5760	0,1510	0,6270	0,3700	0,9200	0,3730
HLD131	7q36.2	TGGGCTTATT	123	0,1992	0,8008	-	0,2683	0,3203	0,0883	0,0510	0,4790	0,2700	0,6800	0,5210
HLD70	6q16.1	AGCA	123	0,0935	0,9065	-	0,1870	0,1702	0,5972	0,0260	0,3040	0,1600	0,6200	0,6960
HLD6	16q13	GCAGGACTGGGCACC	123	0,5488	0,4512	-	0,4959	0,4973	1,0000	0,1840	0,6220	0,3700	0,9900	0,3780
HLD111	17p11.2	CACA	123	0,6382	0,3618	-	0,4634	0,4637	1,0000	0,1570	0,6030	0,3600	0,9300	0,3970
HLD58	5q14.1	AGGA	123	0,8740	0,1260	-	0,2358	0,2212	0,6907	0,0400	0,3730	0,2000	0,6500	0,6270
HLD56	4q25	TAAGT	123	0,4553	0,5447	-	0,5041	0,4980	1,0000	0,1910	0,6190	0,3700	1,0100	0,3810
HLD118	20p11.1	CCCCA	123	0,7642	0,2358	-	0,3740	0,3618	0,8047	0,0990	0,5250	0,3000	0,8000	0,4750
HLD92	11q22.2	GTTT	123	0,6829	0,3171	-	0,3740	0,4348	0,1451	0,0990	0,5970	0,3400	0,8000	0,4030
HLD93	12q22	ACTTT	123	0,5325	0,4675	-	0,5610	0,4999	0,2056	0,2470	0,5870	0,3700	1,1400	0,4130
HLD99	14q23.1	TGAT	123	0,2520	0,7358	0,0122	0,3902	0,3966	0,0399	0,1080	0,5570	0,3300	0,8200	0,4430
HLD88	9q22.32	CCACAAAGA	123	0,3130	0,6870	-	0,4146	0,4318	0,6766	0,1230	0,5870	0,3400	0,8500	0,4130
HLD101	15q26.1	GTAG	123	0,1911	0,8089	-	0,3171	0,3104	1,0000	0,0710	0,4750	0,2600	0,7300	0,5250
HLD67	5q33.2	CTACTGAC	123	0,3211	0,6789	-	0,3984	0,4378	0,4075	0,1130	0,5960	0,3400	0,8300	0,4040
HLD83	8p22	AAGG	123	0,3171	0,6829	-	0,3577	0,4348	0,0597	0,0900	0,5990	0,3400	0,7800	0,4010
HLD114	17p13.3	TCCTATTCTACTCTGAAT	123	0,2480	0,7520	-	0,3659	0,3745	0,8104	0,0940	0,5380	0,3000	0,7900	0,4620
HLD48	2q11.2	GACTT	123	0,2154	0,7846	-	0,3496	0,3394	1,0000	0,0860	0,5040	0,2800	0,7700	0,4960
HLD124	22q12.3	GTGGA	123	0,8496	0,1504	-	0,2683	0,2566	1,0000	0,0510	0,4160	0,2200	0,6800	0,5840
HLD122	21q22.11	GAAGTCTGAGG	123	0,6545	0,3455	-	0,4634	0,4541	0,8444	0,1570	0,5940	0,3500	0,9300	0,4060
HLD125	22q11.23	ATTGCC	123	0,7683	0,2317	-	0,4146	0,3575	0,0804	0,1230	0,5130	0,2900	0,8500	0,4870
HLD64	5q12.3	GACAAA	123	0,2276	0,7724	-	0,3252	0,3531	0,4399	0,0740	0,5180	0,2900	0,7400	0,4820
HLD81	7q21.3	GTAAGCATTGT	123	0,4024	0,5976	-	0,4959	0,4829	0,8520	0,1840	0,6080	0,3700	0,9900	0,3920
HLD136	22q13.1	TGTTT	123	0,1585	0,8415	-	0,2520	0,2679	0,5036	0,0460	0,4240	0,2300	0,6700	0,5760
HLD133	3p22.1	CAACCTGGATT	123	0,5854	0,4146	-	0,4715	0,4874	0,8523	0,1640	0,6230	0,3700	0,9500	0,3770
HLD97	13q12.3	AGAGAAAGCTGAAG	123	0,5976	0,4024	-	0,2683	0,4829	0,0000	0,0510	0,6410	0,3700	0,6800	0,3590
HLD40	1p32.3	GGACAGGTGGCCACTAGGAC	123	0,7520	0,2480	-	0,3659	0,3745	0,8099	0,0940	0,5380	0,3000	0,7900	0,4620
HLD128	1q31.3	ATTAAATA	123	0,2927	0,7073	-	0,3740	0,4157	0,2820	0,0990	0,5780	0,3300	0,8000	0,4220
HLD39	1p22.1	CCTAAACAAAAATGGGAT	123	0,4797	0,5203	-	0,5203	0,5012	0,7200	0,2060	0,6130	0,3700	1,0400	0,3870
HLD84	8q24.12	CTTTC	123	0,2805	0,7033	0,0163	0,3821	0,4282	0,2144	0,1030	0,6030	0,3500	0,8100	0,3970

Ho: Observed heterozigoty; He: Expected heterozigoty; HWE: Hardy-Weinberg equilibrium; PE: Power of exclusion; DP: Discrimination power; PIC: Polymorphic information content; TPI: Typical paternity index; MP: Matching probability

Supplementary Table S3: Allelic frequencies and some statistic and forensic parameters of 30 InDel genetic markers in Mozambique immigrant population.

InDel	Chromosome	Sequence	Allelic frequencies											
Locus	localization		N	Deletion	Insertion	Microvariant	Ho	He	HWE	PE	DP	PIC	TPI	MP
HLD77	7q31.1	TAAG	72	0,5903	0,4097	-	0,4861	0,4871	1,0000	0,1760	0,6150	0,3700	0,9700	0,3850
HLD45	2q31.1	CACG	72	0,5069	0,4931	-	0,4861	0,5034	0,8163	0,1760	0,6320	0,3700	0,9700	0,3680
HLD131	7q36.2	TGGGCTTATT	72	0,4167	0,5833	-	0,5000	0,4895	1,0000	0,1880	0,6110	0,3700	1,0000	0,3890
HLD70	6q16.1	AGCA	72	0,2569	0,7431	-	0,3750	0,3845	1,0000	0,0990	0,5460	0,3100	0,8000	0,4540
HLD6	16q13	GCAGGACTGGGCACC	72	0,5625	0,4375	-	0,5417	0,4956	0,4781	0,2270	0,5940	0,3700	1,0900	0,4060
HLD111	17p11.2	CACA	72	0,5972	0,4028	-	0,4444	0,4845	0,6262	0,1430	0,6290	0,3700	0,9000	0,3710
HLD58	5q14.1	AGGA	72	0,7014	0,2986	-	0,5139	0,4218	0,0889	0,2000	0,5370	0,3300	1,0300	0,4630
HLD56	4q25	TAAGT	72	0,4514	0,5486	-	0,4861	0,4987	1,0000	0,1760	0,6270	0,3700	0,9700	0,3730
HLD118	20p11.1	CCCCA	72	0,6181	0,3819	-	0,5972	0,4754	0,0446	0,2880	0,5340	0,3600	1,2400	0,4660
HLD92	11q22.2	GTTT	72	0,5208	0,4792	-	0,4306	0,5026	0,2453	0,1340	0,6520	0,3700	0,8800	0,3480
HLD93	12q22	ACTTT	72	0,5000	0,5000	-	0,5278	0,5035	0,8133	0,2130	0,6100	0,3800	1,0600	0,3900
HLD99	14q23.1	TGAT	72	0,3750	0,6250	-	0,4722	0,4720	1,0000	0,1640	0,6060	0,3600	0,9500	0,3940
HLD88	9q22.32	CCACAAAGA	72	0,4028	0,5972	-	0,5278	0,4845	0,4723	0,2130	0,5910	0,3700	1,0600	0,4090
HLD101	15q26.1	GTAG	72	0,3611	0,6389	-	0,4444	0,4647	0,7991	0,1430	0,6100	0,3500	0,9000	0,3900
HLD67	5q33.2	CTACTGAC	72	0,3611	0,6389	-	0,5278	0,4647	0,3098	0,2130	0,5710	0,3500	1,0600	0,4290
HLD83	8p22	AAGG	72	0,4792	0,5208	-	0,5694	0,5026	0,3415	0,2560	0,5820	0,3700	1,1600	0,4180
HLD114	17p13.3	TCCTATTCTACTCTGAAT	72	0,3681	0,6319	-	0,4583	0,4684	1,0000	0,1540	0,6080	0,3600	0,9200	0,3920
HLD48	2q11.2	GACTT	72	0,3264	0,6736	-	0,4306	0,4428	1,0000	0,1340	0,5920	0,3400	0,8800	0,4080
HLD124	22q12.3	GTGGA	72	0,5764	0,4236	-	0,4306	0,4917	0,3349	0,1340	0,6410	0,3700	0,8800	0,3590
HLD122	21q22.11	GAAGTCTGAGG	72	0,6458	0,3542	-	0,4861	0,4607	0,7967	0,1760	0,5890	0,3500	0,9700	0,4110
HLD125	22q11.23	ATTGCC	72	0,5556	0,4444	-	0,5278	0,4973	0,6378	0,2130	0,6040	0,3700	1,0600	0,3960
HLD64	5q12.3	GACAAA	72	0,2986	0,7014	-	0,4306	0,4218	1,0000	0,1340	0,5710	0,3300	0,8800	0,4290
HLD81	7q21.3	GTAAGCATTGT	72	0,4514	0,5486	-	0,5139	0,4987	0,8160	0,2000	0,6130	0,3700	1,0300	0,3870
HLD136	22q13.1	TGTTT	72	0,3889	0,6111	-	0,3611	0,4786	0,0476	0,0920	0,6410	0,3600	0,7800	0,3590
HLD133	3p22.1	CAACCTGGATT	72	0,5000	0,5000	-	0,5278	0,5035	0,8130	0,2130	0,6100	0,3800	1,0600	0,3900
HLD97	13q12.3	AGAGAAAGCTGAAG	72	0,5903	0,4097	-	0,4861	0,4871	1,0000	0,1760	0,6150	0,3700	0,9700	0,3850
HLD40	1p32.3	GGACAGGTGGCCACTAGGAC	72	0,6250	0,3750	-	0,5000	0,4720	0,8014	0,1880	0,5940	0,3600	1,0000	0,4060
HLD128	1q31.3	ATTAAATA	72	0,4444	0,5556	-	0,5000	0,4973	1,0000	0,1880	0,6190	0,3700	1,0000	0,3810
HLD39	1p22.1	CCTAAACAAAAATGGGAT	72	0,4861	0,5139	-	0,5556	0,5031	0,4801	0,2410	0,5920	0,3700	1,1300	0,4080
HLD84	8q24.12	CTTTC	72	0,4097	0,5833	0,0069	0,4444	0,4952	0,5257	0,1430	0,6460	0,3800	0,9000	0,3540

Ho: Observed heterozigoty; He: Expected heterozigoty; HWE: Hardy-Weinberg equilibrium; PE: Power of exclusion; DP: Discrimination power; PIC: Polymorphic information content; TPI: Typical paternity index; MP: Matching probability

Supplementary Table S4: Combined forensic parameters studied for 30 InDel markers.

	CMP	CPD	CPE
Angola immigrants	1,98X10-12	>99,99%	99,12%
Mozambique immigrants	8,67X10-13	>99,99%	99,75%
Guinea-Bissau immigrants	2,80X10-11	>99,99%	97,44%

CMP: combined matching probability; CPD: combined power of discrimination; CPE: combined power of exclusion

Supplementary Table S5: Pairwise difference genetic distance (FST) between the studied populations and other populations.

Populations	Angola immigrants	Mozambique immigrants	Guinea-Bissau immigrants	Caucasian immigrants	Cape Verde immigrants
Angola immigrants	-				
Mozambique immigrants	0,0084	-			
Guinea-Bissau immigrants	0,0096	0,0336	-		
Caucasian immigrants	0,0528	0,0203	0,0976	-	
Cape Verde immigrants	0,0034	0,0110	0,0093	0,0516	-

ELECTRONIC SUPPLEMENTARY MATERIAL

Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa

INTERNATIONAL JOURNAL OF LEGAL MEDICINE

Ana Inácio[1,2], Heloísa Afonso Costa[1,6], Cláudia Vieira da Silva[1,6], Teresa Ribeiro[1,6], Maria João Porto[1,6], Jorge Costa Santos[3,4,6],Gilberto Igrejas[2], António Amorim[1,4,5,6]

1 - Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal

2 - Departamento de Genética e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

3 - Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

4 - Instituto Superior de Ciências da Saúde Egas Moniz, Monte de Caparica, Portugal

5 - Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

6 - CENCIFOR - Centro de Ciências Forenses, Coimbra, Portugal

Ana Inácio: anainacio14@gmail.com

Heloísa Afonso Costa: hcosta@dlinml.mj.pt

Cláudia Vieira da Silva: csilva@dlinml.mj.pt

Teresa Ribeiro: tribeiro@dlinml.mj.pt

Maria João Porto: mariajoao.porto@dcinml.mj.pt

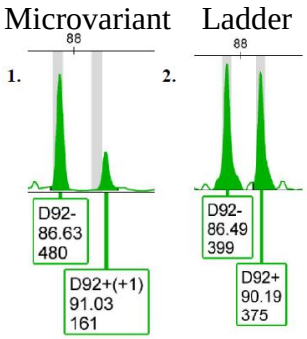
Jorge Costa Santos: costasantosjm@gmail.com

Gilberto Igrejas: gigrejas@utad.pt

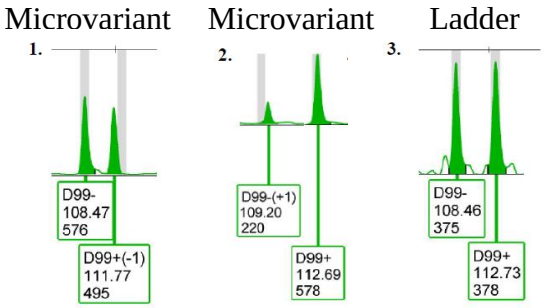
António Amorim: aamorim@dlinml.mj.pt

In this electronic supplementary material part of the manuscript gives additional information on:

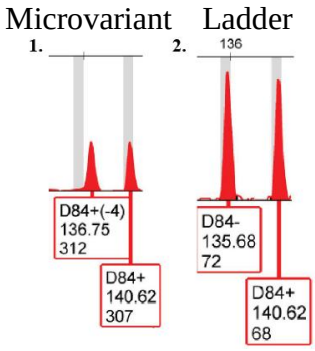
- Supplementary Figure S1, S2 and S3 show the microvariants detected during the analysis of InDel markers.
- In each supplementary figure it's possible to see the peak out off ladder and the corresponding allelic ladder.



Supplementary Figure S1: In **1.** it's possible to observe the peak out off ladder D92 + (+1), detected with a nucleotide more than what was expected as seen in **2.** to allelic ladder.



Supplementary Figure S2: In **1.** can be seen the peak out off ladder D99 + (-1) and in **2.** the peak out off ladder D99- (+1), both detected with one nucleotide less and more, respectively, than the one that was expected as seen in the allelic ladder in **3**.



Supplementary Figure S3: In **1.** you can see the peak out off ladder D84 + (-4) with a nucleotide more than what was expected as seen in **2.** for allelic ladder.

Nos últimos anos, uma nova classe de marcadores genéticos autossómicos tem vindo a suscitar o interesse da genética forense, principalmente pela sua abundância no genoma humano e pelo facto dos métodos de análise que requerem serem semelhantes aos métodos em utilização para STR nos laboratórios forenses (A. Carvalho & Pinheiro, 2013). Designam-se por polimorfismos de inserção e de deleção e estão presentes por todo o genoma humano, sendo referidos como InDel (inserção-deleção) ou DIP (deletion-insertion polymorphisms) e são caraterizados pela presença ou ausência de uma sequência específica de nucleótidos, tipicamente de 1 a 50 pb (W. Chen & Zhang, 2015; LaRue, Ge, King, & Budowle, 2012; Meng et al., 2015; Pepinski et al., 2013; Zidkova, Horinek, Kebrdlova, & Korabecna, 2013). Estes marcadores podem ser utilizados como marcadores de ancestralidade devido às diferenças significativas presentes nas frequências alélicas entre diferentes grupos ou populações que têm sido relatados em alguns estudos (Ferreira Palha et al., 2015; LaRue et al., 2012; Li, Zhang, & Zhao, 2011; Meng et al., 2015; Mills et al., 2006; Neuvonen, Palo, Hedman, & Sajantila, 2012; Pepinski et al., 2013; Seong et al., 2014; Wei, Qin, Dong, Jia, & Li, 2014). Desta forma, utilizamos um painel destes marcadores autossómicos - InDel - para tipagem genética dos imigrantes PALOP que estudamos, com vista a determinar se através do estudo com marcadores autossómicos as relações filogenéticas observadas entre os indivíduos estudados com DNAmt eram reforçadas ou, pelo contrário, observaríamos relações filogenéticas diversas.

Os marcadores do tipo InDel demonstraram, através dos resultados obtidos, uma elevada diferenciação entre as

populações africanas, e entre populações africanas e a população portuguesa.

As populações de Angola, da Guiné-Bissau e de Moçambique, antes do principal fenómeno de migração do povo Bantu, eram essencialmente constituídas por tribos Khoisan. Assim, o principal fenómeno de evolução das populações africanas poderá dever-se à expansão das tribos Bantu para duas regiões diferentes, uma a Este de África, onde Moçambique está incluído, e uma a Oeste de África, onde Angola e Guiné-Bissau estão incluídas (Beleza, Gusmão, Amorim, Carracedo, & Salas, 2005; Berniell-Lee et al., 2009). Relativamente a Cabo-Verde, foi descoberto desabitado e, posteriormente, foi colonizado principalmente por escravos provenientes da costa Oeste africana, no âmbito do tráfico de escravos praticado à data pelos portugueses (Brehm, Pereira, Bandelt, Prata, & Amorim, 2002)]. Assim, não é de estranhar que estas três populações de imigrantes - Angola, Guiné-Bissau e Cabo Verde - sejam mais próximas umas das outras e que Moçambique apresente uma distância genética significante em relação a estas três populações africanas (PEREIRA, GUSMÃO, ALVES, AMORIM, & PRATA, 2002).

Os resultados obtidos mostram que as populações de imigrantes de Angola, Guiné-Bissau e Cabo Verde estão mais próximas entre elas e se encontram separadas significativamente da população portuguesa. Moçambique está mais próximo da população portuguesa que as outras populações em estudo e pode participar num cluster individual. Angola, Guiné-Bissau e Cabo Verde estão incluídos assim noutro cluster, como esperado, tendo em conta a expansão das tribos Bantu ao longo das duas regiões distintas de África - Oeste e Este – como mencionado anteriormente.

Capítulo IX - Resultados e Conclusões Gerais:

Mitochondrial DNA studies of Lisbon immigrants from Portuguese speaking African countries

Amorim A, Afonso Costa H, Vieira da Silva C, Ribeiri T, Porto MJ, Taveira N, Fernandes T. ***Mitochondrial DNA studies of Lisbon immigrants from Portuguese speaking African countries.*** Virus Evolution (2018) 4, Suppl. 1:S24.

serological testing; increased genetic information from diverse sources should enable PCR design that will reliably detect a wide range of strains thus facilitating diagnosis and epidemiological analysis of outbreaks.

A64 Overview of virus metagenomics classification tools

Sam Nooij,[1,2] Dennis Schmitz,[1,2] Annelies Kroneman,[1] Harry Vennema,[1] and Marion Koopmans[1,2]

[1]National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands and [2]Erasmus University Medical Centre (EMC), Rotterdam, The Netherlands

The use of next-generation sequencing for discovery of viruses has yielded vast amounts of known and putative viral reads. The computational analysis of the reads, however, is quite a challenge, in particular the classification of reads to viral taxa. This is illustrated by the many computational tools that have been devised and new tools that appear monthly. These provide opportunities for other researchers, but the large numbers make it hard for virologists to pick a tool that suits their own study. To facilitate this choice and guide users to through the forest of computational pipelines, we have surveyed publications describing fifty tools, inventoried their approaches and scored their methods, user-friendliness, validation, and other performance criteria for diagnostics, outbreak source tracing, virus discovery, and virome profiling. The tools have variable approaches to the classification of viral reads, but they all rely on searching (i.e., homology (44/50 pipelines) and composition search (8/50)) through reference databases (e.g., nucleotide databases, protein databases, or virus-specific databases). Some pipelines include quality control/pre-processing of reads (23/50), filtering non-viral reads (20/50), and de novo assembly (18/50) before the search, and checking and correcting classifications with phylogenetic or statistical methods after the search step (8/50). Furthermore, some are tailor-made for particular studies; others are more generally applicable. Few tools provide a graphical user interface, and when they do they are often online, which increases the ease of use. Reported runtimes vary greatly—from several minutes per sample to days; newer tools are often faster than old ones. Moreover, some tools have been validated in wet-lab experiments or compared to other tools with in silico benchmark tests. The overview of pipelines is presented on the COMPARE website (https://compare.cbs.dtu.dk/inventory#pipeline). A decision tree is provided separately to help virologists with any level of bioinformatics expertise select suitable analysis tools. The next step will be to benchmark the most promising tools with the COMPARE and VIROGENESIS projects to better assess their performance for diagnostics and surveillance studies.

A65 Mitochondrial DNA studies of Lisbon immigrants from Portuguese speaking African countries

A. Amorim,[1,2,3,4] A. Afonso Costa,[1,5] C. Vieira da Silva,[1] T. Ribeiro,[1] M. J. Porto,[1] N. Taveira,[3,6] and T. Fernandes[4]

[1]Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal, [2]Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal, [3]Instituto Superior de Ciências da Saúde Egas Moniz, Almada, Portugal, [4]Escola de Ciências e Tecnologia da Universidade de Évora, Évora, Portugal, [5]Faculdade de Ciências da Saúde da Universidade da Beira Interior, Covilhã, Portugal and [6]Faculdade de Farmácia da Universidade de Lisboa, Lisboa, Portugal

Since the end of the 1970s, Portugal has had an important role in migratory movements, becoming a destination for immigrants of a wide range of nationalities, mainly from African countries. According to PORDATA, until the end of 2014 there were ~40,000 immigrants from Cape Verde, 20,000 from Angola, 18,000 from Guinea-Bissau, and 3,000 from Mozambique living in Portugal, and of those, >80 per cent live in the Lisbon region. This may be one of the main contributors to genetic variation of Lisbon residents in the present and the future. Mitochondrial DNA (mtDNA) has features that make it desirable for forensics, namely, high copy number, lack of recombination, and matrilineal inheritance. These features are also important in evolutionary and population studies. We aim to characterize mtDNA diversity in immigrants from Portuguese Speaking African Countries (PALOP) living in Lisbon and their potential contribution to genetic variation of Lisbon population. Blood samples were collected from 439 PALOP immigrants living in Lisbon, of which 173 immigrants from Angola, 103 immigrants from Cape Verde, eighty-three immigrants from Mozambique and eighty immigrants from Guinea-Bissau, from January 2000 to December 2016. The control region of the mtDNA was amplified using two pairs of primers—L15971/H016 and L16555/H639, and sequenced by BigDye Terminator v.3.1 Cycle Sequence (AB). Sequenced products were detected in a sequencer Genetic Analyzer 3130 (AB). Finally the results were analysed by Sequencing Analysis v.5.2 software and also compared with Revised Cambridge Reference Sequence (rCRS) using SeqScape v.3 (AB) software. The haplogroups were determined based on Phylotree, build 17. Genetic distances and other genetic parameters were calculated with Arlequin software ver.3.5 and analysed and represented with PhyML 3.0. For each sample, the complete sequence of the control region was obtained. The comparison of the sequences obtained with the rCRS, among the 439 analysed individuals, allowed the identification of 319 different haplotypes, corresponding to 164 different haplogroups distributed by ten macrohaplogroups. Macrohaplogroup L was the most common with 386 haplotypes followed by U with fifteen haplotypes, H with twelve haplotypes, M and T with six haplotypes, K with five, R with four, X and J with two and HV with one. PALOP's immigrants presented a high number of unique haplotypes, most of them belonging to macrohaplogroup L, originating from sub-Saharan regions of Africa. This macrohaplogroup is uncommon in European and Portuguese populations. Consistent with this, phylogenetic analysis showed the establishment of two distinct groups, one composed of the Portuguese population and another of the African populations. In comparing the different immigrant populations living in Lisbon, the genetically closest community to the Portuguese population is Mozambique and the furthest is Cape Verde, followed by Guinea-Bissau and Angola. Our results show that the PALOP immigrants living in Lisbon are genetically heterogeneous. The increase in genetic diversity in Lisbon due to immigrants from PALOP countries may have a major impact on haplotypic and allelic frequencies, on which all forensic and medico-legal investigations are based.

A66 Multi-drug-resistant *Klebsiella pneumoniae* strains circulating in hospital setting: Whole-genome sequencing and Bayesian phylogenetic analysis for outbreak investigations

E. Cella,[1,2] M. Ciccozzi,[1,3] M. Fogolari,[1] T. Azarian,[4] M. Prosperi,[5] G. Dicuonzo,[1] M. Salemi,[6] and S. Angeletti[1]

[1]Unit of Clinical Pathology and Microbiology, University Campus Bio-Medico of Rome, Italy, [2]Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy, [3]Department of Infectious Parasitic and Immune-Mediated Diseases, Istituto Superiore di Sanità, Rome, Italy, [4]Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard's T.H. Chan School of Public Health, Boston, MA, [5]Department of Epidemiology, University of Florida, Gainesville, FL, USA and [6]Department of Pathology, Immunology, and

Foram obtidas sequências de DNAmt com cerca de 1 122 pb de todos os 439 imigrantes de PALOP incluídos no estudo. Não contrariando a generalidade dos estudos que apontam a homoplasmia com condição com representação quase absoluta nos indivíduos da espécie humana, nos 439 indivíduos que estudamos não identificamos heteroplasmia de posição em nenhum caso.

Os imigrantes estudados encontam-se distribuídos pelos haplogrupos L0, L1, L2, L3, L4, L5, M, R, J, T, H, K, U e X, conforme informação apresentada na tabela 11.

Tabela 11 - Distribuição dos indivíduos PALOP estudados e dos indivíduos da população de Lisboa por macrohaplogrupos/haplogrupos observados. *Dados retirados de Marques *et al.*, 2015.

	Número de Imigrantes PALOP e de Nativos de Lisboa				
Macro(haplogrupo)	**Cabo Verde**	**Angola**	**Moçambique**	**Guiné-Bissau**	Sul Portugal *
L	95 (92%)	151 (87%)	67 (81%)	73 (91%)	6 (6%)
L0	-	23	21	2	1
L1	18	38	3	8	3
L2	35	32	24	34	1
L3	42	56	17	29	1
L4	-	2	-	-	-
L5	-	-	2	-	-
M	3	1	1	1	2
N	-	-	-	-	1
R	-	4	-	-	12
I	-	-	-	-	9
J	-	1	1	-	7
T	-	5	1	-	9
H	-	5	7	1	30
K	-	3	2	-	2
U	3	3	4	5	17
W	-	-	-	-	1
X	2	-	-	-	4
N. Total Indivíduos	103	173	83	80	100

Nos quatro grupos distintos de imigrantes estudados, oriundos de quatro países distintos, a percentagem de haplótipos e respetivos haplogrupos integrados no macrohaplogrupo L determinada, foi sempre superior a 80% e, em dois dos grupos, foi mesmo superior a 90%.

Relativamente aos imigrantes oriundos de Cabo Verde, cerca de 92% dos indivíduos estudados apresentam haplótipos

integrados nos haplogrupos L1, L2 e L3. Nos indivíduos oriundos de Angola, cerca de 87% apresentam haplótipos integrados nos haplogrupos L0, L1, L2, L3 e L4. Nos indivíduos oriundos de Moçambique, cerca de 81% apresentam haplóticos dos haplogrupos L0, L1, L2, L3 e L5. Nos indivíduos oriundos da Guiné-Bissau, cerca de 91% apresentam haplótipos dos haplogrupos L0, L1, L2 e L3.

Os haplogrupos M, R, J, T, H, K, U e X, associados a regiões euroasiáticas, surgem nos quatro grupos distintos de imigrantes de PALOP com representação sempre inferior a 20%, sendo os grupos de imigrantes onde esta representação é menor os grupos de Cabo Verde e Guiné-Bissau com 8 e 9% de representação, respetivamente.

A presença de reduzida representação de haplótipos europeus nas populações PALOP estudadas poderá refletir, aquando do processo de colonização, a pequena contribuição europeia feminina nas linhagens maternas destes países africanos e uma eventual e preferencial mistura entre indivíduos europeus do sexo masculino e mulheres nativas africanas, aliás, panorama algo comum nas ex-colónias do ex-Império Colonial Português (Brehm et al., 2002; Gonçąalves et al., 2003; L. Pereira et al., 2001; Purps et al., 2014; Trovoada et al., 2004).

Juntamente com a elevada frequência de haplótipos únicos observados e associada diversidade genética, como seria de esperar, as caraterísticas observadas apontam para grupos de indivíduos com caraterísticas típicas de populações de África. Na tabela 12 apresentamos os valores de diversidade genética calculados, o número de diferentes haplótipos observados, o número de locais polimórficos observados, o número de transições observadas, o número de transversões observadas, o número total de substituições

observadas e o número de InDels observado, para todos os imigrantes PALOP estudados.

Globalmente, as caraterísticas observadas nos imigrantes de Cabo Verde são amplamente coincidentes e não apresentam contradições relativamente às caraterísticas descritas em populações locais de Cabo Verde por Brehm em 2002 (Brehm et al., 2002).

Tabela 12 - Valores de Diversidade Genética calculados e número de haplótipos diferentes, locais polimórficos, transições, transversões, substituições e InDels observados para cada grupo de imigrantes PALOP estudado.

	Imigrantes de PALOP estudados distribuídos por País de origem			
	Cabo Verde	**Angola**	**Moçambique**	**Guiné-Bissau**
EMPOP (N. da População)	EMP00616	EMP00662	EMP00681	EMP00704
N. Indivíduos	103	173	83	80
N. de haplótipos diferentes obs.	75	173	83	80
N. de locais polimórficos obs.	103	143	122	105
N. de transições obs.	96	132	110	93
N. de transversões obs.	11	16	11	13
N. de substituições obs.	107	148	121	106
N. de InDels obs.	4	4	6	5
Diversidade Genética	0.9916±0.0031	1.0000±0.0006	1.0000±0.0019	1.0000±0.0020

As caraterísticas observadas nos imigrantes de Angola também são amplamente coincidentes e não apresentam contradições relativamente às caraterísticas descritas em populações locais angolanas por Alves Silva e colaboradores, em 2001 (Bandelt et al., 2001), por Plaza e colaboradores, em 2004 (Plaza et al., 2004), e por Salas e colaboradores, em 2004 (A Salas et al., 2004).

As caraterísticas observadas nos imigrantes de Moçambique são amplamente coincidentes e não apresentam contradições relativamente às caraterísticas descritas em populações locais de Moçambique por Pereira e colaboradores, em 2000 (Alonso et al., 2002), e por Salas e colaboradores, em 2002 (Antonio Salas et al., 2002).

As caraterísticas observadas nos imigrantes da Guiné-Bissau são amplamente coincidentes e não apresentam contradições relativamente às caraterísticas descritas em

populações locais da Guiné-Bissau por Carvalho e colaboradores, em 2011 (M. Carvalho et al., 2011).

Globalmente, os imigrantes PALOP estudados parecem representar as suas populações africanas de origem.

A qualidade das sequências obtidas no nosso estudo e o interesse populacional dos grupos estudados determinaram a sua inserção na EMPOP. É, assim, atualmente, possível a toda a comunidade científica a nível internacional aceder às sequências de DNAmt dos imigrantes de Cabo Verde, Angola, Moçambique e Guiné-Bissau integrados na população de Lisboa em www.empop.org, através dos *accession number* EMP00616, EMP00662, EMP00681 EMP00704, respetivamente.

A EMPOP, presentemente, conta com um total de 34 617 haplótipos de DNAmt, entre os quais 4 361 de indivíduos de regiões de África subsariana.

De entre os 4 361 haplótipos de indivíduos de África subsariana existentes na EMPOP, 129 são de indivíduos de Angola, 79 são de indivíduos da Guiné-Bissau e não existe qualquer representação de indivíduos de Moçambique nem de Cabo Verde. Com o contributo do nosso estudo a EMPOP não integrará indivíduos nativos e residentes em países de África, mas confirmamos que os 439 indivíduos/sequências integrados no nosso estudo têm linhagens maternas com ancestralidade africana, designadamente angolana, cabo verdeana, guineense e moçambicana. Angola que tinha uma representação de 129 sequências de DNAmt na EMPOP passa a contar com mais 173 indivíduos/sequências. Guiné-Bissau que tinha uma representação de 79 sequências de DNAmt na EMPOP passa a contar com mais 80 indivíduos/sequências. Cabo Verde que não tinha qualquer representação na EMPOP passa a contar com 103

indivíduos/sequências. Moçambique que também não tinha qualquer representação na EMPOP passa a contar com 83 indivíduos/sequências.

De entre os 34 617 haplótipos de DNAmt existentes na EMPOP, 10 028 são de indivíduos europeus. De entre estes somente 733 são de indivíduos portugueses e de entre os portugueses só 64 são de indivíduos da região sul de Portugal onde se inclui Lisboa. Com o contributo do nosso estudo a EMPOP passará a contar com mais 439 indivíduos/sequências de DNAmt de Portugal. Não obstante, sabemos que os indivíduos que estudamos têm ancestralidade africana. Seria então importante assegurar uma representação robusta da população portuguesa em geral e da população de Lisboa em particular na EMPOP.

Não podemos, contudo, deixar de sublinhar que muito embora os indivíduos que estudamos tenham linhagens maternas com ancestralidade africana, são efetivamente, na atualidade e do ponto de vista legal, elementos da população de Lisboa. Desta forma, e caso a integração destes grupos de imigrantes no seio dos indivíduos da população de acolhimento se venha a consolidar, será possível num futuro eventualmente não muito longíncuo, termos indivíduos nativos de Lisboa com caraterísticas fenotípicas caucasianas/europeias e simultaneamente DNAmt tipicamente africano e mesmo de haplogrupos L0 bem como indivíduos nativos de Lisboa com caraterísticas fenotípicas africanas e simultaneamente DNAmt tipicamente caucasiano/europeu.

A introdução de linhagens maternas africanas na Europa terá ocorrido maioritariamente como resultado do tráfico Atlântico de escravos empreendido no início do século XVII, tendo Portugal desempenhado um papel fulcral quer no

referido tráfico de escravos, quer na introdução de escravos africanos na Europa via Portugal, quer na dispersão posterior das linhagens africanas a partir do seu território para outras regiões na Europa (Malyarchuk et al., 2008). Por outro lado, a invasão e conquista medievais da Península Ibérica pelos árabes/berberes terá também constituído um importante contributo para a introdução de linhagens africanas de DNAmt Ibéria (A Salas et al., 2004).

As linhagens africanas de DNAmt, globalmente, representam cerca de 1% das linhagens de DNAmt das populações euroasiáticas (Antonio Salas et al., 2002). No entanto, em Portugal, e de acordo com o estudo de Marques e colaboradores, 2015, a representação de linhagens africanas de DNAmt na população portuguesa é superior à descrita para a generalidade das populações europeias (Marques et al., 2015). Neste estudo, que incidiu sobre 231 indivíduos portugueses, foram observadas linhagens africanas de DNAmt com uma frequência média, a nível nacional, na ordem dos 5%, sendo a frequência de linhagens africanas relativamente mais elevada no Sul de Portugal comparativamente ao Norte e Centro do país. No Sul de Portugal a frequência de linhagens africanas chega a ultrapassar os 6%. Os haplótipos africanos que identificaram na população portuguesa distribuem-se pelos haplogrupos L0a, L1b, L2a, L3b, L3d, L3e e L3f.

Além dos haplótipos africanos identificados por Marques e colaboradores, em 2015, na população de Portugal, no nosso estudo, na população de Lisboa, Portugal, identificamos imigrantes africanos, atualmente integrados na população de Lisboa, com haplótipos distribuídos pelos haplogrupos L0d, L1c, L2b, L2c, L3a, L3h, L3k, L4b e L5a, de onde resulta a confirmação do fenómeno migratório como importante,

indiscutível e incontornável contributo para o aumento da diversidade genética da população de Lisboa.

L0 foi a primeira ramificação do macrohaplogrupo L e terá ocorrido há cerca de 140/160 000 anos (Rosa & Brehm, 2011), estando predominantemente associado a regiões Sul, Este e Sudeste de África (Atkinson et al., 2009; Gonder et al., 2007). Subdividiu-se no haplogrupos L0a, L0b, L0d, L0f e L0k (van Oven & Kayser, 2008). O haplogrupo L0d, identificado no nosso estudo (1 imigrante oriundo de Angola e 6 imigrantes de Moçambique), terá sido o primeiro haplogrupo a divergir do haplogrupo L0, há cerca de 100 000 anos, e tem sido associado, quase exclusivamente, a populações Khoisan e identificado em Moçambique, Tanzânia e Angola (Atkinson et al., 2009; Behar et al., 2008; Gonder et al., 2007; Toomas Kivisild et al., 2004; Rosa & Brehm, 2011; A Salas et al., 2004; Schlebusch et al., 2013).

O haplogrupo L1 surgiu há cerca de 140/150 000 anos nas regiões Centro e Ocidente de África e subdivide-se nos haplogrupos L1b e L1c (Atkinson et al., 2009; Gonder et al., 2007; Toomas Kivisild et al., 2004; Rosa & Brehm, 2011). O haplogrupo L1c, identificado no nosso estudo (35 imigrantes de Angola, 4 de Moçambique, 5 de Cabo Verde e 3 da Guiné-Bissau) é um haplogrupo associado a populações Bantu, surgindo a Sudeste e Sudoeste de África em resultado das migrações do povo Bantu (Batini et al., 2007; Beleza et al., 2005; Gonder et al., 2007; Rosa & Brehm, 2011). O haplogrupo L1c encontra-se também no Sul da América, concretamente em indivíduos afroamericanos, em resultado do tráfico transatlântico de escravo (Batini et al., 2007; Plaza et al., 2004; Antonio Salas et al., 2002).

O haplogrupo L2 terá surgido no Oeste de África há cerca de 90/105 000 anos, subdividindo-se nos haplogrupos L2a, L2b,

L2c, L2d e L2e (Rosa & Brehm, 2011; van Oven & Kayser, 2008). Distribui-se por todo o continente de África, tendo a sua dispersão a Sudeste e Centro-Oeste de África ocorrido em consequência do tráfico transariano de escravos (Harich et al., 2010; Silva et al., 2015; Tishkoff et al., 2007). Os haplogrupos L2b e L2c, identificados no nosso estudo (10 imigrantes de Angola, 20 imigrantes de Cabo Verde, 2 imigrantes de Moçambique e 13 imigrantes da Guiné-Bissau) , são haplogrupos também comuns em populações afroamericanas (Derek C. Johnson, Sadeep Shrestha, Howard W. Wiener, Robert Makowsky, Ashish Kurundkar, Craig M. Wilson, 2015).

O haplogrupo L3 surgiu no Este de África, há cerca de 60/75 000 anos, divergindo nos haplogrupos L3a a L3h, estando amplamente distribuído por todo o continente de África (Atkinson et al., 2009; Rosa & Brehm, 2011; Antonio Salas et al., 2002; van Oven & Kayser, 2008). Este haplogrupo originou os haplogrupos M e N que posteriormente originaram todos os restantes haplogrupos não L (Pala et al., 2012; Rosa & Brehm, 2011). L3a, identificado no nosso estudo (1 imigrante de Angola), é um haplogrupo comum no norte de África, identificado designadamente no Sudão, Núbia, Egipto, Marrocos, Algéria e Tunísia (Derek C. Johnson, Sadeep Shrestha, Howard W. Wiener, Robert Makowsky, Ashish Kurundkar, Craig M. Wilson, 2015; González et al., 2006). L3h, identificado no nosso estudo (2 imigrantes de Cabo Verde e 1 imigrante da Guiné-Bissau) é um haplogrupo que foi identificado pela primeira vez, em 2004, na Tunísia (Toomas Kivisild et al., 2004). Relativamente ao haplogrupo L3k, identificado no nosso estudo (3 imigrantes de Cabo Verde), com origem no Norte de África, tem sido associado a

indivíduos das populações do Líbano e Tunísia (Harich et al., 2010).

O haplogrupo L4 terá surgido antes do haplogrupo L3, também na região Este de África, estando representado em países como a Etiópia e populações berberes (Behar et al., 2008; Toomas Kivisild et al., 2004).O haplogrupo L4b, identificado no nosso estudo (2 imigrantes de Angola), atualmente encontra-se distribuído principalmente pelo Este de África e Próximo Oriente (Fernandes et al., 2015).

Relativamente ao haplogrupo L5 terá surgido há cerca de 110 000 anos na região este de África, e apesar de ser muito raro em regiões fora do Este africano, o haplogrupo L5a, identificado no nosso estudo (2 imigrantes da Guiné-Bissau), pode ser observado com alguma frequência em populações indígenas Bantu do Sudeste e em populações de países como em populações de países como Egipto, Sudão, Etiópia, Quénia, Ruanda e Tanzânia (Plaza et al., 2004).

Relativamente às distâncias genéticas observadas entre a população de acolhimento - Lisboa/Sul de Portugal -, e as populações de imigrantes estudadas - Cabo Verde, Angola, Moçambique e Guiné-Bissau -, os resultados obtidos com o estudo de marcadores autossómicos InDel acompanharam os resultados obtidos com o estudo de DNAmt. Os resultados obtidos mostram que as populações de imigrantes de Angola, Guiné-Bissau, Cabo Verde e Moçambique estão mais próximas entre elas e que se encontram separadas significativamente da população portuguesa. De entre as populações de imigrantes PALOP, Moçambique é a que está menos afastada da população portuguesa e pode, eventualmente, constituir um cluster individual. Angola, Guiné-Bissau e Cabo Verde estão assim incluídas noutro cluster, aliás como esperado tendo em

conta a expansão das tribos Bantu ao longo das duas regiões distintas de África - Oeste e Este -. Portanto, a expansão para Oeste associada a Angola, Guiné-Bissau e Cabo Verde e a expansão para Este associada a Moçambique.

Referências Bibliográficas

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., … Torroni, A. (2004). The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. *The American Journal of Human Genetics*, *75*(5), 910–918. https://doi.org/10.1086/425590

Allen, M., Engström, A., Meyers, S., Handt, O., Saldeen, T., von Haeseler, A., … Gyllensten, U. (1998). Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *Journal of Forensic Sciences*, *43*, 453–464.

Alonso, A., Salas, A., Albarrán, C., Arroyo, E., Castro, A., Crespillo, M., … Gómez, J. (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. https://doi.org/10.1016/S0379-0738(01)00602-8

Alvarez, J. C., Johnson, D. L. E., Lorente, J. A., Martinez-espin, E., Martinez-Gonzalez, L. J., Allard, M., … Budowle, B. (2007). Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. *Legal Medicine*, *9*(6), 293–304. https://doi.org/10.1016/j.legalmed.2007.05.005

Atkinson, Q. D., Gray, R. D., & Drummond, A. J. (2009). Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1655), 367–373. https://doi.org/10.1098/rspb.2008.0785

Bandelt, H. J., Alves-Silva, J., Guimarães, P. E. M., Santos, M. S., Brehm, A., Pereira, L., … Pena, S. D. J. (2001). Phylogeography of the human mitochondrial haplogroup L3e: A snapshot of African prehistory and Atlantic slave trade. *Annals of Human Genetics*, *65*(6), 549–563. https://doi.org/10.1046/j.1469-1809.2001.6560549.x

Batini, C., Coia, V., Battaggia, C., Rocha, J., Pilkington, M. M., Spedini, G., … Calafell, F. (2007). Phylogeography of the human mitochondrial L1c haplogroup: Genetic signatures of the prehistory of Central Africa. *Molecular Phylogenetics and Evolution*, *43*(2), 635–644. https://doi.org/10.1016/j.ympev.2006.09.014

Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., … Rosset, S. (2008). The dawn of human matrilineal diversity. *American Journal of Human Genetics*, *82*(5), 1130–1140. https://doi.org/10.1016/j.ajhg.2008.04.002

Beleza, S., Gusmão, L., Amorim, A., Carracedo, A., & Salas, A. (2005). The genetic legacy of western Bantu migrations. *Human Genetics*, *117*(4), 366–375. https://doi.org/10.1007/s00439-005-1290-3

Bermisheva, M. A., Kutuev, I. A., Korshunova, T. Y., Dubova, N. A., Villems, R., & Khusnutdinova, E. K. (2004). Phylogeographic analysis of mitochondrial DNA in the Nogays: A strong mixture of maternal lineages from Eastern and Western Eurasia. *Molecular Biology*, *38*(4), 516–523. https://doi.org/10.1023/B:MBIL.0000037003.28999.45

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., … Comas, D. (2009). Genetic and demographic implications of the bantu expansion: Insights from human paternal lineages. *Molecular Biology and Evolution*, *26*(7), 1581–1589. https://doi.org/10.1093/molbev/msp069

Brandst??tter, A., Zimmermann, B., Wagner, J., G??bel, T., R??ck, A. W., Salas, A., … Parson, W. (2008). Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East. *BMC Evolutionary Biology*, *8*(1), 1–13. https://doi.org/10.1186/1471-2148-8-191

Brehm, A., Pereira, L., Bandelt, H. J., Prata, M. J., & Amorim, A. (2002). Mitochondrial portrait of the Cabo Verde archipelago: The Senegambian outpost of Atlantic slave trade. *Annals of Human Genetics*, *66*(1), 49–60. https://doi.org/10.1017/S0003480001001002

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Adler, C. J., Richards, S. M., … Genographic, T. (2014). Europe PMC Funders Group Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. https://doi.org/10.1038/ncomms2656.Neolithic

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*(1), 119–141. https://doi.org/10.1146/annurev.genom.4.070802.110352

Budowle, B., Wilson, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., Holland, M. M., & Monson, K. L. (1999). Mitochondrial DNA regions HVI and HVII population data. *Forensic Science International*, *103*(1), 23–35. https://doi.org/10.1016/S0379-0738(99)00042-0

Cann, R. L., Stoneking, M., & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature*, *325*(6099), 31–36. https://doi.org/10.1038/325031a0

Carvalho, A., & Pinheiro, M. F. (2013). Population data of 30 insertion/delection polymorphisms from a sample taken in the North of Portugal. *International Journal of Legal Medicine*, *127*(1), 65–67. https://doi.org/10.1007/s00414-012-0693-7

Carvalho, M., Brito, P., Bento, A. M., Gomes, V., Antunes, H., Costa, H. A., … Gusmão, L. (2011). Paternal and maternal lineages in Guinea-Bissau population. *Forensic Science International: Genetics*, *5*(2), 114–116. https://doi.org/10.1016/j.fsigen.2010.10.007

Chen, W., & Zhang, L. (2015). The pattern of DNA cleavage intensity around indels. *Scientific Reports*, *5*, 8333.

https://doi.org/10.1038/srep08333
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. *The American Journal of Human Genetics*, *66*(4), 1362–1383. https://doi.org/10.1086/302848
Coia, V., Cipollini, G., Anagnostou, P., Maixner, F., Battaggia, C., Brisighelli, F., … Zink, A. (2016). Whole mitochondrial DNA sequencing in Alpine populations and the genetic history of the Neolithic Tyrolean Iceman. *Scientific Reports*, *6*(January), 1–7. https://doi.org/10.1038/srep18932
Derek C. Johnson, Sadeep Shrestha, Howard W. Wiener, Robert Makowsky, Ashish Kurundkar, Craig M. Wilson, and B. A. (2015). Mitochondrial DNA diversity in the African American population. *Mitochondrial DNA*, *26*(3), 445–451. https://doi.org/10.1002/nbm.3369.Three
Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x
Fendt, L., Röck, A., Zimmermann, B., Bodner, M., Thye, T., Tschentscher, F., … Parson, W. (2012). MtDNA diversity of Ghana: A forensic and phylogeographic view. *Forensic Science International: Genetics*, *6*(2), 244–249. https://doi.org/10.1016/j.fsigen.2011.05.011
Fernandes, V., Triska, P., Pereira, J. B., Alshamali, F., Rito, T., Machado, A., … Pereira, L. (2015). Genetic stratigraphy of key demographic events in Arabia. *PLoS ONE*, *10*(3), 1–27. https://doi.org/10.1371/journal.pone.0118625
Ferreira Palha, T. D. J. B., Ribeiro Rodrigues, E. M., Cavalcante, G. C., Marrero, A., De Souza, I. R., Seki Uehara, C. J., … Santos, S. (2015). Population genetic analysis of insertion-deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit. *Forensic Science International: Genetics*, *19*, 10–14. https://doi.org/10.1016/j.fsigen.2015.03.015
Finnilä, S., Lehtonen, M. S., & Majamaa, K. (2001). Phylogenetic Network for European mtDNA. *The American Journal of Human Genetics*, *68*(6), 1475–1484. https://doi.org/10.1086/320591
Gill, P., Ivanov, P. L., Kimpton, C., Piercy, R., Benson, N., Tully, G., … Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics*, *6*(2), 130–135. https://doi.org/10.1038/ng0294-130
Goncaşalves, R., Rosa, A., Freitas, A., Fernandes, A., Kivisild, T., Villems, R., & Brehm, A. (2003). Y-chromosomes lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. *Human Genetics*, *113*(6), 467–472. https://doi.org/10.1007/s00439-003-1007-4
Gonder, M. K., Mortensen, H. M., Reed, F. A., De Sousa, A., & Tishkoff, S. A. (2007). Whole-mtDNA genome sequence analysis of ancient african lineages. *Molecular Biology and Evolution*,

24(3), 757–768. https://doi.org/10.1093/molbev/msl209
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*(5), 631–657. https://doi.org/10.1111/j.1469-1809.2006.00259.x
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*(1), 1–18. https://doi.org/10.1186/1471-2148-10-138
Harrub, B., Ph, D., & Thompson, B. (2010). The Demise Of Mitochondrial Eve. *Homo*, (October), 1–7.
Holland, M. M., Fisher, D. L., Mitchell, L. G., Rodriquez, W. C., Canik, J. J., Merril, C. R., & Weedn, V. W. (1993). Mitochondrial DNA Sequence Analysis of Human Skeletal Remains: Identification of Remains from the Vietnam War. *Journal of Forensic Sciences*, *38*(3), 13439J. https://doi.org/10.1520/JFS13439J
Howell, N., Kubacka, I., & Mackey, D. A. (1996). How rapidly does the human mitochondrial genome evolve? *American Journal of Human Genetics*, *59*(3), 501–509.
INE. (2014). *Anuário Estatístico de Portugal - 2014*. *AEP 2014*. https://doi.org/10.1007/s13398-014-0173-7.2
Ingman, M., Kaessmann, H., Paabo, S., & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans, *408*(6813), 708–713.
Ivanov PL, Wadhams MJ, Roby RK, Holland MM, W. V. & P. T. (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet.*, (12), 417–420. https://doi.org/10.1038/ng0496-417
Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investigative Genetics*, *6*(1), 1–10. https://doi.org/10.1186/s13323-015-0022-2
Kivisild, T., Bamshad, M., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., … Villems, R. (1999). Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. *Current Biology*, *9*(22), 1331–1334. https://doi.org/10.1016/S0960-9822(00)80057-3
Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., … Villems, R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *American Journal of Human Genetics*, *75*(5), 752–770. https://doi.org/10.1086/425161
LaRue, B. L., Ge, J., King, J. L., & Budowle, B. (2012). A validation study of the Qiagen Investigator DIPplex kit; An INDEL-based assay for human identification. *International Journal of Legal Medicine*, *126*(4), 533–540. https://doi.org/10.1007/s00414-012-

0667-9
Li, C. T., Zhang, S. H., & Zhao, S. M. (2011). Genetic analysis of 30 InDel markers for forensic use in five different Chinese populations. *Genetics and Molecular Research : GMR*, *10*(2), 964–979. https://doi.org/10.4238/vol10-2gmr1082
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. https://doi.org/10.1016/j.ics.2005.10.055
Lutz S, Weisser HJ, Heizmann J, P. S. (1996). mtDNA as a tool for identification of human remains Identification using mtDNA. *International Journal of Legal Medicine*, *109*, 205–209.
Mairal, Q., Santos, C., Silva, M., Marques, S. L., Ramos, A., Aluja, M. P., … Alvarez, L. (2013). Linguistic isolates in Portugal: insights from the mitochondrial DNA pattern. *Forensic Science International. Genetics*, *7*(6), 618–623. https://doi.org/10.1016/j.fsigen.2013.08.009
Malyarchuk, B. A., 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), 1091–1096. https://doi.org/10.1038/ejhg.2008.70
Marques, S. L., Goios, A., Rocha, A. M., Prata, M. J., Amorim, A., Gusmão, L., … Alvarez, L. (2015). Portuguese mitochondrial DNA genetic diversity—An update and a phylogenetic revision. *Forensic Science International: Genetics*, *15*, 27–32. https://doi.org/10.1016/j.fsigen.2014.10.004
Meng, H. T., Zhang, Y. D., Shen, C. M., Yuan, G. L., Yang, C. H., Jin, R., … Zhu, B. F. (2015). Genetic polymorphism analyses of 30 InDels in Chinese Xibe ethnic group and its population genetic differentiations with other groups. *Scientific Reports*, *5*, 1–6. https://doi.org/10.1038/srep08260
Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., & Devine, S. E. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research*, *16*(9), 1182–1190. https://doi.org/10.1101/gr.4565806
Neuvonen, A. M., Palo, J. U., Hedman, M., & Sajantila, A. (2012). Discrimination power of Investigator DIPplex loci in Finnish and Somali populations. *Forensic Science International: Genetics*, *6*(4), e99–e102. https://doi.org/10.1016/j.fsigen.2011.09.005
Ottoni, C., Primativo, G., Kashani, B. H., Achilli, A., Martínez-Labarga, C., Biondi, G., … Rickards, O. (2010). Mitochondrial haplogroup H1 in north Africa: An early Holocene arrival from Iberia. *PLoS ONE*, *5*(10), e13378. https://doi.org/10.1371/journal.pone.0013378
Page, R. D. M. (1996). Tree View: An application to display phylogenetic trees on personal computers. *Bioinformatics*, *12*(4), 357–358. https://doi.org/10.1093/bioinformatics/12.4.357
Pakendorf, B., & Stoneking, M. (2005). Mitochondrial Dna and Human

Evolution. *Annual Review of Genomics and Human Genetics*, *6*(1), 165–183. https://doi.org/10.1146/annurev.genom.6.080604.162249

Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., … Richards, M. B. (2012). Mitochondrial DNA signals of late glacial recolonization of europe from near eastern refugia. *American Journal of Human Genetics*, *90*(5), 915–924. https://doi.org/10.1016/j.ajhg.2012.04.003

Parson, W., & Dür, A. (2007). EMPOP—A forensic mtDNA database. *Forensic Science International: Genetics*, *1*(2), 88–92. https://doi.org/10.1016/j.fsigen.2007.01.018

Parson, W., Gusmão, L., Hares, D. R. R., Irwin, J. A. A., Mayr, W. R. R., Morling, N., … Parsons, T. J. J. (2014). DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Science International: Genetics*, *13*, 134–142. https://doi.org/10.1016/j.fsigen.2014.07.010

Pepinski, W., Abreu-Glowacka, M., Koralewska-Kordel, M., Michalak, E., Kordel, K., Niemcunowicz-Janica, A., … Konarzewska, M. (2013). Population genetics of 30 INDELs in populations of Poland and Taiwan. *Molecular Biology Reports*, *40*(7), 4333–4338. https://doi.org/10.1007/s11033-013-2521-7

PEREIRA, L., GUSMÃO, L., ALVES, C., AMORIM, A., & PRATA, M. J. (2002). Bantu and European Y-lineages in Sub-Saharan Africa. *Annals of Human Genetics*, *66*(5–6), 369–378. https://doi.org/10.1017/S0003480002001306

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*(5), 439–458. https://doi.org/10.1046/j.1469-1809.2001.6550439.x

Pereira, L., Richards, M., Goios, A., Alonso, A., Albarrán, C., Garcia, O., … Amorim, A. (2005). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Research*, *15*(1), 19–24. https://doi.org/10.1101/gr.3182305

Pereira, V., Gomes, V., Amorim, A., Gusmão, L., Prata, M. J., & João Prata, M. (2010). Genetic characterization of uniparental lineages in populations from Southwest Iberia with past malaria endemicity. *American Journal of Human Biology : The Official Journal of the Human Biology Council*, *22*(5), 588–595. https://doi.org/10.1002/ajhb.21049

Pfeiffer, H., Hühne, J., Ortmann, C., Waterkamp, K., & Brinkmann, B. (1999). Mitochondrial DNA typing from human axillary, pubic and head hair shafts - Success rates and sequence comparisons. *International Journal of Legal Medicine*, *112*(5), 287–290. https://doi.org/10.1007/s004140050251

Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, Á., & Comas, D. (2004). Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human*

Genetics, *115*(5), 439–447. https://doi.org/10.1007/s00439-004-1164-0

Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., … Roewer, L. (2014). A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International: Genetics*, *12*, 12–23. https://doi.org/10.1016/j.fsigen.2014.04.008

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*(6), 531–550.

Richards, M. (2000). Tracing European founder lineage in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.*, *67*, 1251–1276.

Roostalu, U., Kutuev, I., Loogväli, E. L., Metspalu, E., Tambets, K., Reidla, M., … Villems, R. (2007). Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: The Near Eastern and Caucasian perspective. *Molecular Biology and Evolution*, *24*(2), 436–448. https://doi.org/10.1093/molbev/msl173

Rosa, A., & Brehm, A. (2011). African human mtDNA phylogeography at-a-glance. *Journal of Anthropological Sciences*, *89*, 25–58. https://doi.org/10.4436/jass.89006

Rosa, A., Brehm, A., Kivisild, T., Metspalu, E., & Villems, R. (2004). MtDNA profile of West Africa Guineans: Towards a better understanding of the Senegambia region. *Annals of Human Genetics*, *68*(4), 340–352. https://doi.org/10.1046/j.1529-8817.2004.00100.x

Salas, A., Richards, M., De la Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., … Carracedo, A. (2002). The making of the African mtDNA landscape. *American Journal of Human Genetics, The*, *71*(5), 1082–1111. https://doi.org/10.1086/344348

Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., … Carracedo, A. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am. J. Hum. Genet*, *74*(3), 454–465.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schlebusch, C. M., Lombard, M., & Soodyall, H. (2013). MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa.

Seong, K. M., Park, J. H., Hyun, Y. S., Kang, P. W., Choi, D. H., Han, M. S., … Chung, K. W. (2014). Population genetics of insertion-deletion polymorphisms in South Koreans using Investigator DIPplex kit. *Forensic Science International: Genetics*, *8*(1), 80–

83. https://doi.org/10.1016/j.fsigen.2013.06.013
Silva, M., Alshamali, F., Silva, P., Carrilho, C., Mandlate, F., Jesus Trovoada, M., … Soares, P. (2015). 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. *Scientific Reports*, *5*(June), 1–13. https://doi.org/10.1038/srep12526
Stringer, C. (2003). New perspectives on the Neanderthals. *Evolutionary Anthropology: Issues, News, and Reviews*, *11*(S1), 58–59. https://doi.org/10.1002/evan.10057
Sullivan, K. M., Hopgood, R., & Gill, P. (1992). Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *International Journal of Legal Medicine*, *105*(2), 83–86. https://doi.org/10.1007/BF02340829
Teixeira, J. C., Nogueiro, I., Goios, A., Gusmão, L., Amorim, A., & Alvarez, L. (2011). Mitochondrial DNA-control region sequence variation in the NE Portuguese Jewish community. *Forensic Science International: Genetics Supplement Series*, *3*(1), e51–e52. https://doi.org/10.1016/j.fsigss.2011.08.025
Tishkoff, S. A., Gonder, M. K., Henn, B. M., Mortensen, H., Knight, A., Gignoux, C., … Mountain, J. L. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Molecular Biology and Evolution*, *24*(10), 2180–2195. https://doi.org/10.1093/molbev/msm155
Torroni, a, Richards, M., Macaulay, V., Forster, P., Villems, R., Norby, S., … Bandelt, H. J. (2000). mtDNA haplogroups and frequency patterns in Europe. *American Journal of Human Genetics*, *66*(3), 1173–1177. https://doi.org/10.1086/302789
Trovoada, M. J., Pereira, L., Gusmão, L., Abade, A., Amorim, A., & Prata, M. J. (2004). Insights from pattern of mtDNA variation into the genetic history of São Tomé e Prı□ncipe. *International Congress Series*, *1261*(C), 377–379. https://doi.org/10.1016/S0531-5131(03)01633-9
Turchi, C., Stanciu, F., Paselli, G., Buscemi, L., Parson, W., & Tagliabracci, A. (2016). The mitochondrial DNA makeup of Romanians: A forensic mtDNA control region database and phylogenetic characterization. *Forensic Science International: Genetics*, *24*, 136–142. https://doi.org/10.1016/j.fsigen.2016.06.013
van Oven, M., & Kayser, M. (2008). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(September), 386–394. https://doi.org/10.1002/humu.20921
Walsh, P. S., Metzger, D. A., & Higuchi, R. (2013). Chelex 100 as a Medium for Simple Extraction of DNA for PCR-Based Typing from Forensic Material. *BioTechniques*, *54*(3), 506–513. https://doi.org/10.2144/000114018
Wei, Y. L., Qin, C. J., Dong, H., Jia, J., & Li, C. X. (2014). A validation study of a multiplex INDEL assay for forensic use in four Chinese populations. *Forensic Science International: Genetics*, *9*(1), e22–e25. https://doi.org/10.1016/j.fsigen.2013.09.002

Wilson, M. R., DiZinno, J. A., Polanskey, D., Replogle, J., & Budowie, B. (1995). Validation of mitochondrial DNA sequencing for forensic casework analysis. *International Journal of Legal Medicine*, *108*, 68–74. https://doi.org/10.1007/BF01369907

Zidkova, A., Horinek, A., Kebrdlova, V., & Korabecna, M. (2013). Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population. *International Journal of Legal Medicine*, *127*(1), 7–10. https://doi.org/10.1007/s00414-011-0649-3

UNIVERSIDADE DE ÉVORA
INSTITUTO DE INVESTIGAÇÃO
E FORMAÇÃO AVANÇADA

Contactos:

Universidade de Évora

Instituto de Investigação e Formação Avançada - IIFA

Palácio do Vimioso | Largo Marquês de Marialva, Apart. 94

7002-554 Évora | Portugal

Tel: (+351) 266 706 581