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

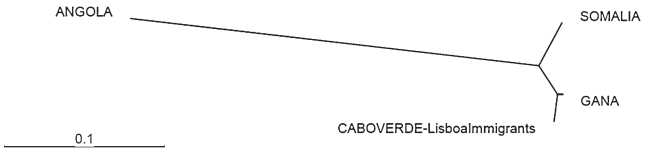

Figure 1. Phylogram derived from the analysis of genetic distances between selected populations from the literature and the studied population.

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