Duarte Nuno Vieira • Anthony Busuttil
Denis Cusack • Philip Beth
Editors

Acta Medicinae Legalis et Socialis

IMPRENSA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

ACADEMIA INTERNATIONALIS MEDICINAE LEGALIS

COIMBRA 2010

M. Carvalho[1], P. Brito[1], A. M. Bento[1], V. Gomes[2,3], H. Antunes[1], H. Afonso Costa[1], V. Lopes[1], A. Serra[1], F. Balsa[1], L. Andrade[1], M. J. Anjos[1], L. Gusmão[2], F. Corte-Real[4,5]

[1] *Genetic Forensic Service, Centre Branch of the National Institute of Legal Medicine, Coimbra, Portugal*
[2] *IPATIMUP – Institute of Molecular Pathology and Immunology, University of Oporto, Portugal*
[3] *Institute of Legal Medicine, Genomics Medicine Group, University of Santiago de Compostela, Spain*
[4] *National Institute of Legal Medicine, Coimbra, Portugal*
[5] *Faculty of Medicine, University of Coimbra, Portugal*

PATERNAL AND MATERNAL LINEAGES IN CABO VERDE ARCHIPELAGO POPULATION

Abstract: The paternal and maternal lineages of Cabo Verde archipelago were characterized using 22 Y-Single Nucleotide Polymorphisms (SNPs) and Y-minimal haplotype (STRs) for paternal lineage and the two hypervariable segments (HVI and HVII) of the mtDNA control region for maternal lineage. A high variability of haplotype and haplogroup composition was found in the studied population. A total of 13 haplogroups was found with Y-SNPs and 24 haplogroups with HVI and HVII mtDNA. Using Y-STR minimal haplotype information, genetic distances were obtained between Cabo Verde and European/African populations. While almost all mitochondrial lineages were of sub-Saharan origin (95%), the Y-chromosome lineages reveal a high diverse composition, with more than 57% of Y lineages of European ancestry.

Introduction

The analysis of mitochondrial DNA (mtDNA) and Single Nucleotide Polymorphisms (SNPs) located on the Y chromosome specific region can be helpful in forensics, since they define haplogroups showing geographic specificity, providing information about the paternal or maternal ancestry of an individual or evidence under investigation. Moreover, the study of lineage markers can be extremely useful in order to evaluate population substructure in admixed populations, which is essential for definition of relevant forensic databases.

The aim of this work was to study the origin of paternal lineage (Y-SNPs) and maternal lineage (mtDNA) of Cabo Verde archipelago population by phylogeographic analysis of the observed haplogroups. This population was genetically characterized for Y-STR minimal haplotype (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393 and DYS385). The genetic distances obtained were compared with those reported to other populations from Europe and Africa, in order to evaluate the contribution of these populations to the genetic pool of Cabo Verde.

Materials and Methods

The population of Cabo Verde was characterized, in a sample of 42 unrelated males, for Y chromosome specific STR loci typing the minimal haplotype DYS19, DYS389 I,

DYS389 II, DYS390, DYS391, DYS392, DYS393 and DYS385. The DYS19, DYS389 I, DYS389 II, DYS390 and DYS393 were amplified as described by Gusmão *et al*., 1999 [1]. The DYS385 amplification conditions complied with the methodology described by Schneider *et al.*, 1998 [2], and multiplex amplification of DYS391, DYS392, DYS393 was carried out according to Kloosterman *et al.*, 1998 [3]. Alleles were designated according to the International Society for Forensic Genetics (ISFG) guidelines for forensic analysis using Y-STRs [4]. Pairwise Rst genetic distances were calculated using Arlequin v. 3.0 software Excoffier *et al.*, 2005 [5], and DYS385 was not considered.

To determine the frequency distribution of the male lineages, 22 Y-SNPs were typed in two PCR-SNaPshot multiplex reactions. The first multiplex included nine Y-SNPs (92R7, M70, M22, Tat, P25, SRY10831, M173, M213 and M9) and the second one included thirteen Y-SNPs (P2, M154, M293, M81, M85, M78, M35, M96, V6, M191, M33, M123 and M2). The first multiplex was performed according to Brión *et al.,* 2004 [6] and the second one by using a newly developed strategy.

The polymorphism of the two hypervariable segments (HVI and HVII) of the mtDNA control region was analyzed in 77 unrelated individuals from Cabo Verde, using the amplification method and *primers* referred by Wilson *et al.,* 1995 [7]. Sequences were obtained with ABI PRISM *Big Dye Terminator and dRhodamine Terminator Cycle Sequencing Ready Reaction Kits,* with *amplitaq DNA polymerase FS,* and were detected in an ABI 3100 Avant sequencer. Haplogroups were classified based on the different polymorphic positions of these two hypervariable regions using the software *mtDNA manager-forensic mtDNA database* [8].

Results and Discussion

The Y-chromosome haplotype and haplogroup and mtDNA haplogroup frequencies observed in Cabo Verde population were presented in Tables 1, 2 and 3.

The pairwise Rst genetic distances analysis obtained among male population sample from Cabo Verde and others from Iberia and Africa sub-Saharan, showed a similar significant differentiation from Iberian populations ($0.13352<RST<0.17027$, $p=0.00000$) [9, 10, 11] and from Africa sub-Saharan populations ($0.13401<RST<0.15268$, $p=0.00000$) [12, 13, 14].

The two most frequent Y-haplogroups were R1b1-P25 with a frequency of 26.19% and E1b1a(xE1b1a4,7)-M2 with 28.57%. The haplogroup R1b1-P25 reaches high frequencies in Weastern Europe, being the most frequent in Iberia [15]. The other haplogroups that are usually represented in Western European populations [6, 15] were also found in our sample, namely E1b1b1a-M78, E1b1b1b-M81, K2-M70, J2-M172, P(xR1)-92R7, I2a2-M26, I(xI2a2)-M170 and G-M201. Altogether, the proportion of the lineages that can be explained by European contributions reaches a frequency of 57% in our sample from Cabo Verde. Haplogroup E1b1a(xE1b1a4,7)-M2 is known to be of sub-Saharan origin, being the most frequent in all Bantu speaking populations [16], and was found to be the second most frequent in our sample.

For mtDNA, the majority of haplogroups (95%) were of sub-Saharan origin with exception of X and X2d that were of West Eurasian origin and D4k and N9b that were of East Asian origin [8].

Conclusion

Our results, like other recent studies [17, 18], confirm a strong male/female asymmetry concerning the European and African contributions to the genetic composition of the nowadays Cabo Verde population. While almost all mitochondrial lineages were of sub-Saharan origin (95%), the Y-chromosomes lineages reveal a high diverse composition, with more than 57% of Y lineages having an European ancestry.

References

[1]. GUSMÃO L., GONZALEZ-NEIRA A., PESTONI C., BRION M., LAREU M.V., CARRACEDO A. Robustness of the Y STRs DYS19, DYS389 I and II, DYS390 and DYS393: optimization of a PCR pentaplex. Forensic Sci. Int., 106, 163-172, 1999.

[2] SCHNEIDER P.M., MEUSER S., WAIYAWUTH W., SEO Y., RITTNER C. Tandem repeat structure of the duplicate chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. Forensic Sci. Int., 97, 61-70, 1998.

[3] KLOOSTERMAN A.D., POUWELS M., DASELAAR P., JANSEN H. J. T. Population genetic study of Y – chromosome specific STR- loci in Dutch Caucasians. Progress in Forensic Genetic: 7, 491 – 493, 1998.

[4] GUSMÃO L., BUTLER J.M., CARRACEDO A., GILL P., KAYSER M., MAYR W.R., MORLING N., PRINZ M., ROEWER L., TYLER-SMITH C., SCHNEIDER P.M. DNA Commission of International Society of Forensic Genetics (ISFG): an update of recommendations on the use of Y-STRs in forensic analysis. Forensic Sci. Int., 157, 187-197, 2006.

[5] EXCOFFIER L., LAVAL G. & SCHNEIDER S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, 1, 47-50, 2005.

[6]. brion M., sobrino b., blanco-verea a., lareu m.v., carracedo a. Hierarchical analysis of 30 Y-chromosome SNPs in European populations, Int. J. Legal Med., 119, 10-15, 2004.

[7]. wilson M.R., dizinno j.a., polanskey d., replogle j., budowle b. Validation of mitochondrial DNA sequencing for forensic casework, Int. J. Legal Med., 108, 68-74, 1995.

[8]. lee H. y., song i., ha e., cho s. – b., yang w.i., shin k. – j. mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control region sequences, BMC Bioinformatics, 9, 483, 2008.

[9]. Sánchez-Diz P., Alves C., Carvalho E., *et al.* Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study, Int. J. Legal Med. 122, 529-533, 2008.

[10]. Martín P., García-Hirschfeld J., García O. *et al.* A Spanish population study of 17 Y-chromosome STR loci, Forensic Sci. Int. 139, 231-235, 2004.

[11]. Bento A.M., Carvalho M., LOPES V., SERRA A., aFONSO-COSTA H., ANDRADE L., BALSA F., OLIVEIRA C., BATISTA L., GAMERO J., ANJOS M.J., GUSMÃO L., Corte Real F. Distribution of Y chromosomal haplotypes in the Central Portuguese population using 17-STRs. Forensic Sci. Int: Genetics, 2009 *(in press).*

[12] Arroyo-Pardo E., Gusmão L., López-Parra A.M., Baeza C., Mesa M.S., Amorim A. Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea (Central Africa). Forensic Sci. Int. 149, 109-113, 2005.

[13] Beleza S. Phylogenetic and demographic history of two human populations revealed by the analysis of two non-recombining segments of the genome: Y-chromosome and mitochondrial DNA. PhD thesis, University of Santiago de Compostela, Spain, 2005.

[14] Alves C., Gusmão L., Barbosa J., Amorim A. Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. Forensic Sci. Int., 134, 126-133, 2003.

[15]. BELEZA S., GUSMÃO L., LOPES A., ALVES C., GOMES I., GIOUZELI M., CALAFELL F., CARRACEDO A., AMORIM A. Microphylogeographic and demographic history of Portuguese male lineages. Ann. Hum. Genet., 70, 181-194, 2006.

[16]. BELEZA S., GUSMÃO L., AMORIM A., CARRACEDO A., SALAS A. The genetic legacy of western Bantu migrations. Hum. Genet., 117, 366-375, 2005.

[17]. gonçalves r., rosa a., freitas a., fernandes a., kivisild T., villems r., brehm a. Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers, Hum. Genet., 113, 467-472, 2003.

[18]. brehm a., pereira l., bandelt h.- j., prata m.j., amorim a. Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade, Ann. Hum. Genet., 66, 49 – 60, 2002.

Frequency	Haplotype*	Haplogroup
3	14/10/26/24/11/13/13/11-14	R1b1-P25
2	13/12/29/24/10/11/13/17-18	E1b1b1a-M78
2	14/13/29/23/11/13/13/11-14	R1b1-P25
2	15/12/29/22/11/11/13/11-18	E1a-M33
2	15/13/30/21/10/11/14/15-16	E1b1a (xE1b1a4,7)-M2
1	13/13/29/24/9/11/13/10-14	E1b1b1b-M81
1	13/13/30/24/11/11/13/16-18	E1b1b1a-M78
1	13/13/31/22/10/11/13/14-16	E1b1b1a-M78
1	13/13/31/24/10/14/13/15-16	P(xR1)-92R7
1	14/12/28/24/10/12/13/11-17	B2b(xB2b3)-M112
1	14/12/30/23/10/11/12/13-17	J2-M172
1	14/13/29/24/11/13/13/10-14	R1b1-P25
1	14/13/29/24/11/14/13/12-14	R1b1-P25
1	14/13/29/24/12/13/14/11-13	R1b1-P25
1	14/13/30/24/11/13/13/11-14	R1b1-P25
1	14/13/31/23/10/15/11/15-18	K2-M70
1	14/14/30/24/11/13/13/11-14	R1b1-P25
1	14/14/31/23/9/11/13/10-14	E1b1b1b-M81
1	15/11/27/22/9/11/13/11-12	A3b2-M13
1	15/12/26/24/11/13/13/12-14	R1b1-P25
1	15/12/29/21/10/11/14/17-17	E1b1a (xE1b1a4,7)-M2
1	15/12/30/25/11/11/13/16-18	E1b1b1a-M78
1	15/13/30/21/10/11/14/14-16	E1b1a (xE1b1a4,7)-M2
1	15/13/30/21/10/11/14/15-15	E1b1a (xE1b1a4,7)-M2

1	15/13/30/21/10/11/14/16-17	E1b1a (xE1b1a4,7)-M2
1	15/13/30/22/10/11/14/13-13	G-M201
1	15/14/29/23/10/11/13/12-12	I2a2-M26
1	15/14/31/22/10/11/13/16-18	E1b1a (xE1b1a4,7)-M2
1	16/12/29/22/10/11/13/15-15	E1a-M33
1	16/12/30/22/11/11/13/15-16	E1a-M33
1	16/14/30/21/10/11/14/18-18	E1b1a (xE1b1a4,7)-M2
1	16/14/33/21/10/11/14/15-17	E1b1a (xE1b1a4,7)-M2
1	17/12/29/21/10/11/14/17-17	E1b1a (xE1b1a4,7)-M2
1	17/12/30/25/11/11/14/13-17	I(xI2a2)-M170
1	17/14/31/21/10/11/13/16-17	E1b1a (xE1b1a4,7)-M2
1	17/14/31/21/10/11/13/17-17	E1b1a (xE1b1a4,7)-M2

*The minimal haplotype: DYS19/DYS389I/II/DYS390/DYS391/DYS392/DYS393/DYS385

Table 1. Y-SNP Haplotype frequencies and haplogroup in Cabo Verde archipelago (N=42)

Haplogroup	Cabo Verde (N=42)
E1b1a (xE1b1a4,7)-M2	28.57
R1b1-P25	26.19
E1b1b1a-M78	11.90
E1a-M33	9.52
E1b1b1b-M81	4.76
K2-M70	2.38
J2-M172	2.38
P(xR1)-92R7	2.38
I2a2-M26	2.38
I(xI2a2)-M170	2.38
A3b2-M13	2.38
B2b(xB2b3)-M112	2.38
G-M201	2.38

Table 2. Y-SNP Haplogroup frequencies (%) in Cabo Verde archipelago

S	H	HVI and HVII Specific Control Region Sequences*	N
H1	*D4k*	73G 93G 146C 150T 152C 182T 195C 198T 204C 309.1C 325T 16192T 16223T 16278T 16390A	1
H2	*L0a1*	93G 152C 189G 200G 236C 247A 309.1C 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T	1
H3	*L1b*	73G 152C 182T 185T 189G 195C 247A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C	1
H4	*L1b1*	73G 152C 182T 185T 195C 247A 357G 16114G 16126C 16187T 16188T 16189C 16223T 16264T 16270T 16278T 16293G 16311C	1

S	H	HVI and HVII Specific Control Region Sequences*	N
H5	*L1b1*	73G 152C 182T 185T 195C 247A 367G 16126C 16145A 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C	1
H6	*L1b1*	73G 152C 182T 185T 195C 247A 16104T 16187T 16189C 16223T 16270T 16278T 16289G 16293G 16311C	1
H7	*L1b1*	73G 152C 182T 185T 195C 247A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C	1
H8	*L1b1*	73G 152C 182T 185T 195C 247A 16126C 16187T 16189C 16223T 16256T 16264T 16278T 16293G 16311C	1
H9	*L1b1*	73G 152C 182T 185T 195C 247A 16114G 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C	1
H10	*L1c*	73G 151T 152C 182T 186A 189C 195C 247A 291T 297G 316A 16129A 16187T 16189C 16223T 16248T 16261T 16278T 16311C 16360T	1
H11	*L1c1*	73G 151T 152C 182T 186A 189C 195C 198T 247A 316A 16129A 16163G 16187T 16189C 16223T 16278T 16293G 16294T 16304C 16311C 16360T	1
H12	*L1c1*	73G 151T 152C 182T 186A 189C 247A 316A 16129A 16163G 16187T 16189C 16223T 16278T 16293G 16294T 16304C 16311C 16360T	1
H13	*L1c1*	73G 151T 152C 182T 186A 189C 195C 198T 247A 316A 16129A 16163G 16187T 16189C 16223T 16278T 16293G 16294T 16304C 16311C 16360T	1
H14	*L1c1*	73G 151T 152C 182T 186A 189C 195C 247A 316A 16129A 16163G 16187T 16189C 16223T 16278T 16293G 16294T 16304C 16311C 16360T	2
H15	*L2a*	73G 146C 152C 195C 16223T 16230G 16278T 16294T 16390A	1
H16	*L2a1*	73G 143A 146C 152C 195C 16111T 16223T 16278T 16294T 16309G 16390A	1
H17	*L2a1*	73G 143A 146C 152C 195C 309.1C 16183C 16189C 16223T 16274A 16278T 16294T 16309G 16390A	1
H18	*L2a1*	73G 143A 146C 152C 195C 198T 16086C 16223T 16278T 16294T 16309G 16390A	1
H19	*L2a1*	73G 143A 146C 152C 195C 309.2C 16093C 16189C 16223T 16264T 16278T 16294T 16309G 16390A	1
H20	*L2a1*	73G 143A 146C 152C 195C 264T 16183C 16189C 16192T 16223T 16278T 16294T 16309G 16390A	1
H21	*L2a1*	73G 143A 146C 152C 195C 16111T 16223T 16278T 16294T 16309G 16390A	2
H22	*L2b*	73G 146C 150T 152C 182T 195C 198T 207A 16093C 16114A 16129A 16213A 16223T 16271C 16278T 16390A	1
H23	*L2b1*	73G 150T 152C 182T 195C 198T 204C 249d 309.1C 16114A 16129A 16213A 16223T 16278T 16355T 16362C 16390A	1
H24	*L2b1*	73G 150T 152C 182T 195C 198T 204C 16114A 16129A 16213A 16223T 16278T 16355T 16362C 16390A	1
H25	*L2c*	73G 93G 146C 150T 152C 182T 195C 325T 16223T 16278T 16311C 16390A	1
H26	*L2c*	73G 89C 93G 146C 150T 152C 182T 195C 198T 325T 16192T 16223T 16278T 16390A	1
H27	*L2c*	73G 93G 146C 150T 152C 182T 195C 198T 325T 16178C 16223T 16278T 16380T 16390A	1
H28	*L2c*	73G 89C 93G 146C 150T 152C 182T 195C 198T 309.1C 325T 16192T 16223T 16261T 16278T 16390A	1
H29	*L2c*	73G 93G 146C 150T 152C 182T 195C 198T 325T 16051G 16223T 16278T 16390A	1
H30	*L2c*	73G 146C 150T 152C 182T 195C 297G 325T 16177G 16223T 16278T 16311C 16390A	1
H31	*L2c*	73G 93G 146C 150T 152C 182T 185C 189G 325T 16223T 16278T	1
H32	*L2c*	73G 93G 146C 150T 152C 182T 183G 195C 198T 199C 204C 325T 16223T 16278T 16390A	1
H33	*L2c*	73G 93G 146C 150T 152C 182T 195C 198T 309.1C 319C 325T 16051G 16223T 16278T 16390A	1

S	H	HVI and HVII Specific Control Region Sequences*	N
H34	*L2c*	73G 146C 150T 152C 182T 195C 198T 325T 16223T 16261T 16278T 16390A	2
H35	*L2c*	73G 93G 146C 150T 152C 182T 195C 198T 204C 325T 16192T 16223T 16278T 16390A	1
H36	*L2c*	73G 93G 146C 150T 152C 182T 195C 198T 309.1C 325T 16223T 16278T 16390A	1
H37	*L2c*	73G 146C 150T 152C 182T 195C 325T 16177G 16223T 16278T 16311C 16390A	2
H38	*L2c*	73G 93G 146C 150T 152C 182T 195C 198T 325T 16223T 16278T 16320T 16390A	1
H39	*L2c2*	73G 93G 146C 150T 152C 182T 195C 198T 325T 16084A 16093C 16220G 16223T 16264T 16278T 16311C 16390A	2
H40	*L2c2*	73G 93G 146C 150T 152C 182T 195C 325T 16093C 16126C 16223T 16264T 16274A 16278T 16390A	1
H41	*L2d1*	73G 146C 150T 152C 195C 310C 16129A 16182G 16183C 16193.1C 16278T 16300G 16354T 16390A	1
H42	*L2d1*	73G 146C 150T 195C 16093C 16129A 16189C 16259T 16278T 16300G 16354T 16390A	1
H43	*L2d2*	73G 146C 151T 152C 182T 185A 189G 16111A 16145A 16184T 16189C 16223T 16239T 16278T 16291T 16292T 16355T 16390A	1
H44	*L3b*	73G 309.1C 16124C 16223T 16234T 16278T 16362C	2
H45	*L3b*	73G 151T 152C 16124C 16223T 16234T 16278T 16362C	3
H46	*L3b*	73G 150T 16124C 16183C 16189C 16214T 16223T 16278T 16362C	1
H47	*L3b*	73G 309.1C 16124C 16188T 16223T 16278T 16362C	1
H48	*L3b*	73G 189G 16124C 16223T 16278T 16355T 16362C	1
H49	*L3d*	73G 152C 189G 195C 207A 16124C 16223T	2
H50	*L3d*	73G 150T 152C 16124C 16223T	2
H51	*L3d*	73G 146C 152C 16093C 16124C 16223T	1
H52	*L3d*	73G 152C 199C 309.1C 16111T 16124C 16223T	1
H53	*L3d1*	73G 146C 152C 195C 16124C 16223T 16319A	1
H54	*L3e2b*	73G 150T 195C 16172C 16183C 16189C 16223T 16259T 16320T	1
H55	*L3e4*	73G 150T 16051G 16223T 16264T 16299G	1
H56	*L3e4*	73G 150T 309.1C 16051G 16223T 16264T 16299G	1
H57	*L3e4*	73G 150T 16051G 16093C 16223T 16247G 16264T 16311C	1
H58	*L3e4*	73G 150T 309.1C 16051G 16223T 16264T	1
H59	*L3e4*	73G 150T 309.1C 16051G 16223T 16257T 16264T	1
H60	*L3e4*	73G 150T 16051G 16148T 16223T 16264T	1
H61	*L3e4*	73G 150T 257G 16051G 16223T 16264T	1
H62	*L3f1*	73G 189G 16153A 16209C 16223T 16230G 16260T 16292T	1
H63	*N9b*	73G 150T 195C 16172C 16182G 16183C 16189C 16223T	1
H64	*U6a*	73G 150T 309.1C 16172C 16183C 16189C 16219G 16278T	2
H65	*X*	73G 146C 152C 185A 189G 16111A 16145A 16183T 16189C 16223T 16239T 16278T 16292T 16355T 16390A	1
H66	*X2d*	73G 146C 182T 195A 207A 316A 16189C 16223T 16259T 16274A 16278T 16390A	1

* All sequences (S) carry 263/315.1C mutations/insertions.

Table 3. mtDNA haplogroup (H) frequencies in Cabo Verde archipelago (N=77).