

Genetic Differentiation of the Cabo Verde Archipelago Population Analysed by STR Polymorphisms

A. T. Fernandes^a, R. Velosa^a, J. Jesus^a, A. Carracedo^b and A. Brehm^{a,*}

^aHuman Genetics Laboratory, Centre of Macaronesian Studies, University of Madeira, Penteada, 9000 Funchal, Portugal

^bInstitute of Legal Medicine, Faculty of Medicine, University of Santiago de Compostela, E-15705 Santiago de Compostela, Spain

Summary

Allele frequencies for 17 STR loci were analyzed in a sample of unrelated males from the Cabo Verde Archipelago. The samples were gathered in such a way that the origin of the subjects was perfectly identified, and they could be included in one of the leeward or windward groups of islands. This study reveals that there are significant differences between both groups of islands, and between Cabo Verdeans and other populations from sub-Saharan Africa including the Guineans, the most probable source population for Cabo Verdeans. This study confirms mtDNA data and, together with HLA and Y chromosome data already published, shows that the Cabo Verde population is sub-structured and atypical, diverging substantially from mainland sub-Saharan populations. Overall these differences are most probably due to admixture between sub-Saharan slaves brought into the islands and other settlers of European origin. In the absence of a clear indication of a different ethnic composition of the first sub-Saharan settlers of Cabo Verde, the differentiation exhibited in both groups of islands can be most probably be attributed to genetic drift.

Keywords: STRs, Cabo Verde, population genetics, Powerplex[®] 16 System Kit, Promega

Introduction

In recent years we have seen an exponential increase of STR (Short Tandem Repeat) data, usually presented as allele frequencies that may be later used as databases for forensic and population genetics purposes (Holt *et al.* 2000, and see references therein). Differences in STR allele frequencies among populations can correctly reveal their genetic relationships in those cases where collection samples are properly gathered, and only then do genetic profiles really represent the populations under study. We present a population study of Cabo Verdeans, where subjects were chosen in such a way that the families of both parents were natives from the same island for at least three generations. In this way we tried to obtain a genetic profile that minimizes the effect of recent population movements between the islands belonging to the archipelago of Cabo Verde.

The Cabo Verde archipelago has a well-documented settlement history. It was discovered by the Portuguese in 1460 and served as a base for exploratory missions to the West African coast. With the onset of the slave trade, Cabo Verde was pivotal in the complex transatlantic commercial network that quickly followed (Russell-Wood, 1998; Godinho, 1965). Slaves were captured mainly in the Senegâmbia region (between Senegal and Guinea), brought into the islands and redistributed elsewhere. The settlement of the islands is well documented and involved a few European males (mainly Iberians) and slaves. It is also known that European women were absent from the settlers, who soon made their own liaisons to slave women. Today's population of Cabo Verde is typically mulatto. Not all the islands were colonized simultaneously. The main island, Santiago, and nearby Fogo and Brava, were colonized first and only at the end of the 16th century and especially during the 17th century did the settlement of the remaining islands occur (Barcellos, 1899).

An early study of sero-anthropology in the Cabo Verde population revealed some differences among

*Author for correspondence: António Brehm, phone +351 291705383, fax +351 291705399, e-mail: brehm@uma.pt

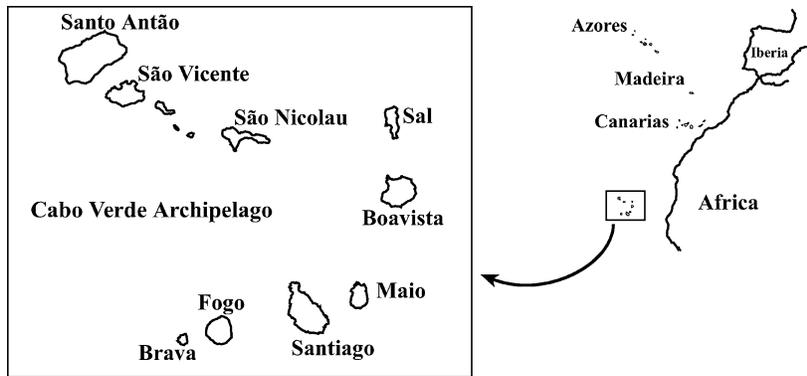


Figure 1 The archipelago of Cabo Verde. Santo Antão, São Vicente and São Nicolau constitute the north-western group of islands (CVN). The south-eastern group includes all the remaining islands (CVS).

islands (Lessa & Ruffié, 1960). The mtDNA genetic structure of the Cabo Verde population has also been recently studied and showed a sharp difference in haplogroup composition between the two groups of islands (Brehm *et al.* 2002) as well as at the HLA level (Spínola *et al.* 2002). Dios *et al.* (1998) presented a study on 4 STRs from autochthonous but randomly selected Cabo Verdeans, but found no sign of within population differentiation. The differences found in mtDNA between the two main groups of islands prompted us to perform a study using a new methodological approach in order to see if, on the basis of STR polymorphisms, the Cabo Verde population was also sub-structured. In the present study we use recently published data on the STR composition of Guiné-Bissau, a West African country and the region of putative origin of the slaves that were introduced to Cabo Verde following its discovery. Strong differences between the population of Cabo Verde and that of Guiné-Bissau are seen at the autosomal level, using both STRs and HLA data. Moreover, recent findings in the Y-chromosome haplogroup composition in these regions support the conclusion that Cabo Verdeans, although being of African origin, show in their gene pool a substantial amount of non-African input.

Materials and Methods

Samples

The study population consisted of unrelated males from Cabo Verde whose ancestors were known to be originally from one of the nine islands of the archipelago at

least back to three generations. Each subject gave oral consent after an interview where the purpose of the study was explained. The total number of individuals sampled varied for each locus between 161 and 328 (see Supplementary Material). Here we followed Brehm *et al.* (2002) and divided the archipelago of Cabo Verde into two island groups based on both historical and geographical grounds: the south-eastern group (CVS) comprising the islands of Brava, Fogo, Santiago, Maio, Boavista and Sal, and the north-western group (CVN) consisting of Santo Antão, São Vicente and São Nicolau (Figure 1). Therefore, population samples belonging to each of these groups were pooled in all data analysis.

DNA Amplification

The leukocyte fraction of whole blood was used to extract DNA by the Chelex method (Lareu *et al.* 1994). STR amplification was achieved using the new PowerPlex[®] 16 system Kit from Promega according to the manufacturer's instructions. This kit allows the amplification of 16 STR loci; one of them, the amelogenin locus, was excluded from the analysis since our samples are all from males. PCR products were run in an ABI-DNA Automated Sequencer (AB Applied Biosystems). The *CD4* and *FES* loci were amplified using primers and conditions described by Edwards *et al.* (1991) and Polymeropoulos *et al.* (1991), respectively, and were visualized by silver staining.

Although the 16 loci are coamplified, a different number of subjects were typed for each locus. This was due to the fact that when performing a big number of runs,

it is normal that by chance, some loci fail to amplify. The strategy in these cases is to amplify a different subject, bringing the number of the missing locus or loci to the number desired but simultaneously adding more and new data to the other loci.

Statistical Analysis

Allele frequencies for each STR were determined by direct gene counting. Expected and observed heterozygosity values, Hardy Weinberg equilibrium (HWE) and parameters of forensic interest were calculated for each marker in both populations: PD (power of discrimination and exclusion, Jones, 1972) and CE (*a priori* chance of exclusion, Ohno *et al.* 1982). Due to the fact that the sample size for some islands was small, no inter-island analysis of molecular variance was performed. The two biggest samples belonged to the two more populated islands, Santiago and Santo Antão. Each of the remaining populations can be unambiguously assigned to one of these two on historical grounds. Therefore, in the subsequent analysis we used only two gene pools corresponding to these two groups of islands. Cabo Verde STR frequencies were also compared with those from 11 other populations whose data sets contain at least 12 common loci (Table 1), including populations which were putatively in the origin of Cabo Verde settlers: Portuguese and Guineans. In order to estimate the relative contribution of Portuguese and Guinean gene

pools to the present day genetic composition of Cabo Verdeans, several admixture coefficients were computed using Admix1_0 (G. Bertorelle, www2.unife.it/genetica/Giorgio/docadmhtml). These estimators are m_Y , described in Bertorelle & Excoffier (1998), which depends on coalescence times in both parental populations and between these and the hybrid population; m_R , a frequency-based estimator described by Roberts & Hiorns (1965); and the maximum-likelihood estimator m_C , proposed by Chakraborty *et al.* (1992). In this case we used a pooled sample of North Portugal and Azores data (p_1) as one parental population, deliberately excluding Madeira to avoid possible bias due to the fact that the island had an important contribution of sub-Saharan slaves in its settlement. The Guinean sample was the other parental population (p_2) and Cabo Verde was naturally the “hybrid” population (h). Sampling errors for these coefficients were estimated using a bootstrap technique computed by randomly resampling the original data as implemented in the program.

Genotypic differentiation for each locus was assessed with G-like tests and for all *loci* by a χ^2 test under Fisher’s method, as implemented in Genepop v3.1d. Exact tests between all pairs of populations are according to Raymond & Rousset (1995) using Arlequin v2.000 (Schneider *et al.* 2000). An AMOVA on genotype frequencies in each population was performed to test the hierarchical partitioning of genetic variation within and

Table 1 Populations used in this study, codes, sample size and references

Population	Code (n)	Reference
Cabo Verde Northwestern	CVN (117)	This study
Cabo Verde Southeastern	CVS (211)	This study
Madeira	MA (100)	Fernandes <i>et al.</i> 2001
Azores	AZ (100)	Velosa <i>et al.</i> 2002
Guiné-Bissau	GU (100)	Gonçalves <i>et al.</i> 2002
North Portugal	PN (427, 304, 419, 78)	Amorim <i>et al.</i> 2001, Gusmão <i>et al.</i> 1997, Perez-Lezaun <i>et al.</i> 2000, Pinheiro <i>et al.</i> 1997
Centre Portugal	PC(146, 153)	Geada <i>et al.</i> 2000, Santos <i>et al.</i> 1996
Afro-American	UA (301, 195, 380)	Budowle <i>et al.</i> 2001, Holt <i>et al.</i> 2000, Hammond <i>et al.</i> 1994
Caucasian American	UC (506, 200, 308)	Budowle <i>et al.</i> 2001, Holt <i>et al.</i> 2000, Hammond <i>et al.</i> 1994
Mozambique	MO (110)	Alves <i>et al.</i> 2001
Berber	BE (100)	Perez-Lezaun <i>et al.</i> 2000
Morocco	MR (94)	Perez-Lezaun <i>et al.</i> 2000
Spain	ES (98)	Perez-Lezaun <i>et al.</i> 2000

n = sample size.

among the populations belonging to the CVN and CVS groups. The relationships among these populations were assessed in two ways using the MVSP v3.12 software package: by performing a Principal Component Analysis (PCA) and Neighbour-Joining analysis with Nei *et al.* (1983) average distances. We performed two NJ analyses, the first including only those populations for which a full set of 15 STR markers was available, and the second including all populations but restricted to the 12 STRs common to all.

Results and Discussion

Allele Diversity

Allele frequencies for the 17 loci analysed in both Cabo Verde populations as well as gene diversity values and forensic efficiency parameters for each locus are deposit as Supplementary Material to this article. The sample size studied for each locus and in each group of islands varied from 79 to 211. Both the Cabo Verde populations are in Hardy-Weinberg equilibrium. Observed and expected heterozygosity values are not statistically different. Nevertheless, CVN and CVS show significant differences at 6 loci (D3S1358, D18S51, D8S1179, TH01, TPO and FES; G-like test, $P < 0.01$), and overall both populations are significantly different from each other (X^2 test, $P < 0.0001$). However, this differentiation was not detected by the AMOVA, which attributed more than 99% to variation within populations and the remaining to variation among the two groups (CVN and CVS) and among populations within groups (data not shown). Table 2 shows the results from pairwise genotypic differentiation tests between CVN and CVS with other populations, including mainland Portugal (PC, PN), Madeira (MA) and the Azores (AZ) archipelagos, as well as a Caucasian (UC) and three African populations: one from US Afro-Americans (UA) and the others from Guiné-Bissau (GU) and Mozambique (MO). CVN and CVS showed remarkable differences with UA, MO and GU.

Interestingly, the previous report on the distribution of 4 STRs in Cabo Verde (TH01, VWA TPO, and CYP19, Dios *et al.* 1998) failed to detect any geographical differentiation within the archipelago. Clearly there

Table 2 Genotypic differentiation exact tests between pairs of populations. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$), NA not available

Populations compared	D3S1358	TH01	D21S11	D18S51	PentaE	D5S818	D13S317	D7S820	D16S539	CSF1PO	PentaD	VWA	D8S1179	TPO	FGA	CD4	FES
CVN/CVN	*	*	**	**	**	**	**	*	*	*	*	*	**	**	**	**	**
CVN/UA	*	**	**	**	**	**	**	*	*	*	*	*	**	**	**	**	**
CVN/MA	*	**	**	**	**	**	**	*	*	*	*	*	**	**	**	**	**
CVN/AZ	*	*	**	**	**	**	*	*	*	*	*	*	**	**	**	**	**
CVN/PN	**	*	**	**	NA	**	**	*	*	*	*	*	**	*	**	**	**
CVN/PC	**	*	**	**	NA	**	**	*	*	*	*	*	**	*	**	**	NA
CVN/UC	**	**	**	**	**	**	**	*	*	*	*	*	**	*	**	**	**
CVN/GU	*	**	**	**	**	**	**	*	*	*	*	*	**	*	**	**	**
CVN/MO	**	**	**	**	NA	**	**	*	*	*	*	*	**	*	**	**	NA
CVS/UA	**	**	**	**	**	**	**	*	*	*	*	*	**	*	**	**	**
CVS/MA	**	**	**	**	**	**	**	*	*	*	*	*	**	*	**	**	**
CVS/AZ	**	**	**	**	**	**	**	*	*	*	*	*	**	*	**	**	**
CVS/PN	**	**	**	**	NA	**	**	*	*	*	*	*	**	*	**	**	NA
CVS/PC	**	**	**	**	NA	**	**	*	*	*	*	*	**	*	**	**	NA
CVS/UC	**	**	**	**	**	**	**	*	*	*	*	*	**	*	**	**	**
CVS/GU	**	**	*	**	**	**	**	*	*	*	*	*	**	*	**	**	NA
CVS/MO	**	**	*	*	NA	**	*	*	*	*	*	*	**	*	**	**	NA

are significant differences between the two groups at several loci (Table 2).

Allele Comparisons

Several interesting features are worth noting when comparing the allele composition of Cabo Verdeans with other populations. Several alleles found in CV do appear in GU but not in other sub-Saharan populations (e.g., allele 13 from D3S1358, allele 29,2 from D21S11). There are numerous examples of alleles that have been found in CVS and GU and other African populations but not in CVN (e.g., allele 10,2 from D18S51, allele 12 from VWA or allele 18 from PentaE). Of course this may simply be regarded as a result of a genetic drift phenomenon, once the populations from CVN were settled by fugitive slaves from CVS (Barcelos, 1899; Carreira, 1983). The Cabo Verde population also presents several alleles not found in other African populations. This is the case for allele 24,2 of D21S11, although it exists in the North of Portugal. Madeira and CVN share allele 15 of the CSF 1PO locus, which until now has only been reported in Caucasian populations. Allele 23,2 of FGA is present in CVS and in both mainland Portugal and Madeira. These alleles, which are mainly found in Caucasian populations and not in sub-Saharans or Afro-Americans, further suggest an

Table 3 Contribution of the Portuguese population to present day Cabo Verde gene pool

	Estimator	Estimated admixture coefficient	Bootstrap average ^a	Bootstrap standard deviation
STRs	m _R	0.427	0.436	0.034
	m _C	0.412	0.424	0.046
	m _Y	0.544	0.530	0.102
HLA ^b	m _R	0.432	0.534	0.087
	m _C	0.558	0.558	0.115
	m _Y	0.449	0.269	1.252

^aComputed as an average over 10,000 random resampling of the original data

^bData from Spínola *et al.* (2002)

input from the European pool involved in the settlement of Cabo Verde. Finally allele 7 of D13S317, which appears in one individual from Santiago Is. (CVS), has only been spotted in Bahamians and Jamaicans (Budowle *et al.* 1999). Results from the admixture analysis show an extremely high percentage of the *p*₁ contribution to the Cabo Verde gene pool (Table 3). The three estimates used gave values ranging from 41% (m_C) to 54% (m_Y). However, looking at the standard deviation of bootstraps, it seems that m_R and m_C may actually give a more accurate measure of *p*₁ than m_Y. This was also observed using HLA data from the same populations (Spínola *et al.* 2002). Both m_R and m_C behaved

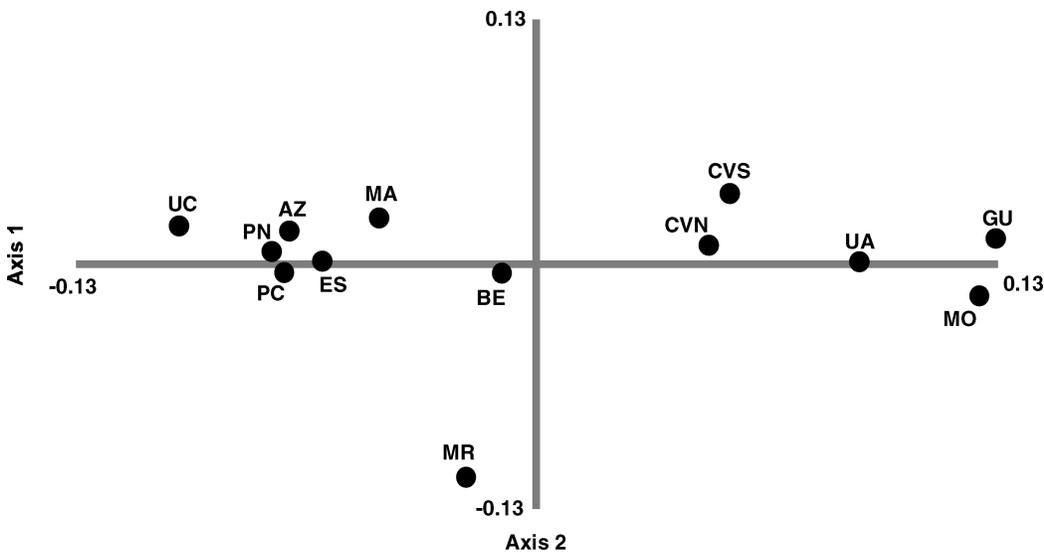


Figure 2 Principal component analysis of the populations included in Table 1, using 12 STR markers. First axis extracted 50.3% of information and the second 10.4%.

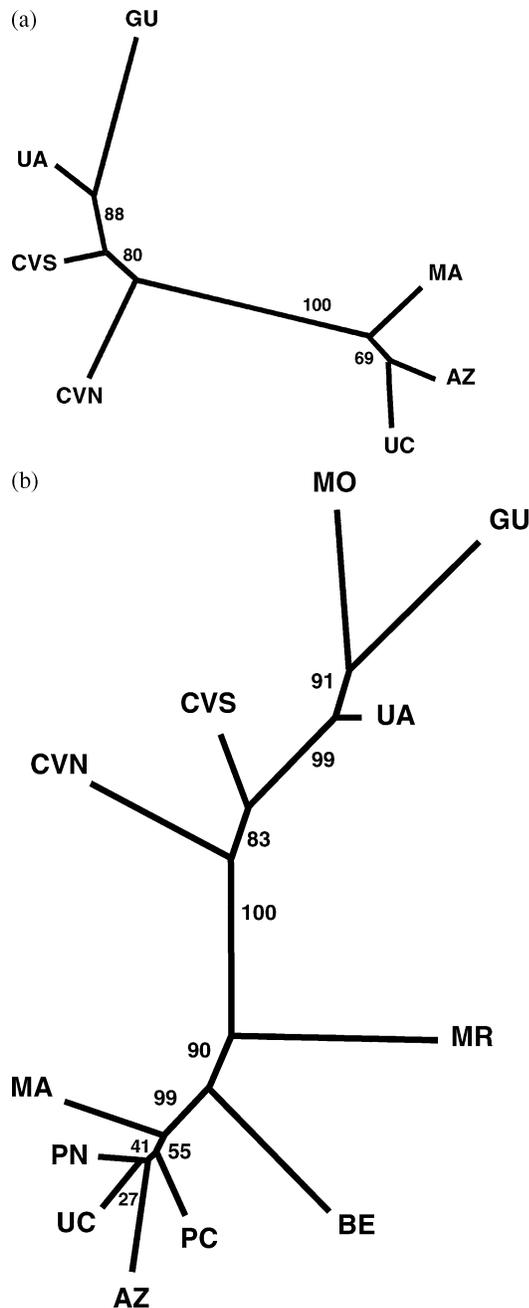


Figure 3 Neighbor-Joining trees based on Nei's DA genetic distances (Nei *et al.* 1983) and constructed using information from 15 (a) and 12 loci (b). Bootstrap values are indicated on branches. Population codes are as in Table 1.

well and gave an independent confirmation of the p_1 component in the Cabo Verde gene pool. In this case, m_Y proved to be inappropriate, with a high variance, most probably due to the fact that in this case only 3 loci were surveyed (Bertorelle & Excoffier, 1998).

Clearly in the latter case conventional estimators behave better.

Population's Comparisons

A principal component analysis of Cabo Verde and other populations using 12 loci discriminates between sub-Saharan and non-sub-Saharan samples (Figure 2). Axis 1 separates both Cabo Verde populations from the typically sub-Saharan ones, but it is axis 2 that is responsible for the discrimination within the archipelago. The differences between CVS and CVN are almost of the same magnitude as those separating GU (north western African coast) and MO (south eastern Africa). Although a large number of alleles are contributing to the spatial ordination of populations, the ones with slightly higher influence belong to TH01, D13S317 and TPO (axis 1), and TH01, D13S317, D3S1358, D21S11, D18S51, D7S820, VWA (axis 2). The two NJ analyses based on, respectively, 15 and 12 loci reveal a basically identical pattern with all nodes well supported by bootstrap values. It is important to note that the first analysis (based on a larger number of loci) clusters CVN with Caucasians, and CVS with sub-Saharans (Figure 3a). Besides the differences between CVN and CVS, it is worth noting the particular position occupied by Madeira (MA), as an intermediate between Cabo Verde and the Azores and Caucasian populations (AZ, UC). The NJ analysis using fewer loci moves CVN to the sub-Saharan cluster but even so, it is well separated from the others (Figure 3b). The cluster further shows that the genetic structure of Guineans (GU) is very different from the other two African populations analysed, MO and UA, suggesting that differences in STR allele diversity and composition do exist among populations from different African regions.

Both groups of islands also show differences in mtDNA haplotype and haplogroup frequencies, which were attributed to a loss of diversity in the CVN group (Brehm *et al.* 2002). In fact, "heterozygosity" values for CVS are similar to those found in mainland Senegal, but the diversities of the CVN pool are comparable to other island populations such as La Gomera and La Palma (Canary Islands, Rando *et al.* 1999). Thus, our data agrees with the findings from mtDNA, in that the CVN population probably represents a sub-set of the CVS pool

from where it may have originated. If, on the basis of their mtDNA content, Cabo Verde is a typically African population with almost null European input, the same is not true for autosomal and Y markers. A recent study with HLA markers found a much larger percentage of alleles uncommon in Africa but present at high frequencies in Caucasian populations in the Cabo Verde population (Spínola *et al.* 2002). Even more striking are the findings for the Y chromosome. We are presently typing Y binary loci in several European and sub-Saharan populations including Cabo Verde. The percentage of typical sub-Saharan Ys [E3a according to the Y Chromosome Consortium (2002)] is 21.1% in CVS but only 8.7% in CVN. Inversely, the percentage of European Ys (haplogroups E*, E1 and E2) accounts for 30.4% in CVN but only 19.8% in CVS. The latter haplogroups are present in Guineans at a mere 16% (Rosa, unpublished data).

The present study shows the relevance of the methodology of sampling for construction of databases to be used in forensic and population genetics studies. Also, population databases constructed with the largest number of loci possible allow a better definition and discrimination of closely related populations belonging to the same continent, and can demonstrate remarkable differences in genetic composition.

Acknowledgements

The authors are grateful to the Chairman of the Joint Chiefs of Staff from the Republic of Cabo Verde and the Ministry of Public Health for allowing collection of samples. We would like also to thank two anonymous reviewers for valuable comments in earlier drafts of this manuscript. This work was supported by an ICCTI (Instituto de Cooperação Científica e Tecnológica Internacional, Lisbon) grant to AB. This work was done under a research partnership with PROMEGA Corp.

References

- Alves, C., Gusmão, L. & Amorim, A. (2001). STR data (AmpF/STR Profiler Plus and Geneprint CTTv) from Mozambique. *Forensic Sci Int* **119**, 131–133.
- Amorim, A., Gusmão, L. & Alves, C. (2001). STR data (AmpF/STR Profiler Plus) from north Portugal. *Forensic Sci Int* **115**, 119–121.
- Barcellos, Ch. J. (1899). Subsídios para a história de Cabo Verde e Guiné. Academia Real das Ciências, Lisboa.
- Bertorelle, G. & Excoffier, L. (1998). Inferring admixture proportions from molecular data. *Mol Biol Evol* **15**, 1298–1311.
- 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. *Ann Hum Genet* **66**, 49–60.
- Budowle, B., Moretti, T., Baumstark, A., Defenbaugh, D. & Keys, K. (1999). Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J Forensic Sciences* **44**, 1277–1286.
- Budowle, B., Masibay, A., Anderson, S., Barna, C., Biega, L., Brenneke, S., Brown, B., Cramer, J., DeGroot, G., Douglas, D., Duceman, B., Eastman, A., Giles, R., Hamill, J., Haase, D., Janssen, D., Kupferschmid, T., Lawton, T., Lemire, C., Llewellyn, B., Moretti, T., Neves, J., Palaski, C., Schueler, S., Sgueglia, J., Sprecher, C., Tomsey, C. & Yet, D. (2001). STR primer concordance study. *Forensic Sci Int* **124**, 47–54.
- Carreira, A. (1983). Cabo Verde. Formação e extinção de uma sociedade escravocrata (1460–1878). Instituto Caboverdeano do Livro, Praia, Santiago.
- Chakraborty, R., Kamboh, M., Nwankwo, M. & Ferrel, R. (1992). Caucasian genes in American Blacks: new data. *Am J Hum Genet* **50**, 145–155.
- Dios, S., Luis, J. R., Teixeira Ribeiro, J. C. & Caeiro, B. (1998). Population database of STRs in West Africa: a genetic study of TPOX, HUMVWA31/A, HUMTH01, and CYP19. *Genetica* **104**, 77–83.
- Edwards, M. C., Clemens, P. R., Tristan, M., Pizzuti, A. & Gibbs, R. A. (1991). Pentanucleotide repeat length polymorphism at the human CD4 locus. *Nucleic Acids Reserch* **19**, 4792.
- Fernandes, A. T., Brehm, A., Alves, C., Gusmão, L. & Amorim, A. (2001). Genetic profile of the Madeira Archipelago population using the new PowerPlex® 16 System kit. *Forensic Sci Int* **125**, 281–283.
- Geada, H., Brito, R., Ribeiro, T. & Espinheira, R. (2000). Portuguese population and paternity investigation studies with a multiplex PCR—the AmpF1STR Profiler Plus™. *Forensic Sci Int* **108**, 31–37.
- Godinho, V. M. (1965). Os Descobrimentos e a Economia Mundial. Arcádia Ed., Vol 1 and 2, Lisboa.
- Gonçalves, R., Jesus, J., Fernandes, A. T. & Brehm, A. (2002). Genetic Profile of a multi-ethnic population from Guiné-Bissau (West African coast) using the new PowerPlex® 16 System kit. *Forensic Sci Int* **129**, 78–80.
- Gusmão, L., Prata, M. J., Amorim, A., Silva, F. & Bessa, I. (1997). Characterization of four tandem repeat loci (TH01, VWA31/A, CD4 and TP53) in North Portugal. *Hum Biol* **69**, 31–41.
- Hammond, H., Jin, L., Zhong, Y., Caskey, C. T. & Chakraborty, R. (1994). Evaluation of 13 short tandem

- repeat *loci* for use in personal identification applications. *Am J Hum Genet* **55**, 175–189.
- Holt, C., Stauffer, C., Wallin, J., Lazaruk, K., Nguyen, T., Budowle, B. & Walsh, P. (2000). Practical applications of genotyping surveys for forensic STR testing. *Forensic Sci Int* **112**, 91–109.
- Jones, D. A. (1972). Blood samples: probability of discrimination. *J Forensic Sci Soc* **12**, 355–359.
- Lareu, M. V., Phillips, C., Carracedo, A., Lincoln, P., Sindercombe, D. & Thompson, J. (1994). Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic Sci Int* **66**, 41–52.
- Lessa, A. & Ruffié, J. (1960). Seroantropologia das Ilhas de Cabo Verde. Mesa redonda sobre o homem Cabo-Verdiano. Junta de Investigações do Ultramar (ed.). Estudos, Ensaios e Documentos, Vol 32. Lisbon.
- Nei, M., Tajima, F. & Tatenos, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* **19**, 153–170.
- Ohno, Y., Sebetan, I. M. & Akaishi, A. (1982). A simple method for calculating the probability of excluding paternity with any number of codominant alleles. *Forensic Sci Int* **19**, 93–98.
- Pérez-Lezaun, A., Calafell, F., Clarimón, J., Bosch, E., Mateu, E., Gusmão, L., Amorim, A., Benchemsi, N. & Bertranpetit, J. (2000). Allele frequencies of 13 short tandem repeats in population samples from the Iberian Peninsula and Northern Africa. *Int J Legal Med* **113**, 208–214.
- Pinheiro, F., Pontes, L., Gené, M., Hugué, E., Pinto da Costa, J. & Moreno, P. (1997). Population study of the HUMTH01, HUMVWA31A, HUM13A1 and HUMFES/FPS STR polymorphisms in the North of Portugal. *J Forensic Sci* **42**, 121–124.
- Polymeropoulos, M. H., Rath, D. S., Xiao, H. & Merrill, C. R. (1991). Tetranucleotide repeat polymorphism at the human *c-fes/fps* proto-oncogene (FES). *Nucleic Acid Research* **19**, 4018.
- Rando, J., Cabrera, V., Larruga, J., Hernández, M., González, A., Pinto, F. & Bandelt, H.-J. (1999). Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* **63**, 413–428.
- Raymond, M. & Rousset, F. (1995). GENEPOP, a population genetics software for exact tests and ecumenicism. *J Hered* **86**, 248–249.
- Roberts, D. & Hiorns, R. (1965). Methods of analysis of genetic composition of a hybrid population. *Hum Biol* **37**, 38–43.
- Russell-Wood, A. J. (1998). The Portuguese empire, 1415–1808. A world on the move. Johns Hopkins Univ. Press. Baltimore.
- Santos, S., Budowle, B., Smerick, J., Keys, K. & Moretti, T. (1996). Portuguese population data on the six short tandem repeat *loci*- CSF1PO, TPOX, TH01, D3S1358, VWA and FGA. *Forensic Sci Int* **83**, 229–235.
- Schneider, S., Kueffer, J.-M., Roessli, D. & Excoffier, L. (2000). Arlequin : a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Spínola, H., Brehm, A., Williams, F., Jesus, J. & Middleton, D. (2002). Distribution of HLA alleles in Portugal and Cabo Verde. Relationships with the slave trade route. *Ann Hum Genet* **66**, 285–296.
- The Y Chromosome Consortium (2002). A nomenclature system for the tree of Human Y-chromosomal binary haplogroups. *Genome Research* **12**, 339–348.
- Velosa, R., Fernandes, A. T. & Brehm, A. (2002). Genetic profile of the Açores Archipelago population using the new PowerPlex® 16 System kit. *Forensic Sci Int* **129**, 68–71.

Supplementary Material is available at www.uma.pt/bioarticles.

Received: 3 July 2002

Accepted: 31 March 2003